Contents

Foreword v

Preface vii

Overview ix

List of Figures xxiii

List of Tables xxvii

Introduction 1

1. Fundamentals 7
1.1 Document models 7
1.1.1 Multimedia 8
1.1.2 Hypermedia 9
1.1.3 Distribution 11
1.2 Web history 12
1.2.1 Organizational history 13
1.2.2 Technical history 14
1.3 Terminology 16
1.4 The Internet 18
1.4.1 The Internet environment 19
1.4.1.1 Connecting to the Internet 19
1.4.1.2 Names and addresses 20
1.4.2 Transferring data 24
1.4.2.1 Underlying networks 26
1.4.2.2 Internet Protocol (IP) 26
1.4.2.3 Transmission Control Protocol (TCP) 28
1.4.2.4 Application protocols 28
1.4.3 Data types 29
1.4.3.1 Identification of data types 29
1.4.3.2 Multipurpose Internet Mail Extensions . . . 30

Part I. Basics

2. Universal Resource Identifier (URI) 35
2.1 The need for resource identification 35
2.1.1 Web identifiers from a user perspective 36
2.1.2 The design of web identifiers 38
2.2 URI syntax 42
2.3 URI semantics 44
2.3.1 Uniform Resource Name (URN) 44
2.3.2 Uniform Resource Locator (URL) 47
2.4 The future of URI 50

3. Hypertext Transfer Protocol (HTTP) 53
3.1 History 54
3.1.1 HTTP/0.9 55
3.1.2 HTTP/1.0 55
3.1.3 HTTP/1.1 56
3.2 Hypertext Transfer Protocol 1.1 (HTTP/1.1) 58
3.2.1 Messages 61
3.2.2 General and entity headers 62
3.2.2.1 General headers 62
3.2.2.2 Entity headers 66
3.2.3 Request 69
3.2.3.1 Request methods 70
3.2.3.2 Request headers 74
3.2.4 Response 80
3.2.4.1 Status code definitions 81
3.2.4.2 Response headers 82
3.2.5 Content negotiation 85
3.2.5.1 Server-driven negotiation 86
3.2.5.2 Agent-driven negotiation 88
3.2.5.3 Transparent negotiation 89
3.2.6 Authentication 90
3.2.6.1 Basic authentication 92
3.2.6.2 Digest access authentication 93
3.2.7 Persistent connections 95
3.2.8 Chunked encoding 98
3.2.9 Caching 99
3.2.9.1 History of HTTP caching 103
3.2.9.2 Expiration model 104
3.2.9.3 Validation model 107

3.2.9.4 Cache control 109
3.2.9.5 Implementation of caching 114
3.3 Security 116
3.3.1 HTTP over SSL (HTTPS) 117
3.3.1.1 Secure Sockets Layer (SSL) 118
3.3.1.2 Combining HTTP and SSL 120
3.3.2 Secure HTTP (S-HTTP) 122
3.4 Cookies 122
3.5 Usage of HTTP 126
3.6 Non-standard HTTP extensions 128
3.6.1 Refreshing web pages 128
3.6.2 Interpage transitions 129
3.7 The future of HTTP 130
3.7.1 Protocol Extension Protocol (PEP) 133
3.7.2 HTTP next generation (HTTP-ng) 134
3.7.3 Multiplexing Protocol (SMUX) 134

4. Standard Generalized Markup Language (SGML) 137
4.1 SGML concepts 138
4.1.1 Content and presentation 139
4.1.2 Using markup for structuring 141
4.1.3 Document classes 144
4.1.4 The presentation of content 147
4.2 The SGML standard 150
4.2.1 SGML declaration 153
4.2.1.1 Defining a concrete syntax 153
4.2.1.2 Used character sets 155
4.2.1.3 Processing capacities 155
4.2.1.4 Features 155
4.2.2 Document Type Definition (DTD) 157
4.2.2.1 Elements 158
4.2.2.2 Model groups 159
4.2.2.3 Text elements 161
4.2.2.4 Exceptions 161
4.2.2.5 Attributes 162
4.2.2.6 Entities 164
4.2.3 SGML documents 166
4.3 Usage of SGML 167
4.3.1 SGML profiles 167
4.3.2 SGML applications 168
4.3.3 SGML validation 169
4.4 The future of SGML 171

