

# Contents

|   |              |
|---|--------------|
| <b>Acknowledgments</b>  | <b>v</b>     |
| <b>Foreword by Dare Obasanjo</b>                                    | <b>xxvii</b> |
| <b>Foreword by Greg Reinacker</b>                                   | <b>xxix</b>  |
| <b>Introduction</b>   | <b>xxxii</b> |
| <br>  |              |
| <b>Part I: Understanding the Issues and Taking Control</b>          | <b>1</b>     |
| <br>  |              |
| <b>Chapter 1: Managing the Flow of Information: A Crucial Skill</b> | <b>3</b>     |
| <hr/>   |              |
| <b>New Vistas of Information Flow</b>                               | <b>4</b>     |
| The Information Well and Information Flow                           | 4            |
| The Information Well  | 4            |
| The Information Flow  | 5            |
| The Information Flood   | 5            |
| <b>Managing Information</b>   | <b>5</b>     |
| What Do You Want to Do with Information?                            | 6            |
| Browse and Discard  | 6            |
| Read  | 6            |
| Study and Keep  | 6            |
| <b>Taking Control of Information</b>                                | <b>7</b>     |
| Determining What Is Important to You                                | 7            |
| Avoiding Irrelevant Information                                     | 7            |
| Determining the Quality of Information                              | 8            |
| <b>Information Flows Other Than the Web</b>                         | <b>8</b>     |
| Books   | 8            |
| Magazines   | 9            |
| Newspapers  | 9            |
| Broadcast Media   | 9            |
| <b>The Web and Information Feeds</b>                                | <b>10</b>    |
| New Information Opportunities                                       | 10           |
| New Information Problems  | 10           |
| The Need to Keep Up-to-Date   | 11           |
| Distractions  | 11           |
| <b>Summary</b>  | <b>11</b>    |
| <b>Exercise</b>   | <b>11</b>    |

|   |           |
|---|-----------|
| <b>Chapter 2: Where Did Information Feeds Start?</b>    | <b>13</b> |
| <b>The Nature of the Web</b>                            | <b>13</b> |
| HTTP  | 14        |
| HTML  | 14        |
| XML   | 14        |
| Polling the Web   | 14        |
| <b>Precursors to RSS</b>                                | <b>17</b> |
| MCF and HotSauce  | 17        |
| Netscape Channels                                       | 17        |
| The Microsoft Channel Definition Format                 | 18        |
| <b>RSS: An Acronym with Multiple Meanings</b>           | <b>18</b> |
| RSS 0.9   | 19        |
| RSS 0.91  | 19        |
| RSS 1.0   | 19        |
| RSS 0.92, 0.93, and 0.94                                | 19        |
| RSS 2.0   | 19        |
| Use of RSS and Atom Versions                            | 20        |
| <b>Summary</b>  | <b>21</b> |
| <b>Exercises</b>  | <b>21</b> |
| <b>Chapter 3: The Content Provider Viewpoint</b>        | <b>23</b> |
| <b>Why Give Your Content Away?</b>                      | <b>23</b> |
| Selling Your Content                                    | 24        |
| Creating Community                                      | 25        |
| <b>Content to Include in a Feed</b>                     | <b>25</b> |
| The Importance of Item Titles                           | 25        |
| Internal Links  | 25        |
| One Feed or Several?                                    | 26        |
| Structuring Topics                                      | 26        |
| <b>Blogging Tools</b>                                   | <b>26</b> |
| Wikis   | 28        |
| <b>Publicizing Your Information Feed</b>                | <b>28</b> |
| Deciding on a Target Audience                           | 28        |
| Registering with Online Sites                           | 28        |
| How Information Feeds Can Affect Your Site's Visibility | 29        |
| <b>Advertisements and Information Feeds</b>             | <b>29</b> |
| Power to the User?                                      | 30        |
| Filtering Out Advertisements                            | 30        |
| <b>Summary</b>  | <b>31</b> |
| <b>Exercise</b>   | <b>31</b> |

