Help for Web Authors

Covers Netscape 2.0 Extensions

HTML

The Definitive Guide

A NUTSHELL®

HANDBOOK

Chuck Musciano & Bill Kennedy

O'Reilly & Associates, Inc.

HTML: The Definitive Guide
by Chuck Musciano and Bill Kennedy

Published by O'Reilly & Associates, Inc., 103 Morris Street, Suite A, Sebastopol, CA 95472.

Editor: Mike Loukides

Production Editor: Mary Anne Weeks Mayo

Printing History:

April 1996: First Edition.

This book is printed on acid-free paper with 85% recycled content, 15% post-consumer waste. O'Reilly & Associates is committed to using paper with the highest recycled content available consistent with high quality.

ISBN: 1-56592-175-5 [5/96]

This book is dedicated to our wives and children,
Cindy, Courtney, and Cole, and
Jeanne, Eva, and Ethan.

Without their love and patience
we never would have had
the time or strength to write.

Preface

Our Function

Learning HTML is like learning any new language, computer or human. Most students first immerse themselves in examples. Think how adept you'd become if Mom, Dad, your brothers, and sisters all spoke fluent HTML. Studying others is a natural way to learn, making learning easy and fun. Our advice to anyone wanting to learn HTML 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' HTML source files 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 well and how to use the language appropriately in many different situations. You could learn 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 HTML syntax correct, or it won't work. Then, too, there is the problem of "standards." Committees of academics and industry experts try to define the proper syntax and usage of a computer language like HTML. The problem is that HTML browser manufacturers like Netscape and Spyglass choose what parts of the standard they will use and which parts they'll ignore. They even make up their own parts, which may eventually become standards.

To be safe, the better way to become HTML fluent is through a comprehensive language reference: a resource that covers the language syntax, semantics, and variations in detail, and helps you distinguish between good and bad usage.

There's one more step leading to fluency in a language. To become a true master of HTML, you need to develop your own style. That means knowing what's not only 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, fully versed in the language's syntax, semantics, and elements of style. We take the natural learning approach with examples: good ones, of course. We cover every element of the currently accepted version (2.0)* of the language in detail, as well as many of the so-called "extensions" the popular HTML browsers support, 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 give you suggestions for style and composition to help you decide how best to use the language and 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, starting with basic syntax and semantics, and finishing with broad style directions that should help you create beautiful, informative, accessible documents you'll be proud to deliver to your browsers.

Our Audience

We wrote this book for anyone interested in learning and using HTML, from the most casual user to a full-time design professional. We don't expect you to have any experience in the language before picking up this book. In fact, we don't even expect that you've ever used the World Wide Web, although we'd be surprised if you haven't at least experimented with this technology. 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 a World Wide Web browser like Netscape, Mosaic, or Internet Explorer for your computer: the very latest version, if possible. Because HTML is stored in a universally accepted format—ASCII text—and because the language is completely independent of any specific computer, we won't even make an assumption about the kind of computer you're using. However, browsers do

* Contrary to popular misbelief, there is no HTML version 3.0. See "Whatever Happened to HTML 3.0?" later in this Preface.

vary by platform and operating system, which means your HTML documents can and often do look quite different depending on the computer and version of browser. We will explain how certain language features are used by various popular browsers as we go through the book, so don't be surprised if your browser gets mentioned here or there.

If you are new to HTML, the World Wide Web, or hypertext documentation in general, you should start by reading Chapter 1, *HTML and the World Wide Web.* This chapter describes how all these technologies come together to create webs of interrelated documents.

If you are already familiar with the Web, but not HTML specifically, or if you are interested in the new features in HTML, start by reading Chapter 2, *HTML 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 the HTML. 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 the `courier` typeface to highlight any literal element of the HTML standard, and tags and attributes. We always use lowercase letters for HTML tags. (Although the language standard is case-insensitive with regard to tag and attribute names, this isn't so for other elements like source filenames, so be careful.) We use the *ttalic* font to indicate new concepts when they are defined and 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 nauseating detail) in a shorthand, quick-reference definition box that looks like the box on the following page.

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. Tags and attributes may also be parenthetically labeled if they are not included in the HTML 2.0 standard (the last official version), but are additions to the language. "Extension," as in the example, means that the nonstandard tag or attribute is supported by more than one of the popular browsers; otherwise the name of the extended browser appears—Mosaic, Netscape, or Internet Explorer—with the word "only" to identify the appropriate exception. HTML 3.0 tags are considered extensions.

