

XML
FOR
DUMMIES®
4TH EDITION

XML
FOR
DUMMIES®
4TH EDITION

XML
FOR
DUMMIES®
4TH EDITION

by Lucinda Dykes and Ed Tittel



WILEY

Wiley Publishing, Inc.

XML For Dummies®, 4th Edition

Published by
Wiley Publishing, Inc.
111 River Street
Hoboken, NJ 07030-5774

www.wiley.com

Copyright © 2005 by Wiley Publishing, Inc., Indianapolis, Indiana

Published by Wiley Publishing, Inc., Indianapolis, Indiana

Published simultaneously in Canada

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 646-8600. Requests to the Publisher for permission should be addressed to the Legal Department, Wiley Publishing, Inc., 10475 Crosspoint Blvd., Indianapolis, IN 46256, (317) 572-3447, fax (317) 572-4355, or online at <http://www.wiley.com/go/permissions>.

Trademarks: Wiley, the Wiley Publishing logo, For Dummies, the Dummies Man logo, A Reference for the Rest of Us!, The Dummies Way, Dummies Daily, The Fun and Easy Way, Dummies.com, and related trade dress are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates in the United States and other countries, and may not be used without written permission. All other trademarks are the property of their respective owners. Wiley Publishing, Inc., is not associated with any product or vendor mentioned in this book.

LIMIT OF LIABILITY/DISCLAIMER OF WARRANTY: THE PUBLISHER AND THE AUTHOR MAKE NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE ACCURACY OR COMPLETENESS OF THE CONTENTS OF THIS WORK AND SPECIFICALLY DISCLAIM ALL WARRANTIES, INCLUDING WITHOUT LIMITATION WARRANTIES OF FITNESS FOR A PARTICULAR PURPOSE. NO WARRANTY MAY BE CREATED OR EXTENDED BY SALES OR PROMOTIONAL MATERIALS. THE ADVICE AND STRATEGIES CONTAINED HEREIN MAY NOT BE SUITABLE FOR EVERY SITUATION. THIS WORK IS SOLD WITH THE UNDERSTANDING THAT THE PUBLISHER IS NOT ENGAGED IN RENDERING LEGAL, ACCOUNTING, OR OTHER PROFESSIONAL SERVICES. IF PROFESSIONAL ASSISTANCE IS REQUIRED, THE SERVICES OF A COMPETENT PROFESSIONAL PERSON SHOULD BE SOUGHT. NEITHER THE PUBLISHER NOR THE AUTHOR SHALL BE LIABLE FOR DAMAGES ARISING HEREFROM. THE FACT THAT AN ORGANIZATION OR WEBSITE IS REFERRED TO IN THIS WORK AS A CITATION AND/OR A POTENTIAL SOURCE OF FURTHER INFORMATION DOES NOT MEAN THAT THE AUTHOR OR THE PUBLISHER ENDORSES THE INFORMATION THE ORGANIZATION OR WEBSITE MAY PROVIDE OR RECOMMENDATIONS IT MAY MAKE. FURTHER, READERS SHOULD BE AWARE THAT INTERNET WEBSITES LISTED IN THIS WORK MAY HAVE CHANGED OR DISAPPEARED BETWEEN WHEN THIS WORK WAS WRITTEN AND WHEN IT IS READ.

For general information on our other products and services, please contact our Customer Care Department within the U.S. at 800-762-2974, outside the U.S. at 317-572-3993, or fax 317-572-4002.

For technical support, please visit www.wiley.com/techsupport.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Library of Congress Control Number: 2005923240

ISBN-13: 978-0-7645-8845-7

ISBN-10: 0-7645-8845-1

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

40/QT/QV/QV/IN



About the Author

Lucinda Dykes started her career in a high-tech area of medicine, but left medicine to pursue her interests in technology and the Web. She has been writing code and developing Web sites since 1994, and also teaches and develops online courses — including the JavaScript courses for the International Webmasters Association/HTML Writers' Guild at www.eclasses.org.

Lucinda has authored, co-authored, edited, and been a contributing author to numerous computer books; the most recent include *Dreamweaver MX 2004 Savvy* (Sybex), *XML for Dummies* (3rd Edition, Wiley), *Dreamweaver MX Fireworks MX Savvy* (Sybex), *XML Schemas* (Sybex), and *Mastering XHTML* (Sybex). When she can manage to move herself away from her keyboard, other interests include holographic technologies, science fiction, and Bollywood movies.

