

The fun and easy way® to get going with XML
— covers the newest specifications!

XML

FOR DUMMIES®

4th Edition

**A Reference
for the
Rest of Us!®**

FREE eTips at dummies.com®

**Lucinda Dykes
Ed Tittel**

Author of PC Magazine® Fighting
Spyware, Viruses, and Malware

Companion Web site
has sample code,
links to trial software,
and more



The fun and easy way® to get going with XML
— covers the newest specifications!

XML

FOR

DUMMIES®

4th Edition

**A Reference
for the
Rest of Us!®**

FREE eTips at dummies.com®

**Lucinda Dykes
Ed Tittel**

Author of PC Magazine® Fighting
Spyware, Viruses, and Malware

Companion Web site
has sample code,
links to trial software,
and more



***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 St.

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

4O/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

[Title](#)

[Introduction](#)

[About This Book](#)

[Conventions Used in This Book](#)

[Foolish Assumptions](#)

[How This Book Is Organized](#)

[Icons Used in This Book](#)

[Where to Go from Here](#)

[Part I : XML Basics](#)

[Chapter 1: Getting to Know XML](#)

[XML \(eXtreMely cooL\)](#)

[Figuring Out What XML Is Good For](#)

[Beyond the Hype: What XML Isn't](#)

[Building XML Documents](#)

[Chapter 2: Using XML for Many Purposes](#)

[Moving Legacy Data to XML](#)

[The Many Faces of XML](#)

[Alphabet Soup: Even More XML](#)

[Chapter 3: Slicing and Dicing Data Categories: The Art of Taxonomy](#)

[Taking Stock of Your Data](#)

[Breaking Down Data in Different Ways](#)

[Developing Your Taxonomy](#)

[Testing Your Taxonomy](#)

[Looking Ahead to Validation](#)

[Part II : XML and the Web](#)

[Chapter 4: Adding XHTML for the Web](#)

[HTML, XML, and XHTML](#)

[Comparing XML and HTML](#)

[XHTML Makes the Move to XML Syntax](#)

[Converting a document from HTML to XHTML](#)

[The Role of DOCTYPE Declarations](#)

[Chapter 5: Putting Together an XML File](#)

[Anatomy of an XML File](#)

[Playing by the Rules: Well-Formed Documents](#)

[Adding Style for the Web](#)

[Seeking Validation with DTD and XML Schema](#)

[Chapter 6: Adding Character\(s\) to XML](#)

[About Character Encodings](#)

[Introducing Unicode](#)

[Character Sets, Fonts, Scripts, and Glyphs](#)

[For Each Character, a Code](#)

[Key Character Sets](#)

[Using Unicode Characters](#)

[Finding Character Entity Information](#)

[Chapter 7: Handling Formatting with CSS](#)

[Viewing XML on the Web with CSS](#)

[Basic CSS Formatting: CSS1](#)

[The Icing on the Cake: CSS2](#)

[Building a CSS Stylesheet](#)

[Adding CSS to XML](#)

[Adding CSS to XSLT](#)

[Part III : Building in Validation with DTDs and Schemas](#)

[Chapter 8: Understanding and Using DTDs](#)

[What's a DTD?](#)

[Inspecting the XML Prolog](#)

[Reading a DTD](#)

[Using Element Declarations](#)

[Declaring Attributes](#)

[Discovering Entities](#)

[Understanding Notations](#)

[Calling a DTD](#)

[Chapter 9: Understanding and Using XML Schema](#)

[What's an XML Schema?](#)

[So Many Datatypes, So Little Time](#)

[XML Prolog](#)

[Document Structures](#)

[Datatype Declarations](#)

[Dealing with Entities, Notations, and More](#)

[Annotations](#)

[Deciding When to Use a Schema](#)

[Referencing XML Schema Documents](#)

[Double-Checking Your Schemas and Documents](#)

[Chapter 10: Building a Custom XML Schema](#)

[Doing the Validity Rag](#)