---

|  |           |
|--|-----------|
| <b>Chapter 4: The Content Recipient Viewpoint</b>                | <b>33</b> |
| <b>Access to Information</b>                                     | <b>34</b> |
| Convenience  | 34        |
| Timeliness of Access   | 35        |
| Timeliness and Data Type   | 35        |
| Timeliness and Data Source                                       | 35        |
| <b>Newsreaders and Aggregators</b>                               | <b>36</b> |
| Aggregating for Intranet Use                                     | 36        |
| Security and Aggregators   | 36        |
| Directories  | 37        |
| <b>Finding Information about Interesting Feeds</b>               | <b>38</b> |
| The Known Sites Approach   | 38        |
| The Blogroll Approach  | 39        |
| The Directory Approach   | 41        |
| <b>Filtering Information Feeds</b>                               | <b>42</b> |
| Filtering Blogs  | 42        |
| <b>Summary</b>   | <b>42</b> |
| <b>Chapter 5: Storing, Retrieving, and Exporting Information</b> | <b>43</b> |
| <b>Storing Information</b>                                       | <b>44</b> |
| Storing URLs   | 44        |
| Storing Content  | 44        |
| Storing Static Files   | 44        |
| Relational Databases   | 45        |
| RDF Triple Stores  | 46        |
| <b>Two Examples of Longer-Term Storage</b>                       | <b>46</b> |
| Onfolio 2.0  | 46        |
| OneNote 2003   | 50        |
| <b>Retrieving Information</b>                                    | <b>52</b> |
| Search in Onfolio 2.0  | 53        |
| Search in OneNote 2003   | 54        |
| <b>Exporting Information</b>                                     | <b>55</b> |
| Exporting in Onfolio 2.0   | 55        |
| Exporting in OneNote 2003  | 55        |
| <b>Summary</b>   | <b>56</b> |

|  |           |
|--|-----------|
| <b>Part II: The Technologies</b>           | <b>57</b> |
| <b>Chapter 6: Essentials of XML</b>        | <b>59</b> |
| <b>What Is XML?</b>                        | <b>60</b> |
| XML Declaration                            | 62        |
| XML Names                                  | 62        |
| XML Elements                               | 62        |
| XML Attributes                             | 63        |
| XML Comments                               | 64        |
| Predefined Entities                        | 64        |
| Character References                       | 65        |
| <b>XML Namespaces</b>                      | <b>66</b> |
| <b>HTML, XHTML, and Feed Autodiscovery</b> | <b>68</b> |
| <b>Summary</b>                             | <b>70</b> |
| <b>Exercises</b>                           | <b>70</b> |
| <b>Chapter 7: Atom 0.3</b>                 | <b>71</b> |
| <b>Introducing Atom 0.3</b>                | <b>72</b> |
| The Atom 0.3 Specification                 | 72        |
| The Atom 0.3 Namespace                     | 72        |
| <b>Atom 0.3 Document Structure</b>         | <b>72</b> |
| The feed Element                           | 73        |
| The title Element                          | 74        |
| The link Element                           | 74        |
| The author Element                         | 74        |
| The id Element                             | 74        |
| The generator Element                      | 75        |
| The copyright Element                      | 75        |
| The info Element                           | 75        |
| The modified Element                       | 75        |
| The tagline Element                        | 75        |
| The entry Element                          | 76        |
| <b>Using Modules with Atom 0.3</b>         | <b>78</b> |
| <b>Summary</b>                             | <b>79</b> |
| <b>Exercises</b>                           | <b>79</b> |
| <b>Chapter 8: RSS 0.91 and RSS 0.92</b>    | <b>81</b> |
| <b>What Is RSS 0.91?</b>                   | <b>82</b> |
| <b>The RSS 0.91 Document Structure</b>     | <b>82</b> |
| The rss Element                            | 83        |
| The channel Element                        | 83        |

---

|  |            |
|--|------------|
| Required Child Elements of channel                       | 83         |
| Optional Child Elements of channel                       | 84         |
| The image Element  | 85         |
| The textInput Element                                    | 85         |
| The item Element   | 86         |
| <b>Introducing RSS 0.92</b>                              | <b>87</b>  |
| <b>The RSS 0.92 Document Structure</b>                   | <b>87</b>  |
| New Child Elements of the item Element                   | 87         |
| The cloud Element  | 88         |
| <b>Summary</b>   | <b>88</b>  |
| <b>Exercises</b>   | <b>88</b>  |
| <br>   |            |
| <b>Chapter 9: RSS 1.0</b>                                | <b>89</b>  |
| <hr/>  |            |
| <b>What Is RSS 1.0?</b>                                  | <b>89</b>  |
| RSS 1.0 Is RDF   | 90         |
| RSS 1.0 Uses XML Namespaces                              | 90         |
| RSS 1.0 Uses Modules                                     | 91         |
| <b>The RSS 1.0 Document Structure</b>                    | <b>91</b>  |
| The channel Element                                      | 93         |
| The items Element  | 93         |
| The image Element  | 94         |
| The item Element   | 94         |
| The textinput Element                                    | 95         |
| <b>Some Real-World RSS 1.0</b>                           | <b>95</b>  |
| <b>Summary</b>   | <b>98</b>  |
| <b>Exercise</b>  | <b>99</b>  |
| <br>   |            |
| <b>Chapter 10: RSS 1.0 Modules</b>                       | <b>101</b> |
| <hr/>  |            |
| <b>RSS Modules</b>                                       | <b>101</b> |
| <b>The RSS 1.0 Official Modules</b>                      | <b>102</b> |
| RDF Parser Compatibility                                 | 103        |
| Module Compatibility                                     | 103        |
| The Content Module                                       | 103        |
| The content:items Element                                | 104        |
| The content:item Element                                 | 104        |
| The Dublin Core Module                                   | 105        |
| The Syndication Module                                   | 107        |
| <b>Including Other Modules in RSS 1.0 Feed Documents</b> | <b>108</b> |
| Adding the Namespace Declaration                         | 108        |
| The Admin Module   | 108        |
| The FOAF Module  | 108        |
| <b>Summary</b>   | <b>109</b> |

