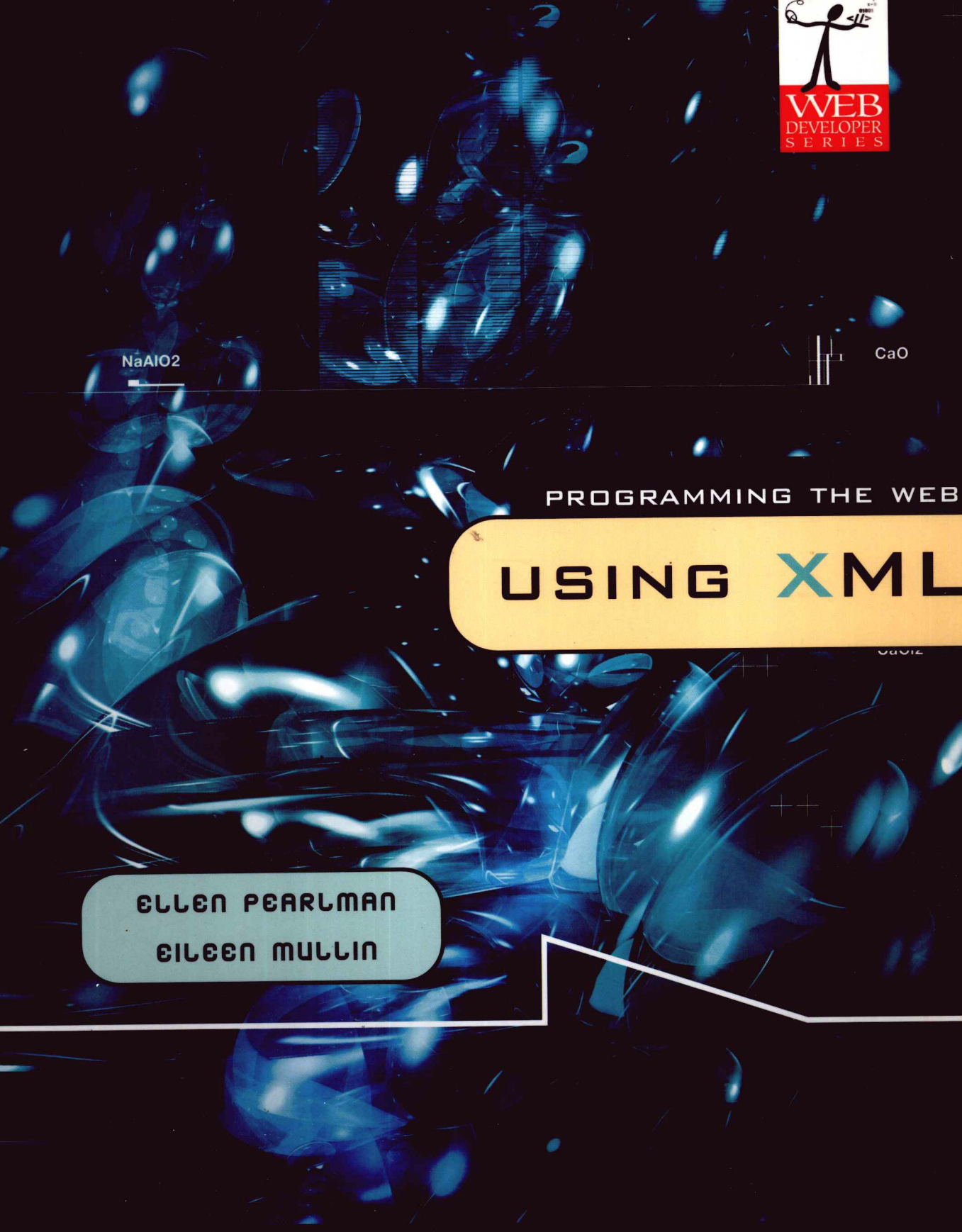
WEB
DEVELOPER
SERIES
NaAlO2
CaO
PROGRAMMING THE WEB
USING XML
ELLEN PEARLMAN
EILEEN MULLIN

Programming the Web Using XML

Ellen Pearlman
Columbia University

Eileen Mullin

Technology Education

Boston Burr Ridge, IL Dubuque, IA Madison, WI New York San Francisco St. Louis
Bangkok Bogotá Caracas Kuala Lumpur Lisbon London Madrid Mexico City
Milan Montreal New Delhi Santiago Seoul Singapore Sydney Taipei Toronto

PROGRAMMING THE WEB USING XML

This book is printed on acid-free paper.

