

# XML<sup>™</sup> Bible



# XML<sup>TM</sup> Bible

**Elliote Rusty Harold**



**IDG Books Worldwide, Inc.**  
**An International Data Group Company**

**Foster City, CA ♦ Chicago, IL ♦ Indianapolis, IN ♦ New York, NY**

## XML™ Bible

Published by

**IDG Books Worldwide, Inc.**

An International Data Group Company

919 E. Hillsdale Blvd., Suite 400

Foster City, CA 94404

www.idgbooks.com (IDG Books Worldwide Web site)

Copyright © 1999 IDG Books Worldwide, Inc. All rights reserved. No part of this book, including interior design, cover design, and icons, may be reproduced or transmitted in any form, by any means (electronic, photocopying, recording, or otherwise) without the prior written permission of the publisher.

ISBN: 0-7645-3236-7

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

10/QV/QY/ZZ/FC

Distributed in the United States by IDG Books Worldwide, Inc.

Distributed by CDG Books Canada Inc. for Canada; by Transworld Publishers Limited in the United Kingdom; by IDG Norge Books for Norway; by IDG Sweden Books for Sweden; by IDG Books Australia Publishing Corporation Pty. Ltd. for Australia and New Zealand; by TransQuest Publishers Pte Ltd. for Singapore, Malaysia, Thailand, Indonesia, and Hong Kong; by Gotop Information Inc. for Taiwan; by ICG Muse, Inc. for Japan; by Norma Comunicaciones S.A. for Colombia; by Intersoft for South Africa; by Eyrolles for France; by International Thomson Publishing for Germany, Austria and Switzerland; by Distribuidora Cuspide for Argentina; by Livraria Cultura for Brazil; by Ediciones ZETA S.C.R. Ltda. for Peru; by WS Computer Publishing Corporation, Inc., for the Philippines; by Contemporanea de Ediciones for Venezuela; by Express Computer Distributors for the Caribbean and West Indies; by Micronesia Media Distributor, Inc. for Micronesia; by Grupo Editorial Norma S.A. for Guatemala; by Chips Computadoras S.A. de C.V. for Mexico; by Editorial Norma de Panama S.A. for Panama; by American Bookshops for Finland. Authorized Sales Agent: Anthony Rudkin Associates for the Middle East and North Africa.

For general information on IDG Books Worldwide's books in the U.S., please call our Consumer Customer Service department at 800-762-2974. For reseller information, including discounts and premium sales, please call our Reseller Customer Service department at 800-434-3422.

For information on where to purchase IDG Books Worldwide's books outside the U.S., please contact our International Sales department at 317-596-5530 or fax 317-596-5692.

For consumer information on foreign language translations, please contact our Customer Service department at 800-434-3422, fax 317-596-5692, or e-mail [rights@idgbooks.com](mailto:rights@idgbooks.com).

For information on licensing foreign or domestic rights, please phone +1-650-655-3109.

For sales inquiries and special prices for bulk quantities, please contact our Sales department at 650-655-3200 or write to the address above.

For information on using IDG Books Worldwide's books in the classroom or for ordering examination copies, please contact our Educational Sales department at 800-434-2086 or fax 317-596-5499.

For press review copies, author interviews, or other publicity information, please contact our Public Relations department at 650-655-3000 or fax 650-655-3299.

For authorization to photocopy items for corporate, personal, or educational use, please contact Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, or fax 978-750-4470.

Library of Congress Cataloging-in-Publication Data  
Harold, Eliote Rusty.

XML bible / Eliote Rusty Harold.

p. cm.

ISBN 0-7645-3236-7 (alk. paper)

1. XML (Document markup language) I. Title.

QA76.76.H94H34 1999

99-31021

005.7'2--dc21

CIP

**LIMIT OF LIABILITY/DISCLAIMER OF WARRANTY: THE PUBLISHER AND AUTHOR HAVE USED THEIR BEST EFFORTS IN PREPARING THIS BOOK. THE PUBLISHER AND AUTHOR MAKE NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE ACCURACY OR COMPLETENESS OF THE CONTENTS OF THIS BOOK AND SPECIFICALLY DISCLAIM ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. THERE ARE NO WARRANTIES WHICH EXTEND BEYOND THE DESCRIPTIONS CONTAINED IN THIS PARAGRAPH. NO WARRANTY MAY BE CREATED OR EXTENDED BY SALES REPRESENTATIVES OR WRITTEN SALES MATERIALS. THE ACCURACY AND COMPLETENESS OF THE INFORMATION PROVIDED HEREIN AND THE OPINIONS STATED HEREIN ARE NOT GUARANTEED OR WARRANTED TO PRODUCE ANY PARTICULAR RESULTS, AND THE ADVICE AND STRATEGIES CONTAINED HEREIN MAY NOT BE SUITABLE FOR EVERY INDIVIDUAL. NEITHER THE PUBLISHER NOR AUTHOR SHALL BE LIABLE FOR ANY LOSS OF PROFIT OR ANY OTHER COMMERCIAL DAMAGES, INCLUDING BUT NOT LIMITED TO SPECIAL, INCIDENTAL, CONSEQUENTIAL, OR OTHER DAMAGES.**

**Trademarks:** All brand names and product names used in this book are trade names, service marks, trademarks, or registered trademarks of their respective owners. IDG Books Worldwide is not associated with any product or vendor mentioned in this book.



is a registered trademark or trademark under exclusive license to IDG Books Worldwide, Inc. from International Data Group, Inc. in the United States and/or other countries.

# ABOUT IDG BOOKS WORLDWIDE

Welcome to the world of IDG Books Worldwide.

IDG Books Worldwide, Inc., is a subsidiary of International Data Group, the world's largest publisher of computer-related information and the leading global provider of information services on information technology. IDG was founded more than 30 years ago by Patrick J. McGovern and now employs more than 9,000 people worldwide. IDG publishes more than 290 computer publications in over 75 countries. More than 90 million people read one or more IDG publications each month.

