Contents

Preface vii

Acknowledgments xv

Part I: Introducing XML 1

Chapter 1: An Eagle's Eye View of XML 3

What Is XML? 3
XML is a meta-markup language 3
XML describes structure and semantics, not formatting 5
Why Are Developers Excited About XML? 6
Domain-specific markup languages 6
Self-describing data 7
Interchange of data among applications 8
Structured data 8
The Life of an XML Document 9
Editors 9
Parsers and processors 10
Browsers and other applications 10
The process summarized 10
Related Technologies 11
HTML 11
CSS 12
XSL 12
URLs and URIs 13
XLinks and XPointers 14
Unicode 15
Putting the pieces together 15
Summary 16

Chapter 2: XML Applications 17

What Is an XML Application? 17
Chemical Markup Language 18
Mathematical Markup Language 19
RSS 22
Classic literature 25
Synchronized Multimedia Integration Language 26

Open Software Description . . . 26
Scalable Vector Graphics . . . 27
MusicXML . . . 29
VoiceXML . . . 34
Open Financial Exchange . . . 36
Extensible Forms Description Language . . . 36
HR-XML . . . 40
XML for XML . . . 43
XSL . . . 43
XLinks . . . 44
Schemas . . . 45
Behind-the-Scene Uses of XML . . . 46
Microsoft Office 2003 . . . 46
Netscape's What's Related . . . 49
UPS . . . 52
Summary . . . 54

Chapter 3: Your First XML Document . . . 55

Hello XML . . . 55
Creating a simple XML document . . . 56
Saving the XML file . . . 56
Loading the XML file into a web browser . . . 57
Exploring the Simple XML Document . . . 58
Meaning in Markup . . . 59
Writing a Style Sheet for an XML Document . . . 60
Attaching a Style Sheet to an XML Document . . . 61
Summary . . . 62

Chapter 4: Structuring Data . . . 63

Examining the Data . . . 64
XMLizing the Data . . . 67
The Advantages of the XML Format . . . 87
Preparing a Style Sheet for Document Display . . . 88
Linking to a style sheet . . . 89
Assigning style rules to the root element . . . 90
Assigning style rules to titles . . . 91
The complete style sheet . . . 98
Summary . . . 101

Chapter 5: Attributes, Empty-Element Tags, and XSL . . . 103

Attributes . . . 103
Attributes versus Elements . . . 109
Structured metadata . . . 109
Meta-metadata . . . 113

What's your metadata is someone else's data . . . 113
Elements are more extensible . . . 114
Good times to use attributes . . . 114
Empty Elements and Empty-Element Tags . . . 116
XSL . . . 119
Templates . . . 120
The body of the document . . . 121
The title . . . 123
Stations . . . 126
Shows . . . 129
Sorting . . . 134
Tables . . . 137
CSS or XSL? . . . 143
Summary . . . 144

Chapter 6: Well-formedness . . . 145

Well-formedness Rules . . . 145
XML Documents . . . 146
The XML declaration . . . 147
Single root element . . . 148
Text in XML . . . 148
Elements and Tags . . . 149
Element names . . . 149
Every start-tag must have a corresponding end-tag . . . 150
Empty-element tags . . . 151
Elements may nest but may not overlap . . . 152
Attributes . . . 154
Attribute names . . . 154
Attribute values . . . 154
Predefined attributes . . . 155
Entity References . . . 160
Comments . . . 162
Processing Instructions . . . 164
CDATA Sections . . . 165
Unicode . . . 167
Character encodings . . . 168
The encoding declaration . . . 168
Numeric character references . . . 169
XML 1.1 . . . 171
Well-formed HTML . . . 173
Rules for HTML . . . 173
Tools . . . 183
Summary . . . 185

Part II: Document Type Definitions 187

Chapter 7: Validity 189

Document Type Definitions 189
Element Declarations 190
DTD Files 192
Document Type Declarations 192
Internal DTDs 194
Internal and external DTD subsets 195
Public DTDs 196
DTDs and style sheets 197
Validating against a DTD 199
Command-line validators 200
Web-based validators 201
Summary 203

Chapter 8: Element Declarations 205

Analyzing the Document 205
ANY 208
#PCDATA 211
Child Elements 213
+ One or More Children 215
? Zero or One Child 215
* Zero or More Children 216
Choices 216
Parentheses 217
Mixed Content 220
Empty Elements 221
Comments in DTDs 223
Summary 227

