

Data-Centric Systems and Applications

Series Editors

M.J. Carey
S. Ceri

Editorial Board

P. Bernstein
U. Dayal
C. Faloutsos
J.C. Freytag
G. Gardarin
W. Jonker
V. Krishnamurthy
M.-A. Neimat
P. Valduriez
G. Weikum
K.-Y. Whang
J. Widom

Vipul Kashyap · Christoph Bussler ·
Matthew Moran

The Semantic Web

Semantics for Data and
Services on the Web

With 61 Figures and 18 Tables

 Springer

Vipul Kashyap
Partners HealthCare System
Clinical Informatics R&D
93 Worcester St, Suite 201
Wellesley, MA 02481
USA
vkashyap1@partners.org

Christoph Bussler
Merced Systems, Inc.
333 Twin Dolphin Drive
Suite 500
Redwood Shores, CA 94065
USA
chbussler@aol.com

Matthew Moran
Nortel
Mervue Industry Park
Galway
Ireland
maitiu_moran@hotmail.com

ISBN 978-3-540-76451-9

e-ISBN 978-3-540-76452-6

DOI 10.1007/978-3-540-76452-6

Library of Congress Control Number: 2008931474

ACM Computing Classification (1998): H.4, I.2, J.3

© 2008 Springer-Verlag Berlin Heidelberg

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Cover design: KünkelLopka GmbH, Heidelberg

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Preface

A decade ago Tim Berners-Lee proposed an extraordinary vision: despite the phenomenal success of the Web, it would not, and could not, reach its full potential unless it became a place where automated processes could participate as well as people. This meant the publication of documents and data to the web in such a way that they could be interpreted, integrated, aggregated and queried to reveal new connections and answer questions, rather than just browsed and searched. Many scoffed at this idea, interpreting the early emphasis on language design and reasoning as AI in new clothes. This missed the point. The Grand Challenge of the Semantic Web is one that needs not only the information structure of ontologies, metadata, and data, but also the computational infrastructure of Web Services, P2P and Grid distributed computing and workflows. Consequently, it is a truly whole-system and multi-disciplinary effort.

This is also an initiative that has to be put into practice. That means a pragmatic approach to standards, tools, mechanisms and methodologies, and real, challenging examples. It would seem self-evident that the Semantic Web should be able to make a major contribution to clinical information discovery. Scientific communities are ideal incubators: knowledge-driven, fragmented, diverse, a range of structured and unstructured resources with many disconnected suppliers and consumers of knowledge. Moreover, the clinicians and biosciences have embraced the notions of annotation and classification using ontologies for centuries, and have demanding requirements for trust, security, fidelity and expressivity.

This book is the first to describe comprehensively the two main characteristics of the Semantic Web – its information and its processes – and to apply it not to toy, artificial examples but to a challenging application that matters, namely translational medicine. As such, it will become a key text for all of those serious about discovering the many facets of the Semantic Web, those who need to understand the current state of the art as it really is in practice, and those who need to be knowledgeable about its future.

Professor Carole Goble

University of Manchester

Acknowledgments

Vipul would like to dedicate this book to his son Ashish and daughter Aastha. A significant portion of this book was written during paternity leave, when Ashish was born! Vipul also expresses his gratitude to his wife Nitu for tolerating the demands on their family time. The support of Vipul's parents, Yogendra Singh and Vibha, during these times was very valuable and helped reduce stress for Vipul and his family. Vipul would like to acknowledge Dr. Tonya Hongsermeier for her spirited evangelism of the Semantic Web at Partners Healthcare and the W3C Healthcare and Life Sciences Interest Group (HCLSIG). She helped create opportunities to work on Semantic-Web-related projects and pointed me to the translational medicine use case, discussed in Chapter 2 of the book. Vipul also valued his interactions with Dr. Howard Goldberg, which helped him understand pragmatic applications of Semantic Web technologies in the healthcare area. Dr. Blackford Middleton's support for work related to Semantic Web technologies at Partners is also gratefully acknowledged. Discussions and debates with Olivier Bodenreider at the National Library of Medicine were very instructive and introduced Vipul to the area of biomedical informatics. The interactions with various colleagues in HCLSIG, such as Eric Neumann, Kei Cheung, Bill Bug, Susie Stephens, Don Doherty, June Kinoshita, Alan Ruttenberg and others, were also very thought provoking and intellectually stimulating. With Stefan Decker, Vipul had the pleasure of co-presenting a tutorial on the Semantic Web at VLDB 2003, on which this book is based. Vipul has also enjoyed tremendously the follow-on collaboration with Christoph Bussler and Matt Moran on writing this book. Finally, Vipul would like to acknowledge Dr. Amit Sheth, who has been a great mentor and was responsible for introducing me to the research area of semantics and the Semantic Web.

