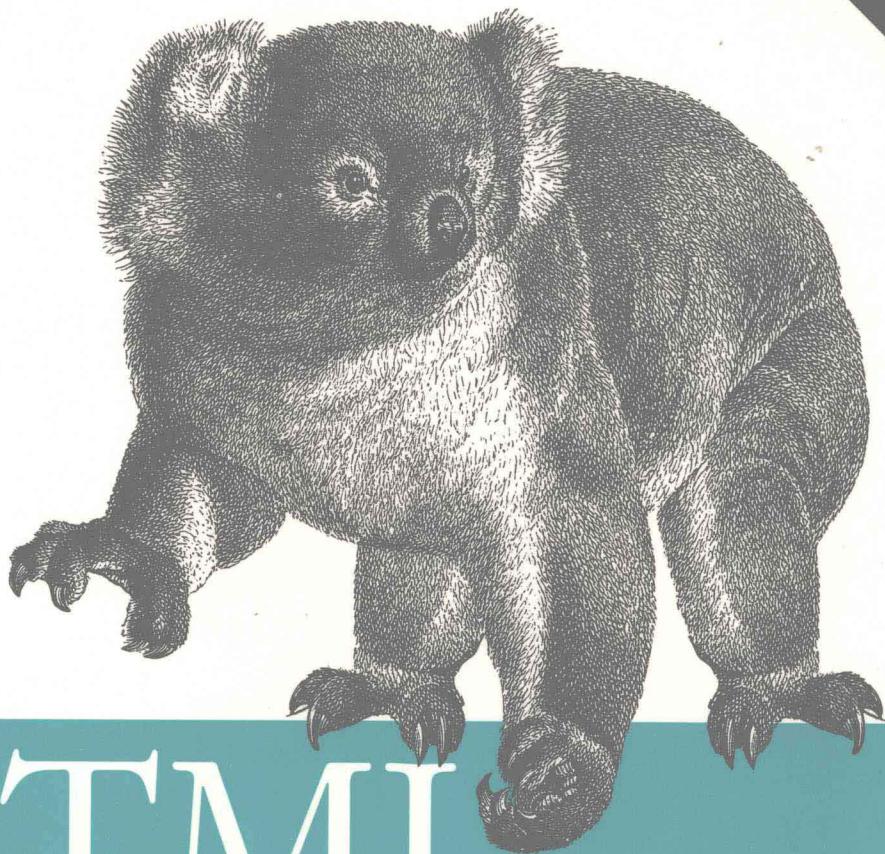


*Creating Effective Web Pages*

**4th Edition**



# HTML & XHTML

*The Definitive Guide*

**O'REILLY®**

*Chuck Musciano & Bill Kennedy*

---

# HTML and XHTML

## *The Definitive Guide*

*Fourth Edition*

Chuck Musciano and Bill Kennedy

**O'REILLY®**

*Beijing · Cambridge · Farnham · Köln · Paris · Sebastopol · Taipei · Tokyo*

## ***HTML and XHTML: The Definitive Guide, Fourth Edition***

by Chuck Musciano and Bill Kennedy

Copyright © 2000 O'Reilly & Associates, Inc. All rights reserved.

Portions of this book previously appeared in *HTML: The Definitive Guide*, Copyright © 1998, 1997, 1996 O'Reilly & Associates, Inc. All rights reserved.

Printed in the United States of America.

Published by O'Reilly & Associates, Inc., 101 Morris Street, Sebastopol, CA 95472.

**Editors:** Mike Loukides and Deb Cameron

**Production Editor:** Colleen Gorman

**Cover Designer:** Edie Freedman

### ***Printing History:***

April 1996:	First Edition.
May 1997:	Second Edition.
August 1998:	Third Edition.
August 2000:	Fourth Edition.