Launched in 1990, IDG Books Worldwide is today the #1 publisher of best-selling computer books in the United States. We are proud to have received eight awards from the Computer Press Association in recognition of editorial excellence and three from Computer Currents' First Annual Readers' Choice Awards. Our best-selling *...For Dummies*® series has more than 50 million copies in print with translations in 31 languages. IDG Books Worldwide, through a joint venture with IDG's Hi-Tech Beijing, became the first U.S. publisher to publish a computer book in the People's Republic of China. In record time, IDG Books Worldwide has become the first choice for millions of readers around the world who want to learn how to better manage their businesses.

Our mission is simple: Every one of our books is designed to bring extra value and skill-building instructions to the reader. Our books are written by experts who understand and care about our readers. The knowledge base of our editorial staff comes from years of experience in publishing, education, and journalism — experience we use to produce books to carry us into the new millennium. In short, we care about books, so we attract the best people. We devote special attention to details such as audience, interior design, use of icons, and illustrations. And because we use an efficient process of authoring, editing, and desktop publishing our books electronically, we can spend more time ensuring superior content and less time on the technicalities of making books.

You can count on our commitment to deliver high-quality books at competitive prices on topics you want to read about. At IDG Books Worldwide, we continue in the IDG tradition of delivering quality for more than 30 years. You'll find no better book on a subject than one from IDG Books Worldwide.



*John J. Kilcullen*

John Kilcullen  
Chairman and CEO  
IDG Books Worldwide, Inc.

*Steven Berkowitz*

Steven Berkowitz  
President and Publisher  
IDG Books Worldwide, Inc.



*Eighth Annual  
Computer Press  
Awards* 1992



*Ninth Annual  
Computer Press  
Awards* 1993



*Tenth Annual  
Computer Press  
Awards* 1994



*Eleventh Annual  
Computer Press  
Awards* 1995

IDG is the world's leading IT media, research and exposition company. Founded in 1964, IDG had 1997 revenues of \$2.05 billion and has more than 9,000 employees worldwide. IDG offers the widest range of media options that reach IT buyers in 75 countries representing 95% of worldwide IT spending. IDG's diverse product and services portfolio spans six key areas including print publishing, online publishing, expositions and conferences, market research, education and training, and global marketing services. More than 90 million people read one or more of IDG's 290 magazines and newspapers, including IDG's leading global brands — Computerworld, PC World, Network World, Macworld and the Channel World family of publications. IDG Books Worldwide is one of the fastest-growing computer book publishers in the world, with more than 700 titles in 36 languages. The *...For Dummies*® series alone has more than 50 million copies in print. IDG offers online users the largest network of technology-specific Web sites around the world through IDG.net (<http://www.idg.net>), which comprises more than 225 targeted Web sites in 55 countries worldwide. International Data Corporation (IDC) is the world's largest provider of information technology data, analysis and consulting, with research centers in over 41 countries and more than 400 research analysts worldwide. IDG World Expo is a leading producer of more than 168 globally branded conferences and expositions in 35 countries including E3 (Electronic Entertainment Expo), Macworld Expo, ComNet, Windows World Expo, ICE (Internet Commerce Expo), Agenda, DEMO, and Spotlight. IDG's training subsidiary, ExecuTrain, is the world's largest computer training company, with more than 230 locations worldwide and 785 training courses. IDG Marketing Services helps industry-leading IT companies build international brand recognition by developing global integrated marketing programs via IDG's print, online and exposition products worldwide. Further information about the company can be found at [www.idg.com](http://www.idg.com).

# Credits

**Acquisitions Editor**

John Osborn

**Development Editor**

Terri Varveris

**Contributing Writer**

Heather Williamson

**Technical Editor**

Greg Guntle

**Copy Editors**

Amy Eoff

Amanda Kaufman

Nicole LeClerc

Victoria Lee

**Production**

IDG Books Worldwide Production

**Proofreading and Indexing**

York Production Services

# About the Author

Elliotte Rusty Harold is an internationally respected writer, programmer, and educator both on the Internet and off. He got his start by writing FAQ lists for the Macintosh newsgroups on Usenet, and has since branched out into books, Web sites, and newsletters. He lectures about Java and object-oriented programming at Polytechnic University in Brooklyn. His Cafe con Leche Web site at <http://metalab.unc.edu/xml/> has become one of the most popular independent XML sites on the Internet.

Elliotte is originally from New Orleans where he returns periodically in search of a decent bowl of gumbo. However, he currently resides in the Prospect Heights neighborhood of Brooklyn with his wife Beth and cats Charm (named after the quark) and Marjorie (named after his mother-in-law). When not writing books, he enjoys working on genealogy, mathematics, and quantum mechanics. His previous books include *The Java Developer's Resource*, *Java Network Programming*, *Java Secrets*, *JavaBeans*, *XML: Extensible Markup Language*, and *Java I/O*.

*For Ma, a great grandmother*



# Preface

---

Welcome to the *XML Bible*. After reading this book I hope you'll agree with me that XML is the most exciting development on the Internet since Java, and that it makes Web site development easier, more productive, and more fun.

This book is your introduction to the exciting and fast growing world of XML. In this book, you'll learn how to write documents in XML and how to use style sheets to convert those documents into HTML so legacy browsers can read them. You'll also learn how to use document type definitions (DTDs) to describe and validate documents. This will become increasingly important as more and more browsers like Mozilla and Internet Explorer 5.0 provide native support for XML.

## About You the Reader

Unlike most other XML books on the market, the *XML Bible* covers XML not from the perspective of a software developer, but rather that of a Web-page author. I don't spend a lot of time discussing BNF grammars or parsing element trees. Instead, I show you how you can use XML and existing tools today to more efficiently produce attractive, exciting, easy-to-use, easy-to-maintain Web sites that keep your readers coming back for more.

This book is aimed directly at Web-site developers. I assume you want to use XML to produce Web sites that are difficult to impossible to create with raw HTML. You'll be amazed to discover that in conjunction with style sheets and a few free tools, XML enables you to do things that previously required either custom software costing hundreds to thousands of dollars per developer, or extensive knowledge of programming languages like Perl. None of the software in this book will cost you more than a few minutes of download time. None of the tricks require any programming.

