

Roberto De Virgilio
Fausto Giunchiglia
Letizia Tanca (Eds.)

Semantic Web Information Management

A Model-Based Perspective



Springer

Contents

1	Introduction	1
	Roberto De Virgilio, Fausto Giunchiglia, and Letizia Tanca	
2	Data and Metadata Management	9
	Paolo Atzeni and Riccardo Torlone	
2.1	Introduction	9
2.2	Databases, Schemas and Dictionaries	10
2.3	Data Models	12
2.4	Management of Multiple Models	14
2.5	Model-generic Schema Translation	18
2.6	Related Work	21
2.7	Conclusion	22
	References	22
3	The Semantic Web Languages	25
	Fausto Giunchiglia, Feroz Farazi, Letizia Tanca, and Roberto De Virgilio	
3.1	Introduction	25
3.2	The Hierarchy of Languages	27
3.3	RDF(S)	29
3.4	OWL	31
3.5	C-OWL	33
3.6	Semantic Web and Databases	34
3.7	Conclusion	36
	Appendix A RDF(S) Constructs	36
	Appendix B OWL Constructs	36
	References	37
Part I Semantic Web Data Storage		
4	Relational Technologies, Metadata and RDF	41
	Yannis Velegrakis	
4.1	Introduction	41
4.2	The Relational Model	44

4.3	Modeling Metadata in Relational Systems	45
4.3.1	Separate Structures	45
4.3.2	Intermixed Context	45
4.3.3	Intensional Associations	47
4.4	RDF	48
4.5	Using Relational Systems for RDF Storage	50
4.5.1	Storing RDF as XML	51
4.5.2	Vertical Table	53
4.5.3	Graph-based Storage	54
4.5.4	Graph Schema–Vertical Data	56
4.5.5	Property Table	58
4.5.6	Vertical Partitioning	60
4.5.7	Smart Indexing	62
4.6	Conclusion	64
	References	64
5	A Metamodel Approach to Semantic Web Data Management	67
	Roberto De Virgilio, Pierluigi Del Nostro, Giorgio Gianforme, and Stefano Paolozzi	
5.1	Introduction	67
5.2	Motivating Example	69
5.3	Management of RDF Data	70
5.3.1	A Conceptual Representation	72
5.3.2	Logical Level	75
5.3.3	Physical Level	77
5.3.4	Query Processing	79
5.4	Related Work	79
5.5	Experimental Results	82
5.5.1	RDF Benchmark	82
5.5.2	Platform Environment	85
5.5.3	Evaluating Results	86
5.6	Conclusion	89
	References	90
6	Managing Terabytes of Web Semantics Data	93
	Michele Catasta, Renaud Delbru, Nickolai Toupikov, and Giovanni Tummarello	
6.1	Introduction	93
6.1.1	Definition and Model for the Web of Data	94
6.2	Providing Services on the Web of Data: A Model for a Large Scale Semantic Data Processing Infrastructure	95
6.3	Semantic Sitemap: A Dataset Publishing Model	97
6.3.1	The Sitemap Protocol and robots.txt	98
6.3.2	The Semantic Sitemaps Extension	98
6.4	Pre-processing: Context-dependent Reasoning	100
6.4.1	Contexts on the Semantic Web	101

6.4.2	Import Closure of RDF Models	102
6.4.3	Deductive Closure of RDF Models	103
6.4.4	Technical Overview	104
6.4.5	Performance Overview	105
6.4.6	Discussion	105
6.5	Indexing: The SIREn Model	106
6.5.1	SIREn Data Model	106
6.5.2	SIREn Query Model	108
6.5.3	Experimental Benchmark	110
6.6	Ranking: The DING Model	112
6.6.1	A Two-layer Model for Web Data	113
6.6.2	DING Algorithm	114
6.6.3	Combining DING and Entity Rank	115
6.7	Leveraging Sindice in SIGMA	115
6.7.1	Sig.ma: Processing Dataflow and Interaction with Sindice	116
6.8	Conclusion	120
	References	122