Ed Tittel is a 23-year veteran of the computing industry. After spending his first seven years in harness writing code, Ed switched to the softer side of the business as a trainer and talking head. A freelance writer since 1986, Ed has written hundreds of magazine and Web articles — and worked on over 100 computer books, including numerous *For Dummies* titles on topics that include several Windows versions, NetWare, HTML, XHTML, and XML.

Ed is also Technology Editor for *Certification Magazine*, writes for numerous TechTarget Web sites, and writes a twice-monthly newsletter, “Must Know News,” for CramSession.com. In his spare time, Ed likes to shoot pool, cook, and spend time with his wife Dina and his son Gregory. He also likes to explore the world away from the keyboard with his trusty Labrador retriever, Blackie. Ed can be contacted at etittel@yahoo.com.

Dedication

To the heroes at the W3C and OASIS, sung and unsung, especially members of the many XML working groups who have made the world (or the Web, at least) a better place through their tireless efforts, and to all those Web pioneers who generously offered help and support to those of us trying to figure out how to make our contribution to the Web in the early '90s.

Author's Acknowledgments

Lucinda Dykes: Thanks to everyone on the scene and behind the scenes who has contributed to making this project possible.

First, I'd like to thank Ed Tittel for giving me not only the opportunity to be involved in this book, but who also played a major role in my entry into the world of technical writing. Ed and I share a long-term interest in language, computers, and markup languages. I'd also like to thank everyone involved in any edition of this book for the excellent foundation they made for this edition to build on.

Next, thanks to the team at Wiley, especially Katie Feltman for her vision and support of this project, Paul Levesque for quiet and steady guidance in addition to excellent editing, Allen Wyatt for insight and outstanding technical editing, and Barry Childs-Helton for superb copy-editing as well as a delightful sense of humor. And thanks to Carole McClendon, my agent at Waterside Productions, who made it possible for me to lead this project.

On a personal note, special thanks to my mother, Doris Dykes, who instilled and supported a lifelong interest in learning and in books. She claims that I'm the first child she lost to the Internet — but that makes me easy to find. Mom: I'll be in front of the nearest computer screen. Thanks and love always to Wali for making it possible for me to spend all these late nights tapping away at the keyboard, and for always making me remember the things that are really important. Thanks to our dear friends, Rose Rowe and Karmin Perless, who walked softly and made room for having a writer around. And finally, thanks to Wendy Fries and Cheryl Kline for great conversation, good advice, and lots of laughter at our monthly writers' session at the Coffee Grove.

Publisher's Acknowledgments

We're proud of this book; please send us your comments through our online registration form located at www.dummies.com/register/.

Some of the people who helped bring this book to market include the following:

Acquisitions, Editorial, and Media Development

Project Editor: Paul Levesque
Acquisitions Editor: Katie Feltman
Copy Editor: Barry Childs-Helton
Technical Editor: Allen Wyatt, Sr.
Editorial Manager: Leah Cameron
Permissions Editor: Laura Moss
Media Development Specialist: Kit Malone
Media Development Manager:
Laura VanWinkle
Media Development Supervisor:
Richard Graves
Editorial Assistant: Amanda Foxworth
Cartoons: Rich Tennant
(www.the5thwave.com)

Composition Services

Project Coordinator: Maridee Ennis
Layout and Graphics: Andrea Dahl,
Stephanie D. Jumper, Julie Trippetti
Proofreaders: Leeann Harney, Joe Niesen,
Carl William Pierce, TECHBOOKS
Production Services
Indexer: TECHBOOKS Production Services

Publishing and Editorial for Technology Dummies

Richard Swadley, Vice President and Executive Group Publisher
Andy Cummings, Vice President and Publisher
Mary Bednarek, Executive Acquisitions Director
Mary C. Corder, Editorial Director

Publishing for Consumer Dummies

Diane Graves Steele, Vice President and Publisher
Joyce Pepple, Acquisitions Director

Composition Services

Gerry Fahey, Vice President of Production Services
Debbie Stailey, Director of Composition Services