Chapter 9: Attribute Declarations 229

What Is an Attribute? 229
Declaring Attributes 230
Declaring Multiple Attributes 231
Alternatives to Default Attribute Values 232
#REQUIRED 232
#IMPLIED 233
#FIXED 234
Attribute Types 235
The CDATA attribute type 235
The NMTOKEN attribute type 236

The NMTOKENS attribute type . . . 236
The enumerated attribute type . . . 237
The ID attribute type . . . 238
The IDREF attribute type . . . 238
The IDREFS attribute type . . . 239
The ENTITY attribute type . . . 240
The ENTITIES attribute type . . . 241
The NOTATION attribute type . . . 241
A DTD for Attribute-Based Television Listings . . . 242
Declaring SCHEDULE attributes . . . 243
Declaring STATION attributes . . . 243
Declaring SHOW attributes . . . 244
Declaring person attributes . . . 245
The complete DTD for the television listings example . . . 246
Summary . . . 248

Chapter 10: Entity Declarations . . . 249

What Is an Entity? . . . 249
Internal General Entities . . . 250
Defining an internal general entity reference . . . 251
Using general entity references in the DTD . . . 254
Predefined general entity references . . . 255
External General Entities . . . 256
Text declarations . . . 258
Nonvalidating parsers . . . 260
Internal Parameter Entities . . . 260
External Parameter Entities . . . 263
Building a Document from Pieces . . . 269
Non-XML Data . . . 274
Notations . . . 275
Unparsed entities . . . 278
Conditional Sections . . . 283
Summary . . . 284

Chapter 11: Namespaces . . . 287

The Need for Namespaces . . . 287
Namespace Syntax . . . 289
Defining namespaces with xmlns attributes . . . 291
Multiple namespaces . . . 294
Attributes . . . 298
Default namespaces . . . 300
Namespaces and Validity . . . 304
Summary . . . 307

Part III: Style Languages 309

Chapter 12: CSS Style Sheets 311

What Are Cascading Style Sheets? 311
A simple CSS style sheet 312
Comments 312
Attaching style sheets to documents 313
DTDs and style sheets 316
CSS1 versus CSS2 316
CSS3 317
Selecting Elements 317
The universal selector 320
Grouping selectors 320
Hierarchy selectors 321
Attribute selectors 323
ID selectors 324
Pseudo-elements 324
Pseudo-classes 326
Inheritance 328
Cascades 329
Different Rules for Different Media 330
Importing Style Sheets 332
Character Sets 333
Summary 333

Chapter 13: CSS Layouts 335

CSS Units 335
Length values 337
URL values 339
Color values 340
System colors 341
Keyword values 343
Strings 343
The Display Property 343
Inline elements 347
Block elements 348
None 348
Compact and run-in elements 350
Marker 350
Tables 350
List items 351
Box Properties 356
Margin properties 356
Border properties 358
Padding properties 363

Size 364
The width and height properties 365
The min-width and min-height properties 367
The max-width and max-height properties 367
The overflow property 368
Positioning 369
The position property 369
Stacking elements with the z-index property 373
The float property 373
The clear property 374
Formatting Pages 375
@page 375
The size property 376
The margin property 376
The mark property 377
The page property 377
Controlling page breaks 377
Widows and orphans 379
Summary 379

Chapter 14: CSS Text Styles 381

Fonts 381
Choosing the font family 382
Choosing the font style 384
Small caps 385
Setting the font weight 385
Setting the font size 386
The font shorthand property 390
Color 392
Text 392
Word spacing 393
The letter-spacing property 394
The text-decoration property 395
The vertical-align property 396
The text-transform property 397
The text-align property 398
The text-indent property 399
The line-height property 400
The white-space property 402
Backgrounds 404
The background-color property 404
The background-image property 405
The background-repeat property 406
The background-attachment property 408
The background-position property 409
The background shorthand property 413

Visibility . . . 413
Content . . . 414
Quotes . . . 415
Attributes . . . 416
URIs . . . 416
Counters . . . 417
Summary . . . 421