Part II Reasoning in the Semantic Web

7	Reasoning in Semantic Web-based Systems	127
	Florian Fischer and Gulay Unel	
7.1	Introduction	127
7.2	Background	128
7.2.1	Logic Programming	128
7.2.2	Description Logics	130
7.3	Standards on the Semantic Web	131
7.3.1	OWL 2	131
7.3.2	WSML	132
7.4	Reasoning Techniques	133
7.4.1	Description Logic Paradigm	133
7.4.2	Rule Paradigm	135
7.5	Reasoning on the Web	139
7.5.1	Expressivity	140
7.5.2	Approximate Reasoning	142
7.6	Conclusion	144
	References	144
8	Modular Knowledge Representation and Reasoning in the Semantic Web	147
	Luciano Serafini and Martin Homola	
8.1	Introduction	147
8.2	Ontologies and Description Logics	149
8.2.1	Description Logic $\mathcal{SHOIQ}_b$	150
8.2.2	Sublanguages of $\mathcal{SHOIQ}_b$	152
8.3	Reference Distributed Ontology Framework	153

8.4	Distributed Description Logics	154
8.4.1	Formalization	154
8.4.2	Reasoning in DDL	156
8.4.3	Modeling with DDL	157
8.5	$\mathcal{E}$ -connections	160
8.5.1	Formalization	160
8.5.2	Reasoning in $\mathcal{E}$ -connections	164
8.5.3	Modeling with $\mathcal{E}$ -connections	165
8.6	Package-based Description Logics	168
8.6.1	Formalization	168
8.6.2	Reasoning in P-DL	170
8.6.3	Modeling with P-DL	170
8.7	Integrated Distributed Description Logics	173
8.7.1	Formalization	173
8.7.2	Reasoning in IDDL	175
8.7.3	Modeling with IDDL	175
8.8	<i>Conclusion</i>	177
	References	179
9	Semantic Matching with S-Match	183
	Pavel Shvaiko, Fausto Giunchiglia, and Mikalai Yatskevich	
9.1	Introduction	183
9.2	State of the Art	184
9.3	Semantic Matching	186
9.3.1	The Tree Matching Algorithm	187
9.3.2	Node Matching Algorithm	192
9.4	Efficient Semantic Matching	193
9.4.1	Conjunctive Concepts at Nodes	194
9.4.2	Disjunctive Concepts at Nodes	195
9.5	The S-Match Architecture	196
9.6	Evaluation	196
9.6.1	Evaluation Set Up	196
9.6.2	Evaluation Results	198
9.7	<i>Conclusion</i>	200
	References	200
10	Preserving Semantics in Automatically Created Ontology Alignments 203	
	Giorgio Orsi and Letizia Tanca	
10.1	Introduction	203
10.2	Related Research Fields	204
10.3	Preliminaries	205
10.3.1	Ontologies	205
10.3.2	Correspondences and Mappings	206
10.3.3	Distributed Ontologies	207
10.3.4	Faults and Diagnoses	208

10.4	Mapping Revision	209
10.4.1	Model-based Mapping Revision	210
10.4.2	Heuristic Mapping Revision	213
10.4.3	Uncertainty-Aware Mapping Revision	218
10.5	Conclusion	220
	References	221
11	tOWL: Integrating Time in OWL	225
	Flavius Frasinca, Viorel Milea, and Uzay Kaymak	
11.1	Introduction	225
11.2	Preliminaries	227
11.2.1	Concrete Domains	227
11.2.2	4D Fluents	228
11.3	tOWL	230
11.3.1	tOWL Overview	230
11.3.2	OWL Schema of tOWL	230
11.4	A tOWL Ontology for the Leveraged Buyouts	234
11.4.1	Leveraged Buyouts	235
11.4.2	TBox	235
11.4.3	ABox	239
11.4.4	Use Cases	241
11.5	Related Work	242
11.5.1	Temporal RDF	242
11.5.2	OWL-Time	243
11.5.3	4D Fluents	244
11.6	Conclusion	245
	References	246

Part III Semantic Web Data Querying

12	Datalog Extensions for Tractable Query Answering over Ontologies	249
	Andrea Cali, Georg Gottlob, and Thomas Lukasiewicz	
12.1	Introduction	249
12.2	Preliminaries	252
12.3	Guarded Datalog [±]	257
12.4	Linear Datalog [±]	262
12.5	Weakly Guarded Datalog [±]	264
12.6	Extensions	268
12.6.1	Negative Constraints	268
12.6.2	Non-conflicting Keys	269
12.7	Ontology Querying	271
12.7.1	DL-Lite _A	271
12.7.2	F-Logic Lite	275
12.8	Conclusion	277
	References	278