Contents at a Glance

<i>Introduction</i>	1
<i>Part I: XML Basics</i>	9
Chapter 1: Getting to Know XML	11
Chapter 2: Using XML for Many Purposes	23
Chapter 3: Slicing and Dicing Data Categories: The Art of Taxonomy	33
<i>Part II: XML and the Web</i>	45
Chapter 4: Adding XHTML for the Web	47
Chapter 5: Putting Together an XML File	65
Chapter 6: Adding Character(s) to XML	83
Chapter 7: Handling Formatting with CSS	95
<i>Part III: Building In Validation with DTDs and Schemas</i>	109
Chapter 8: Understanding and Using DTDs	111
Chapter 9: Understanding and Using XML Schema	135
Chapter 10: Building a Custom XML Schema	157
Chapter 11: Modifying an Existing Schema	173
<i>Part IV: Transforming and Processing XML</i>	195
Chapter 12: Handling Transformations with XSL	197
Chapter 13: The XML Path Language	215
Chapter 14: Processing XML	235
<i>Part V: XML Application Development</i>	245
Chapter 15: Using XML with Web Services	247
Chapter 16: XML and Forms	259
Chapter 17: Serving Up the Data: XML and Databases	271
Chapter 18: XML and RSS	285
<i>Part VI: The Part of Tens</i>	299
Chapter 19: XML Tools and Technologies	301
Chapter 20: Ten Top XML Applications	313
Chapter 21: Ten Ultimate XML Resources	321
<i>Glossary</i>	329
<i>Index</i>	347

Table of Contents

<i>Introduction</i>	1
About This Book	1
Conventions Used in This Book	2
Foolish Assumptions	3
How This Book Is Organized	4
Part I: XML Basics	4
Part II: XML and the Web	4
Part III: Building in Validation with DTDs and Schemas	5
Part IV: Transforming and Processing XML	6
Part V: XML Application Development	6
Part VI: The Part of Tens	7
Glossary	7
Icons Used in This Book	7
Where to Go from Here	8
<i>Part I: XML Basics</i>	9
Chapter 1: Getting to Know XML	11
XML (eXtremeMely cool)	12
Mocking up your own markup	12
Separating data and context	12
Making information portable	13
XML means business	13
Figuring Out What XML Is Good For	14
Classifying information	14
Enforcing rules on your data	15
Outputting information in a variety of ways	16
Using the same data across platforms	17
Beyond the Hype: What XML Isn't	18
It's not just for Web pages anymore	19
It's not a database	20
It's not a programming language	20
Building XML Documents	21
Chapter 2: Using XML for Many Purposes	23
Moving Legacy Data to XML	23
The Many Faces of XML	24
Creating XML-enabled Web pages	24
Print publishing with XML	25

Using XML for business forms	28
Incorporating XML into business processes	29
Serving up XML from a database	31
Alphabet Soup: Even More XML	31

Chapter 3: Slicing and Dicing Data Categories:**The Art of Taxonomy 33**

Taking Stock of Your Data	33
Looking at business practices and partners	34
Gathering some content	34
Checking whether a DTD or schema already exists	35
Searching for a schema repository	36
Breaking Down Data in Different Ways	37
Winnowing out the wheat from the chaff	38
Types of data that can be stored in XML	39
Developing Your Taxonomy	39
Testing Your Taxonomy	41
Using trial and error for the best fit	41
Testing your content analysis	42
Looking Ahead to Validation	43

Part II: XML and the Web 45**Chapter 4: Adding XHTML for the Web 47**

HTML, XML, and XHTML	47
What HTML does best	48
The limits of HTML	49
Comparing XML and HTML	50
Using XML to describe data	51
The benefits of using HTML	53
The benefits of using XML	53
XHTML Makes the Move to XML Syntax	54
Making the switch	55
Every element must be closed	56
Empty elements must be formatted correctly	56
Tags must be properly nested	57
Case makes a difference	57
Attribute values are in quotation marks	58
Converting a document from HTML to XHTML	59
The Role of DOCTYPE Declarations	62

Chapter 5: Putting Together an XML File 65

Anatomy of an XML File	65
The XML declaration	67
Marking up your content	68
Playing by the Rules: Well-Formed Documents	74

