



PROFESSIONAL XML

Covers W3C DOM, SAX, CSS, XSLT, DTDs, XML Schemas, XLink, XPointer, XPath, E-Commerce, BizTalk, B2B, SOAP, WAP, WML

Didier Martin, Mark Birbeck, Michael Kay, Brian Loesgen,
Jon Pinnock, Steven Livingstone, Peter Stark, Kevin Williams,
Richard Anderson, Stephen Mohr, David Baliles, Bruce Peat, Nikola Ozu



Table of Contents

Introduction	1
Chapter 1: Introducing XML	9
Markup Languages	9
Origins and Goals of XML	11
Web Architectures: Past and Future	13
Classic Web Architecture	13
Web Architecture with XML	14
The Nature of XML	15
Self-Describing Data	15
Vocabularies	16
The Main Features of XML Technology	20
Well-Formed XML	21
Document Type Definitions	21
Data Modeling	22
Document Object Model	23
Simple API for XML	23
Namespaces and Schemas	23
Linking and Querying	24
Transforming XML	25
XML and Databases	25
Server to Server	26
eBusiness and XML	26
Styling	27
Wireless Application Protocol and WML	28
An Example Application of XML	28
Summary	29

Chapter 2: XML Syntax	31
Markup Syntax	31
Characters	32
Names	33
Document Parts	34
Elements	35
Start-tags	36
End-tags	36
Empty-Element Tags	37
Tags: A Simple Example	37
Document Element	38
Child Elements	39
Element Nesting	39
Improper Nesting	41
String Literals	42
Character Data	42
Attributes	43
Special Attributes	44
xml:space Attribute	44
xml:lang Attribute	44
White Space	47
End-of-line Handling	48
Character and Entity References	48
Character References	48
Entity References	49
Processing Instructions (PIs)	51
Comments	52
CDATA Sections	52
Document Structure	54
Prolog	54
XML Declaration	54
Document Type Declaration	57
Epilog	59
Summary of XML Syntax	60

Well-formed Documents	61
Parsers	62
Event-driven Parsers	63
Tree-based Parsers	63
Parser Benchmarks	64
The Book Catalog Application	64
The <Book> Element	65
Summary	67

Chapter 3: Document Type Definitions **69**

Why Formal Structure?	70
Document the Domain	70
Validating Documents	71
Writing DTDs: General Principles	71
Associating a DTD with an XML Document	71
The DOCTYPE Tag	71
The Internal DTD Subset	73
External DTDs	73
Basic Markup Declarations	74
Formal DTD Structure	75
Entities	75
Predefined Entities	76
General Entities	76
Parameter Entities	77
Elements	79
Attributes	82
Default Values	83
CDATA	85
ID, IDREF, IDREFS: Relationships within Documents	85
ENTITY, ENTITIES: Replaceable Content	86
NMTOKEN, NMTOKENS: Name Tokens	88
NOTATION: Data Other than XML	88
Enumerations: Choices	89
Conditional Sections	90
The Shortcomings of DTDs	91

A DTD for the Book Catalog Domain	92
Formal Definition of the Book Catalog Domain	93
The Domain Model	93
Books	94
Publishers	95
Threads	96
The Publishing Catalog DTD	96
Solving the Relationship Problem	102
Not Quite There Yet	102
Summary	103

Chapter 4: Data Modeling and XML **105**

Information Modeling	105
Static and Dynamic Models	107
Documents and Data	107
Where Do You Start?	108
The Static Information Model	109
Step 1: Naming Things	109
Step 2: Taxonomy	110
Step 3: Finding Relationships	111
Step 4: Defining Properties	114
Dynamic Modeling: What Happens to the Data?	115
Process and Workflow Models	115
Data Flow Models	115
Object Models	116
Object Life Histories	116
Use Cases	116
Object Interaction Diagrams	117
Choosing a Dynamic Modeling approach	117
Designing XML Documents	117
Two Roles for XML	117
XML for Messages	118
XML For Persistent Data	119
Mapping the Information Model to XML	121
Representing Object Types	121
Representing Relationships	122

