
USING XML

A How-To-Do-It Manual®
and CD-ROM
for Librarians

KWONG BOR NG

HOW-TO-DO-IT MANUALS

NUMBER 154

NEAL-SCHUMAN PUBLISHERS, INC.
New York London

Published by Neal-Schuman Publishers, Inc.
100 William St., Suite 2004
New York, NY 10038

Copyright © 2007 Neal-Schuman Publishers, Inc.

“A How-To-Do-It Manual®” and “A How-To-Do-It Manual for Librarians®” are registered trademarks of Neal-Schuman Publishers, Inc.

All rights reserved. Reproduction of this book, in whole or in part, without written permission of the publisher, is prohibited.

Printed and bound in the United States of America.

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48-1992.

ISBN-13: 978-1-55570-567-1
ISBN-10: 1-55570-567-7

Library of Congress Cataloging-in-Publication Data

Ng, Kwong Bor.

Using XML : a how-to-do-it manual and CD-ROM for librarians / by Kwong Bor Ng.

p. cm. — (How-to-do-it manuals ; no. 154)

Includes bibliographical references and index.

ISBN 978-1-55570-567-1 (alk. paper)

1. XML (Document markup language). 2. Cataloging—Data processing. 3. Machine-readable bibliographic data formats. 4. Metadata. I. Title.

Z678.93.X54N48 2007

006.7'4—dc22

2007008089

CONTENTS

List of Figures	vii
List of Exercises	ix
Preface	xi
Part I: Introducing XML	1
1. What is XML?	3
Markup Languages	3
From GML and SGML to XML	7
Growth of the XML Family	8
2. XML Applications in Libraries	11
Digital Libraries	11
Encoding Text Collections	11
Library Automation	13
Library Technical Services	13
Part II: Basic XML Techniques	15
3. Creating and Organizing an XML Document	17
Sample XML Document	17
The Prolog and XML Declaration	18
Elements	19
Comments	20
Attributes	22
4. Processing an XML Document: Special Characters, Skipping Data, Entities and Entity References	25
PCDATA	25
CDATA	26
Entities and Entity References	28
A Well-formed XML Document	30

5. Viewing XML Documents: Using Cascading Style Sheets (CSS) . . .	33
Viewing XML Documents without Style	
Instructions	33
How Style Sheets Work	34
Cascading Style Sheet Syntax	37
More CSS Rules	45
Caveat on Using the Official CSS2	
Recommendation	48
Part III: Intermediate XML Techniques	49
6. Assigning Structure Using Document Type Declarations	
and Document Type Definitions	51
Document Type Declarations	51
Internal DTDs	52
External DTDs	55
Subsets	57
Narration Mixed with Metadata	57
7. Structuring an XML Document Using Attribute List	
Declarations and Entity Declarations	65
Attribute List Declarations	65
Attribute Type	67
Attribute Default	71
Entity	72
MARC DTD	77
Validity of an XML Document	78
8. Namespaces and the Limitations of DTD	79
DTDs Revisited	79
The Dublin Core Metadata Element Set	79
Encoding DCMES in XML DTD	80
Using the DCMES DTD to Create a Catalog	
Record	81
Namespaces	83
Limitations of DTDs	86

Part IV: Advanced XML Techniques	87
9. Assigning Structure to an XML Document Using W3C XML Schema Language.....	89
Introduction to Schemas.....	89
A Simple Schema.....	89
Element Type Declarations	93
Attributes in Element Type Declarations.....	93
Define Your Own Data Type	95
Anonymous and Named Types	99
Target Namespaces.....	100
Associating an Instance Document with Its Schema	103
A More Flexible Approach to Schema Creation	105
From DTD to Schema	108
10. Transforming XML Documents Using Extensible Stylesheet Language	115
Components of XSL.....	115
XSLT Templates.....	116
Outputting HTML Codes	119
The Value-of Element and the Select Attribute	121
Context Node and Text Node.....	124
Server-Side XML Parsing	127
Appendix 1: Marc DTD—A Simplified Version	131
Appendix 2: The MARC21 Slim Schema	139
Appendix 3: Glossary—Some Basic XML Terms	165
Appendix 4: Table of Common Numeric References	171
Index	175
About the Author	177