domestic 1 2 3 4 5 6 7 8 9 0 DOC/DOC 0 9 8 7 6 5 4 3
international 1 2 3 4 5 6 7 8 9 0 DOC/DOC 0 9 8 7 6 5 4 3

ISBN 0-07-284550-3

Editor in chief: *Bob Woodbury*
Publisher: *Brandon Nordin*
Senior sponsoring editor: *Donald J. Hull*
Developmental editor: *Lisa Chin-Johnson*
Marketing manager: *Andy Bernier*
Executive producer: Media technology: *Mark Christianson*
Project manager: *Kristin Puscas*
Senior production supervisor: *Sesha Bolisetty*
Lead designer: *Pam Verros*
Senior supplement producer: *Rose M. Range*
Senior digital content specialist: *Brian Nacik*
Cover design: *Adam Rooke*
Typeface: *10/12 New Baskerville*
Compositor: *Black Dot Group*
Printer: *R. R. Donnelley*

Library of Congress Cataloging-in-Publication Data

Pearlman, Ellen.
Programming the Web using XML / Ellen Pearlman, Eileen Mullin.
p. cm. – (Web developer series)
ISBN 0-07-284550-3 (alk. paper) – ISBN 0-07-121504-2 (international : alk. paper)
1. XML (Document markup language) 2. Internet programming. 3. World Wide Web. I. Mullin, Eileen. II. Title, III. Series.
QA76.76.H94P43 2004
006.7'6–dc22

2003066462

INTERNATIONAL EDITION ISBN 0-07-121504-2

www.mhhe.com

McGraw-Hill Technology Education

At McGraw-Hill Technology Education, we publish instructional materials for the technology education market, in particular computer instruction in post-secondary education—from introductory courses in traditional four-year universities to continuing education and proprietary schools. McGraw-Hill Technology Education presents a broad range of innovative products—texts, lab manuals, study guides, testing materials, and technology-based training and assessment tools.

We realize that technology has created and will continue to create new mediums for professors and students to use in managing resources and communicating information to one another. McGraw-Hill Technology Education provides the most flexible and complete teaching and learning tools available as well as offer solutions to the changing world of teaching and learning. McGraw-Hill Technology Education is dedicated to providing the tools for today's instructors and students, which will enable them to successfully navigate the world of Information Technology.

- McGraw-Hill/Osborne—This division of The McGraw-Hill Companies is known for its best-selling Internet titles, Harley Hahn's *Internet & Web Yellow Pages*, and the *Internet Complete Reference*. For more information, visit Osborne at www.osborne.com.
- Digital Solutions—Whether you want to teach a class online or just post your "bricks-n-mortar" class syllabus, McGraw-Hill Technical Education is committed to publishing digital solutions. Taking your course online doesn't have to be a solitary adventure, nor does it have to be a difficult one. We offer several solutions that will allow you to enjoy all the benefits of having your course material online.
- Packaging Options—For more information about our discount options, contact your McGraw-Hill sales representative at 1-800-338-3987 or visit our web site at **www.mhhe.com/it**.

McGraw-Hill Technology Education is dedicated to providing the tools for today's instructors and students.

Preface

To the Student

This book gives a simple but extensive introduction to XML with its many derivations. The textbook contains everything necessary to learn XML and build on that basic understanding to understand how many different aspects of the Web and other display and nondisplay devices XML encompasses. If you are a Web developer, database administrator, graphic designer, system administrator, or even a computer-science pundit, this book is essential for you to grasp both the subtle and wide reach of XML in today's dynamic and changing technology environment.

Initially the Web dealt primarily with plain text and images through the use of HTML. The necessity for working with data in any form other than straight pages of typed information was an afterthought. XML has taken the art of data in its many forms to a new level. All student levels should be able to comprehend how XML newly defines data and how that data can be put to new and varied uses. After comparing HTML, XHTML, and XML, this book looks at creating Document Type Definitions and schemas and examines checking that information through the use of XML parsers and the expanded role of the multilingual capabilities of Unicode. It examines the role of the Document Object Model and then shows how to style basic XML though Cascading Style Sheets, eXtensible Style Sheets, and the linking of documents. The last section of the book examines the more complex subjects of Scalable Vector Graphics, Synchronized Multimedia Integration Language, databases, and, finally, the expanding use of Web Services. After completing all the chapters, exercises, and projects, you not only should be knowledgeable about XML, but also should be able to begin implementing it in your own projects.