Representing Properties	124
Elements or Attributes?	124
Encoding Property Values	127
Using XML Entities	131
Schema Languages and Notations	132
The Role of a Schema	132
Why Do We Need Schemas?	132
The Schema as a Set of Constraints	132
The Schema as an Explanation	133
The DTD as a Schema	134
What Constraints can you Express in a DTD?	134
Can you Rely on DTD Validation?	134
Writing DTDs	135
DTDs and Namespaces	135
Deriving a DTD from the Information Model	135
Defining the Content of Elements	136
Elements with EMPTY Content	136
Elements with ANY Content	137
Elements with Element Content	137
Elements with Mixed Content	137
Elements with PCDATA Content	138
Modeling the Object Type Hierarchy	139
When Several Properties Have the Same Data Type	139
Defining Attributes in the DTD	139
Choosing an Attribute Type	140
The XML Schema proposal	140
Summary	142

Chapter 5: The Document Object Model **145**

What is the Document Object Model?	145
XML Document Structure	146
Why use the DOM?	148
The DOM Ensures Proper Grammar and Well-Formedness	148
The DOM Abstracts Content Away from Grammar	148
The DOM Simplifies Internal Document Manipulation	148
The DOM Closely Mirrors Typical Hierarchical and Relational Database Structures	148
The DOM specification	149

DOM Level 1	149
DOM Level 2	150
Understanding IDL and Bindings	151
The DOM in the real world	151
The DOM in Current Browsers	151
Grace and the Art of Degradation	152
The HTML DOM as a Special Case of the XML DOM	152
HTML Isn't XML	152
Problems with the HTML DOM	153
The Future of the Internet: XHTML	153
Working with the DOM	155
DOM APIs	155
What is an API?	156
The DOM is not an API	156
Programming for XML Data Structures	156
Client Side and Server Side	156
DOM on the Server Side	157
Document Interchange	157
Archiving	157
DOM on the Client Side	158
Flexible Client Rendering	158
Client Data Entry	158
Scenarios for Using the DOM in the Publishing Process	158
The DOM and Databases	158
Using the DOM to Create Complex XML Documents	160
Sample Application Using DOM and XML	161
A Simple Client Side Example	162
Using the DOM to Let the User Modify the Page	163
A More Complex Programmatic Example	170
Creating XML Documents from Text Using the DOM	171
Modifying Information Using the DOM	178
Deleting Elements Using the DOM	179
Generation of an XML Stream, Revisited	181
Future of the DOM and XML	182
Work by the W3C	182
Applications	182
Databases, Schemas, and the DOM	182
Summary	183

An Event-Based Interface	186
Why Use an Event-Based Interface?	187
The Benefits of SAX	187
The Drawbacks of SAX	188
The Origins of SAX	189
The Structure of SAX	190
The Basic Structure	190
A Simple Example	191
The DocumentHandler Interface	195
Document Events	195
Element Events	195
Character Data	196
Processing Instructions	197
Error Handling	198
Handling XML errors	198
Application-Detected Errors	199
Identifying Where the Error Occurred	199
Another Example: Using Character Data and Attributes	200
Maintaining Context	205
Advanced SAX features	208
Alternative Input Sources	208
Taking Input from a Byte Stream or Character Stream	209
Specifying a Filename rather than a URL	210
Input from Non-XML Sources	211
Handling External Entities	211
Unparsed Entities and Notations	213
Choosing a Parser	213
Some SAX Design Patterns	215
The Filter Design Pattern	215
The ParserFilter class	216
An Example ParserFilter: an Indenter	216
The End of the Pipeline: Generating XML	220
Other ParserFilters	224
The Rule-Based Design Pattern	225
SAX 2.0	232
The Configurable Interface	232

Core Features and Properties	234
Summary	235

Chapter 7: Namespaces and Schemas **237**

Mixing Vocabularies	239
Segmenting a Problem	239
Reuse	240
Ambiguity and Name Collision	240
Namespaces	242
Using and Declaring Namespaces	243
Declaring a Namespace	243
Qualified Names	244
Scope	245
Default	245
Qualified	246
Using Namespaces in a Well-Formed Book Example	247
Schemas	248
The Problem with DTDs	249
An Abundance of Help Creating Schemas	251
Resource Description Framework	252
XML Data	255
Document Content Description	257
Finding the Right Balance	258
W3C Work on XML Schemas	258
DTD vs. XML Schema: A Contrast	259
Structures	260
Writing Schemas	260
Deriving Type Definitions	267
Composition	268
Annotating Schemas	268
Datatypes	269
Primitive Types	270
Generated and User Defined Types	271
XML Data – Reduced	272
What is XML Data – Reduced?	273