## **Chapter 11: RDF: The Resource Description Framework** **111**

---

### **What Is RDF?** **112**

Simple Metadata 112

Simple Facts Expressed in RDF 112

The RDF Triple 113

Using URIs in RDF 113

Directed Graphs 114

How RDF and XML Are Related 115

What RDF Is Used For 115

RDF and RSS 1.0 116

### **RDF Vocabularies** **117**

Dublin Core 117

FOAF 117

### **RDF Toolkits** **118**

Jena 118

Redland 118

RDFLib 118

rdflib.org 118

### **Summary** **118**

## **Chapter 12: RSS 2.0: Really Simple Syndication** **121**

---

### **What Is RSS 2.0?** **121**

XML Namespaces in RSS 2.0 122

New Elements in RSS 2.0 122

### **The RSS 2.0 Document Structure** **122**

The rss Element 122

The channel Element 123

The image Element 124

The cloud Element 125

The textinput Element 125

The item Element 126

An Example RSS 2.0 Document 126

### **RSS 2.0 Extensions** **127**

The blogChannel RSS Module 127

### **Summary** **128**

## **Chapter 13: Looking Forward to Atom 1.0** **129**

---

### **Why Another Specification?** **130**

Aiming for Clarity 130

Archiving Feeds 130

---

|   |            |
|---|------------|
| The RDF Issue   | 130        |
| The Standards Body Aspect                               | 131        |
| <b>What Is Atom?</b>                                    | <b>131</b> |
| Further Information about Atom 1.0                      | 131        |
| Atom 1.0 Document Structure                             | 132        |
| The XML Declaration                                     | 132        |
| The feed Element  | 132        |
| The head Element  | 133        |
| The entry Element                                       | 133        |
| Extending Atom 1.0 Documents                            | 134        |
| Other Aspects of Atom 1.0                               | 134        |
| The Atom Publishing Protocol                            | 134        |
| An Atom Entry Document                                  | 135        |
| The Atom Notification Protocol                          | 135        |
| Atom Feed Autodiscovery                                 | 136        |
| <b>Summary</b>  | <b>136</b> |
| <br>  |            |
| <b>Part III: The Tools</b>                              | <b>137</b> |
| <br>  |            |
| <b>Chapter 14: Feed Production Using Blogging Tools</b> | <b>139</b> |
| <hr/>   |            |
| <b>Movable Type</b>                                     | <b>140</b> |
| <b>WordPress</b>  | <b>146</b> |
| <b>Blojsom</b>  | <b>147</b> |
| <b>Summary</b>  | <b>154</b> |
| <br>  |            |
| <b>Chapter 15: Aggregators and Similar Tools</b>        | <b>155</b> |
| <hr/>   |            |
| <b>Overview of Desktop Aggregators</b>                  | <b>156</b> |
| Managing Subscriptions                                  | 156        |
| Updating Feeds  | 156        |
| Viewing Web Pages                                       | 156        |
| <b>Individual Desktop Aggregators</b>                   | <b>157</b> |
| Ablon   | 157        |
| ActiveRefresh   | 159        |
| Aggreg8   | 160        |
| BlogBridge  | 161        |
| Firefox Live Bookmarks                                  | 162        |
| Google News Alerts                                      | 162        |
| Habari Xenu   | 163        |
| NewzCrawler   | 164        |
| Newsgator   | 165        |
| Omea  | 166        |