LIST OF FIGURES

Figure 1-1: Markup Languages: Examples	6
Figure 5-1: Display of Example 4-2 (filename “ex04B.xml”) in Mozilla Firefox	34
Figure 5-2: After the minus signs in front of the three record elements are clicked.	35
Figure 5-3: Display of an XML file with CSS style sheet.	37
Figure 5-4: Display of Example 5-3 with CSS style sheet from Example 5-4	44
Figure 5-5: Display of Example 5-3 with CSS style sheet from Example 5-5	47
Figure 10-1: Display of the file “ex10B.xml” in Internet Explorer	118
Figure 10-2: Display of the file “ex10B.xml” in Internet Explorer when there is no match between the template rule and the elements in the document.	119
Figure 10-3: Display of the file “ex10E.xml” in Internet Explorer	121
Figure 10-4: Display of the file “ex10G.xml” in Internet Explorer	123
Figure 10-5: Display of the file “ex10I.xml” in Internet Explorer	125
Figure 10-6: MARC21 SLIM Schema.	157
Figure 10-7: A Bibliographic Catalog Using the MARC21 SLIM Schema	162
Figure 10-8: An Authority Record Using the MARC21 SLIM Schema	164

LIST OF EXERCISES

Example 3-1: A simple XML document	17
Example 3-2: An XML document of two record elements and comments	21
Example 3-3: An XML document with attributes attached to some elements	23
Example 4-1: An XML Document with a CDATA section	27
Example 4-2: An XML Document with a predefined entity reference	29
Example 5-1: A very simple CSS file	35
Example 5-2: An XML Document associated with a style sheet . . .	36
Example 5-3: Andersen's <i>The Little Match-Seller</i>	38
Example 5-4: A very simple CSS file	41
Example 5-5: A CSS file with more rules	46
Example 6-1: An XML document with an internal DTD	52
Example 6-2: An external DTD file.	55
Example 6-3: An XML document that uses external DTD	56
Example 6-4: Andersen's <i>The Little Match Seller</i> , encoded in XML with DTD.	57
Example 6-5: An external DTD file for fairytale.	62
Example 6-6: Using an external DTD for Andersen's <i>The Little Match Seller</i> (only the prolog is shown).	63
Example 7-1: An XML document with attribute list declarations in its DTD	65
Example 7-2: An XML document with ID-type and IDREF-type attributes.	69
Example 7-3: An XML document with entity declaration	73

Example 7-4:	An XML file to be used as an external entity	74
Example 7-5:	An XML document that uses an external entity	75
Example 7-6:	An external DTD file with a parameter entity.	77
Example 8-1:	A DTD file for a simple DCMES record	81
Example 8-2:	A simple Dublin Core record of Andersen’s <i>The Emperor’s New Suit</i>	82
Example 8-3:	A simple Dublin Core record of Andersen’s <i>The Emperor’s New Suit</i> , with namespace specification.	84
Example 9-1:	An instance document	90
Example 9-2:	An XML schema	90
Example 9-3:	An XML schema with namespace prefixes.	91
Example 9-4:	An XML schema with a named type.	100
Example 9-5:	An (incomplete) XML schema with a target namespace	101
Example 9-6:	An XML schema with a target namespace	102
Example 9-7:	An instance XML document conforming to the schema “ex09F.xsd”	104
Example 9-8:	Another approach to creating our schema.	105
Example 9-9:	A schema with more global elements	106
Example 9-10:	An XML document that conforms to Example 9-9.	107
Example 9-11:	An external DTD file for a fairytale	108
Example 9-12:	The fairytale DTD converted to a schema.	109
Example 9-13:	An instance document that conforms to the fairytale schema.	111
Example 10-1:	A very simple XSL file	116
Example 10-2:	Use the simple XSL file	117
Example 10-3:	An XSL file that outputs HTML codes.	120
Example 10-4:	An XSLT style sheet with two templates	121
Example 10-5:	An XSLT style sheet with multiple templates.	124
Example 10-6:	An XSLT style sheet with special output for the ft:speech element	126
Example 10-7:	A simple PHP script	128

PREFACE

The acronym XML stands for eXtensible Markup Language, today's standard method of presenting content on the Web and in other online arenas. This amazingly versatile language has many library applications.