Adding Style for the Web	76
Seeking Validation with DTD and XML Schema	78
Why describe XML documents?	79
Choosing between DTD and XML Schema	80
Chapter 6: Adding Character(s) to XML	83
About Character Encodings	84
Introducing Unicode	85
Character Sets, Fonts, Scripts, and Glyphs	87
For Each Character, a Code	88
Key Character Sets	89
Using Unicode Characters	91
Finding Character Entity Information	93
Chapter 7: Handling Formatting with CSS	95
Viewing XML on the Web with CSS	96
Basic CSS Formatting: CSS1	97
The Icing on the Cake: CSS2	98
Building a CSS Stylesheet	98
Adding CSS to XML	99
A simple CSS stylesheet for XML	101
Dissecting a simple CSS stylesheet	102
Linking CSS and XML	106
Adding CSS to XSLT	107
 <i>Part III: Building In Validation with DTDs and Schemas</i>	 109
Chapter 8: Understanding and Using DTDs	111
What's a DTD?	112
When to use a DTD	113
When NOT to use a DTD	113
Inspecting the XML Prolog	114
Examining the XML declaration	115
Discovering the DOCTYPE	116
Understanding comments	116
Processing instructions	117
How about that white space?	117
Reading a DTD	118
Using Element Declarations	119
Using the EMPTY element type and the ANY element type	120
Adding mixed content	121
Using element content models	122
Declaring Attributes	123
Discovering Entities	125
General entities	126
Parameter entities	128

Understanding Notations	130
Calling a DTD	131
Internal DTDs	131
External DTDs	132
When to use an internal or external DTD	133
Chapter 9: Understanding and Using XML Schema	135
What's an XML Schema?	136
So Many Datatypes, So Little Time	138
XML Prolog	139
Document Structures	141
Element declarations	141
</confirmOrder> Attribute declarations	144
Attribute groups	144
What about that white space?	145
Datatype Declarations	148
Simple datatypes	148
Complex datatypes	149
Defining constraints and value checks	149
Dealing with Entities, Notations, and More	150
Annotations	151
Deciding When to Use a Schema	152
Referencing XML Schema Documents	153
The inside view: Referencing a schema in an XML document	153
Calling for outside support: Referencing external schemas in your schema	153
Double-Checking Your Schemas and Documents	155
Chapter 10: Building a Custom XML Schema	157
Doing the Validity Rag	157
Step 1: Understanding Your Data	159
Step 2: Being the Root of All Structure: Elements	159
Step 3: Building Content Models	161
Step 4: Using Attributes to Shed Light on Data Structure	163
Step 5: Using Datatype Declarations to Define What's What	164
Tricks of the Trade	167
Creating a Simple Schema	168
Using a Schema with an XML File in Word 2003	170
Chapter 11: Modifying an Existing Schema	173
Trading Control for Flexibility	174
Eliciting Markup from an XML Schema	174
Modifying a Schema	176
Using Datatypes Effectively	177
Using datatypes with data-intensive content	177
Using datatypes with text-intensive content	179

Making Elements Work Wisely and Well	180
Creating crafty content models	180
A matter of selection	181
Mixing up the order	183
Using Complex Datatypes	183
When XML Schemas Collide: Namespaces	185
Including External Data	188
Including/Excluding Document Content	188
Converting DTDs to Schemas	190

***Part IV: Transforming and Processing XML* 195**

Chapter 12: Handling Transformations with XSL197

The Two Faces of XSL	198
XSLT	198
XSL-FO	200
XSL Stylesheets Are XML Documents	201
A Simple Transformation Using XSLT	202
An XSLT Stylesheet for Converting XML to HTML	202
The pieces of the stylesheet puzzle	205
Processing element content	207
Dealing with repeating elements	209
Creating an XSLT Stylesheet with XSLT Editors	210

Chapter 13: The XML Path Language 215

Why Do You Need Directions?	216
XPath document trees	217
Understanding XPath nodes	218
XPath Directions and Destinations	220
XPath Syntax	221
Some simple location paths	222
Adding expressions	223
Taking steps along the XPath	223
Looking at attributes	224
Going backward	224
Reversing direction	225
Null results	225
Getting back to your roots	226
XPath functions	226
Using XPath with XMLSpy	226
The Short Version	228
Child-axis abbreviations	229
Attribute-axis abbreviation	229
Predicate and expression abbreviations	229
Some more abbreviations	230
What's New in XPath 2.0?	231
Where to Now?	233