# Contents

---

|   |            |
|---|------------|
| Onfolio   | 168        |
| RSS Bandit  | 169        |
| Thunderbird 1.0                                     | 170        |
| <b>Summary</b>                                      | <b>171</b> |
| <b>Exercise</b>                                     | <b>172</b> |
| <b>Chapter 16: Long-Term Storage of Information</b> | <b>173</b> |
| <b>Choosing an Approach to Long-Term Storage</b>    | <b>174</b> |
| The Use Case for Long-Term Storage                  | 174        |
| Storage Options                                     | 174        |
| Choosing Information to Store                       | 174        |
| Determining Who Will Access the Data                | 175        |
| Ease of Storage                                     | 175        |
| Choosing a Backup Strategy                          | 176        |
| <b>Characteristics of Long-Term Storage</b>         | <b>176</b> |
| Scalability   | 176        |
| Availability  | 176        |
| <b>Software to Support Long-Term Storage</b>        | <b>177</b> |
| Onfolio   | 177        |
| Microsoft OneNote                                   | 178        |
| Microsoft SQL Server 2005                           | 179        |
| <b>Summary</b>                                      | <b>180</b> |
| <b>Chapter 17: Online Tools</b>                     | <b>181</b> |
| <b>Advantages and Disadvantages of Online Tools</b> | <b>181</b> |
| Registration  | 182        |
| Interface Issues                                    | 182        |
| Online Tools When Mobile                            | 182        |
| Cost of Online Tools                                | 182        |
| Stability of Access                                 | 183        |
| Performance   | 183        |
| Notifications                                       | 183        |
| Cookies   | 183        |
| Using Multiple Machines                             | 183        |
| Making a Choice                                     | 184        |
| <b>Choosing Between Individual Online Tools</b>     | <b>184</b> |
| Bloglines   | 184        |
| Feedster  | 185        |
| Furl  | 186        |
| Google News Alerts                                  | 187        |
| Kinja   | 188        |

|   |            |
|---|------------|
| Newsgator Online  | 190        |
| Rollup.org  | 191        |
| Syndic8.com   | 192        |
| Technorati  | 193        |
| PubSub.com  | 194        |
| <b>Summary</b>  | <b>194</b> |
| <b>Chapter 18: Language-Specific Developer Tools</b>            | <b>195</b> |
| <b>Python Tools</b>   | <b>195</b> |
| PyXML   | 196        |
| RSS.py  | 196        |
| Universal Feed Parser   | 196        |
| xpath2rss   | 196        |
| Chumpologica  | 196        |
| <b>PHP Tools</b>  | <b>196</b> |
| lastRSS   | 197        |
| MagpieRSS   | 197        |
| <b>Java Tools</b>   | <b>197</b> |
| ROME  | 197        |
| Jena  | 197        |
| <b>The Redland RDF Application Framework</b>                    | <b>198</b> |
| <b>XSLT Tools</b>   | <b>198</b> |
| <b>Perl Tools</b>   | <b>198</b> |
| XML::RSS  | 199        |
| Support for XSLT  | 199        |
| The rss2html Module   | 199        |
| The Regular Expressions Approach                                | 200        |
| <b>Miscellaneous Tools</b>                                      | <b>200</b> |
| <b>Useful Sources</b>   | <b>200</b> |
| <b>Summary</b>  | <b>201</b> |
| <b>Part IV: The Tasks</b>                                       | <b>203</b> |
| <b>Chapter 19: Systematic Overview</b>                          | <b>205</b> |
| <b>Before You Start</b>   | <b>206</b> |
| Python  | 206        |
| Recommended Downloads   | 206        |
| <b>States and Messages: Exchanges Between Client and Server</b> | <b>207</b> |
| Resources and Representations                                   | 207        |
| States and Statelessness  | 208        |
| RPC vs. Document-Oriented Messaging                             | 208        |

# Contents

---

|   |            |
|---|------------|
| <b>Communication Layers</b>                                 | <b>210</b> |
| Web Languages   | 210        |
| The HTTP Client   | 210        |
| <b><i>Try It Out: Client Requests</i></b>                   | <b>210</b> |
| Server Response   | 212        |
| Acronym City  | 214        |
| Application Layers  | 214        |
| <b>Server Producer: Producing Feeds</b>                     | <b>214</b> |
| The HTTP Server   | 214        |
| Sockets and Ports   | 215        |
| <b><i>Try It Out: Serving an RSS Feed</i></b>               | <b>216</b> |
| Getting Flexible  | 219        |
| <b><i>Try It Out: Improved Version with Python</i></b>      | <b>220</b> |
| Serving and Producing                                       | 221        |
| <b>Client Consumer: Viewing Feeds</b>                       | <b>222</b> |
| Single-Page View  | 223        |
| Three-Pane View   | 223        |
| Syndication, the Protocol                                   | 224        |
| Polling   | 225        |
| Reducing the Load   | 226        |
| <b><i>Try It Out: HTTP Client with gzip Compression</i></b> | <b>227</b> |
| The Trade Off   | 230        |
| <b>Client Producer: Blogging APIs</b>                       | <b>230</b> |
| XML-RPC   | 231        |
| RESTful Communications                                      | 232        |
| <b>Server Consumer</b>                                      | <b>233</b> |
| Weblog Comments   | 233        |
| Trackback   | 233        |
| <b>Architectural Approaches</b>                             | <b>234</b> |
| <b>Summary</b>  | <b>234</b> |
| <b>Exercise</b>   | <b>234</b> |
| <br>  |            |
| <b>Chapter 20: Modeling Feed Data</b>                       | <b>237</b> |
| <br>  |            |
| <b>Model or Syntax?</b>                                     | <b>237</b> |
| <b>The Conceptual Feed</b>                                  | <b>238</b> |
| The Feed-Oriented Model                                     | 238        |
| Source or Styled?   | 239        |
| <b><i>Try It Out: RSS with CSS</i></b>                      | <b>243</b> |
| Drawbacks of a Feed-Oriented Model                          | 245        |
| Item-Oriented Models  | 247        |
| Twenty-First Century Press Clippings                        | 247        |
| Microcontent and Memes                                      | 247        |

