



XPath™ 2.0

Programmer's Reference

Michael Kay



Updates, source code, and Wrox technical support at www.wrox.com

Contents

Acknowledgments	xiv
Introduction	xvi

Chapter 1: XPath 2.0 in Context **1**

What Is XPath?	1
Path Expressions	3
Composability	4
What's New in XPath 2.0?	4
XPath 2.0 Processors	9
Where XPath fits in the XML Family	15
XPath and XSLT	15
XPath and the InfoSet	16
XML Namespaces	17
XPath and XPointer	18
XPath and XQuery	19
XPath and XML Schemas	20
XPath, the DOM, and Java	21
XPath 2.0 as a Language	22
The Syntax of XPath	22
An Embedded Language	23
A Language for Processing Sequences	24
Types Based on XML Schema	25
Summary	26

Chapter 2: The Data Model **27**

Changes in 2.0	27
Sequences	28
Atomic Values	30
Nodes and Trees	33
XML as a Tree	33
Completing the UML Class Diagram	45
Names and Namespaces	45
IDs	50

Contents

Characters in the Data Model	52
What Does the Tree Leave Out?	53
From Textual XML to a Data Model	56
Controlling Serialization	57
Document Order	59
Summary	60
Chapter 3: The Type System	61
What Is a Type System?	61
XML Schema: An Overview	63
Simple Type Definitions	63
Elements with Attributes and Simple Content	65
Elements with Mixed Content	66
Elements with Element-Only Content	67
Defining a Type Hierarchy	69
Substitution Groups	73
Atomic Types	76
The Major Atomic Types	78
The Minor Atomic Types	93
Derived Numeric Types	97
Derived String Types	100
Untyped Atomic Values	103
NMTOKENS, IDREFS, and ENTITIES	104
Schema Types and XPath Types	105
The Type Matching Rules	107
Static and Dynamic Type Checking	109
Summary	113
Chapter 4: The Evaluation Context	115
The Static Context	115
XPath 1.0 Compatibility Mode	116
In-Scope Namespaces	117
Default Namespaces	118
In-Scope Schema Definitions	119
In-Scope Variables	121
In-Scope Functions	122
Collations	123
Base URI	125
Statically Known Documents and Collections	125

The Dynamic Context	126
The Focus	126
Dynamic Variables	129
Function Implementations	129
Current Date and Time	129
Implicit Timezone	130
Available Documents and Collections	131
Summary	132
 Chapter 5: Basic Constructs	 133
Notation	134
Where to Start	136
Expressions	136
Examples	139
Lexical Constructs	140
Comments	141
Numeric Literals	142
String Literals	144
Names	146
Operators	150
Primary Expressions	152
Examples	153
Variable References	153
Usage	153
Examples	155
Parenthesized Expressions	155
Changes in XPath 2.0	156
Context Item Expressions	156
Changes in XPath 2.0	157
Usage	157
Function Calls	158
Identifying the Function to be Called	159
Converting the Arguments and the Result	160
Changes in XPath 2.0	162
Side Effects	163
Examples	164
Conditional Expressions	165
Changes in XPath 2.0	166
Examples	167
Summary	167

Chapter 6: Operators on Items	169
Arithmetic Operators	169
Syntax	170
Type Promotion	170
Changes in XPath 2.0	171
Effect	172
Arithmetic Using Numbers	173
Examples of Numeric Arithmetic	175
Arithmetic Using Durations	176
Value Comparisons	181
Permitted Operand Types	182
Type Checking for Value Comparisons	186
Examples of Value Comparisons	188
General Comparisons	188
Changes in XPath 2.0	189
Rules for General Comparisons	190
Comparing Sequences	191
Examples of General Comparisons	195
Node Comparisons	196
The «is» Operator	196
The operators «<<» and «>>»	197
Changes in XPath 2.0	198
Boolean Expressions	198
Examples	199
Summary	200
 Chapter 7: Path Expressions	 201
Examples of Path Expressions	201
Changes in XPath 2.0	203
Full Path Expressions	204
Syntax	204
The Root Expression «/»	205
Absolute Paths	206
Relative Paths	208
Steps	211
Syntax of Steps	212
Axes	215
Node Tests	220
Name Tests	221
Kind Tests	224