## What You Need to Know

XML does build on HTML and the underlying infrastructure of the Internet. To that end, I will assume you know how to use ftp files, send email, and load URLs in your Web browser of choice. I will also assume you have a reasonable knowledge of HTML at about the level supported by Netscape 1.1. On the other hand, when I discuss newer aspects of HTML that are not yet in widespread use like cascading style sheets, I will cover them in depth.

To be more specific, in this book I assume that you can:

- ◆ Write a basic HTML page including links, images, and text using a text editor.
- ◆ Place that page on a Web server.

On the other hand, I do not assume that you:

- ◆ Know SGML. In fact, this preface is almost the only place in the entire book you'll see the word SGML used. XML is supposed to be simpler and more widespread than SGML. It can't be that if you have to learn SGML first.
- ◆ Are a programmer, whether of Java, Perl, C, or some other language, XML is a markup language, not a programming language. You don't need to be a programmer to write XML documents.

## What You'll Learn

This book has one primary goal; to teach you to write XML documents for the Web. Fortunately, XML has a decidedly flat learning curve, much like HTML (and unlike SGML). As you learn a little you can do a little. As you learn a little more, you can do a little more. Thus the chapters in this book build steadily on each other. They are meant to be read in sequence. Along the way you'll learn:

- ◆ How an XML document is created and delivered to readers.
- ◆ How semantic tagging makes XML documents easier to maintain and develop than their HTML equivalents.
- ◆ How to post XML documents on Web servers in a form everyone can read.
- ◆ How to make sure your XML is well-formed.
- ◆ How to use international characters like `&` and `&` in your documents.
- ◆ How to validate documents with DTDs.
- ◆ How to use entities to build large documents from smaller parts.
- ◆ How attributes describe data.
- ◆ How to work with non-XML data.
- ◆ How to format your documents with CSS and XSL style sheets.
- ◆ How to connect documents with XLinks and Xpointers.
- ◆ How to merge different XML vocabularies with namespaces.
- ◆ How to write metadata for Web pages using RDF.

In the final section of this book, you'll see several practical examples of XML being used for real-world applications including:

- ◆ Web Site Design
- ◆ Push
- ◆ Vector Graphics
- ◆ Genealogy

## How the Book Is Organized

This book is divided into five parts and includes three appendixes:

- I. Introducing XML
- II. Document Type Definitions
- III. Style Languages
- IV. Supplemental Technologies
- V. XML Applications

By the time you're finished reading this book, you'll be ready to use XML to create compelling Web pages. The five parts and the appendixes are described below.

### Part I: Introducing XML

Part I consists of Chapters 1 through 7. It begins with the history and theory behind XML, the goals XML is trying to achieve, and shows you how the different pieces of the XML equation fit together to create and deliver documents to readers. You'll see several compelling examples of XML applications to give you some idea of the wide applicability of XML, including the Vector Markup Language (VML), the Resource Description Framework (RDF), the Mathematical Markup Language (MathML), the Extensible Forms Description Language (XFDL), and many others. Then you'll learn by example how to write XML documents with tags you define that make sense for your document. You'll see how to edit them in a text editor, attach style sheets to them, and load them into a Web browser like Internet Explorer 5.0 or Mozilla. You'll even learn how you can write XML documents in languages other than English, even languages that aren't written remotely like English, such as Chinese, Hebrew, and Russian.

## Part II: Document Type Definitions

Part II consists of Chapters 8 through 11, all of which focus on document type definitions (DTDs). An XML document may optionally contain a DTD that specifies which elements are and are not allowed in an XML document. The DTD specifies the exact context and structure of those elements. A validating parser can read a document and compare it to its DTD, and report any mistakes it finds. This enables document authors to make sure that their work meets any necessary criteria.

In Part II, you'll learn how to attach a DTD to a document, how to validate your documents against their DTDs, and how to write your own DTDs that solve your own problems. You'll learn the syntax for declaring elements, attributes, entities, and notations. You'll see how you can use entity declarations and entity references to build both a document and its DTD from multiple, independent pieces. This allows you to make long, hard-to-follow documents much simpler by separating them into related modules and components. And you'll learn how to integrate other forms of data like raw text and GIF image files in your XML document.

## Part III: Style Languages

Part III consists of Chapters 12 through 15. XML markup only specifies what's in a document. Unlike HTML, it does not say anything about what that content should look like. Information about an XML document's appearance when printed, viewed in a Web browser, or otherwise displayed is stored in a style sheet. Different style sheets can be used for the same document. You might, for instance, want to use a style sheet that specifies small fonts for printing, another one that uses larger fonts for on-screen use, and a third with absolutely humongous fonts to project the document on a wall at a seminar. You can change the appearance of an XML document by choosing a different style sheet without touching the document itself.

Part III describes in detail the two style sheet languages in broadest use on the Web, Cascading Style Sheets (CSS) and the Extensible Style Language (XSL).

CSS is a simple style-sheet language originally designed for use with HTML. CSS exists in two versions: CSS Level 1 and CSS Level 2. CSS Level 1 provides basic information about fonts, color, positioning, and text properties, and is reasonably well supported by current Web browsers for HTML and XML. CSS Level 2 is a more recent standard that adds support for aural style sheets, user interface styles, international and bi-directional text, and more. CSS is a relatively simple standard that applies fixed style rules to the contents of particular elements.

XSL, by contrast, is a more complicated and more powerful style language that cannot only apply styles to the contents of elements but can also rearrange elements, add boilerplate text, and transform documents in almost arbitrary ways. XSL is divided into two parts: a transformation language for converting XML trees to alternative trees, and a formatting language for specifying the appearance of the elements of an XML tree. Currently, the transformation language is better supported by most tools

than the formatting language. Nonetheless, it is beginning to firm up, and is supported by Microsoft Internet Explorer 5.0 and some third-party formatting engines.