13	On the Semantics of SPARQL	281
	Marcelo Arenas, Claudio Gutierrez, and Jorge Pérez	
13.1	Introduction	281
13.2	The W3C Syntax of SPARQL	283
13.2.1	Basic Definitions	284
13.2.2	Basic Structures	284
13.2.3	More Complex Queries	286
13.2.4	Final Remarks	288
13.3	An Algebraic Syntax for SPARQL	289
13.3.1	Translating SPARQL into the Algebraic Formalism	291
13.4	Semantics of SPARQL	293
13.4.1	Blank Nodes in Graph Patterns	299
13.4.2	Bag Semantics of SPARQL	300
13.5	On the Complexity of the Evaluation Problem	301
13.6	Related Work	305
13.7	Conclusion	306
	References	307
14	Labeling RDF Graphs for Linear Time and Space Querying	309
	Tim Furche, Antonius Weinzierl, and François Bry	
14.1	Introduction	310
14.1.1	Contributions	311
14.2	Motivating Example	312
14.3	Preliminaries—RDF as Graphs	314
14.3.1	Queries on RDF Graphs	315
14.3.2	Triple Patterns and Adjacency	318
14.4	Labeling Schemes for RDF Graphs	319
14.4.1	Foundation: Tree Labeling	319
14.4.2	Reachability in Graphs	322
14.5	CIG-labeling Scheme	323
14.5.1	Label Size	325
14.5.2	Adjacency & Reachability Test	326
14.5.3	Optimal CIG-labeling	327
14.6	CIG-labeling on Trees and CIGs	328
14.6.1	CIGs: Sharing-Limited Graphs	328
14.6.2	Labeling CIGs and Trees	330
14.6.3	Properties of CIG-labelings on CIGs and Trees	330
14.6.4	Limitations and Extensions	331
14.7	Processing SPARQL with CIG-labelings	332
14.7.1	Processing 1-SPARQL	332
14.7.2	Processing A-SPARQL	333
14.7.3	Towards Full SPARQL	335
14.7.4	Towards Path Expressions	336
14.8	Conclusion	336
	References	337

15	SPARQLog: SPARQL with Rules and Quantification	341
	François Bry, Tim Furche, Bruno Marnette, Clemens Ley, Benedikt Linse, and Olga Poppe	
15.1	Introduction	342
15.2	Preliminaries	344
15.3	SPARQL Rule Languages	347
15.3.1	SPARQL and Rule Extensions of SPARQL	347
15.3.2	Other Rule-based RDF Query Languages	348
15.3.3	Quantifier Alternation in Data Exchange	349
15.4	SPARQLog: SPARQL with Rules and Quantification	350
15.4.1	SPARQLog Syntax	352
15.4.2	Denotational Semantics for SPARQLog	353
15.4.3	Relational Operational Semantics for SPARLog	355
15.5	Properties of SPARQLog	360
15.5.1	Designing Tractable Fragments of SPARQLog	360
15.5.2	Expressiveness of Quantifier Alternation in SPARQLog	364
15.5.3	Experimental Comparison with SPARQL Engines	367
15.6	Conclusion	369
	References	369
16	SP²Bench: A SPARQL Performance Benchmark	371
	Michael Schmidt, Thomas Hornung, Michael Meier, Christoph Pinkel, and Georg Lausen	
16.1	Introduction	371
16.2	Benchmark Design Decisions	374
16.3	The SP ² Bench Data Generator	375
16.3.1	Characteristics of DBLP Data	375
16.3.2	Data Generator Implementation and RDF Scheme	379
16.4	The SP ² Bench Benchmark Queries	382
16.4.1	RDF Characteristics	382
16.4.2	SPARQL Characteristics	384
16.4.3	Discussion of Benchmark Queries	384
16.5	Benchmark Metrics	390
16.6	Conclusion	391
	References	392