|   |            |
|---|------------|
| Including Item Provenance                       | 250        |
| Element Order                                   | 252        |
| <b>Common Features Among Formats</b>            | <b>252</b> |
| Entities and Relationships                      | 252        |
| Entities in Feeds                               | 253        |
| Relationships in Feeds                          | 253        |
| Identity  | 254        |
| Uniform and Unique                              | 254        |
| GUID  | 255        |
| Crisis, What Crisis?                            | 255        |
| Parallel Elements                               | 256        |
| Groups of Feeds                                 | 256        |
| Extending Structures                            | 256        |
| <b>An XML Model</b>                             | <b>257</b> |
| <i>Try It Out: Reverse Engineering a Schema</i> | <b>259</b> |
| <b>An Object-Oriented Model: XML to Java</b>    | <b>260</b> |
| <i>Try It Out: Creating a Feed</i>              | <b>262</b> |
| Reading a Feed                                  | 264        |
| <i>Try It Out: A Demo Channel</i>               | <b>264</b> |
| Problems of Autogeneration                      | 266        |
| <b>The RDF Models</b>                           | <b>266</b> |
| There Is No Syntax!                             | 266        |
| <i>Try It Out: Yet Another Syntax</i>           | <b>268</b> |
| RSS 1.0 RDF Schema                              | 268        |
| Groups of Feeds in RDF                          | 272        |
| <b>A Relational Model</b>                       | <b>274</b> |
| Syndication Applications of Databases           | 274        |
| Content Storage                                 | 274        |
| Metadata Management                             | 274        |
| User Tracking                                   | 274        |
| Query and Search                                | 275        |
| Requirements for Storing Feed Data              | 275        |
| Storing Items                                   | 275        |
| Feed Lists and State Management                 | 275        |
| The Relational Model, Condensed                 | 275        |
| Relations and Tables                            | 275        |
| Tables, Take Two                                | 276        |
| Tables, Take Three                              | 277        |
| Keys  | 279        |
| Extensibility in Databases                      | 279        |
| Tables, Take Four                               | 280        |
| <b>Summary</b>                                  | <b>281</b> |
| <b>Exercises</b>                                | <b>282</b> |

|  |            |
|--|------------|
| <b>Chapter 21: Storing Feed Data</b>               | <b>285</b> |
| <b>The Document Object Model</b>                   | <b>285</b> |
| A (Forgetful) DOM-Based Weblog                     | 289        |
| XML in PHP   | 290        |
| Source Code  | 290        |
| Representing Feed Lists in OPML                    | 301        |
| <b>Creating a SQL Database for Feeds</b>           | <b>302</b> |
| Download Requirements                              | 302        |
| <i>Try It Out: Working with SQL</i>                | <b>302</b> |
| <b>RDF in Relational Databases</b>                 | <b>308</b> |
| A Minimal Triple Store                             | 308        |
| Filling the Store                                  | 310        |
| RSS 1.0 Parser                                     | 310        |
| Connecting to a Database                           | 313        |
| <i>Try It Out: Interfacing with a Database</i>     | <b>313</b> |
| Aggregating Triples                                | 316        |
| Code Overview                                      | 316        |
| Store Wrapper                                      | 316        |
| Connecting Parser to Store                         | 321        |
| The Joy of SQL VIEWS                               | 322        |
| <i>Try It Out: Using the Triplestore</i>           | <b>323</b> |
| Customizing for RSS                                | 324        |
| Utility Methods                                    | 324        |
| A View of Items                                    | 325        |
| Starting with a Blogroll                           | 326        |
| <i>Try It Out: Aggregating with a Triple Store</i> | <b>327</b> |
| Pros and Cons of Using an RDF Store                | 329        |
| <b>Summary</b>                                     | <b>330</b> |
| <b>Exercises</b>                                   | <b>330</b> |
| <b>Chapter 22: Consuming Feeds</b>                 | <b>333</b> |
| <b>Consumer Strategies</b>                         | <b>333</b> |
| Parsing Data                                       | 335        |
| Scanning for Tokens                                | 335        |
| Parsing for Structure                              | 336        |
| Interpretation for Semantics                       | 336        |
| XML Parsing  | 336        |
| Semantics in Feeds                                 | 337        |
| A More Practical Consumer                          | 337        |
| All Kinds of XML                                   | 338        |
| Markup in Content                                  | 339        |
| Other Approaches                                   | 339        |