Schema Support	273
Schemas	273
Elements and Attributes	274
Groups	277
Comments	277
Data Types	278
Namespaces and Schema Support in MSXML	279
Namespaces	279
Validation	280
Datatypes	280
Variations on the Book Catalog	281
Why Bother?	281
Segmentation	281
Additional Expression	282
Metadata Discovery	282
Recasting the DTD	282
Author Schema	283
Catalog Schema	284
Publisher	284
Thread	286
Book	286
Schema Concordance	288
Finding the Elements	288
Processing an Element Declaration	289
Displaying the Results	291
Summary	292

Chapter 8: Linking and Querying 295

XML Information Set	296
Information Types	296
Importance of the Information Set	303
Summary	306
Linking	306
What is Linking?	306
The Distinction Between Conceptual Linking and Rendering	306
The Problem with HTML Links	306
The W3C Specification: XLink	307

XLink Declarations	308
Link Types	308
Simple links	309
Simple Link Examples	310
Extended Links	311
Inline Extended Links vs. Out-Of-Line Extended Links	316
Out-Of-Line Extended Links – Using Linkbases	316
Some Examples of Extended XLinks	318
Link Databases	318
Marking Up Read-Only Documents	321
XLink Summary	322

XPointer	322
----------	-----

HTML Pointers	322
---------------	-----

The Problem With HTML Pointers	323
--------------------------------	-----

The XPointer Specification	323
----------------------------	-----

Specifying XPointers in a URI	324
-------------------------------	-----

How May Fragment Identifiers be Specified?	324
--	-----

The W3C XPath Recommendation	326
------------------------------	-----

Location Steps	326
----------------	-----

Axis	326
------	-----

Node Tests	327
------------	-----

Predicates	327
------------	-----

Putting It All Together	331
-------------------------	-----

Abbreviations	332
---------------	-----

XPointer Extensions to XPath	333
------------------------------	-----

Points	333
--------	-----

Ranges	333
--------	-----

Additional Functions	333
----------------------	-----

XPointer Errors	334
-----------------	-----

Syntax Error	334
--------------	-----

Resource Error	334
----------------	-----

Sub-Resource Error	334
--------------------	-----

Summary	334
---------	-----

XML Fragment Interchange	335
--------------------------	-----

What are Document Fragments?	335
------------------------------	-----

Some uses for document fragments	336
----------------------------------	-----

The Problem: Bare Document Fragments Aren't Always Enough	338
---	-----

The Solution: Context Information	340
What is Context Information?	340
What is Allowed to be Part of Context Information?	341
How are Fragments Represented?	341
Examples, Revisited	342
What Do I Describe?	342
Using IDREF and IDREFS	343
Validating Processors	344
How Fragments May Be Transmitted	344
The Separate File Mechanism	344
The Proposed Package Mechanism	345
Summary	346
Querying	347
What Is a Query Language?	347
The Differences Between Relational Databases and XML Documents	349
The History of Query Languages for XML	353
XML-QL	353
XQL	354
XSLT and XPath	356
Using XPath and XSLT to Query XML Documents	356
Testing XSLT and XPath Queries With XT	356
The Future of Query Languages	367
Summary	367

Chapter 9: Transforming XML **369**

Why Transform XML?	370
Translating Between Different Vocabularies	371
Dynamic Transformations	372
Different Browsers	372
XSL	373
XSLT	373
How Does the XSL Processor Transform the Source Document?	375
Using XSLT	378
XSLT Style Sheet Structure	386

Creating Templates	387
Dealing with White Space	387
Output Format	387
Combining Style Sheets	388
Embedding Style Sheets	389
Examples of Using XSLT	390
Structural Transformations	390
Repetition	397
Sorting	398
Conditional Processing	400
Named templates	403
Numbering	404
Copying	407
Transformation of an XML Document with the DOM	408
Structural Transformations with the DOM	408
Modifying an XSLT Document at Runtime	412
Comparing XSL Transformations with DOM Transformations	417
Summary	417