Nutshell Handbook, the Nutshell Handbook logo, and the O'Reilly logo are registered trademarks of O'Reilly & Associates, Inc. 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 O'Reilly & Associates, Inc. was aware of a trademark claim, the designations have been printed in caps or initial caps. The association between the image of a koala and the topic of HTML and XHTML is a trademark of O'Reilly & Associates, Inc.

While every precaution has been taken in the preparation of this book, the publisher assumes no responsibility for errors or omissions, or for damages resulting from the use of the information contained herein.

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

Musciano, Chuck.

HTML & XHTML, the definitive guide / Chuck Musciano & Bill Kennedy--4th ed.  
p. cm.

Rev. ed. of HTML, the definitive guide. 2nd ed. c1997.

ISBN 0-596-00026-X

1. HTML (Document markup language) 2. XHTML (Document markup language)  
I. Kennedy, Bill, 1951- II. Title.

QA76.76.H94 M875 2000

005.07'2--dc21

99-085714

ISBN: 0-596-00026-X

[12/00]

[M]

---

## *Preface*

Learning Hypertext Markup Language (HTML) and Extensible Hypertext Markup Language (XHTML) is like learning any new language, computer or human. Most students first immerse themselves in examples. Studying others is a natural way to learn, making learning easy and fun. Our advice to anyone wanting to learn HTML and XHTML is to get out there on the World Wide Web with a suitable browser and see for yourself what looks good, what's effective, what works for you. Examine others' documents and ponder the possibilities. Mimicry is how many of the current webmasters have learned the language.

Imitation can take you only so far, though. Examples can be both good and bad. Learning by example will help you talk the talk, but not walk the walk. To become truly conversant, you must learn how to use the language appropriately in many different situations. You could learn all that by example, if you live long enough.

Remember, too, that computer-based languages are more explicit than human languages. You've got to get the language syntax correct or it won't work. Then, too, there is the problem of "standards." Committees of academics and industry experts define the proper syntax and usage of a computer language like HTML. The problem is that browser manufacturers like Netscape Communications Corporation (now an America Online company) and Microsoft Corporation choose the parts of the standard they will use and which parts they will ignore. They even make up their own parts, which may eventually become standards.

Standards change, too. As we write this current edition, HTML is undergoing a conversion into XHTML, making it an application of the Extensible Markup Language (XML). HTML and XHTML are so similar that we often refer to them as a single language. But there are key differences; more about this later in the preface.

To be safe, the way to become fluent in HTML and XHTML is through a comprehensive, up-to-date language reference that covers the language syntax, semantics, and variations in detail to help you distinguish between good and bad usage.

There's one more step leading to fluency in a language. To become a true master of the language, you need to develop your own style. That means knowing not only what is appropriate, but what is effective. Layout matters. A lot. So does the order of presentation within a document, between documents, and between document collections.

Our goal in writing this book is to help you become fluent in HTML and XHTML, fully versed in their syntax, semantics, and elements of style. We take the natural learning approach, using examples: good ones, of course. We cover every element of the currently accepted versions (HTML 4.01 and XHTML 1.0) of the languages in detail, as well as all of the current extensions supported by the popular browsers, explaining how each element works and how it interacts with all the other elements.

And, with all due respect to Strunk and White, throughout the book we will give you suggestions for style and composition to help you decide how best to use HTML and XHTML to accomplish a variety of tasks, from simple online documentation to complex marketing and sales presentations. We'll show you what works and what doesn't, what makes sense to those who view your pages, and what might be confusing.

In short, this book is a complete guide to creating documents using HTML and XHTML, starting with basic syntax and semantics, and finishing with broad style guidelines to help you create beautiful, informative, accessible documents that you'll be proud to deliver to your browsers.

## *Our Audience*

We wrote this book for anyone interested in learning and using the language of the Web, from the most casual user to the full-time design professional. We don't expect you to have any experience in HTML or XHTML before picking up this book. In fact, we don't even expect that you've ever browsed the World Wide Web, although we'd be very surprised if you haven't at least experimented with this technology by now. Being connected to the Internet is not necessary to use this book, but if you're not connected, this book becomes like a travel guide for the homebound.