---

|  |            |
|--|------------|
| <b>Reading Feed Data</b>                           | <b>341</b> |
| The Polite Client                                  | 341        |
| Push and Pull                                      | 342        |
| HTTP Status Codes                                  | 342        |
| Timing Issues                                      | 344        |
| XML-Level Timing Hints                             | 344        |
| HTTP Compression                                   | 345        |
| Conditional GET                                    | 345        |
| A Reasonably Polite Client                         | 346        |
| <i>Try It Out: Downloading a Feed</i>              | <b>350</b> |
| Parsing XML with SAX                               | 356        |
| <i>Try It Out: Reading an RDF/XML Channel List</i> | <b>359</b> |
| <b>Feed/Connection Management Subsystem</b>        | <b>362</b> |
| A Fetcher That Fetches the Feeds                   | 364        |
| <i>Try It Out: Running the Consumer</i>            | <b>375</b> |
| Implementation Notes                               | 379        |
| <b>Summary</b>                                     | <b>379</b> |
| <b>Exercises</b>                                   | <b>380</b> |
| <br>   |            |
| <b>Chapter 23: Parsing Feeds</b>                   | <b>381</b> |
| <hr/>  |            |
| <b>A Target Model</b>                              | <b>382</b> |
| <b>Parsing Feeds as XML</b>                        | <b>384</b> |
| <i>Try It Out: Parsing RSS with SAX</i>            | <b>387</b> |
| <b>The Trouble with Feeds</b>                      | <b>392</b> |
| The Syndication Stack                              | 393        |
| The Postel Paradox                                 | 393        |
| The Bozo Bit                                       | 394        |
| <i>Try It Out: Draconian Parsing</i>               | <b>394</b> |
| Tag Soup   | 395        |
| Soup Parser  | 395        |
| Different Parser, Same Handler                     | 406        |
| <i>Try It Out: Parsing Tag Soup</i>                | <b>408</b> |
| HTTP, MIME, and Encoding                           | 409        |
| Encoding Declaration in XML                        | 410        |
| Encoding Declaration in HTTP                       | 410        |
| The Bad News                                       | 411        |
| The Good News                                      | 411        |
| Gluing Together the Consumer                       | 411        |
| <i>Try It Out: A Consumer Subsystem</i>            | <b>417</b> |
| One More Thing...                                  | 419        |
| A Universal Feed Consumer?                         | 420        |
| <b>Summary</b>                                     | <b>420</b> |
| <b>Exercises</b>                                   | <b>421</b> |
|  | xix        |

# Contents

---

## **Chapter 24: Producing Feeds** **423**

---

### **Content Management Systems** **423**

|                              |     |
|------------------------------|-----|
| Content and Other Animals    | 424 |
| Metadata                     | 424 |
| Applications and Services    | 424 |
| Accessibility                | 424 |
| Device Independence          | 425 |
| Personalization              | 425 |
| Security                     | 425 |
| Inter-System Communication   | 425 |
| Legacy Integration           | 426 |
| Managing Content: Many Tiers | 426 |

### **SynWiki: A Minimal CMS** **427**

|  |            |
|--|------------|
| What Is PHP?                                     | 427        |
| Installing PHP                                   | 428        |
| Open Web   | 428        |
| Basic Operation                                  | 428        |
| System Operation                                 | 431        |
| System Requirements                              | 432        |
| Files  | 432        |
| Data Store                                       | 432        |
| <i>Try It Out: Interacting with the Database</i> | <b>435</b> |
| Clients Are Clients                              | 439        |
| Utilities  | 439        |
| Authoring Page                                   | 442        |
| Viewing Page                                     | 444        |
| Making Modules                                   | 447        |
| URL Rewriting                                    | 448        |
| Blog Module                                      | 448        |
| Reverse-Chrono Here We Go                        | 450        |
| Producing RSS 1.0                                | 452        |
| Alternative Techniques                           | 455        |
| Steps in Producing RSS from a Content Database   | 456        |

### **Summary** **456**

### **Exercise** **456**

## **Chapter 25: Queries and Transformations** **459**

---

### **Paths and Queries** **459**

|                                  |            |
|----------------------------------|------------|
| XQuery Runner                    | 460        |
| Source Code                      | 460        |
| <i>Try It Out: XPath Queries</i> | <b>465</b> |