Chapter 15: XSL Transformations . . . 423

What Is XSL? . . . 423
Overview of XSL Transformations . . . 424
Trees . . . 424
XSLT style sheet documents . . . 428
Where does the XML transformation happen? . . . 430
Using xsltproc . . . 430
Browser display of XML files with XSLT style sheets . . . 433
XSL Templates . . . 434
The xsl:apply-templates element . . . 435
The select attribute . . . 437
Computing the Value of a Node with xsl:value-of . . . 437
Processing Multiple Elements with xsl:for-each . . . 439
Patterns for Matching Nodes . . . 440
Matching the root node . . . 440
Matching element names . . . 442
Wildcards . . . 444
Matching children with / . . . 445
Matching descendants with // . . . 446
Matching by ID . . . 447
Matching attributes with @ . . . 447
Matching comments with comment() . . . 449
Matching processing instructions with processing-instruction() . . . 450
Matching text nodes with text() . . . 451
Using the or operator | . . . 451
Testing with [ ] . . . 452
XPath Expressions for Selecting Nodes . . . 454
Node axes . . . 455
Expression types . . . 461
The Default Template Rules . . . 471
The default rule for elements . . . 471
The default rule for text nodes and attributes . . . 472
The default rule for processing instructions and comments . . . 472
Implications of the default rules . . . 473
Attribute Value Templates . . . 473
Deciding What Output to Include . . . 475
Inserting elements into the output with xsl:element . . . 475
Inserting attributes into the output with xsl:attribute . . . 476
Defining attribute sets . . . 477

Generating processing instructions with xsl:processing-instruction . . . 478
Generating comments with xsl:comment . . . 479
Generating text with xsl:text . . . 479
Copying the Context Node with xsl:copy . . . 480
Counting Nodes with xsl:number . . . 482
Default numbers . . . 483
Number to string conversion . . . 485
Sorting Output . . . 486
Modes . . . 490
Defining Constants with xsl:variable . . . 492
Named Templates . . . 493
Passing Parameters to Templates . . . 495
Stripping and Preserving White Space . . . 496
Making Choices . . . 497
xsl:if . . . 498
xsl:choose . . . 498
Merging Multiple Style Sheets . . . 499
Importing with xsl:import . . . 499
Inclusion with xsl:include . . . 500
Output Methods . . . 500
XML Declaration . . . 501
Document Type Declaration . . . 502
Indentation . . . 503
CDATA sections . . . 504
Summary . . . 504

Chapter 16: XSL Formatting Objects . . . 507

Formatting Objects and Their Properties . . . 507
Formatting properties . . . 510
Transforming to formatting objects . . . 514
Using FOP . . . 516
Page Layout . . . 518
The root element . . . 518
Simple page masters . . . 518
Regions . . . 520
Page sequences . . . 522
Page sequence masters . . . 530
Content . . . 533
Block-level formatting objects . . . 534
Inline formatting objects . . . 534
Table formatting objects . . . 535
Out-of-line formatting objects . . . 536
Leaders and Rules . . . 536
Graphics . . . 539
fo:external-graphic . . . 539
fo:instream-foreign-object . . . 541
Graphic properties . . . 544

Links . . . 545
Lists . . . 547
Tables . . . 551
Inlines . . . 556
Footnotes . . . 557
Floats . . . 558
Formatting Properties . . . 559
The id property . . . 559
The language property . . . 559
Paragraph properties . . . 560
Character properties . . . 563
Sentence properties . . . 565
Area properties . . . 567
Summary . . . 575

Part IV: Supplemental Technologies 577

Chapter 17: XLinks . . . 579

XLinks versus HTML Links . . . 579
Linking Elements . . . 580
Declaring XLink Attributes in Document Type Definitions . . . 582
Descriptions of the Remote Resource . . . 584
Link Behavior . . . 585
The xlink:show attribute . . . 585
The xlink:actuate attribute . . . 587
Extended Links . . . 589
Extended Link Syntax . . . 590
Arcs . . . 595
Out-of-Line Links . . . 601
XML Base . . . 607
Summary . . . 614

Chapter 18: XPointers . . . 617

Why XPointers? . . . 617
XPointer Examples . . . 618
A Concrete Example . . . 621
Location Paths, Steps, and Sets . . . 623
The Root Node . . . 625
Axes . . . 626
The child axis . . . 627
The descendant axis . . . 628
The descendant-or-self axis . . . 628
The parent axis . . . 628
The self axis . . . 628
The ancestor axis . . . 629

The ancestor-or-self axis . . . 629
The preceding axis . . . 629
The following axis . . . 629
The preceding-sibling axis . . . 630
The following-sibling axis . . . 630
The attribute axis . . . 630
The namespace axis . . . 630
Node Tests . . . 631
Predicates . . . 633
Functions That Return Node-Sets . . . 635
id() . . . 636
here() . . . 637
origin() . . . 638
Points . . . 638
Ranges . . . 640
Range functions . . . 640
String ranges . . . 641
Child Sequences . . . 643
Summary . . . 644