The only things we ask you to have are a computer, a text editor that can create simple ASCII text files, and copies of the latest leading web browsers—preferably Netscape Navigator and Internet Explorer. Because HTML and XHTML documents are stored in a universally accepted format—ASCII text—and because the languages

are completely independent of any specific computer, we won't even make an assumption about the kind of computer you're using. However, browsers do vary by platform and operating system, which means that your HTML or XHTML documents can look quite different depending on the computer and version of browser. We will explain how the various browsers use certain language features, paying particular attention to how they are different.

If you are new to HTML, the World Wide Web, or hypertext documentation in general, you should start by reading Chapter 1, *HTML, XHTML, and the World Wide Web*. In it, we describe how all the World Wide Web technologies come together to create webs of interrelated documents.

If you are already familiar with the Web, but not with HTML or XHTML specifically, or if you are interested in the new features in the latest standard version of HTML and XHTML, start by reading Chapter 2, *Quick Start*. This chapter is a brief overview of the most important features of the language and serves as a roadmap to how we approach the language in the remainder of the book.



Subsequent chapters deal with specific language features in a roughly top-down approach to HTML and XHTML. Read them in order for a complete tour through the language, or jump around to find the exact feature you're interested in.

## Text Conventions

Throughout the book, we use a **constant-width** typeface to highlight any literal element of the HTML/XHTML standards, tags, and attributes. We always use lower-case letters for tags.\* We use *italic* to indicate new concepts when they are defined and for those elements you need to supply when creating your own documents, such as tag attributes or user-defined strings.

We discuss elements of the language throughout the book, but you'll find each one covered in depth (some might say in nauseating detail) in a shorthand, quick-reference definition box that looks like the following box. The first line of the box contains the element name, followed by a brief description of its function. Next, we list the various attributes, if any, of the element: those things that you may or must specify as part of the element.

We use the following symbols to identify tags and attributes that are not in the HTML 4.01 or XHTML 1.0 standards, but are additions to the languages:

-  Netscape Navigator extension to the standards
-  Internet Explorer extension to the standards

---

\* HTML is case-insensitive with regard to tag and attribute names, but XHTML is case-sensitive. And some HTML items like source filenames, are case-sensitive, so be careful.

## ***<html>***

**Function:**

Delimits a complete HTML document

**Attributes:**

DIR  
VERSION  
LANG

**End tag:**

`</html>`; may be omitted in HTML

**Contains:**

*head\_tag, body\_tag, frames*

The description also includes the ending tag, if any, for the element, along with a general indication whether or not the end tag may be safely omitted in general use with HTML. With the few tags that do not have an end tag in HTML, but for which XHTML requires one, the language lets you indicate that ending with a forward slash (/) at the end of the tag, such as `<br />`. In these cases, the tag also may contain attributes, indicated with an intervening ellipsis, such as `<br ... />`.

“Contains” names the rule in the HTML grammar that defines the elements to be placed within this tag. Similarly, “Used in” lists those rules that allow this tag as part of their content. These rules are defined in Appendix A, *HTML Grammar*.

Finally, HTML and XHTML are fairly intertwined languages. You will occasionally use elements in different ways depending on context, and many elements share identical attributes. Wherever possible, we place a cross-reference in the text that leads you to a related discussion elsewhere in the book. These cross-references, like the one at the end of this paragraph, serve as a crude paper model of hypertext documentation, one that would be replaced with a true hypertext link should this book be delivered in an electronic format. [tag syntax, 3.3.1]

We encourage you to follow these references whenever possible. Often, we’ll only cover an attribute briefly and expect you to jump to the cross-reference for a more detailed discussion. In other cases, following the link will take you to alternative uses of the element under discussion or to style and usage suggestions that relate to the current element.

## ***Versions and Semantics***

The latest HTML standard is Version 4.01, but most updates and changes to the language standard were made in Version 4.0. Therefore, throughout the book, we



generally refer to the HTML standard as HTML 4, encompassing all Versions 4.0 and later. We explicitly state the “dot” version number only when it is relevant.