Part IV Semantic Web Applications

17	Using OWL in Data Integration	397
	Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, Riccardo Rosati, and Marco Ruzzi	
17.1	Introduction	398
17.2	The Data Integration Framework	400
17.3	Computational Characterization of Query Answering	406
17.4	Data Integration Using OWL 2 QL	409
17.5	Related Work	414

17.6 Conclusion	419
References	419
18 Service Knowledge Spaces for Semantic Collaboration in Web-based Systems	425
Devis Bianchini, Valeria De Antonellis, and Michele Melchiori	
18.1 Introduction	425
18.2 Semantic Collaboration in Networked Web-based Systems	426
18.3 Service Knowledge Space	429
18.3.1 Preliminary Notions	429
18.3.2 Model-based Service Description	430
18.3.3 Ontological Infrastructure	431
18.4 Distributed Semantic Service Management	437
18.4.1 Distributed Service Registry Structure	438
18.4.2 Distributed Service Registry Maintenance	439
18.5 Semantic Collaboration with Distributed Service Registry	440
18.5.1 Experimental Evaluation	442
18.6 Related Work	443
18.7 Conclusion	446
References	447
19 Informative Top-k Retrieval for Advanced Skill Management	449
Simona Colucci, Tommaso Di Noia, Azzurra Ragone, Michele Ruta, Umberto Straccia, and Eufemia Tinelli	
19.1 Introduction	450
19.2 Preliminaries	451
19.3 Human Resources Retrieval	453
19.3.1 Ontology Component	455
19.3.2 Database Component	457
19.3.3 Query Process (by Example)	460
19.3.4 Match Explanation	462
19.4 Experiments	466
19.5 Related Work	468
19.5.1 Research Approaches	469
19.5.2 Commercial Tools	470
Appendix A Ontology Axioms Excerpts	470
Appendix B Example Candidate Profiles Set	473
References	475

Part V Engineering Semantic Web Systems

20 MIDST: Interoperability for Semantic Annotations	479
Paolo Aizen, Pierluigi Del Nostro, and Stefano Paolozzi	
20.1 Introduction	479
20.2 State of the Art	481
20.2.1 From Database to Ontology	482

20.2.2	From Ontology to Database	483
20.2.3	Translation Between Ontologies and Databases	483
20.3	Model Independent Schema and Data Translation	485
20.3.1	Basic Translations	486
20.3.2	More Complex Translations	487
20.4	OWL and Relational Database Interoperability	488
20.4.1	Relational Data Model	488
20.4.2	OWL Data Model	490
20.5	An Extended Supermodel	492
20.5.1	Management of Intersections	493
20.5.2	Management of Restrictions	494
20.5.3	Classes Equivalence	494
20.5.4	Properties Equivalence	494
20.5.5	Object and Datatype Properties Generalization	495
20.5.6	Functional Datatype Properties	495
20.5.7	Symmetric, Transitive and Inverse Object Properties	495
20.6	Translation Rules	495
20.7	Conclusion	498
	References	499
21	Virtuoso: RDF Support in a Native RDBMS	501
	Orri Erling and Ivan Mikhailov	
21.1	Introduction	501
21.2	State of the Art	502
21.3	Triple Storage	504
21.4	SPARQL and SQL	506
21.5	Clustering and Scalability	509
21.5.1	Query Execution Model	510
21.5.2	Performance	511
21.6	Mapping Relational Data into RDF for SPARQL Access	512
21.7	Applications and Benchmarks	515
21.8	Future Directions	516
21.9	Conclusion	517
	References	518
22	Hera: Engineering Web Applications Using Semantic Web-based Models	521
	Kees van der Sluijs, Geert-Jan Houben, Erwin Leonardi, and Jan Hidders	
22.1	Introduction	521
22.2	Method	523
22.2.1	Data Modeling	524
22.2.2	Application Modeling	525
22.2.3	Presentation Modeling	531
22.3	User Modeling	532
22.4	Aspect Orientation	533

22.5	Tool Support	535
22.5.1	Hera Studio	535
22.5.2	Other Tools	537
22.6	Related Work	538
22.6.1	WebML	539
22.6.2	OOHDM	539
22.6.3	UWE	540
22.6.4	OOWS	541
22.7	Conclusion	542
	References	543
Index		545