

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.eclases.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

Introduction 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 |

Part I: XML Basics 9

Chapter 1: Getting to Know XML 11

| | |
|---|----|
| XML (eXtreMely 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 |

Part III: Building In Validation with DTDs and Schemas 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 |