Chapter 14: Processing XML235

Frankly, My Dear, I Don't Give a DOM	235
Keeping in touch with the family	238
Understanding DOM structure	238
What Goes In Must Come Out: Processing XML	240
So many processors, so little time	242
Which processor is right for you?	243

Part V: XML Application Development245**Chapter 15: Using XML with Web Services247**

What's Up with Web Services?	248
A Web Services Architecture	251
Transport: Moving XML messages	252
Packaging/Extensions: Managing information exchange	253
Description: Specifying services and related components	254
Discovery: Finding what's available	255
Where Will Web Services Lead?	256

Chapter 16: XML and Forms259

Collecting Information with Forms: The Basics	260
HTML Forms	260
XML Forms	261
XForms	261
InfoPath	267

Chapter 17: Serving Up the Data: XML and Databases271

Using Databases with XML	272
Text-intensive XML	272
Data-intensive XML	273
Creating XML from Database Files	273
Using Word 2003	274
Using InfoPath	275
Using XMLSpy	278
Using XML with Access 2003	281

Chapter 18: XML and RSS285

Introducing RSS	286
Sorting Out the Versions	286
RSS 0.9x	287
RSS 2.0/2.01	290
RSS 1.0	291
Validating an RSS Feed	295
Creating RSS Feeds	296
Get Syndicated!	297
Using an RSS Reader	298

Part VI: The Part of Tens299**Chapter 19: XML Tools and Technologies301**

Creating Documents with Authoring Tools	301
Epic Editor	302
Turbo XML v2.4.1	303
XMetaL Author 4.5	303
XML Pro v2.0.1	303
XML Spy 2005	304
Checking Documents with Parser Tools	304
Ælfred	305
expat	306
Lark	306
Viewing with XML Browsers	307
Amaya	307
Internet Explorer 6	307
Mozilla	308
Firefox 1.0	308
Opera	308
Using XML Parsers and Engines	309
XML C Library for Gnome	309
Java XML Pack	310
Xerces	310
Employing Conversion Tools	311
HTML Tidy	311
Extensible Programming Script (XPS)	311
The Ultimate XML Grab Bag and Goodie Box	312
Microsoft does XML, too!	312
webMethods automates XML excellence	312

Chapter 20: Ten Top XML Applications313

XHTML = XML + HTML	314
XML Style Is a Matter of Application	314
Wireless Markup Language (WML)	314
DocBook, Anyone?	315
Mathematical Markup Language (MathML)	315
Scalable Vector Graphics (SVG)	316
Resource Description Framework (RDF)	316
Synchronized Multimedia Integration Language (SMIL)	317
Servin' Up Web Services	317
XQuery	318
Create XML Applications with Zope	319

Chapter 21: Ten Ultimate XML Resources321

XML's Many and Marvelous Specs	321
An XML Nonpareil	322
Top XML Tutorial Sites	322

XML in the Mail	323
Excellent XML Examples at zvon.org	323
XML News and Information	323
XML Training Options	324
Building a Bodacious XML Bookshelf	325
Studying XML for Certification	326
Serious Searches Lead to Success	327
<i>Glossary</i>	329
<i>Index</i>	347

Introduction

Welcome to the latest frontier of Web technology. In *XML For Dummies*, 4th Edition, we introduce you to the mysteries of eXtensible Markup Language (*XML*). XML is helping developers capture, manipulate, and exchange all kinds of documents and data, ranging from news feeds to financial transactions. In fact, many experts believe XML represents a kind of “lingua franca” that can represent information in just about any imaginable form, more accessibly than ever before — not only to human readers, but also to all kinds of computer applications and services.

We take a practical and straightforward approach to telling you about XML and what it can do for your data and document capture, management, and exchange efforts. We try to keep the amount of technobabble to a minimum and stick to plain English as much as possible. We also try to keep the focus on practical applications of XML technology, including desktop applications such as Office 2003. We have carefully chosen what we feel are the most relevant XML technologies for developers today. Besides plain talk about XML — and the many special-purpose applications that XML supports for document designers and authors, graphics developers, and many other communities of technical and business interests — we include lots of sample markup to help you put XML to work in your organization, business, or personal life. (No personal life is quite complete without a little XML.)



The Web page for this book is available at www.dummies.com/go/xmlfd4e. This Web page includes all the XML example files from this book, as well as numerous XML authoring tools, parsers, development kits, and other goodies for you to download. We hope you'll find it helpful for your own projects!