|  |            |
|--|------------|
| XQuery and Syndication Formats                       | 467        |
| <b><i>Try It Out: Querying Live Feeds</i></b>        | <b>467</b> |
| XQuery Is More Than XPath                            | 469        |
| Getting Selective                                    | 470        |
| <b><i>Try It Out: A Selective XQuery</i></b>         | <b>470</b> |
| Introducing XSLT                                     | 472        |
| Templates  | 472        |
| More Templates                                       | 474        |
| XSLT Processors                                      | 476        |
| <b><i>Try It Out: Transforming XML to Text</i></b>   | <b>479</b> |
| Output Format  | 481        |
| XSLT in Syndication                                  | 481        |
| Transforming to RSS                                  | 481        |
| Transforming from RSS                                | 483        |
| Going Live   | 485        |
| <b>Graphic Display: SVG in a Very Small Nutshell</b> | <b>486</b> |
| Graphic RSS  | 489        |
| Handling Text Blocks                                 | 489        |
| Multiformat Handling                                 | 494        |
| SVG Output   | 496        |
| <b>Summary</b>                                       | <b>502</b> |
| <b>Exercise</b>                                      | <b>502</b> |

---

**Chapter 26: The Blogging Client** **503**

|  |            |
|--|------------|
| <b>Talking to the Server with a Variety of Clients</b> | <b>503</b> |
| Browser Scripts  | 504        |
| The Fat Client   | 505        |
| <b>The Language of HTTP</b>                            | <b>506</b> |
| HTML's Legacy  | 506        |
| Authentication   | 507        |
| HTTP Basic Authentication                              | 508        |
| HTTP Digest  | 508        |
| HTTPS  | 509        |
| Base64 Encoding  | 509        |
| Hashes and Digests                                     | 509        |
| Nonces   | 509        |
| <b>Protocols for Blogging</b>                          | <b>509</b> |
| The Blogger API and XML-RPC                            | 510        |
| Command-Line Tool                                      | 513        |
| Handling Account Details                               | 515        |
| Minimal XML-RPC Client                                 | 518        |
| <b><i>Try It Out: Running the Blogger Client</i></b>   | <b>520</b> |

# Contents

---

|  |            |
|--|------------|
| Adding Post Support                              | 523        |
| MetaWeblog and other APIs                        | 525        |
| Atom Protocol                                    | 525        |
| Atom Protocol, SOAP and WSDL                     | 526        |
| Atom Protocol — RESTful HTTP                     | 529        |
| Building Your Own Atomic Device                  | 531        |
| Other Formats, Transports, and Protocols         | 535        |
| Jabber/XMPP                                      | 535        |
| WebDAV   | 535        |
| Annotea  | 536        |
| InterWiki  | 536        |
| Usenet, E-mail, and Friends                      | 536        |
| <b>Summary</b>                                   | <b>536</b> |
| <b>Exercise</b>                                  | <b>536</b> |
| <br>   |            |
| <b>Chapter 27: Building Your Own Planet</b>      | <b>537</b> |
| <hr/>  |            |
| <b>Why Aggregate?</b>                            | <b>537</b> |
| <b>Planet Core</b>                               | <b>538</b> |
| <b>Implementing a Planet</b>                     | <b>539</b> |
| Extending the Entry                              | 540        |
| Entry Collection                                 | 543        |
| Top-Level Runner                                 | 546        |
| Serializing Output                               | 550        |
| Templates  | 555        |
| <b>Running the Planet</b>                        | <b>556</b> |
| <i>Try It Out: Aggregating the Planet</i>        | 557        |
| <b>Summary</b>                                   | <b>560</b> |
| <br>   |            |
| <b>Chapter 28: Building a Desktop Aggregator</b> | <b>561</b> |
| <hr/>  |            |
| <b>Desktop Aggregator Overview</b>               | <b>562</b> |
| <b>System Construction</b>                       | <b>563</b> |
| Feed Data Model                                  | 563        |
| Event Notification                               | 568        |
| User Interface                                   | 569        |
| DOM-Based Implementation                         | 575        |
| Using an RDF Model                               | 579        |
| Automating Updates                               | 581        |
| <b>Summary</b>                                   | <b>582</b> |

---

|  |            |
|--|------------|
| <b>Chapter 29: Social Syndication</b>              | <b>583</b> |
| <b>What Is Society on the Web?</b>                 | <b>583</b> |
| <b>Social Software and Syndication</b>             | <b>584</b> |
| Avoiding Overload                                  | 584        |
| Counting Atoms                                     | 584        |
| Friends as Filters                                 | 585        |
| <b>A Personal Knowledgebase</b>                    | <b>585</b> |
| Kowari RDF Store                                   | 585        |
| RDF Queries  | 585        |
| <i>Try It Out: An iTQL Session</i>                 | <b>585</b> |
| Going Programmatic                                 | 589        |
| <i>Try It Out: Kowari Blogroll Channels</i>        | <b>591</b> |
| In an Ideal World...                               | 595        |
| Fetching and Normalizing                           | 595        |
| RDF to RDF   | 596        |
| XML to RDF   | 596        |
| Tag Soup to RDF                                    | 596        |
| Coupling the Store                                 | 596        |
| Supervisor   | 597        |
| <i>Try It Out: Fetching and Normalizing Feeds</i>  | <b>604</b> |
| Loading the Data Store                             | 605        |
| <i>Try It Out: Adding Normalized Data to Store</i> | <b>607</b> |
| <b>Summary</b>                                     | <b>608</b> |
| <b>Exercise</b>                                    | <b>608</b> |
| <b>Chapter 30: Additional Content</b>              | <b>609</b> |
| <b>Syndicated Broadcasting</b>                     | <b>609</b> |
| Publishing Audio                                   | 610        |
| Digital Audio in Brief                             | 610        |
| Compression Issues                                 | 611        |
| An Imp for the Web                                 | 612        |
| User Interface of Imp                              | 612        |
| Application Classes                                | 613        |
| AudioTool Interface                                | 614        |
| Audio in Java                                      | 614        |
| Lines and Audio Formats                            | 615        |
| Audio Recorder Class                               | 615        |
| Audio Player Class                                 | 619        |
| Imp User Interface                                 | 621        |