Abbreviations	226
Defaulting the Axis Name in a Step	226
The «@» Abbreviation	227
The «.» Abbreviation	227
The «//» Abbreviation	228
Predicates	230
Combining Sets of Nodes	234
Syntax	235
Examples	236
Usage	236
Set Intersection and Difference in XPath 1.0	237
Summary	237
 Chapter 8: Sequence Expressions	 239
The Comma Operator	240
Examples	242
Numeric Ranges: The «to» Operator	242
Examples	243
Filter Expressions	244
Examples	246
The «for» Expression	247
Mapping a Sequence	248
Examples	249
For Expressions and Path Expressions	249
Combining Multiple Sequences	250
Example	251
Examples in XMLSpy	252
The «some» and «every» Expressions	254
Examples	257
Quantification and the «=» Operator	257
Errors in «some» and «every» Expressions	258
Summary	259
 Chapter 9: Type Expressions	 261
Converting Atomic Values	262
Converting between Primitive Types	264
Converting between Derived Types	276
Sequence Type Descriptors	277
Atomic Types	278

Contents

Matching Nodes	279
Matching Elements and Attributes	281
The «instance of» operator	287
The «treat as» Operator	288
Summary	290

Chapter 10: XPath Functions

A Word about Naming	292
Functions by Category	292
Boolean Functions	292
Numeric Functions	292
String Functions	292
Date and Time Functions	293
Duration Functions	293
Aggregation Functions	293
Functions on URIs	293
Functions on QNames	293
Functions on Sequences	293
Functions that Return Properties of Nodes	293
Functions that Find Nodes	294
Functions that Return Context Information	294
Diagnostic Functions	294
Functions that Assert a Static Type	294
Notation	294
Function Definitions	296
abs	296
adjust-date-to-timezone, adjust-dateTime-to-timezone, adjust-time-to-timezone	297
avg	301
base-uri	302
boolean	304
ceiling	306
codepoints-to-string	308
collection	309
compare	310
concat	312
contains	314
count	316
current-date, current-dateTime, current-time	318
data	320
day-from-date, day-from-dateTime	322
days-from-duration	323

deep-equal	323
default-collation	326
distinct-values	327
doc	329
document-uri	332
empty	333
ends-with	334
error	336
escape-uri	337
exactly-one	340
exists	341
expanded-QName	342
false	343
floor	344
hours-from-dateTime, hours-from-time	346
hours-from-duration	347
id	347
idref	350
implicit-timezone	352
index-of	353
in-scope-prefixes	354
Insert-before	356
lang	357
last	359
local-name	363
local-name-from-QName	365
lower-case	366
matches	368
max	370
min	371
minutes-from-dateTime, minutes-from-time	373
minutes-from-duration	374
month-from-date, month-from-dateTime	374
months-from-duration	375
name	376
namespace-uri	379
namespace-uri-for-prefix	381
namespace-uri-from-QName	382
nilled	383
node-name	384
normalize-space	386
normalize-unicode	388

Contents

not	391
number	393
one-or-more	395
position	396
remove	399
replace	400
resolve-QName	403
resolve-uri	405
reverse	408
root	408
round	409
round-half-to-even	411
seconds-from-dateTime, seconds-from-time	413
seconds-from-duration	414
starts-with	415
string	416
string-join	418
string-length	419
string-to-codepoints	421
subsequence	422
substring	423
substring-after	425
substring-before	427
subtract-dates, subtract-dateTimes	429
sum	431
timezone-from-date, timezone-from-dateTime, timezone-from-time	433
tokenize	434
trace	436
translate	437
true	439
unordered	440
upper-case	442
year-from-date, year-from-dateTime	443
years-from-duration	443
zero-or-one	444
Summary	445
 Chapter 11: Regular Expressions	 447
Branches and Pieces	448
Quantifiers	448
Atoms	449

Character Groups	450
Character Ranges	450
Character Class Escapes	451
Character Blocks	453
Character Categories	455
Disallowed Constructs	457
Summary	457
Appendix A: XPath 2.0 Syntax Summary	459
Appendix B: Operator Precedence	467
Appendix C: Compatibility with XPath 1.0	469
Appendix D: Error Codes	475
Glossary	487
Index	505