About This Book

Think of this book as your friendly, approachable guide to using XML for all kinds of interesting purposes. Using XML is a bit trickier than using HTML, so this book is organized to make it easier to grapple with XML's fundamentals, wrestle them to the ground, and use them well. We also document voluminous additional sources of information, both online and offline. Here are some of the topics we include:

- ✓ An overview of XML's capabilities, terminology, and technologies
- ✓ Tips for styling XML with CSS and XSLT

- ✓ Hands-on practice in developing DTDs and XML Schema for validating XML documents
- ✓ A beginner's guide to XPath
- ✓ An introduction to XForms and InfoPath
- ✓ A guide to XML application development, including Web services, databases, and news feeds



Because XML is essentially a markup language used to create other XML-based markup languages — or what we also call XML applications — it's not exactly accurate to call a document based on one particular XML application or another an “XML document.” It really makes more sense to call it an “XML-based document” because the document itself contains markup defined using XML. But for brevity's sake, we call such documents *XML documents* in this book. After all, such documents must adhere to the rules of XML syntax and structure if they are to work properly. We could get all fussy and always refer to them (more correctly) as “XML-based documents” or “documents based on such-and-such an XML application.” But that makes us squirm too.

Although you might think that using XML requires years of training and advanced technical wizardry, we don't think that's true. If you can tell someone how to drive across town, you can certainly use XML to build documents that do what you want them to. The purpose of this book isn't to turn you into a true-blue geek, complete with pocket protector. Rather, *XML For Dummies*, 4th Edition shows you which design and technical elements you need so you can get a practical handle on what XML is and how it works. We also provide numerous examples and case studies to illustrate how XML behaves, so you can gain the know-how and confidence to use XML to good effect!

Conventions Used in This Book

Throughout this book, you see lots and lots of markup. All XML markup appears in monospace type, like this:

```
<Greeting>Hello, world!</Greeting>..
```

When you type XML tags or other related information, be sure to copy the information exactly as you see it between the angle brackets (< and >), because that's part of the magic that makes XML work. Other than that, we tell you how to marshal and manage the content that makes your pages special, and we tell you exactly what you need to do to mix the elements of XML with your own work.



Because the margins in this book can't accommodate some long lines of XML markup and still stay legible, sometimes we have to break lines of code. That tends to happen in designations for Web sites (called *URLs*, for *Uniform Resource Locators*) or special XML identifiers for namespaces and other information objects (called *URIs*, or *Uniform Resource Identifiers*) and also in the odd monstrously long line of markup that wraps to the next line. On your computer, these wrapped lines would appear on-screen as a single line of XML or as a single URL or URI — so don't insert a hard return when you see any such lines wrap in the book. Here are some examples of wrapped lines:

```
www.infomagic.austin.com/nexus/plexus/lexus/praxis/  
this_is_deliberately_long.html
```

and

```
<Item>Scientists have developed a robot that "learns" to walk like a toddler,  
improving its step and balance with every stride.</Item>
```



XML is sensitive to how element text is entered. If you're following our examples from the comfort of your living room, keep in mind that you have to use uppercase, lowercase, or other characters exactly as they appear in the book (or, more important, as they're defined in the document description that governs any well-formed, valid XML document — be it an XML Schema or a Document Type Definition, or *DTD*). To make your work look like ours as much as possible, enter all element text exactly as it appears in this book. Better yet, download the file from the Web page for the book (www.dummies.com/go/xmlfd4e)!

Foolish Assumptions

Someone once said that making assumptions makes a fool out of the person who makes them and the person who is their subject. Even so, we're going to make a few assumptions about you, our gentle reader:

- ✓ You're already familiar with text files and know how to use a text editor.
- ✓ You have a working connection to the Internet.
- ✓ You're hip to the difference between a Web browser and a Web server.
- ✓ You want to build your own XML documents for fun, for profit, or because it's part of your job.

Also, we assume that you have a modern Web browser — one that can support XML directly. As we write this, that elite includes Internet Explorer 5.5 (and higher), Netscape Navigator 6 (and later), Opera, Firefox, Mozilla, and

Amaya — all have decent XML parsing and rendering capabilities. Don't worry, though, if you don't have such a browser. Part of what you find in these pages and on the Web page for the book is a collection of pointers to help you obtain the tools you need to work directly with XML on your own computer. You don't need to be a master logician or a programming whiz to work with XML; all you need are the time required to discover its ins and outs and the determination to understand its intricacies and capabilities.