[Step 1: Understanding Your Data](#)

[Step 2: Being the Root of All Structure: Elements](#)

[Step 3: Building Content Models](#)

[Step 4: Using Attributes to Shed Light on Data Structure](#)

[Step 5: Using Datatype Declarations to Define What's What](#)

[Tricks of the Trade](#)

[Creating a Simple Schema](#)

[Using a Schema with an XML File in Word 2003](#)

[Chapter 11: Modifying an Existing Schema](#)

[Trading Control for Flexibility](#)

[Eliciting Markup from an XML Schema](#)

[Modifying a Schema](#)

[Using Datatypes Effectively](#)

[Making Elements Work Wisely and Well](#)

[Using Complex Datatypes](#)

[When XML Schemas Collide: Namespaces](#)

[Including External Data](#)

[Including/Excluding Document Content](#)

[Converting DTDs to Schemas](#)

[Part IV : Transforming and Processing XML](#)

[Chapter 12: Handling Transformations with XSL](#)

[The Two Faces of XSL](#)

[XSL Stylesheets Are XML Documents](#)

[A Simple Transformation Using XSLT](#)

[An XSLT Stylesheet for Converting XML to HTML](#)

[Creating an XSLT Stylesheet with XSLT Editors](#)

[Chapter 13: The XML Path Language](#)

[Why Do You Need Directions?](#)

[XPath Directions and Destinations](#)

[XPath Syntax](#)

[Using XPath with XMLSpy](#)

[The Short Version](#)

[What's New in XPath 2.0?](#)

[Where to Now?](#)

[Chapter 14: Processing XML](#)

[Frankly, My Dear, I Don't Give a DOM](#)

[What Goes In Must Come Out: Processing XML](#)

[Part V : XML Application Development](#)

[Chapter 15: Using XML with Web Services](#)

[What's Up with Web Services?](#)

[A Web Services Architecture](#)

[Where Will Web Services Lead?](#)

[Chapter 16: XML and Forms](#)

[Collecting Information with Forms: The Basics](#)

[HTML Forms](#)

[XML Forms](#)

[Chapter 17: Serving Up the Data: XML and Databases](#)

[Using Databases with XML](#)

[Creating XML from Database Files](#)

[Using XML with Access 2003](#)

[Chapter 18: XML and RSS](#)

[Introducing RSS](#)

[Sorting Out the Versions](#)

[Validating an RSS Feed](#)

[Creating RSS Feeds](#)

[Get Syndicated!](#)

[Using an RSS Reader](#)

[Part VI : The Part of Tens](#)

[Chapter 19: XML Tools and Technologies](#)

[Creating Documents with Authoring Tools](#)

[Checking Documents with Parser Tools](#)

[Viewing with XML Browsers](#)

[Using XML Parsers and Engines](#)

[Employing Conversion Tools](#)

[The Ultimate XML Grab Bag and Goodie Box](#)

[Chapter 20: Ten Top XML Applications](#)

[XHTML = XML + HTML](#)

[XML Style Is a Matter of Application](#)

[Wireless Markup Language \(WML\)](#)

[DocBook, Anyone?](#)

[Mathematical Markup Language \(MathML\)](#)

[Scalable Vector Graphics \(SVG\)](#)

[Resource Description Framework \(RDF\)](#)

[Synchronized Multimedia Integration Language \(SMIL\)](#)

[Servin' Up Web Services](#)

[XQuery](#)

[Create XML Applications with Zope](#)

[Chapter 21: Ten Ultimate XML Resources](#)

[XML's Many and Marvelous Specs](#)

[An XML Nonpareil](#)

[Top XML Tutorial Sites](#)

[XML in the Mail](#)

[Excellent XML Examples at zvon.org](#)

[XML News and Information](#)

[XML Training Options](#)

[Building a Bodacious XML Bookshelf](#)

[Studying XML for Certification](#)

[Serious Searches Lead to Success](#)

[Glossary](#)

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.