

The L^AT_EX Web Companion

Integrating T_EX, HTML, and XML

Michel Goossens
CERN, Geneva, Switzerland

Sebastian Rahtz
Elsevier Science Ltd., Oxford, United Kingdom

with Eitan M. Gurari, Ross Moore, and
Robert S. Sutor



ADDISON-WESLEY

Boston • San Francisco • New York • Toronto • Montreal
London • Munich • Paris • Madrid
Capetown • Sydney • Tokyo • Singapore • Mexico City

Contents

List of Figures	xi
List of Tables	xv
Preface	xvii
1 The Web, its documents, and \LaTeX	1
1.1 The Web, a window on the Internet	3
1.1.1 The Hypertext Transport Protocol	4
1.1.2 Universal Resource Locators and Identifiers	5
1.1.3 The Hypertext Markup Language	6
1.2 \LaTeX in the Web environment	11
1.2.1 Overview of document formats and strategies	12
1.2.2 Staying with DVI	14
1.2.3 PDF for typographic quality	15
1.2.4 Down-translation to HTML	16
1.2.5 Java and browser plug-ins	20
1.2.6 Other \LaTeX -related approaches to the Web	21
1.3 Is there an optimal approach?	23
1.4 Conclusion	24
2 Portable Document Format	25
2.1 What is PDF?	26
2.2 Generating PDF from \TeX	27
2.2.1 Creating and manipulating PDF	28

2.2.2	Setting up fonts	29
2.2.3	Adding value to your PDF	33
2.3	Rich PDF with \LaTeX : The hyperref package	35
2.3.1	Implicit behavior of hyperref	36
2.3.2	Configuring hyperref	38
2.3.3	Additional user macros for hyperlinks	45
2.3.4	Acrobat-specific commands	47
2.3.5	Special support for other packages	49
2.3.6	Creating PDF and HTML forms	50
2.3.7	Designing PDF documents for the screen	59
2.3.8	Catalog of package options	62
2.4	Generating PDF directly from \TeX	67
2.4.1	Setting up pdf \TeX	67
2.4.2	New primitives	74
2.4.3	Graphics and color	80
3	The \LaTeX2HTML translator	83
3.1	Introduction	83
3.1.1	A few words on history	84
3.1.2	Principles for Web document generation	84
3.2	Required software and customization	86
3.2.1	Running \LaTeX 2HTML on a \LaTeX document	87
3.2.2	Installation	92
3.2.3	Customizing the local installation	98
3.2.4	Extension mechanisms and \LaTeX packages	100
3.3	Mathematics modes with \LaTeX 2HTML	101
3.3.1	An overview of \LaTeX 2HTML's math modes	102
3.3.2	Advanced mathematics with the math extension	105
3.3.3	Unicode fonts and named entities, in expert mode	108
3.3.4	HTML 4.0 and style sheets	110
3.3.5	Large images and HTML 2.0	112
3.3.6	Future use of MathML	114
3.4	Support for different languages	115
3.4.1	Titles and keywords	116
3.4.2	Character-set encodings	118
3.4.3	Multilingual documents using babel	119
3.4.4	Images using special fonts	120
3.4.5	Converting transliterations using preprocessors	120
3.5	Extending \LaTeX sources with hypertext commands using the html package	124
3.5.1	Hyperlinks to external documents	126
3.5.2	Enhancements appropriate for HTML	128
3.5.3	Alternative text for hyperlinks	132
3.5.4	Conditional environments	135

3.5.5	Navigation and layout of HTML pages	137
3.5.6	Example of linking various external documents	141
3.5.7	Advanced features	145
4	Translating L^AT_EX to HTML using T_EX4ht	155
4.1	Using T _E X4ht	156
4.1.1	Package options	156
4.1.2	Picture representation of special content	159
4.2	A complete example	160
4.3	Manual creation of hypertext elements	164
4.3.1	Raw hypertext code	164
4.3.2	Hypertext pages	166
4.3.3	Hypertext links	167
4.3.4	Cascading Style Sheets	167
4.4	How T _E X4ht works	169
4.4.1	From L ^A T _E X to DVI	169
4.4.2	From DVI to HTML	169
4.4.3	Other matters	170
4.5	Extended customization of T _E X4ht	170
4.5.1	Configuration files	170
4.5.2	Tables of contents	172
4.5.3	Parts, chapters, sections, and so on	175
4.5.4	Defining sectioning commands	177
4.5.5	Lists	178
4.5.6	Environments	179
4.5.7	Tables	180
4.5.8	Small details	182
4.6	The inner workings of T _E X4ht	184
4.6.1	The translation process	184
4.6.2	Running L ^A T _E X	185
4.6.3	Running the tex4ht program	186
4.6.4	A look at t4ht	187
4.6.5	From DVI to GIF	188
4.6.6	A taste of the lg file	189
4.6.7	The font control files	190
4.6.8	The control file	193
5	Direct display of L^AT_EX on the Web	195
5.1	IBM techexplorer Hypermedia Browser	196
5.1.1	Basic formatting issues	198
5.1.2	Your browser and techexplorer	200
5.1.3	Adding hypertext links	204
5.1.4	Popping up windows and footnotes	208
5.1.5	Using images, sound, and video	210

