

Content

<i>Introduction</i>	<i>page ix</i>
Part 1 Modeling Web Data	
1 Data Model	3
1.1 Semistructured Data	3
1.2 XML	5
1.3 Web Data Management with XML	16
1.4 The XML World	18
1.5 Further Reading	28
1.6 Exercises	29
2 XPath and XQuery	32
2.1 Introduction	32
2.2 Basics	32
2.3 XPath	42
2.4 FLWOR Expressions in XQuery	54
2.5 XPath Foundations	62
2.6 Further Reading	67
2.7 Exercises	69
3 Typing	72
3.1 Motivating Typing	72
3.2 Automata	75
3.3 Schema Languages for XML	80
3.4 Typing Graph Data	89
3.5 Further Reading	91
3.6 Exercises	92

Cambridge University Press

978-1-107-01243-1 - Web Data Management

Serge Abiteboul, Ioana Manolescu, Philippe Rigaux, Marie-Christine Rousset and Pierre Senellart

Table of Contents

[More information](#)

vi	Content	
4	XML Query Evaluation	95
4.1	Fragmenting XML Documents on Disk	97
4.2	XML Node Identifiers	99
4.3	XML Query Evaluation Techniques	103
4.4	Further Reading	112
4.5	Exercises	113
5	Putting into Practice: Managing an XML Database with eXIST	116
5.1	Prerequisites	116
5.2	Installing eXIST	117
5.3	Getting Started with eXIST	118
5.4	Running XPath and XQuery Queries with the Sandbox	120
5.5	Programming with eXIST	123
5.6	Projects	127
6	Tree Pattern Evaluation Using SAX	131
6.1	Tree-Pattern Dialects	131
6.2	CTP Evaluation	134
6.3	Extensions to Richer Tree Patterns	138
Part 2 Web Data Semantics and Integration		
7	Ontologies, RDF, and OWL	143
7.1	Introduction	143
7.2	Ontologies by Example	145
7.3	RDF, RDFS, and OWL	148
7.4	Ontologies and (Description) Logics	159
7.5	Further Reading	169
7.6	Exercises	170
8	Querying Data Through Ontologies	171
8.1	Introduction	171
8.2	Querying RDF Data: Notation and Semantics	172
8.3	Querying Through RDFS Ontologies	176
8.4	Answering Queries Through DL-LITE Ontologies	179
8.5	Further Reading	194
8.6	Exercises	195
9	Data Integration	196
9.1	Introduction	196
9.2	Containment of Conjunctive Queries	199
9.3	Global-as-View Mediation	200
9.4	Local-as-View Mediation	204
9.5	Ontology-Based Mediators	215
9.6	Peer-to-Peer Data Management Systems	222
9.7	Further Reading	229
9.8	Exercises	229

Cambridge University Press

978-1-107-01243-1 - Web Data Management

Serge Abiteboul, Ioana Manolescu, Philippe Rigaux, Marie-Christine Rousset and Pierre Senellart

Table of Contents

[More information](#)

	Content	vii
10 Putting into Practice: Wrappers and Data Extraction with XSLT	231	
10.1 Extracting Data from Web Pages	232	
10.2 Restructuring Data	234	
11 Putting into Practice: Ontologies in Practice (by Fabian M. Suchanek)	236	
11.1 Exploring and Installing Yago	236	
11.2 Querying Yago	237	
11.3 Web Access to Ontologies	238	
12 Putting into Practice: Mashups with YAHOO! PIPES and XProc	240	
12.1 YAHOO! PIPES: A Graphical Mashup Editor	240	
12.2 XProc: An XML Pipeline Language	241	
Part 3 Building Web Scale Applications		
13 Web Search	247	
13.1 The World Wide Web	248	
13.2 Parsing the Web	250	
13.3 Web Information Retrieval	257	
13.4 Web Graph Mining	272	
13.5 Hot Topics in Web Search	280	
13.6 Further Reading	281	
13.7 Exercises	283	
14 An Introduction to Distributed Systems	287	
14.1 Basics of Distributed Systems	288	
14.2 Failure Management	295	
14.3 Required Properties of a Distributed System	299	
14.4 Particularities of P2P Networks	303	
14.5 Case Study: A Distributed File System for Very Large Files	305	
14.6 Further Reading	308	
15 Distributed Access Structures	310	
15.1 Hash-Based Structures	311	
15.2 Distributed Indexing: Search Trees	325	
15.3 Further Reading	336	
15.4 Exercises	337	
16 Distributed Computing with MAPREDUCE and PIG	339	
16.1 MAPREDUCE	341	
16.2 PIG	348	
16.3 Further Reading	359	
16.4 Exercises	361	

Cambridge University Press

978-1-107-01243-1 - Web Data Management

Serge Abiteboul, Ioana Manolescu, Philippe Rigaux, Marie-Christine Rousset and Pierre Senellart

Table of Contents

[More information](#)viii **Content**

17	Putting into Practice: Full-Text Indexing with LUCENE (by Nicolas Travers)	364
17.1	Preliminary: A LUCENE Sandbox	364
17.2	Indexing Plain Text with LUCENE – A Full Example	364
17.3	Put It into Practice!	371
17.4	LUCENE – Tuning the Scoring (Project)	372
18	Putting into Practice: Recommendation Methodologies	374
18.1	Introduction to Recommendation Systems	374
18.2	Prerequisites	375
18.3	Data Analysis	377
18.4	Generating Some Recommendations	380
18.5	Projects	385
19	Putting into Practice: Large-Scale Data Management with HADOOP	387
19.1	Installing and Running HADOOP	388
19.2	Running MAPREDUCE Jobs	391
19.3	PIGLATIN Scripts	395
19.4	Running in Cluster Mode (Optional)	395
19.5	Exercises	397
20	Putting into Practice: COUCHDB, a JSON Semistructured Database	400
20.1	Introduction to the COUCHDB Document Database	401
20.2	Putting COUCHDB into Practice!	417
20.3	Further Reading	419
	<i>Bibliography</i>	421
	<i>Index</i>	431