## Part IV: Supplemental Technologies

Part IV consists of Chapters 16 through 19. It introduces some XML-based languages and syntaxes that layer on top of basic XML. XLinks provides multi-directional hypertext links that are far more powerful than the simple HTML `<A>` tag. XPointers introduce a new syntax you can attach to the end of URLs to link not only to particular documents, but to particular parts of particular documents. Namespaces use prefixes and URLs to disambiguate conflicting XML markup languages. The Resource Description Framework (RDF) is an XML application used to embed meta-data in XML and HTML documents. Meta-data is information about a document, such as the author, date, and title of a work, rather than the work itself. All of these can be added to your own XML-based markup languages to extend their power and utility.

## Part V: XML Applications

Part V, which consists of Chapters 20–23, shows you four practical uses of XML in different domains. XHTML is a reformulation of HTML 4.0 as valid XML. Microsoft's Channel Definition Format (CDF), is an XML-based markup language for defining channels that can push updated Web site content to subscribers. The Vector Markup Language (VML) is an XML application for scalable graphics used by Microsoft Office 2000 and Internet Explorer 5.0. Finally, a completely new application is developed for genealogical data to show you not just how to use XML tags, but why and when to choose them.

## Appendixes

This book has two appendixes, which focus on the formal specifications for XML, as opposed to the more informal description of it used throughout the rest of the book. Appendix A provides detailed explanations of three individual parts of the XML 1.0 specification: XML BNF grammar, well-formedness constraints, and the validity constraints. Appendix B contains the official W3C XML 1.0 specification published by the W3C. The book also has a third appendix, Appendix C, which describes the contents of the CD-ROM that accompanies this book.

## What You Need

To make the best use of this book and XML, you need:

- ◆ A PC running Windows 95, Windows 98, or Windows NT
- ◆ Internet Explorer 5.0
- ◆ A Java 1.1 or later virtual machine

Any system that can run Windows will suffice. In this book, I mostly assume you're using Windows 95 or NT 4.0 or later. As a longtime Mac and Unix user, I somewhat regret this. Like Java, XML is supposed to be platform independent. Also like Java, the reality is somewhat short of the hype. Although XML code is pure text that can be written with any editor, many of the tools are currently only available on Windows.

However, although there aren't many Unix or Macintosh native XML programs, there are an increasing number of XML programs written in Java. If you have a Java 1.1 or later virtual machine on your platform of choice, you should be able to make do. Even if you can't load your XML documents directly into a Web browser, you can still convert them to XML documents and view those. When Mozilla is released, it should provide the best XML browser yet across multiple platforms.

## How to Use This Book

This book is designed to be read more or less cover to cover. Each chapter builds on the material in the previous chapters in a fairly predictable fashion. Of course, you're always welcome to skim over material that's already familiar to you. I also hope you'll stop along the way to try out some of the examples and to write some XML documents of your own. It's important to learn not just by reading, but also by doing. Before you get started, I'd like to make a couple of notes about grammatical conventions used in this book.

Unlike HTML, XML is case sensitive. `<FATHER>` is not the same as `<Father>` or `<father>`. The `father` element is not the same as the `Father` element or the `FATHER` element. Unfortunately, case-sensitive markup languages have an annoying habit of conflicting with standard English usage. On rare occasion this means that you may encounter sentences that don't begin with a capital letter. More commonly, you'll see capitalization used in the middle of a sentence where you wouldn't normally expect it. Please don't get too bothered by this. All XML and HTML code used in this book is placed in a monospaced font, so most of the time it will be obvious from the context what is meant.

I have also adopted the British convention of only placing punctuation inside quote marks when it belongs with the material quoted. Frankly, although I learned to write in the American educational system, I find the British system is far more logical, especially when dealing with source code where the difference between a comma or a period and no punctuation at all can make the difference between perfectly correct and perfectly incorrect code.

## What the Icons Mean

Throughout the book, I've used *icons* in the left margin to call your attention to points that are particularly important.



Note

Note icons provide supplemental information about the subject at hand, but generally something that isn't quite the main idea. Notes are often used to elaborate on a detailed technical point.



Tip

Tip icons indicate a more efficient way of doing something, or a technique that may not be obvious.



On the  
CD-ROM

CD-ROM icons tell you that software discussed in the book is available on the companion CD-ROM. This icon also tells you if a longer example, discussed but not included in its entirety in the book, is on the CD-ROM.



Caution

Caution icons warn you of a common misconception or that a procedure doesn't always work quite like it's supposed to. The most common purpose of a Caution icon in this book is to point out the difference between what a specification says should happen, and what actually does.



Cross-  
Reference

The Cross Reference icon refers you to other chapters that have more to say about a particular subject.

## About the Companion CD-ROM

The inside back cover of this book contains a CD-ROM that holds all numbered code listings that you'll find in the text. It also includes many longer examples that couldn't fit into this book. The CD-ROM also contains the complete text of various XML specifications in HTML. (Some of the specifications will be in other formats as well.) Finally, you will find an assortment of useful software for working with XML documents. Many (though not all) of these programs are written in Java, so they'll run on any system with a reasonably compatible Java 1.1 or later virtual machine. Most of the programs that aren't written in Java are designed for Windows 95, 98, and NT.

For a complete description of the CD-ROM contents, you can read Appendix C. In addition, to get a complete description of what is on the CD-ROM, you can load the file `index.html` onto your Web browser. The files on the companion CD-ROM are not compressed, so you can access them directly from the CD.

## Reach Out

The publisher and I want your feedback. After you have had a chance to use this book, please take a moment to complete the IDG Books Worldwide Registration Card (in the back of the book). Please be honest in your evaluation. If you thought a particular chapter didn't tell you enough, let me know. Of course, I would prefer to receive comments like: "This is the best book I've ever read", "Thanks to this book, my Web site won Cool Site of the Year", or "When I was reading this book on the beach, I was besieged by models who thought I was super cool", but I'll take any comments I can get :-).