The XHTML standard is currently in its first iteration, 1.0. For the most part, XHTML 1.0 is identical to HTML 4.01; we detail their differences in Chapter 16, *XHTML*. Throughout the book, we specifically note cases where XHTML handles a feature or element differently than the original language, HTML.

The HTML and XHTML standards make very clear the distinction between “element types” of a document and the markup “tags” that delimit those elements. For example, the standard refers to the paragraph element type, which is not the same as the `<p>` tag. The paragraph element consists of the accepted element-type name within the starting tag (`<p>`), intervening content, and the ending paragraph (`</p>`) tag. The `<p>` tag is the starting tag for the paragraph element, and its contents, known as attributes, ultimately affect the paragraph element type’s contents.

Although these are important distinctions, we’re pragmatists. It is the markup tag that authors apply in their documents and that affects the intervening content, if any. Accordingly, throughout the book, we relax the distinction between element types and tags, most often talking about tags and all related contents, not necessarily using the term element-type when it would be technically appropriate to make the distinction. Forgive us the transgression, but we do so for the sake of clarity.

## *Is HTML Going Away?*

Heavens, no. Why would we even think such a thing?

Well, actually, the language has reached middle age in standard Version 4.01 and is not expected to change again. Rather, HTML is being subsumed and modularized as part of Extensible Markup Language (XML). Its new name is XHTML, Extensible Hypertext Markup Language.

The emergence of XHTML is just another chapter in the often tumultuous history of HTML and the World Wide Web, where confusion for authors is the norm, not the exception. At the worst point, the elders of the World Wide Web Consortium (W3C) responsible for accepted and acceptable uses of the language—i.e., standards—lost control of the language in the browser “wars” between Netscape Communications and Microsoft. The abortive HTML+ standard never got off the ground, and HTML 3.0 became so bogged down in debate that the W3C simply shelved the entire draft standard. HTML 3.0 never happened, despite what some opportunistic marketers claimed in their literature. Instead, by late 1996, the browser manufacturers convinced the W3C to release HTML standard Version 3.2, which for all intents and purposes simply standardized most of the leading browser’s (Netscape’s) HTML extensions.



Fortunately for those of us who appreciate and strongly support standards, the W3C took back its primacy role with HTML 4.0, which stands today as HTML Version 4.01, released in December 1999. The standard is clearer and cleaner than any previous ones, establishes solid implementation models for consistency across browsers and platforms, provides strong supports and incentives for the companion Cascading Style Sheets (CSS) standard for HTML-based displays, and makes provisions for alternative (non-visual) user-agents, as well as for more universal language supports.

Cleaner and clearer aside, the W3C realized that HTML could never keep up with the demands of the web community for more ways to distribute, process, and display documents. HTML only offers a limited set of document creation primitives and is hopelessly incapable of handling non-traditional content like chemical formulae, musical notation, or mathematical expressions. Nor can it well support alternative display media, such as handheld computers or intelligent cellular phones, for instance.

To address these demands, the W3C developed the Extensible Markup Language (XML) standard. XML provides the way to create new, standards-based markup languages that don't take an act of the W3C to implement. XML-compliant languages deliver information that can be parsed, processed, displayed, sliced, and diced by the many different communication technologies that have emerged since the Web sparked the digital communication revolution a decade ago. XHTML is HTML reformulated to adhere to the XML standard. It is the foundation language for the future of the Web.

Why not just drop HTML for XHTML? For many reasons. First and foremost, don't expect everyone to just drop everything and start using XHTML standards (Version 1.0 just got recommended in January 2000). There's just too much current investment in HTML-based documentation and expertise for that to happen anytime soon. Besides, XHTML is HTML 4.01 reformulated as an application of XML. Know HTML 4 and you're all ready for the future.\*

The paradox in all this is that even the HTML 4.01 standard is not the definitive resource. There are many more features of HTML in popular use and supported by the popular browsers than are included in the latest language standard. And there are many parts of the standards that are ignored. We promise you, things can get downright confusing when you're trying to sort it all out.