Chapter 10: XML and Databases **421**

Storing XML	422
The Problem of Persistence	422
Limitations of File Systems	422
Reposing the Problem	424
Summary	427
Document Management	428
RDBMS Solutions	428
OODBMS Solutions	428
Conclusion	429
XML Storage and Databases	430
Object-oriented Databases	430
Relational Databases	435
Conclusion	446
XML for Interchange	447

Standard for Data	447
Standard for Querying	450
Querying XML vs Querying the Database	451
XML Query Languages	452
Optimization	457
Conclusion	459
Standards You Might Need	459
Fragment Interchange	459
SOAP and XML-RPC	460
Book Catalog Sample	460
Handling Existing Data and Applications	460
The Book Catalog Schema	466
Automating Database Creation	483
Conclusion	494

Chapter 11: Server to Server 497

Transporting XML	497
1. XML Over FTP	498
2. XML and Message Queuing	499
3. XML Over HTTP	500
4. XML Over SMTP	500
XML as a Distributed Component Model	500
Message Transfer and Serialization	500
Serialization With XML-RPC and SOAP	502
Coins (Combining XML & Java)	502
Serialization With WDDX	502
Serializing With XMOP	504
KOALA Serialization	505
Tightly Coupled Vs Loosely Coupled Systems	505
Methods of Communication	507
Remote Calls	508
Data Exchange	508
XML-RPC	508
Why Use XML-RPC?	509
Where Does XML-RPC Fit In?	511
The XML-RPC Specification – A Technical Overview	512

Header Requirements	512
Request Payload Format	512
Specifying Parameters	513
Response Header	517
Response Format	517
Response Faults	518
Implementing XML-RPC	518
Sample Book Application	522
Where to go from here	542
ASP	542
COM	542
Java	543
SOAP	543
XML-RPC++	544
Data	545
Invocation	552
Implementation	557
Module Structure	557
Conclusion	572
WebDAV	573
Summary	574

Chapter 12: eBusiness and XML **577**

What is E-Commerce?	578
Selling Direct to the Customer	579
Business to Business Transactions	580
Information Sharing and Content Syndication	580
EDI – Electronic Data Interchange	581
Applying XML in eBusiness	587
Message Headers	588
Message Body	593
Common Misconceptions	595
Perceptions Surrounding the Use of XML in eBusiness	598
Looking to the Future	601
Discovery & Mediation	602

Repositories	605
Agents	608
Templates	609
Wrapping Up The Theory	610
eBusiness Solutions	610
Industry Solutions	611
Key Vertical Industry Solutions	611
Travel	612
Human Resources	614
HR-XML	614
Healthcare	623
cXML – A Detailed Example of Vertical Industries	626
Why cXML?	627
cXML Protocol Specification	627
Message Transfer	632
First Steps – Horizontal Industries	638
ASC X12	638
XML-EDIFACT	640
Electronic Business XML Working Group (ebXML)	641
First Horizontal Steps	642
The BizTalk Tag Specification	643
Goals and Current Constraints of the Specification	643
Document Structure	644
The BizTalk Document Root	645
The Document Header	646
The Document Body	650
A BizTalk Document for Wrox Books	650
Implementing BizTalk Solutions	654
The BizTalk JumpStart Kit Overview	655
BizTalk Server	656
BizTalk Editor	656
BizTalk Mapper	657
Configuring Workflow	658
Summary	659
Helpful Links:	660

Rendition location	664
Server side XML transformation	664
Client side XML transformation	665
Rendition Models	665
Visual Rendition	665
Aural Rendition	667
How to Associate a Style Sheet with an XML Document	668
Rule-Based Languages	668
CSS	669
The CSS Atom: the Box	669
How an XML Document's Elements are Linked to CSS Rules	670
CSS Visual Model	671
CSS Rule structure	674
The Pattern Match Selector	674
Media Types and Modular Style Sheets	677
A Sample CSS Style Script	678
The display Property	681
The position Property	684
Z-Order	687
The float Property	688
Interactive Behavior	690
Concluding Thoughts on CSS	692
XSL	692
XSL Transformations	693
XSL Formatting Objects	693
How XSLF Works	693
Rendering XML using XSL	694
A Sample XSLT Style Sheet with HTML Visual Rendition Objects	695
A Sample XSLT Style Sheet with VOXML Aural Rendition Objects	699
A Sample XSLT Style Sheet with Formatting Objects	704
Concluding Thoughts on XSL	711
DSSSL	712
A Sample Style Script	713