Feel free to send me specific questions regarding the material in this book. I'll do my best to help you out and answer your questions, but I can't guarantee a reply. The best way to reach me is by email:

`elharo@metalab.unc.edu`

Also, I invite you to visit my Cafe con Leche Web site at <http://metalab.unc.edu/xml/>, which contains a lot of XML-related material and is updated almost daily. Despite my persistent efforts to make this book perfect, some errors have doubtless slipped by. Even more certainly, some of the material discussed here will change over time. I'll post any necessary updates and errata on my Web site at <http://metalab.unc.edu/xml/books/bible/>. Please let me know via email of any errors that you find that aren't already listed.

**Elliotte Rusty Harold**

`elharo@metalab.unc.edu`

<http://metalab.unc.edu/xml/>

**New York City, June 1999**

# Acknowledgments

---

The folks at IDG have all been great. The acquisitions editor, John Osborn, deserves special thanks for arranging the unusual scheduling this book required to hit the moving target XML presents. Terri Varveris shepherded this book through the development process. With poise and grace, she managed the constantly shifting outline and schedule that a book based on unstable specifications and software requires. Amy Eoff corrected many of my grammatical shortcomings. Susan Parini and Ritchie Durdin, the production coordinators, also deserve special thanks for managing the production of this book and for dealing with last-minute figure changes.

Steven Champeon brought his SGML experience to the book, and provided many insightful comments on the text. My brother Thomas Harold put his command of chemistry at my disposal when I was trying to grasp the Chemical Markup Language. Carroll Bellau provided me with parts of my family tree, which you'll find in Chapter 17.

I also greatly appreciate all the comments, questions, and corrections sent in by readers of my previous book, *XML: Extensible Markup Language*. I hope that I've managed to address most of those comments in this book. They've definitely helped make *XML Bible* a better book. Particular thanks are due to Alan Esenther and Donald Lancon Jr. for their especially detailed comments.

WandaJane Phillips wrote the original version of Chapter 21 on CDF that is adapted here. Heather Williamson, in addition to performing yeoman-like service as technical editor, wrote Chapter 13, *CSS Level 2*, and parts of Chapters 18, 19, and 22. Her help was instrumental in helping me almost meet my deadline. (Blame for this *almost* rests on my shoulders, not theirs.) Also, I would like to thank Piroz Mohseni, who also served as a technical editor for this book.

The agenting talents of David and Sherry Rogelberg of the Studio B Literary Agency (<http://www.studiob.com/>) have made it possible for me to write more or less full-time. I recommend them highly to anyone thinking about writing computer books. And as always, thanks go to my wife Beth for her endless love and understanding.



# Contents at a Glance



Preface .....	ix
Acknowledgments .....	xvii
<b>Part I: Introducing XML .....</b>	<b>1</b>
Chapter 1: An Eagle's Eye View of XML .....	3
Chapter 2: An Introduction to XML Applications .....	17
Chapter 3: Your First XML Document .....	49
Chapter 4: Structuring Data .....	59
Chapter 5: Attributes, Empty Tags, and XSL .....	95
Chapter 6: Well-Formed XML Documents .....	
Chapter 7: Foreign Languages and Non-Roman Text .....	161
<b>Part II: Document Type Definitions .....</b>	<b>189</b>
Chapter 8: Document Type Definitions and Validity .....	191
Chapter 9: Entities and External DTD Subsets .....	247
Chapter 10: Attribute Declarations in DTDs .....	283
Chapter 11: Embedding Non-XML Data .....	307
<b>Part III: Style Languages .....</b>	<b>321</b>
Chapter 12: Cascading Style Sheets Level 1 .....	323
Chapter 13: Cascading Style Sheets Level 2 .....	389
Chapter 14: XSL Transformations .....	433
Chapter 15: XSL Formatting Objects .....	513
<b>Part IV: Supplemental Technologies .....</b>	<b>569</b>
Chapter 16: XLinks .....	571
Chapter 17: XPointers .....	591
Chapter 18: Namespaces .....	617
Chapter 19: The Resource Description Framework .....	631
<b>PartV: XML Applications .....</b>	<b>655</b>
Chapter 20: Reading Document Type Definitions .....	657
Chapter 21: Pushing Web Sites with CDF .....	775
Chapter 22: The Vector Markup Language .....	805
Chapter 23: Designing a New XML Application .....	833

Appendix A: XML Reference Material .....	863
Appendix B: The XML 1.0 Specification .....	921
Appendix C: What's on the CD-ROM .....	971
Index .....	975
End-User License Agreement .....	1018
CD-ROM Installation Instructions .....	1022

# Contents

Preface .....	ix
Acknowledgments .....	xvii

## Part I: Introducing XML 1

<b>Chapter 1: An Eagle's Eye View of XML .....</b>	<b>3</b>
What Is XML? .....	3
XML Is a Meta-Markup Language .....	3
XML Describes Structure and Semantics, Not Formatting .....	4
Why Are Developers Excited about XML? .....	6
Design of Domain-Specific Markup Languages .....	6
Self-Describing Data .....	6
Interchange of Data Among Applications .....	7
Structured and Integrated Data .....	8
The Life of an XML Document .....	8
Editors .....	9
Parsers and Processors .....	9
Browsers and Other Tools .....	9
The Process Summarized .....	10
Related Technologies .....	10
Hypertext Markup Language .....	10
Cascading Style Sheets .....	11
Extensible Style Language .....	12
URLs and URIs .....	12
XLinks and XPointers .....	13
The Unicode Character Set .....	14
How the Technologies Fit Together .....	14
<b>Chapter 2: An Introduction to XML Applications .....</b>	<b>17</b>
What Is an XML Application? .....	17
Chemical Markup Language .....	18
Mathematical Markup Language .....	19
Channel Definition Format .....	22
Classic Literature .....	22
Synchronized Multimedia Integration Language .....	24
HTML+TIME .....	25
Open Software Description .....	26
Scalable Vector Graphics .....	27
Vector Markup Language .....	29
MusicML .....	30
VoxML .....	32