We've managed to sort things out, so you don't have to sweat over what works with what browser and what doesn't work. This book, therefore, is the definitive guide to HTML and XHTML. We give details for all the elements of the HTML 4.01

---

\* We plumb the depths of XML and XHTML in Chapter 15, *XML*, and Chapter 16, *XHTML*.

and XHTML 1.0 standards, plus the variety of interesting and useful extensions to the language—some proposed standards—that the popular browser manufacturers have chosen to include in their products, such as:

- Cascading Style Sheets
- Java and JavaScript
- Layers
- Multiple columns

And while we tell you about each and every feature of the language, standard or not, we also tell you which browsers or different versions of the same browser implement a particular extension and which don't. That's critical knowledge when you want to create web pages that take advantage of the latest version of Netscape Navigator versus pages that are accessible to the larger number of people using Internet Explorer or even Lynx, a once-popular text-only browser for Unix systems.

In addition, there are a few things that are closely related but not directly part of HTML. For example, we touch, but do not handle, CGI and Java programming. CGI and Java programs work closely with HTML documents and run with or alongside browsers, but are not part of the language itself, so we don't delve into them. Besides, they are comprehensive topics that deserve their own books, such as *CGI Programming with Perl*, by Scott Guelich, Shishir Gundavaram, and Gunther Birznieks, and *Java in a Nutshell*, by David Flanagan, both published by O'Reilly & Associates.

This is your definitive guide to HTML and XHTML as they are and should be used, including every extension we could find. Some extensions aren't documented anywhere, even in the plethora of online guides. But, if we've missed anything, certainly let us know and we'll put it in the next edition.

## *We'd Like to Hear from You*

We have tested and verified all of the information in this book to the best of our ability, but you may find that features have changed (or even that we have made mistakes!). Please let us know about any errors you find, as well as your suggestions for future editions, by writing:

O'Reilly & Associates, Inc.  
101 Morris Street  
Sebastopol, CA 95472  
800-998-9938 (in the U.S. or Canada)  
(707) 829-0515 (international/local)  
(707) 829-0104 (FAX)

To ask technical questions or comment on the book, send email to:

*bookquestions@oreilly.com*

Since the HTML and XHTML standards and browser additions to the languages are evolving so rapidly, some of the information in this book may be slightly out of date by the time you read it. We have a web site for the book, where we'll list errata and plans for future editions. Here you'll also find all the source code from the book available for download so you don't have to type it all in:

*<http://www.oreilly.com/catalog/html4/>*

For more information about our books, conferences, software, Resource Centers, and the O'Reilly Network, see our web site at:

*<http://www.oreilly.com/>*

## *Acknowledgments*

We did not compose, and certainly could not have composed, this book without generous contributions from many people. Our wives, Jeanne and Cindy, and our children, Eva, Ethan, Courtney, and Cole (they happened *before* we started writing), formed the front lines of support. And there are numerous neighbors, friends, and colleagues who helped by sharing ideas, testing browsers, and letting us use their equipment to explore HTML. You know who you are, and we thank you all.

In addition, we thank our technical reviewers, Robert Eckstein, Kane Scarlett, Eric Raymond, and Chris Tacy, for carefully scrutinizing our work. We took most of your keen suggestions. We especially thank Mike Loukides, our editor, who had to bring to bear his vast experience in book publishing to keep us two mavericks corralled. And special thanks to Deb Cameron for her perseverance and insight in bringing the Fourth Edition to fruition.