Chris wants to acknowledge a few people he met during his Semantic Web Services journey in Ireland. First there is Dieter Fensel, who in a nutshell was instrumental in opening up the field of semantics to him during many years of collaboration. Then there is the WSMX team that started making Semantic Web Services a reality in various research projects and through a still ongoing open source implementation that started to make its way into standards and industry: David Aiken, Emilia Cimpian, Enrica Dente, Doug Foxvog, Armin Haller, Thomas Haselwanter, Juan Miguel Gomez, Mick Kerrigan, Edward Kilgariff, Adrian Mocan, Matt Moran, Manuel Ohlendor, Eyal Oren, Ina O'Murchu, Brahmananda Sapkota, Laurentiu Vasiliu, Jana Viskova, Tomas Vitvar, Maciej Zaremba, and Michal Zaremba. In context of this book, Vipul and Matt were great co-authors and

Chris enjoyed working together with them a lot during this project. Last but not least he would like to give a huge thanks to his family, Barbara and Vernon, for joining him and supporting him during their time in Ireland; the memories about that time shall remain with them forever.

Matt wants to thank the folks who have encouraged him in his transition into research over the last four years. The WSMX team that Chris has already listed has been fantastic to work with. Matt enjoyed being in the team and thinks its energy and positive approach has been a strong factor in keeping up the momentum in Semantic Web Services research. He would like to particularly thank Tomas Vitvar and Maciej Zaremba with whom he has established a great relationship over the last two years. He is very grateful to Chris and Dieter Fensel, who gave him the chance to step into Semantic Web Services research with DERI. He has learned a lot from them both and hopes to keep learning from them in the future. Writing this book with Vipul and Chris has been a great experience for Matt. The back-and-forth conversations have been generous and enlightening. Finally, he would like to thank his sisters Maeve, Claire, Joanne and Mags, and his parents, for their biased praise and keeping him supplied with Irish stew and wild salmon through the course of his working on this book.

The three authors would like to thank the team at Springer Verlag for its support, especially Ralf Gerstner, who was always engaged, supportive, patient as well as available for discussion and advice.

June 2008

Vipul Kashyap, Boston, MA, USA
Christoph Bussler, San Jose, CA, USA
Matthew Moran, Galway, Ireland

Contents

Part I

Preliminaries 1

1 Introduction 3

- 1.1 Motivation: Why Semantic Web? 4
- 1.2 A Framework for Semantic Web 5
- 1.3 Use Case: Translational Medicine Clinical Vignette 7
- 1.4 Scope and Organization 9

2 Use Case and Functional Requirements 11

- 2.1 Detailed Clinical Use Case 12
- 2.2 Stakeholders and Information Needs 13
- 2.3 Conceptual Architecture 15
- 2.4 Functional Requirements 17
- 2.5 Research Issues 18
- 2.6 Summary 19

Part II

Information Aspects of the Semantic Web 21

3 Semantic Web Content 23

- 3.1 Nature of Web Content 23
- 3.2 Nature of Semantic Web Content 24
- 3.3 Metadata 25
 - 3.3.1 Metadata Usage in Various Applications 26
 - 3.3.2 Metadata: A Tool for Describing and Modeling Information 27
- 3.4 Ontologies: Vocabularies and Reference Terms for Metadata 30
- 3.5 Summary 33

4 Metadata Frameworks 35

- 4.1 Examples of Metadata Frameworks 35
 - 4.1.1 XML-Based Metadata Framework 36
 - 4.1.2 RDF-Based Metadata Framework 36
 - 4.1.3 OWL-Based Metadata Framework 37
 - 4.1.4 WSMO-Based Metadata Framework 37
- 4.2 Two Perspectives: Data Models and Model-Theoretic Semantics 38