Even if you were one of those who fled English Composition in school and hid out in the computer lab, take heart: If you can write a sentence and you know the difference between a heading and a paragraph, you can build and publish your own XML documents. If you have an imagination and the ability to communicate what's important to you in an organized manner, you've already mastered the ingredients necessary to build useful, information-rich XML documents and data collections. The rest is details — and we help you with those!

How This Book Is Organized

This book contains six major parts; each part contains three or more chapters; each chapter has (in all modesty) lots of good stuff. Any time you need help or information, pick up the book and start anywhere you like, or use the table of contents and index to locate specific topics or keywords. This section of your friendly intro offers a preview of the six parts and what you find in each one.

Part I: XML Basics

Part I sets the stage. It begins with an overview of XML's special capabilities and discusses what XML is and what XML is not. We tempt you toward the XML side of the Force (hopefully) by exploring the many uses for XML — and checking out the applications to which it's so well suited. We also briefly discuss the relationships between and among the many XML languages and let you know which ones we think are particularly useful for today's developer. We conclude Part I with a look at techniques for analyzing and classifying your data so that you can make XML documents meet your data requirements. You also get to see how XML documents gain their structure and content — from a thorough analysis of requirements and examples.

Part II: XML and the Web

In Part II, you find out all about displaying XML content on Web pages. First, we cover what's involved in converting HTML to its XML-based equivalent, XHTML, as a way of introducing XML's syntax and structure.

Chapter 5 picks up that thread, and you find out how to construct an XML document piece by piece while playing by the rules of XML. We show you how to create well-formed documents and discuss how XML documents and data can be made subject to formal descriptions (a great way to define a set of rules that humans and computers can follow with equal ease). You find out why you might (or might not) want to validate your XML documents with a DTD or XML Schema.

In Chapter 6 we explore character sets and related entities that XML depends on to represent content and explain how to use them in your documents.

We conclude Part II with an explanation of what's involved in bringing XML documents to the Web and talk about the best ways to use styles to make their contents more presentable. To that end, we explore ways to use Cascading Style Sheets (CSS) to make native XML documents (or XML content transformed into HTML) easier to read and appreciate online.

Part III: Building in Validation with DTDs and Schemas

In Part III, we explain the purpose and functions that Document Type Definitions (DTDs) can play in describing XML documents. We use a DTD to teach you about the XML markup that it enables. We explain how to read a DTD to recognize the elements, attributes, and content models it contains.

After that, we look at an “all-XML, all the time” alternative to DTDs called XML Schema — an application that provides even more capabilities to describe, use, and control XML documents. One part of XML Schema's appeal derives from its basis in XML itself. Because XML Schema is just another XML application (albeit one that allows you to describe other XML applications), you've got a leg up if you already have a working knowledge of XML: You can apply that knowledge to describing XML applications without having to learn yet another markup language. DTDs (on the other hand) are based on SGML, not XML; you have to have XML under your belt before you can use, customize, or create DTDs that describe XML applications. Another major part of XML Schema's appeal derives from its broad selection of built-in datatypes and support for user-derived datatypes; you can be as specific as you want (or need) to be in describing your data.

We explain how to create elements, attributes, datatypes, and content models to work in XML Schemas. We provide details on how to construct a valid XML Schema document and show you how to use this document to create new XML documents in Word 2003.

We conclude Part III by explaining how to combine XML Schemas and how to mix and match XML Schema contents or components to maximize this technology. We also introduce XML namespaces and take a look at converting DTDs to XML Schemas.

The four chapters in this part represent some of the most important nuts and bolts in the entire book.

Part IV: Transforming and Processing XML

In Part IV, we jump into the ins and outs of the eXtensible Stylesheet Language (*XSL*) that can be used to turn XML-based data or documents into just about any form or format imaginable. After that, we explore the details of transforming an XML document into different formats — and dispel the mysteries involved in putting XSL to work for you when you change things around.

Next, we show you how to use XPath to describe the precise location of elements, attributes, and their values in an XML document.