5.1.6	Defining and using pop-up menus	211
5.1.7	Using color in your documents	215
5.1.8	Building a document hierarchy	218
5.1.9	Running applications	220
5.1.10	Alternating between two displayed expressions	220
5.1.11	Printing from techexplorer	221
5.1.12	Searching in a document	222
5.1.13	Optimizing your documents for techexplorer	222
5.1.14	Scripting techexplorer from Java and JavaScript	223
5.2	WebEQ	224
5.2.1	An introduction to Web \TeX	225
5.2.2	Adding interactivity	229
5.2.3	Using the APPLET tag with WebEQ	230
5.2.4	Preparing HTML pages via the WebEQ Wizard	232
5.3	Embedded content problems and future developments	234
5.3.1	Expression size	235
5.3.2	Ambient style	236
6	HTML, SGML, and XML: Three markup languages	239
6.1	Will HTML lead to the downfall of the Web?	239
6.2	HTML 4: A richer and more coherent language	241
6.2.1	HTML 4 goodies	242
6.2.2	HTML 4, the end of the old road	243
6.3	Why SGML?	243
6.3.1	Different types of markup	244
6.3.2	Generalized logical markup	245
6.3.3	SGML to HTML and XML	247
6.4	Extensible Markup Languages	248
6.4.1	What is XML?	249
6.4.2	The components of XML	251
6.4.3	Declaring document elements	255
6.5	The detailed structure of an XML document	256
6.5.1	XML is truly international	257
6.5.2	XML document components	258
6.5.3	The XML declaration	258
6.5.4	The document type declaration	259
6.5.5	Document elements	270
6.6	XML parsers and tools	271
6.6.1	Emacs and psgml	272
6.6.2	The perlSGML programs	275
6.6.3	The DTDParse tool	277
6.6.4	The Language Technology Group XML toolbox	277
6.6.5	Validating documents with XML parsers	281

7	CSS, DSSSL, and XSL: Doing it with style	289
7.1	Style sheet languages: A short history	289
7.2	Programming or style sheets, which is better?	291
7.3	Formatting with Perl	292
7.3.1	Principles of operation	293
7.3.2	Generating a \LaTeX instance	294
7.4	Cascading Style Sheets	297
7.4.1	The basic structure of a CSS style sheet	298
7.4.2	Associating style sheets with a document	302
7.4.3	A quick look at CSS properties	303
7.4.4	CSS style sheets for formatting XML documents	306
7.4.5	The invitation example revisited	309
7.4.6	Generating HTML with another document instance	311
7.5	Document Style Semantics and Specification Language	312
7.5.1	The components of DSSSL	313
7.5.2	Creating style sheets with DSSSL	315
7.5.3	Introducing Jade	318
7.5.4	The \TeX back-end for Jade and the Jade \TeX macros	325
7.5.5	The Jade SGML transformation interface	331
7.5.6	Formatting real-life documents with DSSSL	335
7.6	Extensible Stylesheet Language	337
7.6.1	XPath for addressing parts of an XML document	338
7.6.2	The XSL Transformation Language	343
7.6.3	Formatting objects and their properties	349
7.6.4	XSL processors and tools	350
7.6.5	Using XSL to generate HTML or \LaTeX	351
7.6.6	Using XSL to generate formatting objects	355
7.6.7	XML, XSL and databases	360
8	MathML, intelligent math markup	367
8.1	Introduction to MathML	368
8.1.1	MathML, Unicode, and XML entities	371
8.2	MathML software	372
8.2.1	Equation editors	373
8.2.2	Web browser support for MathML	376
8.2.3	Converting \LaTeX to MathML	379
8.2.4	Typesetting MathML	387
A	Example files	391
A.1	An example \LaTeX file and its translation to XML	391
A.1.1	The \LaTeX source	391
A.1.2	\LaTeX converted to XML	393
A.1.3	Document Type Definition for XML version	396
A.2	Scripting examples for techexplorer	399

A.2.1	teched.html	399
A.2.2	teched.java	400
B	Technical appendixes	403
B.1	The HyperTeX standard	403
B.2	Configuring T _E X4ht to produce XML	404
B.2.1	Starting from scratch	404
B.2.2	Adding XML tags	407
B.2.3	Getting deeper for extra configurations	410
B.3	XML namespaces	415
B.4	Examples of important DTDs	417
B.4.1	The DocBook DTD	417
B.4.2	The AAP effort and ISO 12083	419
B.4.3	Text Encoding Initiative	420
B.4.4	A DTD for BIB _T E _X	421
B.4.5	L ^A T _E X-like markup, from DTD to printed document	433
B.5	Transforming HTML into XML	450
B.5.1	HTML in XML	452
B.5.2	The Extensible HyperText Markup Language	454
B.6	Java event-based interface	459
B.6.1	The SAX Java classes	459
B.6.2	Running a SAX application	460
C	Internationalization issues	465
C.1	Codes for languages, countries, and scripts	465
C.2	The Unicode standard	475
C.2.1	Character codes and glyphs	477
C.2.2	Unicode and ISO/IEC 10646-1	477
C.2.3	UTF-8 and UTF-16 encodings	479
C.3	Foreign languages in XML	480
C.3.1	Latin-based encodings	480
C.3.2	Handling non-Latin encodings with UTF-8	483
	Glossary	489
	URL catalog	499
	Bibliography	513
	Index	517