---

# Table of Contents

<i>Preface</i> .....	<i>xi</i>
<b>1. HTML, XHTML, and the World Wide Web</b> .....	<b>1</b>
1.1. The Internet, Intranets, and Extranets .....	1
1.2. Talking the Internet Talk .....	5
1.3. HTML: What It Is .....	8
1.4. XHTML: What It Is .....	9
1.5. HTML and XHTML: What They Aren't .....	10
1.6. Nonstandard Extensions .....	11
1.7. Tools for the Web Designer .....	13
<b>2. Quick Start</b> .....	<b>16</b>
2.1. Writing Tools .....	17
2.2. A First HTML Document .....	17
2.3. Embedded Tags .....	18
2.4. HTML Skeleton .....	19
2.5. The Flesh on an HTML or XHTML Document .....	20
2.6. Text .....	21
2.7. Hyperlinks .....	26
2.8. Images Are Special .....	29
2.9. Lists, Searchable Documents, and Forms .....	32
2.10. Tables .....	34
2.11. Frames .....	35
2.12. Style Sheets and JavaScript .....	36
2.13. Forging Ahead .....	37

<b>3. Anatomy of an HTML Document</b>	<b>38</b>
3.1. Appearances Can Deceive	38
3.2. Structure of an HTML Document	40
3.3. Tags and Attributes	40
3.4. Well-Formed Documents and XHTML	45
3.5. Document Content	45
3.6. HTML Document Elements	47
3.7. The Document Header	50
3.8. The Document Body	55
3.9. Editorial Markup	56
3.10. The <code>&lt;bdo&gt;</code> Tag	59
 <b>4. Text Basics</b>	 <b>61</b>
4.1. Divisions and Paragraphs	61
4.2. Headings	69
4.3. Changing Text Appearance	76
4.4. Content-Based Style Tags	77
4.5. Physical Style Tags	84
4.6. HTML's Expanded Font Handling	89
4.7. Precise Spacing and Layout	96
4.8. Block Quotes	109
4.9. Addresses	114
4.10. Special Character Encoding	116
 <b>5. Rules, Images, and Multimedia</b>	 <b>118</b>
5.1. Horizontal Rules	118
5.2. Inserting Images in Your Documents	126
5.3. Document Colors and Background Images	154
5.4. Background Audio	162
5.5. Animated Text	164
5.6. Other Multimedia Content	168
 <b>6. Links and Webs</b>	 <b>171</b>
6.1. Hypertext Basics	171
6.2. Referencing Documents: The URL	172
6.3. Creating Hyperlinks	189
6.4. Creating Effective Links	198
6.5. Mouse-Sensitive Images	203
6.6. Creating Searchable Documents	214

6.7. Relationships .....	218
6.8. Supporting Document Automation .....	223
<b>7. Formatted Lists .....</b>	<b>227</b>
7.1. Unordered Lists .....	227
7.2. Ordered Lists .....	230
7.3. The <li> Tag .....	234
7.4. Nesting Lists .....	237
7.5. Definition Lists .....	239
7.6. Appropriate List Usage .....	243
7.7. Directory Lists .....	244
7.8. Menu Lists .....	245
<b>8. Cascading Style Sheets .....</b>	<b>248</b>
8.1. The Elements of Styles .....	249
8.2. Style Syntax .....	258
8.3. Style Classes .....	263
8.4. Style Properties .....	268
8.5. Tag-less Styles: The <span> Tag .....	299
8.6. Applying Styles to Documents .....	300
<b>9. Forms .....</b>	<b>304</b>
9.1. Form Fundamentals .....	305
9.2. The <form> Tag .....	305
9.3. A Simple Form Example .....	314
9.4. Using Email to Collect Form Data .....	315
9.5. The <input> Tag .....	317
9.6. The <button> Tag .....	329
9.7. Multiline Text Areas .....	331
9.8. Multiple Choice Elements .....	333
9.9. General Form Control Attributes .....	338
9.10. Labeling and Grouping Form Elements .....	341
9.11. Creating Effective Forms .....	345
9.12. Forms Programming .....	349
<b>10. Tables .....</b>	<b>356</b>
10.1. The Standard Table Model .....	356
10.2. Table Tags .....	358
10.3. Newest Table Tags .....	377
10.4. Beyond Ordinary Tables .....	388

