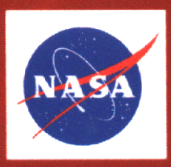

Semantic Web

Concepts, Technologies and Applications

Karin K. Breitman
Marco Antonio Casanova
Walter Truszkowski

BIBLIOTHEQUE DU CERIST

NASA Monographs in Systems
and Software Engineering

BIBLIOTHEQUE DU CERIST

The ***NASA Monographs in Systems and Software Engineering*** series addresses cutting-edge and groundbreaking research in the fields of systems and software engineering. This includes in-depth descriptions of technologies currently being applied, as well as research areas of likely applicability to future NASA missions. Emphasis is placed on relevance to NASA missions and projects.

Also in this series:

Christopher A. Rouff, Michael Hinchey, James Rash, Walter Truszkowski and Diana Gordon-Spears (Eds)
Agent Technology from a Formal Perspective
1-85233-947-0

BIBLIOTHEQUE DU CERIST

K.K. Breitman, M.A. Casanova and
W. Truszkowski

Semantic Web: Concepts, Technologies and Applications

BIBLIOTHEQUE DU CERIST

Springer

Karin Koogan Breitman, MSc, DSc
Marco Antonio Casanova, PhD
Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro, Brazil

Walter Truszkowski, MA, BA
NASA Goddard Space Flight Center, USA

Series Editor
Professor Michael Hinchey

BIBLIOTHEQUE DU CERIST

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Control Number: 2006932966

NASA Monographs in Systems and Software Engineering ISSN 1860-0131
ISBN-10: 1-84628-581-X e-ISBN: 978-1-84628-710-7
ISBN-13: 978-1-84628-581-3

Printed on acid-free paper

© Springer-Verlag London Limited 2007

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms of licences issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers.
The use of 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 laws and regulations and therefore free for general use.
The publisher makes no representation, express or implied, with regard to the accuracy of the information contained in this book and cannot accept any legal responsibility or liability for any errors or omissions that may be made.

9 8 7 6 5 4 3 2 1

Springer Science+Business Media, LLC
springer.com

BIBLIOTHEQUE DU CERIST

Preface

As the volume of Web resources grows exponentially, researchers from industry, government, and academia are now exploring the possibility of creating a Semantic Web in which meaning is made explicit, allowing machines to process and integrate Web resources intelligently. How will this Web of the future be effectively built? No one really knows. Although guesses vary from author to author, some themes are recurrent, most of which are covered in this book. In general, emerging technologies will allow semantics to be added to existing Web resources, so as to make the Semantic Web vision come true.

This book provides a succinct account of this new Web, its principles, concepts, and related tools. Its main contribution lies in the ability to demonstrate how Semantic Web technologies may be integrated and realized in several application domains.

Organization of the Book

The text is divided into four parts. Part I begins with an educated forecast for the future of the current Web. This sets the foundation for the rest of the book. Chapter 1 provides a well-paced introduction to the Semantic Web, from motivations to requirements and guidelines on how to realize this concept.

Part II introduces the fundamental building blocks one should master to grasp the full meaning of the Semantic Web. Chapter 2 summarizes the various uses of the term ontology in computer science. Chapter 3 covers knowledge representation in description logic and provides the background for a better understanding of ontology description languages and tools. Chapters 4 and 5 introduce the Resource Description Framework (RDF), the RDF Vocabulary Description Language 1.0:

RDF Schema, and the Web Ontology Language (OWL), the knowledge representation languages proposed as standards for the Semantic Web. Chapter 4 also contains a very brief introduction to some essential XML concepts. Chapter 6 summarizes four rule languages, Datalog, the Rule Markup Language (RuleML), the Semantic Web Rule Language (SWRL) and TRIPLE, designed to manipulate knowledge bases. Chapter 7 provides a brief description of Web services and introduces OWL-S, a service upper ontology for Web services.

Part III focuses on emerging technologies that provide the necessary application development infrastructure and guidelines to develop Semantic Web applications. Chapter 8 surveys several ontology development methods, addressing their background and applicability. Chapter 9 discusses metadata standards, upper ontologies, and ontology libraries that are relevant to the indexing of resources in the Semantic Web, including the Dublin Core, the Warwick Framework, the Suggested Upper Merged Ontology (SUMO), the Knowledge Representation (KR) ontology, CYC, and WordNet. Chapter 10 covers a small selection of Semantic Web tools, with no intention of producing a complete survey.

Part IV illustrates how Semantic Web technologies, discussed throughout the book, can be consistently applied to four distinct application domains. Chapter 11 discusses software agents in the context of information technology systems and the role that ontologies play in their construction. Chapter 12 exemplifies the use of Semantic Web technologies in computing applications called semantic desktops, which enhance personal information management, software application usage, and collaboration. Chapter 13 reviews standardization efforts that facilitate the development of applications devoted to the cataloguing and dissemination of data about cultural collections. It also describes an application that combines a metadata schema with controlled vocabularies to create semantic annotations for still images of works of art. This profitable combination—metadata schemas, controlled vocabularies, and standardization efforts—is repeated in the next chapter in the context of geospatial applications, setting a pattern for other application areas. Chapter 14 then overviews the technologies that facilitate the development of the Geospatial Semantic Web, emphasizing the role of standard proposals. Each technology is first discussed from a broad perspective and then illustrated with implemented applications.

