

Application development

XML

Elements of Style

Suranaree University of Technology



31051000645693

Simon St. Laurent

Contents

Part 1 Elementary Rules of Usage

| | |
|--|-----------|
| 1 XML Documents: An Overview | 3 |
| Unicode Text as the Foundation | 4 |
| Conformance: Well-formed and Valid Documents | 4 |
| Inside XML Documents | 5 |
| A Sample XML Document | 5 |
| 2 Starting the Document: The Prolog | 7 |
| The XML Declaration | 7 |
| The Document Type Declaration | 9 |
| 3 Foundations: Elements, Attributes, and Namespaces | 15 |
| "Sculpting" and "Designing" XML | 15 |
| Elements | 16 |
| Attributes | 18 |
| Naming Your Elements and Attributes | 19 |
| Elements and/vs. Attributes (and Other Alternatives) | 20 |
| Namespaces | 23 |
| 4 Documents, "Reality," and Modeling Models | 25 |
| Structures, Reality, and Back Again | 25 |
| Top-down or Bottom-up? | 26 |
| The Document and Data Debate | 27 |
| How to Proceed? | 28 |
| 5 Processing Models and Document Structures | 29 |
| Transformations and Their Importance | 29 |
| Keeping Information Meaningful | 30 |
| Middleware Possibilities | 32 |
| "Terminal" Vocabularies | 32 |

| | |
|--|---------------|
| 6 Common Document Patterns | 35 |
| Hierarchies | 35 |
| Lists | 37 |
| Tables | 39 |
| Extended Tables | 41 |
| Messages | 42 |
| Objects | 42 |
| Chaos: "Document Documents" | 42 |
| 7 Describing Foundations: Element Type Declarations | 45 |
| Describing Containers | 45 |
| Basic Syntax | 46 |
| The EMPTY Content Model | 47 |
| The ANY Content Model | 47 |
| Structured-Element Content Models | 48 |
| Mixed-Content Models | 51 |
| Data Models | 52 |
| 8 Describing Foundations: Attribute List Declarations | 53 |
| Describing Annotation | 53 |
| Basic Syntax | 54 |
| Naming Attributes | 55 |
| Attribute Types | 55 |
| Attribute Defaults | 57 |
| Putting It All Together | 58 |
| Overriding Attribute List Declarations in the Internal Subset (and Elsewhere) | 60 |
| 9 Extras: Comments, Processing Instructions, and CDATA Sections | 63 |
| Comments | 63 |
| Processing Instructions | 65 |
| CDATA Sections | 66 |
| 10 White Space | 69 |
| An Introduction to White Space Handling | 69 |
| White Space in Declarations and Markup | 70 |

| | |
|---|-----------|
| White Space in Element Content | 71 |
| White Space in Mixed Content | 72 |
| A Way to Sneak in #xD | 72 |
| xml:space | 73 |
| White Space in Attribute Content | 73 |
| White Space on the Fringes: The Prolog and Epilogue | 73 |
| Implications for Applications and Stylesheets | 74 |
| 11 Data Typing in XML 1.0 | 75 |
| Extra Information | 75 |
| Data Type Descriptors | 79 |

Part 2 Elementary Principles of Composition

| | |
|--|------------|
| 12 Using Character References and Predefined Entities | 83 |
| Identifying Characters—Character References | 83 |
| Avoiding Collisions with Markup—Predefined Entities | 84 |
| Multiple Interpretation of Character References | 86 |
| 13 Using General Entities for Reusable Document Content | 87 |
| Types of Entities | 87 |
| Basic Syntax | 88 |
| Using General Entities | 90 |
| Including Inclusions | 92 |
| Using General Entities | 93 |
| Issues of Inclusion | 94 |
| 14 Using Parameter Entities for Reusable Declarations | 97 |
| Basic Syntax | 97 |
| Using Parameter Entities | 100 |
| 15 Using Notations and Unparsed Entities | 107 |
| Basic Syntax | 107 |
| Using Notations | 110 |
| Using Unparsed Entities | 113 |

| | |
|--|------------|
| 16 Using Conditional Sections: INCLUDE and IGNORE | 117 |
| Basic Syntax | 117 |
| Using Conditional Sections | 119 |