<b>11. Frames</b>	<b>390</b>
11.1. An Overview of Frames	390
11.2. Frame Tags	391
11.3. Frame Layout	393
11.4. Frame Contents	399
11.5. The <noframes> Tag	402
11.6. Inline Frames	404
11.7. Named Frame or Window Targets	406
<b>12. Executable Content</b>	<b>412</b>
12.1. Applets and Objects	412
12.2. Embedded Content	416
12.3. JavaScript	433
12.4. JavaScript Style Sheets	441
<b>13. Dynamic Documents</b>	<b>449</b>
13.1. An Overview of Dynamic Documents	449
13.2. Client-Pull Documents	450
13.3. Server-Push Documents	455
<b>14. Netscape Layout Extensions</b>	<b>460</b>
14.1. Creating Whitespace	461
14.2. Multicolumn Layout	466
14.3. Layers	472
<b>15. XML</b>	<b>484</b>
15.1. Languages and Metalanguages	485
15.2. Documents and DTDs	488
15.3. Understanding XML DTDs	489
15.4. Element Grammar	493
15.5. Element Attributes	498
15.6. Conditional Sections	500
15.7. Building an XML DTD	501
15.8. Using XML	503
<b>16. XHTML</b>	<b>506</b>
16.1. Why XHTML?	506
16.2. Creating XHTML Documents	508
16.3. HTML Versus XHTML	512
16.4. Should You Use XHTML?	517



<b>17. Tips, Tricks, and Hacks</b> .....	<b>521</b>
17.1. Top of the Tips .....	521
17.2. Trivial or Abusive? .....	524
17.3. Custom Bullets .....	524
17.4. Tricks with Tables .....	525
17.5. Transparent Images .....	532
17.6. Tricks with Windows and Frames .....	535
<b>A. HTML Grammar</b> .....	<b>539</b>
<b>B. HTML/XHTML Tag Quick Reference</b> .....	<b>551</b>
<b>C. Cascading Style Sheet Properties Quick Reference</b> .....	<b>582</b>
<b>D. The HTML 4.01 DTD</b> .....	<b>590</b>
<b>E. The XHTML 1.0 DTD</b> .....	<b>608</b>
<b>F. Character Entities</b> .....	<b>627</b>
<b>G. Color Names and Values</b> .....	<b>633</b>
<b>Index</b> .....	<b>637</b>

---

*In this chapter:*

- *The Internet, Intranets, and Extranets*
- *Talking the Internet Talk*
- *HTML: What It Is*
- *XHTML: What It Is*
- *HTML and XHTML: What They Aren't*
- *Nonstandard Extensions*
- *Tools for the Web Designer*

# 1

## *HTML, XHTML, and the World Wide Web*

Though it began as a military experiment and spent its adolescence as a sandbox for academics and eccentrics, recent events have transformed the worldwide network of computer networks—also known as the *Internet*—into a rapidly growing and wildly diversified community of computer users and information vendors. Today, you can bump into Internet users of nearly any and all nationalities, of any and all persuasions, from serious to frivolous individuals, from businesses to non-profit organizations, and from born-again Christian evangelists to pornographers.

In many ways, the World Wide Web—the open community of hypertext-enabled document servers and readers on the Internet—is responsible for the meteoric rise in the network's popularity. You, too, can become a valued member by contributing: writing HTML and XHTML documents and then making them available to web surfers worldwide.

Let's climb up the Internet family tree to gain some deeper insight into its magnificence, not only as an exercise of curiosity, but to help us better understand just who and what it is we are dealing with when we go online.

### *1.1. The Internet, Intranets, and Extranets*

Although popular media accounts are often confused and confusing, the concept of the Internet really is rather simple. It's a worldwide collection of computer networks—a network of networks—sharing digital information via a common set of networking and software protocols. Nearly anyone can connect a computer to the