BIBLIOTHEQUE DU CERIST

Possible Uses for the Book

This book is intended as a reference text on the Semantic Web for software engineers, database and information technology students, academics, and practitioners. It differs from other books on the topic because, rather than presenting the Semantic Web as a quilt of disconnected topics and examples, it provides a coherent body of concepts and technologies. The book reflects the authors' experience in research, graduate and undergraduate teaching, corporate training, and industry and government projects using Semantic Web technologies.

The text may be used as the main reference for an upper-level undergraduate course or an entry-level graduate course. At the graduate level, the text may be covered at a faster pace and complemented with additional material touching, for

example, on description logic, ontology tools, or different application areas, as suggested at the end of each chapter. The second format was tested in several one-semester offerings at the Department of Informatics of the Pontifical Catholic University of Rio de Janeiro, during 2003–2006. Portions of the text were also the subject of conference tutorials and short courses given by the authors.

Acknowledgments

The authors wish to thank Beverley Ford, Helen Desmond, Lesley Poliner, and Frank Ganz, from Springer, for their support throughout the preparation of the manuscript. The authors wish to extend their thanks to Michael G. Hinchey, editor of the *NASA Monograph on Software*, who gave them the opportunity to include this text in the series. Finally, the authors wish to thank Alice A. Casanova for her support in producing the final manuscript.

Karin K. Breitman
Marco A. Casanova
Walter Truszkowski

September 2006

BIBLIOTHEQUE DU CERIST

BIBLIOTHEQUE DU CERIST

Contents

Preface
Organization of the Book v
Possible Uses for the Book vi
Acknowledgements vii

Part I — Introduction

Chapter 1 — The Future of the Internet
1.1 Introduction 3
1.2 The Syntactic Web 4
1.3 The Semantic Web 6
1.4 How the Semantic Web Will Work 6
1.5 What the Semantic Web Is Not 11
1.6 What Will Be the Side Effects of the Semantic Web 13
Recommended Reading 13
References 13
Web sites 14

Part II — Concepts

Chapter 2 — Ontology in Computer Science

2.1 Defining the Term Ontology **17**
2.2 Differences Among Taxonomies, Thesauri, and Ontologies **20**
2.2.1 Taxonomies Versus Ontologies 20
2.2.2 Thesauri Versus Ontologies 22
2.3 Classifying Ontologies **26**
2.3.1 Classifying Ontologies According to a Semantic Spectrum 26
2.3.2 Classifying Ontologies According to Their Generality 27
2.3.3 Classifying Ontologies According to the Information Represented 28
2.4 Web Ontology Description Languages **29**
2.5 Ontologies, Categories, and Intelligence **31**
References **33**

Chapter 3 — Knowledge Representation in Description Logic

3.1 Introduction **35**
3.2 An Informal Example **36**
3.3 The Family of Attributive Languages **43**
3.3.1 Concept Descriptions 43
3.3.2 Terminologies 49
3.3.3 Assertions 51
3.4 Inference Problems **52**
3.4.1 Inference Problems for Concept Descriptions 52
3.4.2 Inference Problems for Assertions 53
Recommended Reading **54**
References **54**

Chapter 4 — RDF and RDF Schema

4.1 Introduction **57**
4.2 XML Essentials **58**
4.2.1 Elements and Attributes 58
4.2.2 URIs and Namespaces 59
4.3 RDF **62**
4.3.1 RDF Statements and Vocabularies 62
4.3.2 RDF Triples and Graphs 65
4.3.3 RDF/XML 67
4.4 RDF Schema **71**
4.4.1 Classes 71
4.4.2 Properties 74
4.4.3 Individuals 75
4.5 A Summary of the RDF/RDF Schema Vocabulary **77**
Recommended Reading **79**
References **79**

BIBLIOTHEQUE DU CERIST

BIBLIOTHEQUE DU CERIST

Chapter 5 — OWL

5.1 Introduction **81**
5.2 Requirements for Web Ontology Description Languages **82**
5.3 Header Information, Versioning, and Annotation Properties **85**
5.4 Properties **86**
5.4.1 Datatype and Object Properties 86
5.4.2 Property Characteristics 87
5.5 Classes **88**
5.5.1 Class Descriptions 89
5.5.2 Class Axioms 95
5.6 Individuals **97**
5.7 Datatypes **98**
5.8 A Summary of the OWL Vocabulary **100**
Recommended Reading **102**
References **103**

Chapter 6 — Rule Languages

6.1 Introduction **105**
6.2 Usage Scenarios for Rule Languages **106**
6.3 Datalog **107**
6.4 RuleML **109**
6.5 SWRL **114**
6.6 TRIPLE **121**
Recommended Reading **124**
References **124**