5. Hypertext Markup Language (HTML) 173
5.1 History 174
5.1.1 Early versions 175
5.1.2 HTML 2.0 175
5.1.3 HTML 3.2 176
5.1.4 HTML 4.0 176
5.2 Hypertext Markup Language 4.0 (HTML 4.0) 177
5.2.1 HTML 4.0 DTDs 178
5.2.2 Common attributes 179
5.2.3 Basic structure of an HTML document 184
5.2.3.1 Document head 186
5.2.3.2 Document body 195
5.2.4 Simple and structured text 197
5.2.4.1 Text 197
5.2.4.2 Lists 204
5.2.4.3 Tables 206
5.2.5 Images and image maps 211
5.2.5.1 Images 213
5.2.5.2 Image maps 214
5.2.6 Links 221
5.2.6.1 Source anchors 223
5.2.6.2 Destination anchors 223
5.2.7 Frames and forms 224
5.2.7.1 Frames 225
5.2.7.2 Forms 234
5.2.8 Dynamic documents 245
5.2.8.1 Scripts 245
5.2.8.2 Objects 246
5.3 Publishing with HTML 254
5.4 Usage of HTML 259
5.5 The future of HTML 261

Part II. Advanced

6. Cascading Style Sheets (CSS) 265
6.1 Principles of CSS 265
6.2 History 269
6.3 Cascading Style Sheets, level 1 (CSS1) 270
6.3.1 Inheritance 271
6.3.2 Selectors 274
6.3.2.1 Type selectors 274

6.3.2.2 Attribute selectors 275
6.3.2.3 Contextual selectors 278
6.3.2.4 External selectors 278
6.3.2.5 New elements 282
6.3.2.6 Combining selector types 283
6.3.3 Units 283
6.3.3.1 Colors 284
6.3.3.2 Lengths 285
6.3.4 Declarations 286
6.3.4.1 Fonts 286
6.3.4.2 Space 291
6.3.4.3 Images 301
6.3.4.4 Colors and backgrounds 303
6.3.4.5 Classification properties 307
6.3.5 Using multiple style sheets 311
6.3.5.1 Cascading 312
6.3.5.2 Conflicts 316
6.3.5.3 Conflict resolution 317
6.4 Related approaches 318
6.5 Converting to CSS 319
6.6 The future of CSS 320
6.6.1 Cascading Style Sheets Positioning (CSS-P) 320
6.6.2 Cascading Style Sheets, level 2 (CSS2) 321

7. Extensible Markup Language (XML) 325
7.1 Principles of XML 326
7.1.1 Creating document types 327
7.1.2 Formatting of XML documents 329
7.1.2.1 Focusing on content 329
7.1.2.2 Using style sheet languages 331
7.1.3 Conversion 333
7.1.3.1 Other formats to XML 334
7.1.3.2 XML to XML 335
7.1.3.3 XML to other formats 336
7.1.3.4 HTML and XML 336
7.2 Extensible Markup Language 1.0 (XML 1.0) 338
7.2.1 Differences to SGML 340
7.3 XML Linking Language (XLink) 341
7.3.1 Linking concepts 342
7.3.1.1 Link location 343
7.3.1.2 Link topology 345
7.3.2 Link information 347

7.3.3 Link types 349
7.3.3.1 Simple links 349
7.3.3.2 Extended links 349
7.3.3.3 Extended link groups 350
7.3.4 Attributes 350
7.4 XML Pointer Language (XPointer) 354
7.4.1 Concepts 355
7.4.2 Location terms 356
7.4.2.1 Absolute term 356
7.4.2.2 Relative term 357
7.4.2.3 Spanning term 361
7.4.2.4 Attribute term 361
7.4.2.5 String term 361
7.4.3 Spanning locators 362
7.4.4 Persistence 362
7.5 Extensible Style Language (XSL) 364
7.5.1 Concepts 365
7.5.1.1 Tree construction 366
7.5.1.2 Formatting 367
7.5.2 Style sheets 368
7.6 XML and other markup languages 369
7.6.1 Comparison of XML and HTML 369
7.6.2 Comparison of XML and SGML 370
7.6.2.1 Restrictions of XML 371
7.6.2.2 Converting SGML documents to XML 371
7.6.2.3 Converting SGML environments to XML . . 371
7.7 The future of XML 372

8. Scripting and Programming 375
8.1 Scripting languages 375
8.1.1 ECMAScript 377
8.1.2 VBScript 378
8.2 Programming languages 378
8.2.1 Java 379
8.3 Distributed programming 381
8.3.1 Remote Method Invocation (RMI) 382
8.3.2 Common Object Request Broker Architecture 382
8.3.3 Distributed Component Object Model (DCOM) . . . 384
8.3.4 Interworking 384