To conclude Part IV, our final stop is inside the machinery that makes XML usable, as we explore what's involved when a computer reads and absorbs an XML document and list what kinds of capabilities the necessary software (usually called an XML processor) can deliver.

Part V: XML Application Development

In Part V, we explore what you can do with XML when you've got some ready to work with — and show you many possible ways to get things done with a little help from XML.

First, we take a look at an exciting set of XML-based applications designed to advertise, locate, and use so-called “Web services” — a software and messaging architecture that enables service providers to advertise their services on the Web and users to locate and use such services. Web services can involve anything from access to proprietary databases, remote storage or processing, or even access to basic productivity applications (word processing, spreadsheets, e-mail, and so forth) that users normally see on their own desktops but often show up running elsewhere on the Internet. There's plenty of hype and hope for the future of Web services, and you explore the reasons why this is the case.

Next, you find out all about using forms to collect XML data and take a close look at two very different ways to use forms with XML: XForms, the W3C's “next generation” of Web forms, and InfoPath, Microsoft's visual XML forms editor.

In Chapter 17, you explore using XML with databases and how to import and export XML data using Word, InfoPath, XMLSpy, and Access.

To conclude Part V, we explain how to use XML on the Web for syndicating content with RSS news feeds. You get the word on how to create an RSS file, as well as how to validate your file and submit it for syndication.

Part VI: The Part of Tens

Part VI introduces our picks of the best XML tools, applications and resources. We begin this part with a brief survey of popular, widely used XML tools and technologies. These include special-purpose XML editors and authoring tools, XML-based management tools, XML-capable browsers, parsers and engines, and conversion tools.

In Chapter 20, you have a chance to observe some of the best and brightest uses of XML and to understand why a certain set of XML applications is of such great interest to so many content designers and developers. Finally, in Chapter 21, you can read about some of the most appealing and useful sources of information about XML and related applications known to man and woman.

Glossary

In the glossary, you can find definitions for all terms that make you go “Huh?” We did our best to choose the ones that really need an explanation and to define them in a way that’s easy to understand.



The materials on the *XML For Dummies*, 4th Edition Web site (www.dummies.com/go/xmlfd4e/) are designed to help you match up the markup and examples that appear within the pages of the book to their electronic counterparts on the Web site. In addition, we’ve provided links to as comprehensive a collection of tools and programs for XML as we could gather here for your delectation and use.

Icons Used in This Book



This icon signals technical details that are informative and interesting but not critical to writing XML. Skip these if you want (but please, for the sake of your inner geek, come back and read them later).



This icon flags useful information that demystifies (and helps un-complicate) XML markup, Web-page design, or other important stuff.



This icon points out information that you shouldn't pass by — don't overlook these gentle reminders (the life you save could be your own).



Be cautious when you see this icon. It warns you of things you shouldn't do; the bomb emphasizes that the consequences of ignoring these bits of wisdom can be severe.

Where to Go from Here

To keep up with the latest version of these references, please visit the related *XML For Dummies* site at www.dummies.com/go/xmlfd4e/. Here, you find the results of our best efforts to keep the information in the book current and a list of errata to straighten out any mistakes, boo-boos, or gotchas that we weren't able to root out before the book went to publication. We hope you find this a convincing demonstration that our hearts are in the right place (we already know we're not perfect).

Please share your feedback with us about the book. We can't claim that we'll follow every suggestion or react to every comment, but you can be pretty certain that suggestions that occur repeatedly — or that add demonstrable value to the book — will find a place in the next edition!

Good luck on your journey, and don't forget to keep your eyes on the road and your hands on the wheel as you cruise the information highway.

Enjoy!

Part I

XML Basics

The 5th Wave

By Rich Tennant



"They're moving on to the XML APIs section.
That should daze and confuse them enough
for us to finish changing the tire."

In this part . . .

Here you get a gentle-but-formal introduction to the eXtensible Markup Language, also known as XML. Starting in Chapter 1, you get a look at XML's capabilities, strengths, and versatility. You get tips on the best uses of XML, and draw a bead on the other pieces that may be necessary for an XML solution. In Chapter 2, we introduce you to the options for XML output — including Web pages, print documents, forms, spreadsheets, and databases. Then the wide variety of XML languages comes to light. Finally, Chapter 3 rounds out your basic toolkit with a close look at how to develop and test a classification scheme for your data.