4.2.1	Data Models	38
4.2.2	Multiple Syntaxes for RDF: A Short Note	47
4.2.3	Model-Theoretic Semantics	48
4.3	Query Languages	51
4.3.1	Query Languages for XML Data	51
4.3.2	Query Languages for RDF Data	62
4.3.3	Extending Query Languages with Reasoning and Entailment	73
4.4	Clinical Scenario Revisited	74
4.4.1	Semantic Web Specifications: LIMS and EMR Data	74
4.4.2	Linking data from Multiple Data Sources	76
4.4.3	Advantages and Disadvantages of using Semantic Web Specifications	78
4.5	Summary	78
5	Ontologies and Schemas	79
5.1	What is an Ontology?	79
5.2	Ontology Representation Languages	84
5.2.1	XML Schema	84
5.2.2	RDF Schema	92
5.2.3	Web Ontology Language	100
5.2.4	The Web Service Modeling Ontology (WSMO)	112
5.2.5	Comparison of Ontology Representation Languages	118
5.3	Integration of Ontology and Rule Languages	122
5.3.1	Motivation and Requirements	122
5.3.2	Overview of Languages and Approaches	123
5.3.3	Semantic Web Rules Language	124
5.4	Clinical Scenario Revisited	126
5.4.1	A Domain Ontology for Translational Medicine	126
5.4.2	Integration of Ontologies and Rules for Clinical Decision Support	130
5.4.3	Advantages and Disadvantages of using Semantic Web Specifications	135
5.5	Summary	135
6	Ontology Authoring and Management	137
6.1	Ontology Building Tools	137
6.1.1	Ontology Editors: Brief Descriptions	138
6.1.2	Ontology Editors: A Comparative Evaluation	143
6.2	Ontology Bootstrapping Approaches	148
6.3	Ontology Merge and Integration Tools	150
6.3.1	Ontology Merge and Integration Tools: A Brief Description	151
6.3.2	Evaluation of Ontology Merge and Integration Tools	152
6.4	Ontology Engines and Reasoners	154

6.5	Clinical Scenario Revisited	157
6.6	Summary	158
7	Applications of Metadata and Ontologies	161
7.1	Tools and Techniques for Metadata Annotation	161
7.1.1	Requirements for Metadata Annotation	162
7.1.2	Tools and Technologies for Metadata Annotation	163
7.1.3	Comparative Evaluation	168
7.2	Techniques for Schema/Ontology Mapping	173
7.2.1	A Classification of Schema-matching Approaches	173
7.2.2	Schema-matching Techniques: Overview	179
7.3	Ontology Driven Information Integration	183
7.3.1	The Role of Ontologies in Information Integration	183
7.3.2	Ontology Representations Used in Information Integration	187
7.3.3	The Role of Mapping in Information Integration	188
7.3.4	The Role of Ontology Engineering in Information Integration	190
7.4	Summary	192
Part III		
Process Aspects of the Semantic Web		193
8	Communication	195
8.1	Communication Concepts	195
8.1.1	Fundamental Types	196
8.1.2	Formats and Protocols (FAP)	197
8.1.3	Separation of Interface and Logic	198
8.1.4	Communicating Parties	199
8.1.5	Mediation	201
8.1.6	Non-functional Aspects	202
8.2	Communication Paradigms	203
8.2.1	Client/Server (C/S)	204
8.2.2	Queueing	204
8.2.3	Peer-to-Peer (P2P)	205
8.2.4	Blackboard	205
8.2.5	Web Services	206
8.2.6	Representational State Transfer (REST)	207
8.2.7	Agents	207
8.2.8	Tuple Spaces	208
8.2.9	Co-location	208
8.2.10	Summary	209
8.3	Long-Running Communication	209
8.3.1	Business-to-Business (B2B) Protocols	210
8.3.2	Application-to-Application (A2A) Protocols	211

8.4	Web Services	211
8.5	Clinical Use Case	212
8.6	Summary	214
9	State of the Art in Web Services	215
9.1	History	215
9.2	Traditional Web Services	216
9.2.1	WSDL	217
9.2.2	SOAP	218
9.2.3	UDDI	219
9.2.4	Summary	219
9.3	Emerging Web Service Specifications (WS*-Stack)	220
9.3.1	Standards	220
9.3.2	Web Service Standards	221
9.3.3	Semantic-Web-Service-Related Standards	222
9.4	Service-oriented Architecture (SOA)	223
9.4.1	Service Paradigm	223
9.4.2	SOA and Web Services	224
9.4.3	Open Issues and Technical Challenges	224
9.5	Semantics and Web Services	226
9.5.1	Semantics, What Semantics?	227
9.5.2	Data Semantics	228
9.5.3	Process Semantics	229
9.5.4	Selection Semantics	229
9.5.5	Other Types of Semantics	230
9.6	Clinical Use Case	231
9.7	Summary	232
10	Web Service Composition	233
10.1	Composition	233
10.1.1	Motivation	233
10.1.2	Definition of Composition	235
10.1.3	Web Services and Composition	237
10.1.4	Choreography and Orchestration	238
10.2	Dynamic Composition	239
10.3	Business-to-Business Communication	240
10.4	Application-to-Application Communication	241
10.5	Complex Business Logic	242
10.6	Standards and Technologies	243
10.6.1	Web Services Business Process Execution Language (WS-BPEL)	244
10.6.2	Business Process Modeling Notation (BPMN)	245
10.6.3	Web Service Choreography Description Language (WS-CDL)	245
10.6.4	Java Business Integration (JBI)	246