9. HTTP Servers 387
9.1 Server performance 388
9.2 Server configuration 389
9.2.1 Server start 389
9.2.1.1 Started on request 390
9.2.1.2 Permanently running 391
9.2.2 Virtual hosts 391
9.2.2.1 IP-based virtual hosts 392
9.2.2.2 Non IP-based virtual hosts 393
9.2.3 Origin servers 394
9.2.3.1 General file system layout 395
9.2.3.2 User specific web pages 396
9.2.3.3 Handling requests 397
9.2.3.4 CGI scripts 399
9.2.4 Proxies 400
9.2.4.1 Caching 400
9.2.4.2 Security 401
9.3 The Apache HTTP server 403
9.3.1 Installation 404
9.3.1.1 Program installation 404
9.3.1.2 Startup installation 405
9.3.2 Configuration 406
9.3.2.1 Virtual hosts 406
9.3.2.2 Authentication 407
9.3.2.3 Content negotiation 408
9.3.2.4 Proxies 409
9.3.2.5 Server-Side Includes (SSI) 409
9.4 Common Gateway Interface (CGI) 411
9.4.1 Communication with the server 413
9.4.1.1 Environment variables 413
9.4.1.2 Command line 416
9.4.1.3 Standard input 416
9.4.1.4 Standard output 417
9.4.2 Forms and CGI 418
9.4.3 Server-Side Includes (SSI) 419

10. Miscellaneous 421
10.1 Server-side technologies 421
10.1.1 FastCGI 422
10.1.2 Java servlets 423
10.2 Browsers 423
10.2.1 Integration 424

10.2.2 Prefetching 426
10.2.2.1 Reducing user-perceived latency 427
10.2.2.2 Reducing network load 427
10.3 Search engines 428
10.4 New content types 430
10.4.1 Virtual Reality Modeling Language (VRML) 430
10.4.2 Mathematical Markup Language (MathML) 432
10.4.3 Portable Network Graphics (PNG) 434
10.4.4 Vector graphics 436
10.4.5 RealMedia Architecture (RMA) 437
10.5 New architectural components 437
10.5.1 Platform for Internet Content Selection (PICS) . . . 438
10.5.1.1 Services and ratings 439
10.5.1.2 Label distribution 439
10.5.1.3 PICSRules 440
10.5.2 Digital Signature Initiative (DSig) 441
10.5.3 Resource Description Framework (RDF) 442
10.5.4 Dynamic fonts 443
10.5.4.1 Embedded TrueType 443
10.5.4.2 TrueDoc 444
10.5.5 Document Object Model (DOM) 445
10.5.6 Synchronized Multimedia Integration Language . . . 446

11. Related Technology 449
11.1 Electronic mail 450
11.1.1 Sending electronic mail 451
11.1.1.1 Simple Mail Transfer Protocol (SMTP) . . . 452
11.1.1.2 Extended SMTP (ESMTP) 452
11.1.2 Receiving electronic mail 453
11.1.2.1 Post Office Protocol (POP) 454
11.1.2.2 Internet Message Access Protocol (IMAP) . 455
11.2 Directory services 456
11.2.1 Directory Access Protocol (DAP) 459
11.2.2 Lightweight Directory Access Protocol (LDAP) . . . 459
11.3 Usenet news 460
11.3.1 Network News Transfer Protocol (NNTP) 461

Part III. Appendices

A. HTTP/1.1 Definitions 465
A.1 Messages 465
A.1.1 General definitions 465
A.1.2 Request 466
A.1.3 Response 467
A.2 Status codes 467
A.2.1 Informational (1xx) 467
A.2.2 Successful (2xx) 468
A.2.3 Redirection (3xx) 468
A.2.4 Client error (4xx) 468
A.2.5 Server error (5xx) 469
A.3 Warn codes 469
A.3.1 Temporary warnings (Ixx) 469
A.3.2 Persistent warnings (2xx) 470

B. HTML 4.0 Definitions 471
B.1 SGML declaration 471
B.2 DTD Entities 473
B.3 HTML 4.0 table DTD 485
B.4 HTML 4.0 form DTD 487

C. XML 1.0 Definitions 491
C.1 SGML declaration 491
C.2 XML DTD for SMIL 495

References 501

Glossary 515

Index 577