The basic structure of XML and the enthusiasm with which the software development community has welcomed it makes XML certain to be around for many years. Whether you are working with only raw data or sophisticated media presentations, XML is certain to be involved with some part of your project planning. XML statements can be written in a simple text editor or in a sophisticated editor such as XML Spy. Many software vendors are integrating XML into their new releases in new and unexpected ways. The examples, exercises, projects, and code samples in this book will give you a solid foundation to understand any new uses of XML that will show up in the future, knowledge that will not become irrelevant after just a year or two.

To the Teacher

The students do not need a programming background to begin to grasp the concepts and uses of XML, though they do need to understand how to navigate and use the Web. After the core components of elements and attributes are understood, a student can begin to make rudimentary statements. Because XML touches on so many areas, students who are not computer science majors can greatly benefit from this book as well as those starting out in that field.

Each chapter breaks down into fundamental programming concepts about a specific aspect of XML, something that was unimaginable just five years ago, when XML was just emerging as a standard. Using the coding examples, students will very quickly learn how to code and correct their mistakes. The Hands On Projects at the end of the chapters build on the core concepts developed in the chapter and real-world examples are also used to illustrate more sophisticated uses in a business or enterprise environment. This book is an overview of many aspects of XML but does not delve too deeply into the technical side, as there are other texts designed to do that. What it does do is give a thorough introduction to XML and allow students to use that knowledge to build on.

Each chapter leads to the next and there are always examples from real life to reinforce the lessons. Each chapter highlights key vocabulary terms and also asks relevant questions, points out technical tips, and supplies specialized alerts and advice.

Following is a short outline describing the contents of all of the chapters.

- Chapters 1, 2, and 3 go over the basic structure of XML and show how to write simple XML statements. The chapters then go on to compare and contrast XML, HTML, and XHTML. Chapter 1 discusses the history of XML and how it works with multiple platforms and devices to share content. Chapter 2 shows how XHTML documents differ from HTML and investigates namespaces. Chapter 3 shows the overall structure of an XML document; looks at the difference between well-formed and valid XML; and examines elements, attributes, entities, and comments within the context of Document Type Definitions. After completing these chapters, students will be familiar with SGML, XHTML, DTDs, well-formed and valid XML, simple DTDs, elements, attributes, entities, and comments.
- Chapters 4, 5, and 6 cover how to build a Document Type Definition and related XML Schemas and how these files are parsed though different editors. They also look at XML's multilanguage capabilities though the use of Unicode. Chapter 4 shows the basic function and syntax of a Document Type Definition and its internal and external subsets, how elements work, and when to use them with attributes. Although the book does not

set up universal standards for Document Type Definitions, it enables the student to quickly comprehend and read these statements. Chapter 5 expands on the structure of an XML statement and explores schemas, an expanded alternative to Document Type Definitions. It defines schema namespaces and shows the difference between `simpleType` schema data and `complexType` schema data. It then shows how child elements and minimum and maximum occurrences of those elements are set. Finally, it explains how to make intelligent choices when designing a schema. Chapter 6 illustrates how an XML statement is parsed and looks at a variety of different programs and how they can treat the exact same statement differently. It briefly looks at the multilingual capabilities of Unicode and explores the different character sets and typefaces that are available to use with it. After completing these chapters students will be familiar and comfortable with markup, structure, and editing XML, and will understand the importance of rigorous coding. Students also will have a simple framework to build XML statements and a clear picture of the importance of step-by-step coding.

- Chapters 7, 8, and 9 are primarily concerned with the simple design and linkage of XML, and show how some of the same principles that apply to HTML can be relevant for styling XML. Chapter 7 covers CSS stylesheets and how they are used with XML documents and explains the difference between CSS and XSL formats. Chapter 8 shows how XSL can transform XML documents into HTML and how templates govern these transformations. It also examines the three components of XSL: XSLT, XPath, and XSL Formatting Objects. Chapter 9 explains how to link specific parts of XML documents using XLink and how single-direction and multidirectional links can be built. Finally, it shows how to use XPointer to point to any section of a target document on either a local or remote Web Server. After completion of these chapters, students will have a basic understanding of how to style XML documents with coding and templates, and how to link those documents to each other and to servers. In addition, students will understand how basic styling of XML documents works and how these documents can be linked to themselves and other documents.

- Chapters 10, 11, and 12 increase in complexity and cover scripting with the Document Object Model, as well as introducing graphic and media uses of XML vis-à-vis SVG and SMIL. Chapter 10 shows how elements and attributes can be represented as objects and how XML data can be loaded and displayed using simple JavaScript statements to make a site work with on-the-fly interactive capabilities. Chapter 11 shows the benefits of coding visual objects in SVG to save bandwidth and make editing of visual information simpler than editing in Flash. Chapter 12 shows how SMIL is now being deployed in multimedia presentations and discusses the critical differences between SMIL 1.0 and SMIL 2.0. After completion of these chapters, students will be ready to design, program, and view simple visual examples of the visual representations of XML.