Chapter 7 — Semantic Web Services

7.1 Introduction **127**
7.2 Web Service Essentials **129**
7.2.1 Basic Components of a Web Service 129
7.2.2 Web Service Security Standards 131
7.2.3 Web Service Standardization Organizations 133
7.2.4 Potential Benefits and Criticism 134
7.3 OWL-S Service Ontology **135**
7.3.1 Overview 135
7.3.2 Service Profile (What It Does) 138
7.3.3 Service Model (How It Does) 139
7.3.4 Service Grounding (How to Access) 142
7.4 An OWL-S Example **143**
7.4.1 Scenario Description 143
7.4.2 Informal Process Definition 145
7.4.3 OWL-S Process Definition 146
References **151**

BIBLIOTHEQUE DU CERIST

Part III — Technologies

Chapter 8 — Methods for Ontology Development

8.1 Introduction **155**
8.2 Uschold and King Ontology Development Method **156**
8.3 Toronto Virtual Enterprise Method **158**
8.4 Methontology **159**
8.5 KACTUS Project Ontology Development Method **163**
8.6 Lexicon-Based Ontology Development Method **163**
8.7 Simplified Methods **167**
8.7.1 Ontology Development 101 167
8.7.2 Horrocks Ontology Development Method 171
Recommended Reading **172**
References **172**

Chapter 9 — Ontology Sources

9.1 Introduction **175**
9.2 Metadata **176**
9.2.1 What is Metadata? 176
9.2.2 Dublin Core 178
9.2.3 Warwick Framework 179
9.2.4 PICS 181
9.2.5 vCards 181
9.2.6 FOAF 182
9.3 Upper Ontologies **184**
9.3.1 SUMO 184
9.3.2 KR Ontology 188
9.3.3 CYC 193
9.3.4 WordNet 195
9.4 Other Ontologies of Interest **195**
9.5 Ontology Libraries **197**
References **198**

Chapter 10 — Semantic Web Software Tools

10.1 Introduction **201**
10.2 Metadata and Ontology Editors **202**
10.2.1 Dublin Core Metadata Editor 202
10.2.2 OilEd 203
10.2.3 OntoEdit 203
10.2.4 Protégé Ontology Editor 203
10.2.5 Protégé Plugins and APIs 207
10.3 Reasoners **209**
10.4 Other Tools **211**
References **214**

BIBLIOTHEQUE DU CERIST

Part IV — Applications

Chapter 11 — Software Agents

11.1 Introduction **219**
11.2 Agent Forms **220**
11.3 Agent Architecture **223**
11.4 Agents in the Semantic Web Context **225**
11.4.1 The Role of Ontologies 225
11.4.2 Software Agent Communication in the Semantic Web 225
References **227**

Chapter 12 — Semantic Desktop

12.1 Introduction **229**
12.2 Semantic Desktop Metadata **230**
12.3 Semantic Desktop Ontologies **233**
12.4 Semantic Desktop Architecture **235**
12.5 Semantic Desktop Related Applications **237**
References **238**

Chapter 13 — Ontology Applications in Art

13.1 Introduction **241**
13.2 Ontologies for the Description of Works of Art **242**
13.2.1 The Art & Architecture Thesaurus 242
13.2.2 The Union List of Artist Names 246
13.2.3 A Visual Annotation Ontology for Art Images 249
13.3 Metadata Schemas for the Description of Works of Art **253**
13.3.1 The Categories for the Description of Works of Art 253
13.3.2 The ISO 21127 — A Reference Ontology for the Interchange of Cultural Heritage Information 255
13.3.3 The Visual Resources Association Core Categories 256
13.4 Semantic Annotation of Art Images **260**
References **263**

Chapter 14 — Geospatial Semantic Web

14.1 Introduction **265**
14.2 Basic Geospatial Concepts **266**
14.3 Classifying Geospatial Features **267**
14.3.1 Geospatial Feature Type Thesauri 267
14.3.2 Geospatial Feature Ontologies 271
14.4 Gazetteers **277**
14.4.1 Examples of Gazetteers 277
14.4.2 Standards for Gazetteers 278
14.5 Geospatial Metadata **280**
14.5.1 Geospatial Metadata Standards 280
14.5.2 The FGDC Metadata Standard 280

14.5.3 The ISO 19115:2003 Metadata Standard 285
14.5.4 The ISO 19119:2005 Service Metadata Standard 290
14.6 The OGC Catalogue Specification 293
14.7 Geospatial Web Services 295
14.7.1 The OGC Services Framework 295
14.7.2 A Geospatial Web Services Architecture 297
14.7.3 Example of an Integrated Collection of Geospatial Web Services 298
14.8 Examples of Spatial Data Infrastructures 300
14.8.1 INSPIRE 300
14.8.2 NSDI 302
14.9 Example of a Metadata Catalogue for Earth Science Data 303
14.9.1 Overview of the GCMD 303
14.9.2 GCMD and Other Data Sources 305
Recommended Reading 308
References 308

Index 313

BIBLIOTHEQUE DU CERIST