Open Financial Exchange .....	34
Extensible Forms Description Language .....	36
Human Resources Markup Language .....	38
Resource Description Framework .....	40
XML for XML .....	42
XSL .....	42
XLL .....	43
DCD .....	43
Behind-the-Scene Uses of XML .....	44
<b>Chapter 3: Your First XML Document .....</b>	<b>49</b>
Hello XML .....	49
Creating a Simple XML Document .....	50
Saving the XML File .....	50
Loading the XML File into a Web Browser .....	51
Exploring the Simple XML Document .....	52
Assigning Meaning to XML Tags .....	54
Writing a Style Sheet for an XML Document .....	55
Attaching a Style Sheet to an XML Document .....	56
<b>Chapter 4: Structuring Data .....</b>	<b>59</b>
Examining the Data .....	59
Batters .....	60
Pitchers .....	62
Organization of the XML Data .....	62
XMLizing the Data .....	65
Starting the Document: XML Declaration and Root Element .....	65
XMLizing League, Division, and Team Data .....	67
XMLizing Player Data .....	69
XMLizing Player Statistics .....	70
Putting the XML Document Back Together Again .....	72
The Advantages of the XML Format .....	80
Preparing a Style Sheet for Document Display .....	81
Linking to a Style Sheet .....	82
Assigning Style Rules to the Root Element .....	84
Assigning Style Rules to Titles .....	85
Assigning Style Rules to Player and Statistics Elements .....	88
Summing Up .....	89
<b>Chapter 5: Attributes, Empty Tags, and XSL .....</b>	<b>95</b>
Attributes .....	95
Attributes versus Elements .....	101
Structured Meta-data .....	102
Meta-Meta-Data .....	105
What's Your Meta-data Is Someone Else's Data .....	106
Elements Are More Extensible .....	106
Good Times to Use Attributes .....	107

Empty Tags .....	108
XSL .....	109
XSL Style Sheet Templates .....	110
The Body of the Document .....	111
The Title .....	113
Leagues, Divisions, and Teams .....	115
Players .....	120
Separation of Pitchers and Batters .....	122
CSS or XSL? .....	130
<b>Chapter 6: Well-Formed XML Documents .....</b>	<b>133</b>
#1: The XML declaration must begin the document .....	144
#2: Use Both Start and End Tags in Non-Empty Tags .....	144
<b>Chapter 7: Foreign Languages and Non-Roman Text .....</b>	<b>161</b>
Non-Roman Scripts on the Web .....	161
Scripts, Character Sets, Fonts, and Glyphs .....	166
A Character Set for the Script .....	166
A Font for the Character Set .....	167
An Input Method for the Character Set .....	167
Operating System and Application Software .....	168
Legacy Character Sets .....	169
The ASCII Character Set .....	169
The ISO Character Sets .....	172
The MacRoman Character Set .....	175
The Windows ANSI Character Set .....	176
The Unicode Character Set .....	177
UTF 8 .....	182
The Universal Character System .....	182
How to Write XML in Unicode .....	183
Inserting Characters in XML Files with Character References .....	183
Converting to and from Unicode .....	184
How to Write XML in Other Character Sets .....	185

## Part II: Document Type Definitions

189

<b>Chapter 8: Document Type Definitions and Validity .....</b>	<b>191</b>
Document Type Definitions .....	191
Document Type Declarations .....	192
Validating Against a DTD .....	195
Listing the Elements .....	201
Element Declarations .....	208
ANY .....	209
#PCDATA .....	209
Child Lists .....	212
Sequences .....	214
One or More Children .....	215

Zero or More Children .....	215
Zero or One Child .....	216
The Complete Document and DTD .....	217
Choices .....	223
Children with Parentheses .....	224
Mixed Content .....	227
Empty Elements .....	228
Comments in DTDs .....	229
Sharing Common DTDs Among Documents .....	234
DTDs at Remote URLs .....	241
Public DTDs .....	241
Internal and External DTD Subsets .....	243
<b>Chapter 9: Entities and External DTD Subsets .....</b>	<b>247</b>
What Is an Entity? .....	247
Internal General Entities .....	249
Defining an Internal General Entity Reference .....	249
Using General Entity References in the DTD .....	251
Predefined General Entity References .....	252
External General Entities .....	253
Internal Parameter Entities .....	256
External Parameter Entities .....	258
Building a Document from Pieces .....	264
Entities and DTDs in Well-Formed Documents .....	274
Internal Entities .....	274
External Entities .....	276
<b>Chapter 10: Attribute Declarations in DTDs .....</b>	<b>283</b>
What Is an Attribute? .....	283
Declaring Attributes in DTDs .....	284
Declaring Multiple Attributes .....	285
Specifying Default Values for Attributes .....	286
#REQUIRED .....	286
#IMPLIED .....	287
#FIXED .....	288
Attribute Types .....	288
The CDATA Attribute Type .....	289
The Enumerated Attribute Type .....	289
The NMTOKEN Attribute Type .....	290
The NMTOKENS Attribute Type .....	291
The ID Attribute Type .....	292
The IDREF Attribute Type .....	292
The ENTITY Attribute Type .....	293
The ENTITIES Attribute Type .....	294
The NOTATION Attribute Type .....	294
Predefined Attributes .....	295
xml:space .....	295
xml:lang .....	297

A DTD for Attribute-Based Baseball Statistics .....	300
Declaring SEASON Attributes in the DTD .....	301
Declaring LEAGUE and DIVISION Attributes in the DTD .....	301
Declaring TEAM Attributes in the DTD .....	302
Declaring PLAYER Attributes in the DTD .....	302
The Complete DTD for the Baseball Statistics Example .....	304
<b>Chapter 11: Embedding Non-XML Data .....</b>	<b>307</b>
Notations .....	307
Unparsed External Entities .....	311
Declaring Unparsed Entities .....	311
Embedding Unparsed Entities .....	312
Embedding Multiple Unparsed Entities .....	315
Processing Instructions .....	315
Conditional Sections in DTDs .....	319