<body>

Function:
Defines the document body

Attributes:
ALINK (Netscape only)
BACKGROUND (extension)
BGCOLOR (extension)
BGPROPERTIES (Internet Explorer only)
LINK (extension)
TEXT (extension)
TOPMARGIN (Internet Explorer only)
VLINK (extension)

End Tag:
</body> may be omitted

Contains:
body_content

Used in:
html_tag

The description also includes the ending tag, if any, for the tag, along with a general indication if the end tag may be safely omitted in general use.

"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 is a fairly "intertwined" language: Elements occasionally are used 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 alternate uses of the element under discussion, or to style and usage suggestions that relate to the current element.

Whatever Happened to HTML 3.0?

Depending on our mood, when people ask us about HTML 3.0, we respond with a groan, a bemused smile, or uproarious laughter. HTML 3.0 doesn't exist. Period. Anyone who tells you otherwise is sadly misinformed or just plain lying.

HTML 3.0 would be the next version of the hypertext markup language standard if the members and committees of the World Wide Web Consortium responsible for such things could and would come to some agreement and formalize its contents. They won't in the near future, if ever. The standard currently is "deprecated," "on the back burner," "still under consideration": you label the dodge. All we know is that, for the foreseeable future, there is only one HTML standard: version 2.0.

Frankly, the browser manufacturers are the tail wagging the HTML dog. They have implemented a variety of extensions to the language standard, some of which are based on proposed features for HTML 3.0. However, claiming HTML 3.0 "compliance" is a shameless marketing ploy. Yes, there is an HTML 3.0 draft, which is what vendors are usually referring to (at least right now) when they claim HTML 3.0 compliance. The sad paradox is that this unapproved draft looks like a standard only because the people whose job it is to ratify the standard couldn't agree on it and gave up.

This book is indeed a definitive guide for HTML, not any specific version of it. We give details for all the elements of the HTML 2.0 standard as well as the variety of interesting and useful extensions to the language that the popular browser manufacturers have chosen to include in their products, such as:

- Tables
- Frames
- Inline multimedia
- Dynamic documents
- Java
- Text font size and face controls

We even tell you, when necessary, which version of a browser implements a particular extension, and which doesn't, including the latest versions of the market-dominating Netscape Navigator, Internet Explorer, Mosaic, and even Lynx, a popular text-only browser for UNIX systems. In case you're worried, everything important that's in HTML 3.0 has found a home somewhere. There has even been talk about an HTML 3.5, but that's an even more shameless ploy by vendors who want to tack a bigger number on their product.

And, too, there are a few things that are closely related, but not directly part of HTML. We touch, but do not handle, for example, 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, and so we don't delve into them. Besides, they are comprehensive topics that deserve their own books, such as *Java in a Nutshell* and *CGI Programming on the World Wide Web* from O'Reilly & Associates, for instance.

In short, this book is your definitive guide to HTML as it is and should be used, including every extension we could find. Many 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.

Acknowledgments

We did not and certainly could not have composed this book without generous contributions from many people. Our wives Jeanne and Cindy (with whom we've just become reacquainted) and our young 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. (Ed Bond, we'll be over soon to repair your Windows.)

We also thank our technical reviewers, Kane Scarlett, Eric Raymond, and Chris Tacy, for carefully scrutinizing our work. We took most of your keen suggestions. And 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, finally, we thank the many people at O'Reilly & Associates who poked our words into sensibility and put them onto these pages. These folks include: Mary Anne Weeks Mayo, project manager/copyeditor for the book; Len Muellner, Ellen Siever, and Erik Ray converted the book from Word for Windows to SGML and contributed their tool-tweaking prowess; Chris Reilley created the excellent figures; Edie Freedman designed the cover; Nancy Priest designed the interior layout; Hanna Dyer designed the back cover; Seth Maislin prepared the index; and Sheryl Avruch, Clairemarie Fisher O'Leary, and Kismet McDonough Chan did the final quality control on the book.

Table of Contents

Figures

In This Chapter:

- *The Internet*
- *Talking the Internet Talk*
- *HTML: What It Is*
- *HTML: What It Isn't*
- *Nonstandard Extensions*
- *Tools for the HTML Designer*

1

HTML 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—aka 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 nonprofit organizations, and from born-again 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 documents and 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

Although popular media accounts often are confused and confusing, the concept of the *Internet* really is rather simple. It's a collection of networks—a network of networks—computers sharing digital information via a common set of networking and software protocols. Nearly anyone can connect their computer to the Internet and immediately communicate with other computers and users on the Net.