Part 3 A Few Matters of Form

| | |
|---|------------|
| 17 Types of Markup: Semantic to Application Specific | 125 |
| Semantics, Semantics, and More Semantics | 125 |
| Focus on Meaning | 127 |
| Focus on Structure | 129 |
| Focus on Applications | 131 |
| 18 Planning for Transformation and Presentation | 135 |
| When Transformation is Useful | 135 |
| Different Kinds of Transformation Tools | 136 |
| Annotative Presentation | 140 |
| Transformative Presentation | 140 |
| Staying Flexible | 141 |
| 19 Building Modular DTDs | 143 |
| Manageability vs. Reusability | 143 |
| Limitations | 144 |
| Naming Conventions and Namespaces | 145 |
| Documentation | 146 |
| Mapping | 146 |
| 20 Leaving Room for the Future: Extensible DTDs | 149 |
| Easy Extensions | 149 |
| Valid XML's Closed World | 150 |
| Opening the Doors: ANY and Omitted Declarations | 150 |
| Notes on Notations | 151 |
| Dependencies and Other Structural Issues | 152 |
| Documenting Extensibility | 153 |
| Shrinkability | 153 |

| | | |
|-----------|---|------------|
| 21 | Developing for an International Audience | 155 |
| | Unicode 2.0 and ISO10646 | 155 |
| | Beyond Character Encodings: Identifying Languages | 159 |
| | Beyond XML: Style Sheets and More | 161 |
| 22 | The Importance of Documentation | 171 |
| | Documenting DTDs | 171 |
| | Documenting Documents | 175 |
| | Documentation Outside of XML 1.0 | 176 |
| | Documenting for the Machine—PIs as Comments | 177 |
| 23 | Creating Webs of Documents: XLink | 179 |
| | Modeling Relationships | 180 |
| | The XLink Approach | 181 |
| | Simple Links | 182 |
| | Extended Links | 184 |
| | Identifying XML Resource Fragments with XPointer and XPath | 189 |
| | Extended Link Groups | 193 |
| | Limitations | 193 |

Part 4 Tools and Features Commonly Misused

| | | |
|-----------|--|------------|
| 24 | Integrating Namespaces with XML 1.0 | 197 |
| | The Problem of Changing Prefixes | 197 |
| | Defaulting (and Declaring) Namespace Declarations | 200 |
| | Disappearing Prefixes | 200 |
| | Validating Qualified Names | 201 |
| | An Argument for Well-Formedness | 201 |
| 25 | Describing Document Types Precisely | 203 |
| | The Hazards of Self-Identifying Documents (DTDs and Namespaces) | 203 |
| | MIME Types and Content Negotiation | 205 |
| | Building Better Packages | 207 |

| | |
|--|------------|
| 26 Processing Instructions: Working Outside Content | 211 |
| Abusing Comments | 211 |
| Abusing Processing Instructions | 214 |
| Using Processing Instructions | 216 |
| Creating Generalized Processing Instructions | 218 |
| The Need for an Application Framework | 219 |
| 27 Troubles with Non-validating Parsers | 221 |
| The Problem | 221 |
| Solution 1: The Standalone Declaration | 223 |
| Solution 2: Controlling Your Applications | 223 |
| Solution 3: Document and Double-Check | 225 |
| Solution 4: Don't Use the Features | 225 |
| The Future? Layers of Processing? | 226 |
| 28 XML Processing Architectures: Trees and Streams | 229 |
| Reading Event Streams | 230 |
| Writing (and Modifying) Event Streams | 232 |
| Reading Trees | 232 |
| Writing Trees | 235 |
| Choosing | 235 |
| 29 Waiting for the Future: | |
| XML Schemas and Data Typing | 237 |
| The Problems | 238 |
| Examples | 239 |
| Advanced Features | 245 |
| Planning the Transition | 245 |

Part 5 An Approach to Style

| | |
|-------------------------------|------------|
| 30 Seeking Consistency | 249 |
| Consistent Criteria | 249 |
| Consistent Communications | 250 |
| Consistent Conventions | 252 |
| Consistent Decision-Making | 253 |

| | |
|--|------------|
| Consistent Application Design | 253 |
| 31 Building on Past Works | 255 |
| History Can Be Useful | 255 |
| Remodeling or Demolition and Rebuilding? | 256 |
| Providing an Upward (and Downward) Path: Transformations | 257 |
| Preserving Compatibility | 257 |
| Thinking Ahead | 258 |
| 32 Maximizing Information Quantity and Quality | 261 |
| Including and Excluding Information | 261 |
| Choosing Conventions | 262 |
| Working Toward Multiple Purposes | 263 |
| Document Your Meanings | 264 |
| 33 Designing for Large-Scale Interchange | 265 |
| Collecting Input | 265 |
| Managing Expectations | 266 |
| Focus on Meaning | 267 |
| Testing, Testing, Testing | 267 |
| Communicating Outside XML | 268 |
| 34 XML and Data Integration | 271 |
| Documents as Data | 271 |
| Data as Documents | 272 |
| Extending XML to Support Data | 273 |
| Middleware Solutions | 273 |
| Moving Forward | 274 |
| 35 Maximizing Flexibility: Radical Extensibility | 277 |
| Don't Count on Consistency | 277 |
| Machine Reading | 279 |
| Plan for Change | 280 |
| Model Jobs, Not Data | 280 |
| Build Up Your Toolset | 282 |
| Moving Toward an Adaptable Web | 282 |
| Index | 285 |