# Contents

---

|  |            |
|--|------------|
| Publishing to a Feed                           | 628        |
| Syndication and Enclosures                     | 628        |
| Supporting the Enclosure Element               | 630        |
| Serializing an Entry                           | 631        |
| Publisher Class                                | 631        |
| FTP Upload                                     | 635        |
| Building the Imp                               | 636        |
| <i><b>Try It Out: Running the Imp</b></i>      | <b>637</b> |
| Troubleshooting                                | 637        |
| Developments in Syndicated Media               | 638        |
| <b>Summary</b>                                 | <b>639</b> |
| <br>   |            |
| <b>Chapter 31: Loose Ends, Loosely Coupled</b> | <b>641</b> |
| <br>   |            |
| <b>Finding Feeds</b>                           | <b>642</b> |
| Autodiscovery                                  | 642        |
| Practical Approaches                           | 643        |
| Consommé Code                                  | 643        |
| <i><b>Try It Out: Parsing HTML</b></i>         | <b>646</b> |
| From Browser to Aggregator                     | 647        |
| A Scheme Handler                               | 648        |
| Registering a URL Scheme in Windows            | 648        |
| Writing a Bookmarklet                          | 648        |
| <i><b>Try It Out: Using a Bookmarklet</b></i>  | <b>649</b> |
| Registering a URI Scheme for Real              | 650        |
| Using MIME Types                               | 651        |
| <b>GRDDL</b>                                   | <b>651</b> |
| Robots, Crawlers, Harvesters, and Scutters     | 652        |
| <b>Inter-Site Communication</b>                | <b>654</b> |
| Clouds   | 654        |
| Weblog Pings                                   | 654        |
| Publish/Subscribe Innovations                  | 656        |
| Feed Deltas                                    | 656        |
| Trackback                                      | 657        |
| Right Back to the Track                        | 657        |
| More Autodiscovery                             | 660        |
| Pulling Out the Comments                       | 661        |
| <b>Defining Extension Modules</b>              | <b>664</b> |
| RSS, Atom, and XML Namespaces                  | 665        |
| Expressing Things Clearly                      | 666        |
| <b>Other Resources</b>                         | <b>667</b> |
| <b>Summary</b>                                 | <b>667</b> |

---

|   |            |
|---|------------|
| <b>Chapter 32: What Lies Ahead in Information Management</b>    | <b>669</b> |
| <b>Filtering the Flood of Information</b>                       | <b>670</b> |
| Flexibility in Subscribing                                      | 670        |
| Finding the Best Feeds  | 670        |
| Filtering by Personality  | 670        |
| Finding the Wanted Non-Core Information                         | 671        |
| Point to Point or Hub and Spoke                                 | 671        |
| New Technologies  | 671        |
| <b>A User-Centric View of Information</b>                       | <b>672</b> |
| Information for Projects  | 672        |
| Automated Publishing Using Information Feeds                    | 672        |
| Ownership of Data   | 672        |
| <b>How Users Can Best Access Data</b>                           | <b>673</b> |
| How to Store Relevant Information                               | 673        |
| How to Retrieve Information from Storage                        | 673        |
| <b>Revisions of Information Feed Formats</b>                    | <b>674</b> |
| <b>Non-Text Information Delivered Through Information Feeds</b> | <b>674</b> |
| Product Support Uses  | 674        |
| Educational Uses  | 675        |
| <b>Summary</b>  | <b>675</b> |
| <br>  |            |
| <b>Appendix A: Answers to Exercises</b>                         | <b>677</b> |
| <br>  |            |
| <b>Appendix B: Useful Online Resources</b>                      | <b>697</b> |
| <br>  |            |
| <b>Appendix C: Glossary</b>                                     | <b>701</b> |
| <br>  |            |
| <b>Index</b>  | <b>705</b> |