10.7	Clinical Use Case	247
10.8	Summary	247
11	Semantic Web Services	249
11.1	Semantics of Web Services	249
11.1.1	Why Semantic Web Services?	249
11.1.2	Interface vs. Implementation	251
11.1.3	Modeling of State	251
11.2	Alternatives for Capturing Semantics of Web Services	253
11.2.1	Finite State Machines	253
11.2.2	Statechart Diagrams	254
11.2.3	Petri Nets	254
11.2.4	Process Algebras	256
11.3	Semantic Web Service Approaches	259
11.3.1	OWL-S	259
11.3.2	SWSF	261
11.3.3	WSDL-S	266
11.3.4	SAWSDL	268
11.3.5	WSMO, WSML and WSMX	269
11.4	Reasoning with Web Service Semantics	276
11.4.1	Discovery	276
11.4.2	Semantic Web Service Composition	281
11.4.3	Mediation	283
11.5	Clinical Use Case	285
11.6	Summary	286
Part IV		
Standards		287
12	Semantic Web Standards	289
12.1	Relevant Standards Organization	289
12.1.1	International Organization for Standardization (ISO)	289
12.1.2	International Electrotechnical Commission (IEC)	290
12.1.3	Organization for the Advancement of Structured Information Standards (OASIS)	290
12.1.4	World Wide Web Consortium (W3C)	290
12.1.5	International Engineering Task Force (IETF)	291
12.1.6	National Institute of Standards and Technology (NIST)	291
12.1.7	The Object Modeling Group (OMG)	291
12.1.8	Semantic Web Services Initiative (SWSI)	292
12.1.9	United States National Library of Medicine (NLM)	292
12.2	Semantic Web Content Standardization Efforts	293
12.2.1	Standard Generalized Markup Language (SGML)	293
12.2.2	eXtensible Markup Language (XML)	293
12.2.3	eXtensible Stylesheet Transformation Language (XSLT) ...	294

12.2.4	XPath	294
12.2.5	XQuery	294
12.2.6	XML Schema	294
12.2.7	Resource Description Framework (RDF)	295
12.2.8	SPARQL	295
12.2.9	RDF Schema	295
12.2.10	Web Ontology Language (OWL)	296
12.2.11	Rule-ML	296
12.2.12	Semantic Web Rules Language (SWRL)	296
12.2.13	Ontology Definition Metamodel (ODM)	296
12.2.14	Unified Modeling Language (UML)	297
12.2.15	Knowledge Interchange Format (KIF)	297
12.2.16	Open Knowledge Base Connectivity Protocol (OKBC)	297
12.2.17	DIG Description Logics Interface	297
12.2.18	OWL API	298
12.2.19	Standardized Vocabularies and Ontologies	298
12.3	Semantic Web Services Standardization Efforts	300
12.3.1	ISO-18629 Process Specification Language (PSL)	301
12.3.2	W3C Semantic Annotations for the Web Services Description Language (SAWSDL)	302
12.3.3	OWL-S	303
12.3.4	Web Services Modeling Ontology (WSMO)	303
12.3.5	Semantic Web Services Framework (SWSF)	304
12.3.6	WSDL-S	304
12.3.7	OASIS Semantic Execution Environment (SEE)	304
12.3.8	OASIS Service-Oriented Architecture Reference Model (SOA RM)	305
12.3.9	Semantic Web Services Architecture (SWSA)	306
12.3.10	Semantic Web Services Interest Group (SWS-IG)	307
12.4	Summary	307

Part V

Putting it All Together and Perspective 309

13 A Solution Approach to the Clinical Use Case 311

13.1	Service Discovery, Composition and Choreography	312
13.1.1	Specification of Clinical Workflow using WSMO	313
13.1.2	Data Structures in Data Flow	316
13.1.3	Data Mediation	319
13.1.4	Goal Definition	328
13.1.5	Discovery	331
13.1.6	Orchestration/Service Composition	333
13.1.7	Process and Protocol Mediation	339
13.2	Data and Knowledge Integration	342

13.2.1	Data Integration Services: WSMO/WSML Specification	343
13.2.2	Semantic Data Integration Architecture	344
13.2.3	A Domain Ontology for Translational Medicine	346
13.2.4	Use of RDF to represent Genomic and Clinical Data	351
13.2.5	The Integration Process	353
13.3	Decision Support	356
13.3.1	Decision Support Services: WSMO/WSML Specification ..	357
13.3.2	Architecture	358
13.3.3	Business Object Model Design	359
13.3.4	Rule Base Design	360
13.3.5	Definitions vs. Actions: Ontology Design	360
13.4	Knowledge Maintenance and Provenance	365
14	Outlook: The Good, the Bad and the Ugly?	369
14.1	The Good - Progress and Impact	369
14.2	The Bad - Major Obstacles to Overcome	371
14.3	The Ugly - Possible Prohibitors	372
Part VI		
References and Index		375
References		377
Index		405