Concluding Thoughts on DSSSL	715
Omnimark	715
A Simple Omnimark Script	715
Concluding Thoughts on Omnimark	716
Summary	717

Chapter 14: Wireless Application Protocol 719

Introducing the New Client	720
Building Sites from The Ground up	721
Getting Data to a Phone	722
WAP to the Rescue	723
Understanding the Wireless Environment	724
Services	724
Wireless Networks	726
How WAP Solves Problems with Wireless Networks	728
Introducing WML	730
How to Get the First Document to the Phone	731
The Structure of a WML Document	731
Common Attributes	734
What's Inside WML?	734
Meta Information	735
Basic Text, Tables, and Presentation	736
Navigation	738
Using Timers	738
Event Handlers	739
Using Variables	740
Tasks and Menus	743
Client-side Templates	748
Form Data	749
Images	752
Generating WML on the Server	753
The WROX WML Application	755
WML Script	759
Basics	759

Byte Code	759
How to Call Script Functions from WML	760
Libraries	760
Manipulating Browser's State – The WMLBrowser Library	761
Using Dialogs	762
Language Specific Functions	762
Floating Point Numbers	763
Strings	764
URLs	765
Where to Get More Information	768
Summary	768

Chapter 15: Case Study 1 – Data Duality **771**

Business Need	772
System Requirements	772
Design Time: Let the Games Begin	772
Implementation Overview	773
Setting Up the Database	775
XML-providing ASP	775
Client-side Page	779
The "view-by-number" XSL Style Sheet	787
The "view by name" XSL Style Sheet	789
Invoking the XSL Style Sheets	790
Problem: Keeping the Tree in Sync	791
Summary	794

Chapter 16: Case Study 2 – XML and Distributed Applications **797**

Where Are Our Present Weaknesses?	798
Five Principles for Building Networked Applications	799
The Business Case	806
Application Design	806

Implementation	812
Entering and Editing Programmers	815
Complete Client Code for Programmer Insertion	820
Complete Server Code for Programmer Insertion	820
Searching for Staff Programmers	823
Clearing the Programmer Form	826
Entering Programmer Performance Reports	827
Clearing the Performance Report History Form	829
Retrieving a Programmer's Performance History	830
Lessons Learned	833
Summary	835

Chapter 17: Case Study 3 – Book Catalog Information Service 837

The Book Catalog Information Service	838
System Overview	839
XML File Formats	840
Subscribers/Users	842
Visual Basic SAX Parser	848
The Transformation Engine Details	850
Creating the BCIS	851
SAX XPath Filter Interface – ITemplate	851
Pattern Matching	855
Load Subscriber Information from Users.XML	867
The XML to XML Transformation	870
The XML to HTML Transformation	878
Scan Directory Functionality	880
The Direct Mail Component	881
Summary	884

Chapter 18: Case Study 4 – SOAP 887

To the Ends of the Earth	888
A SOAP Opera	892

The OPERA client	894
Defining our Interface	894
Building the Client	894
The Exchange() Method	897
The Helper Methods	900
The OPERA server	903
Creating the Project	903
The Main Socket Class	903
The Client Socket Class	906
The Main Dialog	912
A Quick Test	916
The Order Entry Application	916
Objects? What Objects?	920
Summary	934

Appendix A: Extensible Markup Language (XML)	
1.0 Specification	937
Appendix B: IE5 XML Document Object Model	983
Appendix C: SAX 1.0: The Simple API for XML	1041
Appendix D: XML Schemas and Data Types	1073
Appendix E: IE5 XSL Reference	1085
Appendix F: CSS Properties	1101
Appendix G: Installing XT	1117
Appendix H: Support and Errata	1121
Index	1129