Chapter 19: XInclude . . . 647

Use Cases for XInclude . . . 648
Non-Solutions . . . 651
DTDs . . . 651
Embedded XLinks . . . 652
Server-side includes . . . 653
The xinclude:include Element . . . 655
Validating Documents That Use XInclude . . . 658
XPointers in XInclude . . . 660
Unparsed Text . . . 662
Fallbacks . . . 663
Summary . . . 665

Chapter 20: Schemas . . . 667

What's Wrong with DTDs? . . . 667
What Is a Schema? . . . 669
The W3C XML Schema Language . . . 671
Hello Schemas . . . 671
The greeting schema . . . 672
Validating the document against the schema . . . 673
Complex Types . . . 675
minOccurs and maxOccurs . . . 677
Element content . . . 680
Sharing content models . . . 682
Anonymous types . . . 683
Mixed content . . . 685

Grouping . . . 686
The xsd:all Group . . . 688
Choices . . . 689
Sequences . . . 690
Simple Types . . . 690
Numeric data types . . . 692
Time data types . . . 695
XML data types . . . 696
String data types . . . 697
Binary types . . . 698
Miscellaneous data types . . . 699
Deriving Simple Types . . . 700
Deriving by restriction . . . 700
Facets . . . 702
Facets for strings: length, minLength, maxLength . . . 702
The whiteSpace facet . . . 704
Facets for decimal numbers: totalDigits and fractionDigits . . . 705
The enumeration facet . . . 705
The pattern facet . . . 707
Unions . . . 714
Lists . . . 714
Empty Elements . . . 715
Attributes . . . 716
Namespaces . . . 720
Schemas for default namespaces . . . 721
Multiple namespaces, multiple schemas . . . 724
Annotations . . . 727
Summary . . . 728

Part V: XML Applications 731

Chapter 21: XHTML . . . 733

Why Validate HTML? . . . 733
Moving to XHTML . . . 735
Making the document well-formed XML . . . 738
Making the document valid . . . 745
The strict DTD . . . 753
The frameset DTD . . . 767
HTML Tidy . . . 767
Setting the MIME media type . . . 771
What's New in XHTML . . . 772
Character references . . . 772
Custom entity references defined in DTD . . . 776
Encoding declarations . . . 781
The xml:lang attribute . . . 782
CDATA sections . . . 783
Summary . . . 785

Chapter 22: Modular XHTML 787

The Modules of XHTML 787
A Sample DTD Module 790
Element names 793
Element-specific content models 794
Generic content models 795
Generic attribute models 796
INCLUDE and IGNORE blocks 796
Using XHTML entities in other applications 797
The Framework 798
The notations framework module 802
The data types framework module 804
The namespace-qualified names module 805
The common attributes module 807
The character entity modules 808
The Driver DTD 809
The Document Model 816
The XHTML Basic document model 817
A minimal document model 823
A Sample Schema Module 828
Summary 831

Chapter 23: The Resource Directory Description Language 833

What Does a Namespace URL Locate? 834
The Solution 835
The resource Element 836
Natures 843
Purposes 845
Summary 847

Chapter 24: Scalable Vector Graphics 849

What Is SVG? 850
Scalability 851
Vector versus bitmapped graphics 852
A Simple SVG Document 852
Embedding SVG Pictures in Web Pages 855
Simple Shapes 858
The rect element 859
The circle element 861
The ellipse element 862
The line element 863
Polygons and polylines 864
Paths 866
Arcs 869
Curves 872

Text . . . 873
Strings . . . 874
Text on a path . . . 876
Fonts and text styles . . . 878
Text spans . . . 879
Bitmapped Images . . . 879
Coordinate Systems and Viewports . . . 881
The viewport . . . 882
Coordinate systems . . . 883
Grouping Shapes . . . 888
Referencing Shapes . . . 889
Transformations . . . 892
Linking . . . 899
Metadata . . . 900
SVG Editors . . . 903
Summary . . . 904

Chapter 25: Designing a New XML Application . . . 907

Organization of the Data . . . 907
Listing the elements . . . 909
Identifying the fundamental elements . . . 909
Establishing relationships among the elements . . . 912
Choosing a Namespace . . . 914
Persons . . . 915
A sample person . . . 915
The person DTD . . . 918
The person schema . . . 924
Families . . . 932
The family DTD . . . 933
The Family Schema . . . 935
Sources . . . 937
The Family Tree . . . 940
The Family Tree DTD . . . 946
The family tree schema . . . 948
Modularizing the DTDs . . . 950
Designing a Style Sheet for Family Trees . . . 960
A RDDL document for family trees . . . 968
Summary . . . 980

Index . . . 983