## Part III: Style Languages

321

<b>Chapter 12: Cascading Style Sheets Level 1 .....</b>	<b>323</b>
What Is CSS? .....	323
Attaching Style Sheets to Documents .....	324
Selection of Elements .....	327
Grouping Selectors .....	328
Pseudo-Elements .....	328
Pseudo-Classes .....	330
Selection by ID .....	332
Contextual Selectors .....	332
STYLE Attributes .....	333
Inheritance .....	334
Cascades .....	335
The @import Directive .....	336
The !important Declaration .....	336
Cascade Order .....	337
Comments in CSS Style Sheets .....	337
CSS Units .....	338
Length values .....	339
URL Values .....	341
Color Values .....	342
Keyword Values .....	343
Block, Inline, and List Item Elements .....	344
List Items .....	347
The whitespace Property .....	350
Font Properties .....	352
The font-family Property .....	352
The font-style Property .....	354
The font-variant Property .....	355
The font-weight Property .....	356

The font-size Property .....	356
The font Shorthand Property .....	359
The Color Property .....	360
Background Properties .....	361
The background-color Property .....	361
The background-image Property .....	362
The background-repeat Property .....	363
The background-attachment Property .....	364
The background-position Property .....	365
The Background Shorthand Property .....	369
Text Properties .....	369
The word-spacing Property .....	370
The letter-spacing Property .....	371
The text-decoration Property .....	371
The vertical-align Property .....	372
The text-transform Property .....	373
The text-align Property .....	374
The text-indent Property .....	375
The line-height Property .....	375
Box Properties .....	377
Margin Properties .....	378
Border Properties .....	379
Padding Properties .....	382
Size Properties .....	383
Positioning Properties .....	384
The float Property .....	385
The clear Property .....	386
<b>Chapter 13: Cascading Style Sheets Level 2 .....</b>	<b>389</b>
What's New in CSS2? .....	389
New Pseudo-classes .....	390
New Pseudo-Elements .....	391
Media Types .....	391
Paged Media .....	391
Internationalization .....	391
Visual Formatting Control .....	391
Tables .....	391
Generated Content .....	392
Aural Style Sheets .....	392
New Implementations .....	392
Selecting Elements .....	393
Pattern Matching .....	393
The Universal Selector .....	394
Descendant and Child Selectors .....	395
Adjacent Sibling Selectors .....	396
Attribute Selectors .....	396
@rules .....	397
Pseudo Elements .....	402

Pseudo Classes .....	403
Formatting a Page .....	405
Size Property .....	405
Margin Property .....	405
Mark Property .....	405
Page Property .....	406
Page-Break Properties .....	407
Visual Formatting .....	407
Display Property .....	407
Width and Height Properties .....	410
Overflow Property .....	411
Clip Property .....	411
Visibility Property .....	412
Cursor Property .....	412
Color-Related Properties .....	413
Font Properties .....	416
Text Shadow Property .....	419
Vertical Align Property .....	419
Boxes .....	420
Outline Properties .....	420
Positioning Properties .....	422
Counters and Automatic Numbering .....	424
Aural Style Sheets .....	425
Speak Property .....	426
Volume Property .....	426
Pause Properties .....	427
Cue Properties .....	427
Play-During Property .....	428
Spatial Properties .....	428
Voice Characteristics Properties .....	429
Speech Properties .....	431
<b>Chapter 14: XSL Transformations .....</b>	<b>433</b>
What Is XSL? .....	433
Overview of XSL Transformations .....	435
Trees .....	435
XSL Style Sheet Documents .....	437
Where Does the XML Transformation Happen? .....	439
How to Use XT .....	440
Direct Display of XML Files with XSL Style Sheets .....	442
XSL Templates .....	444
The <code>xsl:apply-templates</code> Element .....	445
The <code>select</code> Attribute .....	447
Computing the Value of a Node with <code>xsl:value-of</code> .....	448
Processing Multiple Elements with <code>xsl:for-each</code> .....	450
Patterns for Matching Nodes .....	451
Matching the Root Node .....	451
Matching Element Names .....	452

Matching Children with / .....	454
Matching Descendants with // .....	455
Matching by ID .....	456
Matching Attributes with @ .....	456
Matching Comments with comment() .....	458
Matching Processing Instructions with pi() .....	459
Matching Text Nodes with text() .....	460
Using the Or Operator   .....	460
Testing with [ ] .....	461
Expressions for Selecting Nodes .....	463
Node Axes .....	463
Expression Types .....	470
The Default Template Rules .....	480
The Default Rule for Elements .....	480
The Default Rule for Text Nodes .....	480
Implication of the Two Default Rules .....	481
Deciding What Output to Include .....	481
Using Attribute Value Templates .....	482
Inserting Elements into the Output with xsl:element .....	484
Inserting Attributes into the Output with xsl:attribute .....	484
Defining Attribute Sets .....	485
Generating Processing Instructions with xsl:pi .....	486
Generating Comments with xsl:comment .....	487
Generating Text with xsl:text .....	487
Copying the Current Node with xsl:copy .....	488
Counting Nodes with xsl:number .....	490
Default Numbers .....	491
Number to String Conversion .....	493
Sorting Output Elements .....	494
CDATA and < Signs .....	497
Modes .....	499
Defining Constants with xsl:variable .....	501
Named Templates .....	502
Parameters .....	503
Stripping and Preserving Whitespace .....	505
Making Choices .....	506
xsl:if .....	507
xsl:choose .....	507
Merging Multiple Style Sheets .....	508
Import with xsl:import .....	508
Inclusion with xsl:include .....	508
Embed Style Sheets in Documents with xsl:stylesheet .....	509
<b>Chapter 15: XSL Formatting Objects .....</b>	<b>513</b>
Overview of the XSL Formatting Language .....	513
Formatting Objects and Their Properties .....	514
The fo Namespace .....	517
Formatting Properties .....	518

Transforming to Formatting Objects .....	522
Using FOP .....	524
Page Layout .....	526
Master Pages .....	526
Page Sequences .....	529
Content .....	535
Block-level Formatting Objects .....	535
Inline Formatting Objects .....	537
Table-formatting Objects .....	538
Out-of-line Formatting Objects .....	538
Rules .....	539
Graphics .....	540
Links .....	540
Lists .....	542
Tables .....	543
Characters .....	546
Sequences .....	546
Footnotes .....	547
Floats .....	547
XSL Formatting Properties .....	548
Units and Data Types .....	549
Informational Properties .....	551
Paragraph Properties .....	551
Character Properties .....	554
Sentence Properties .....	556
Area Properties .....	559
Aural Properties .....	565