- Chapters 13 and 14 take XML into even more complex terrain by showing how XML is used with databases and Web Services. Chapter 13 explores how XML is used with a relational database and discusses when to use a native XML database. It gives the student an overview of how database vendors use XML. Chapter 14 explains how XML underlies Web Services and looks at the three components of Web Services—SOAP, UDDI, and WSDL—and shows how they are used to "publish, find, and bind." The student will learn in these chapters how to go on to the next level of whatever subgroup of XML they intend to specialize in, and will have a thorough understanding of how to integrate XML within their own programming specializations.
- There are also lists of questions, exercises, and code examples from all the chapters as well as alerts and advice, tech tips, and hands on projects. The appendices provide additional information to supplement your lessons.

Acknowledgments

My eternal thanks to the McGraw-Hill team: Lisa, Dan, and all the other editors and diligent production staff who worked so thoroughly to make sure this book got out on time. Special thanks go to Dan Silverburg, who initially acquired and shepherded the project, and Lisa Chin-Johnson, who took it over and skillfully managed it. The entire team, from tech editing to proofreading to production, was extraordinarily helpful. The academic reviewers and professional programmers who reviewed each chapter offered many suggestions that made this a better textbook and showed infinite patience in the process.

Hats off to Dennis Green, Director of the CTA (Computer Technology and Applications) Program at Columbia University, for letting me first teach XML before it achieved its raging popularity, as well as Sven Travis at Parsons School of Design, M.A. Program in Digital Design for allowing initial forays into SVG and SMIL and Michael Randazzo and Zee Dempster at the New School University Computing Center for taking the XML leap in 1999.

Bob Albert, of RealNetworks, Inc. Strategic Relations Group for helping with conversion codes as well as Erika Shaffer, Director PR, RealNetworks for taking me through the maze, and Tina Eisinger, Marketing Manager at Altova GmbH, the XML Spy Company for complete, for copies of XML SPY.

To the members of the "Web Cabal" a.k.a. Link Tank, for thoughts on future uses of XML and beyond. Enormous and eternal thanks to Ken Jordan for being a great friend and initiating the collaborations with the VAN (Virtual Arts Network), the Walker Art Museum and the Ford and Rockefeller Foundations for having the foresight to fund the VAN studies.

Also to Jonathan Spira and David Goldes of Basex, Inc. experts in collaborative business knowledge for supplying underlying technologies and expertise. Finally, eternally and always thanks to David Fugate at Waterside Productions, my agent who, through the entire hand holding is, simply and always, the greatest!

—Ellen Pearlman

For Ross, from four to forever

"Let us dedicate ourselves to the task of simplifying the Internet's interfaces and to educating all that are interested in its use."

—Vinton Cerf, The Internet Is for Everyone

I'm most grateful to my family, friends, and colleagues who help me dedicate my time in a small way to these tasks.

I'm obliged to the students (kids and grownups alike) and parents at GenuineClass who continually remind me of the joys of teaching and learning about technology. Instructors Russell Dale, Josh Kaden, and Chippy deserve special thanks for helping me do what I do by doing what they do.

I'm grateful to Adam Chng, Titu Sarder, Russell Sarder, and the extended team at Netcom Information Technologies for their support and resources as I juggled writing, teaching, and entrepreneurial angst.

Warmest thanks go to my parents, Eileen and Jim, my siblings and their spouses, and my nieces and nephews, who graciously understood that I wouldn't be using their photos this time for the book. I'm grateful to my extended family—including Irene, Nat, Jared, Bari, and the whole Terry clan—for always lending a collective sympathetic ear.

I owe a debt of thanks to my dear friends Neil deMause, Mich Nelson, Lee Steele, Jonathan Cott, and my sister Kathleen for always letting me vent and then making me laugh.

My sincerest thanks go to Developmental Editor Lisa Chin-Johnson for her eagle-eyed attention and supportive e-mail during the writing and editing stages. I'm also grateful to Sponsoring Editor Don Hull, Project Manager Kristin Puscas, Production Supervisor Gina Hangos, Technical Editor Sal Rodriguez, and the copyediting and design teams for the final good-looking results.

The biggest family-sized bucket of thanks goes to Ross, my husband and true companion. The Kodak moments and major life events we went through together while this book was being written really take my breath away, just like he always does.

—Eileen Mullin

Programming the Web Using XML

Brief Contents

Contents