Before opening these pages, you may have already tried to learn XML by following free online tutorials or reading general introductory texts. Though the explanations may make sense in theory, they don't address libraries' specialized needs. Applying the knowledge you glean from simple tutorials to tasks such as creating a catalog record may not work. Why not? Because in the real world, you must take into account library-specific constraints. For instance, the Machine Readable Catalog Record (MARC) format has its own data structure, which nonlibrary sources do not explain. Learning XML is not difficult, but you must start with the proper tools. I designed *Using XML: A How-To-Do-It Manual for Librarians* as an introductory resource specific to libraries.

Once you recognize what XML looks like, you'll quickly see that it's already present in databases, catalogs, indexes, and other tools that you and your patrons consult daily. XML also underlies many interlibrary loan, Web design, and digital library applications. In any of these areas, XML knowledge can make tasks easier. It can handle the special characters and non-Roman scripts often found in bibliographic records. For digital libraries and archival systems, it also offers the greatest promise of data longevity. Technical services librarians, metadata librarians, system librarians, and Webmasters will particularly benefit from XML knowledge.

After reading these pages and completing the exercises, you will have a fundamental education in XML-based resource description and bibliographic data management. Using XML covers general XML syntax; explains schema language, which allows you to impose a predetermined structure on a document; and introduces presentational language for content display and delivery. You will also discover multiple vocabulary schemes and practice applying auxiliary technologies to existing standards.

Step-by-step examples illustrate XML principles. Work all the exercises in order, without skipping, so you don't miss important concepts. You can easily check your work using the CD-ROM, which reproduces the examples.

ORGANIZATION

Part I, “Introducing XML,” covers the basics of the language and its primary uses.

- Chapter 1, “What is XML?” explains the concept of markup languages and XML’s developmental history.
- Chapter 2, “XML Applications in Libraries,” summarizes the most common functions, including text collections encoding, automation, digital libraries, and technical services.

In Part II, “Basic XML Techniques,” tutorials help you develop your own documents.

- Chapter 3, “Creating and Organizing an XML Document,” focuses on the basic document’s building blocks, covering prologs and declaration, elements, naming, comments, and attributes.
- In Chapter 4, “Processing an XML Document,” shows how to create special characters, skip data, define entities, and use entity reference.
- Chapter 5, “Viewing XML Documents: Using Cascading Style Sheets (CSS),” discusses the rules and syntax of CSS, displaying a document on the Web, and when and why to use a style sheet.

Part III, “Intermediate XML Techniques,” explains Document Type Declaration (DTD).

- Chapter 6, “Assigning Structure Using Document Type Declarations and Document Type Definitions,” covers the basics of defining document structure through several exercises.
- Chapter 7, “Structuring an XML Document Using Attribute List Declarations and Entity Declarations,” outlines attribute list declaration and entity declaration; provides a short introduction to MARC DTD (discussed further in Appendices 1 and 2) and defines validity in an XML document.
- Chapter 8, “Namespaces and Limitations of DTD,” identifies ways to correct for the format’s weaknesses. Other

topics include the Dublin Core Metadata Element Set (DCMES), the use of DCMES DTD to create a catalog record, and the legal scope of names.

In Part IV, “Advanced XML Techniques,” you will take your knowledge to the next level using schema language and Extensible Stylesheet Language.

- Chapter 9, “Assigning Structure to an XML Document Using W3C XML Schema Language,” addresses another way of defining document attributes. You will also learn how to convert documents from using DTD to using schema.
- Chapter 10, “Transforming XML Documents Using Extensible Stylesheet Language,” shows how to deliver XML documents through the Web and create a clean and pleasing document display using Internet Explorer.

Three appendices provide additional reference information. Appendices 1 and 2 cover a DTD and a slim schema of MARC. Appendix 3 defines basic XML terms.

HOW TO USE THE CD-ROM

The CD-ROM includes complete copies of all the in-chapter exercises. (The book abbreviates some longer exercises.) If your document does not display properly, compare your work to the CD-ROM example. The CD also holds several documents originally published by W3C. Refer to these if you want more information about a W3C standard or protocol.

Current trends suggest that XML use, already ubiquitous, will continue to expand, making XML knowledge one of today’s most important technical skills for librarians. XML’s inherent flexibility means that it will evolve constantly and new approaches will regularly emerge. Understanding XML fundamentals will help you adapt so your library can use XML to the greatest advantage for many years to come.