## Part IV: Supplemental Technologies

569

<b>Chapter 16: XLinks .....</b>	<b>571</b>
XLinks versus HTML Links .....	571
Simple Links .....	572
Descriptions of the Local Resource .....	574
Descriptions of the Remote Resource .....	575
Link Behavior .....	576
Extended Links .....	580
Out-of-Line Links .....	583
Extended Link Groups .....	584
An Example .....	585
The steps Attribute .....	587
Renaming XLink Attributes .....	588
<b>Chapter 17: XPointers .....</b>	<b>591</b>
Why Use XPointers? .....	591
XPointer Examples .....	592
Absolute Location Terms .....	594

id() .....	597
root() .....	598
html() .....	598
Relative Location Terms .....	598
child .....	600
descendant .....	601
ancestor .....	601
preceding .....	601
following .....	601
psibling .....	602
fsibling .....	602
Relative Location Term Arguments .....	602
Selection by Number .....	603
Selection by Node Type .....	606
Selection by Attribute .....	610
String Location Terms .....	611
The origin Absolute Location Term .....	612
Spanning a Range of Text .....	614
<b>Chapter 18: Namespaces .....</b>	<b>617</b>
What Is a Namespace? .....	617
Namespace Syntax .....	620
Definition of Namespaces .....	620
Multiple Namespaces .....	622
Attributes .....	624
Default Namespaces .....	625
Namespaces in DTDs .....	628
<b>Chapter 19: The Resource Description Framework .....</b>	<b>631</b>
What Is RDF? .....	631
RDF Statements .....	632
Basic RDF Syntax .....	634
The root Element .....	634
The Description Element .....	634
Namespaces .....	635
Multiple Properties and Statements .....	637
Resource Valued Properties .....	638
XML Valued Properties .....	641
Abbreviated RDF Syntax .....	642
Containers .....	643
The Bag container .....	643
The Seq Container .....	646
The Alt Container .....	646
Statements about Containers .....	647
Statements about Container Members .....	650
Statements about Implied Bags .....	652
RDF Schemas .....	652

**Part V: XML Applications****655**

<b>Chapter 20: Reading Document Type Definitions .....</b>	<b>657</b>
The Importance of Reading DTDs .....	658
What Is XHTML? .....	659
Why Validate HTML? .....	659
Modularization of XHTML Working Draft .....	660
The Structure of the XHTML DTDs .....	660
XHTML Strict DTD .....	662
XHTML Transitional DTD .....	669
The XHTML Frameset DTD .....	676
The XHTML Modules .....	679
The Common Names Module .....	680
The Character Entities Module .....	684
The Intrinsic Events Module .....	686
The Common Attributes Modules .....	689
The Document Model Module .....	695
The Inline Structural Module .....	704
Inline Presentational Module .....	706
Inline Phrasal Module .....	709
Block Structural Module .....	711
Block-Presentational Module .....	712
Block-Phrasal Module .....	714
The Scripting Module .....	716
The Stylesheets Module .....	718
The Image Module .....	719
The Frames Module .....	720
The Linking Module .....	723
The Client-side Image Map Module .....	725
The Object Element Module .....	726
The Java Applet Element Module .....	728
The Lists Module .....	730
The Forms Module .....	733
The Table Module .....	737
The Meta Module .....	742
The Structure Module .....	743
Non-Standard modules .....	746
The XHTML Entity Sets .....	746
The XHTML Latin-1 Entities .....	747
The XHTML Special Character Entities .....	752
The XHTML Symbol Entities .....	754
Simplified Subset DTDs .....	761
Techniques to Imitate .....	768
Comments .....	768
Parameter Entities .....	770

<b>Chapter 21: Pushing Web Sites with CDF</b> .....	<b>775</b>
What Is CDF? .....	775
How Channels Are Created .....	776
Determining Channel Content .....	776
Creating CDF Files and Documents .....	777
Description of the Channel .....	780
Title .....	780
Abstract .....	781
Logos .....	782
Information Update Schedules .....	783
Precaching and Web Crawling .....	787
Precaching .....	787
Web Crawling .....	788
Reader Access Log .....	789
The BASE Attribute .....	791
The LASTMOD Attribute .....	792
The USAGE Element .....	794
DesktopComponent Value .....	795
Email Value .....	796
NONE Value .....	797
ScreenSaver Value .....	798
SoftwareUpdate Value .....	800
<b>Chapter 22: The Vector Markup Language</b> .....	<b>805</b>
What Is VML? .....	805
Drawing with a Keyboard .....	808
The shape Element .....	808
The shapetype Element .....	811
The group Element .....	813
Positioning VML Shapes with Cascading Style Sheet Properties .....	814
The rotation Property .....	817
The flip Property .....	817
The center-x and center-y Properties .....	820
VML in Office 2000 .....	821
Settings .....	821
A Simple Graphics Demonstration of a House .....	822
A Quick Look at SVG .....	830
<b>Chapter 23: Designing a New XML Application</b> .....	<b>833</b>
Organization of the Data .....	833
Listing the Elements .....	834
Identifying the Fundamental Elements .....	835
Establishing Relationships Among the Elements .....	838
The Person DTD .....	840
The Family DTD .....	845
The Source DTD .....	847

The Family Tree DTD .....	848
Designing a Style Sheet for Family Trees .....	855
<b>Appendix A: XML Reference Material .....</b>	<b>863</b>
<b>Appendix B: The XML 1.0 Specification .....</b>	<b>921</b>
<b>Appendix C: What's on the CD-ROM .....</b>	<b>971</b>
<b>Index .....</b>	<b>975</b>
<b>End-User License Agreement .....</b>	<b>1021</b>
<b>CD-ROM Installation Instructions .....</b>	<b>1022</b>

