

Heiner Stuckenschmidt  
Frank van Harmelen

# Information Sharing on the Semantic Web

 Springer

581281

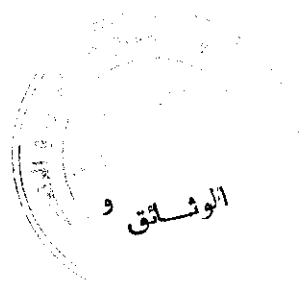


الوثائق

Heiner Stuckenschmidt · Frank van Harmelen

# Information Sharing on the Semantic Web

With 51 Figures and 13 Tables



 Springer

To absent friends.



---

## Preface

### About the book

#### The success of the information society

The rapid progress of the “information society” in the past decade has been made possible by the removal of many technical barriers. Producing, storing, and transporting information in large quantities are no longer significant problems.

*Producing* on-line, digitized information is no longer a problem. Ever more of our commercial, scientific and personal information exchanges happen on-line in digital form. In the professional domain, near 100% of all office documents are produced in digital form (even if afterwards they are distributed in paper form), large parts of the scientific discourse are now taking place in digital form (with physics, computer science and astronomy taking a leading role). In the public domain, newspapers are available on-line, an increasing number of radio and television stations offer their material on-line in streaming form and e-government is an important theme for public administration. Even in the personal area, information is rapidly moving on-line: sales of digital cameras are now higher than for analogue cameras, e-mail and on-line chat have become important channels for maintaining social relations and for personal entertainment the digital DVD is rapidly replacing the analogue video tape. Compact disk (itself already digital) is under serious pressure from on-line music in MP3 format from a variety of sources. In short: production of on-line information is now the norm in virtually all areas of our life.

*Storing* such information in the required volumes is also no longer a problem. The drive of my laptop has truly become an on-line archive, both professionally and personally. As my professional archive, it stores the sources of around 100 scientific papers I have written (and the full sources of three books), all the Master theses of the dozens of students I have supervised, the

The main thesis of this book is that the problem of information sharing (i.e. finding pieces of information and meaningfully relating them with other pieces) is only solvable by giving the computer better access to the *semantics* of the information. Thus, for a document, we do not only need to store such obvious metadata as its author, title, creation date, etc. but we must also make available in a machine-accessible way the important concepts that are discussed in the document, the relation of these concepts with those in other documents, relating these concepts to general background knowledge, etc. Similarly, for digital images, we would not only want to store format and size, but also that it is a satellite image of a specific area of land, where that area is located (e.g. by referencing a vocabulary of geographic locations), etc.

If computers had access to such *metadata* about the information items, they would be able to support us in finding relevant items, and in combining multiple items into a coherent answer to our questions. In this book we discuss active research on exactly this topic:

- how can the semantics of our information items be made available in a machine-accessible form?
- how can such metadata be exploited in retrieving and integrating information?

Of course it is crucial that the intended meaning of the metadata is shared between the different parties involved (e.g. those creating the metadata, and those using it). It is here that *ontologies* play a crucial role: shared formalized models of a particular domain, whose intended semantics is both shared between different parties and machine-interpretable (because it is “formalized”).

It has been argued that ontologies are a key technology for resolving the open problem of meaningful information sharing. However, most approaches rely on the existence of well-established data structures that can be used to analyze and exchange information. This book investigates ontology-based approaches for resolving semantic heterogeneity *weakly* structured environments, and in particular the World Wide Web. In doing this, we have to provide solutions for the following problems that arise from the nature of the Web:

*Missing conceptual models:* On the Web, we have no access to the conceptual model of an information source or the resulting logical data model. This lack of structure makes it difficult to refer to the context of information items, which is necessary for stating context transformation rules.

*Unclear system boundaries:* On the Web, it is not possible to clearly determine which information has to be taken into account, because information sources are added, removed or changed frequently. Therefore, we cannot rely on a fixed set of context-transformation rules.

*Part I*

introduces the general problem of information sharing and the need for explicit representations of information semantics in order to share information in a meaningful way. Further, it introduces the notion of ontology as a way of representing information semantics that has proven its value in different application domains. We also introduce the Web Ontology Language OWL as a standard for representing ontologies on the Semantic Web.

*Part II*

covers the creation of explicit representations of the information semantics. This includes the development of ontology encoded in OWL based on a given information sharing problem and the mostly automatic annotation of information sources with metadata that uses terms from ontologies to describe the content of an information source. We describe the basic methods for creating ontologies and metadata and describe experiments with real data and integration problems.

*Part III*

describes the use of the representational infrastructure created using the methods described in Part II for the purpose of information sharing. We discuss the semantic integration of terminologies used by different information sources and the integrated retrieval of information from multiple sources based on the result of the integration. Special attention is paid to the use of conjunctive queries that contain terms from ontologies. After discussing basic notions, we report the use of Semantic Web technologies for retrieving statistical information that revealed the need to take spatial relevance into account. We summarize with a description of the functionality of existing systems for information sharing and explain how the different aspects discussed in this part of the book are implemented in these systems.

*Part IV*

takes us back to some more fundamental questions concerning the use of ontologies for information sharing in a distributed environment such as the Semantic Web. In particular, we re-consider situations where the ontology itself is distributed across the Web. We extend the import mechanism of the Web Ontology Language by introducing the notion of modular ontology. We define a non-standard semantics for modular ontologies and compare the expressiveness of the model with OWL. We study the evolution of a modular ontology, in particular the impact of changes in a modular ontology, characterize changes according to their impact on other modules and define an update strategy that guarantees consistency of the overall model.

- IDEA Group Publishing

Some of the material reported here is the result of joint work with colleagues not mentioned as authors. We would like to thank the following persons for the fruitful cooperation in the past that led to the results reported in this book, as well as their permission to use material of joint papers and tutorials:

- Grigoris Antoniou, ITC-FORTH, Greece (Chap. 1)
- Fausto Giunchiglia, DIT, University of Trento, Italy (Chap. 6)
- Jens Hartmann, AIFB, University of Karlsruhe, Germany (Chap. 5)
- Catholijn Jonker, Vrije Universiteit Amsterdam, the Netherlands (Chap. 7)
- Michel Klein, Vrije Universiteit Amsterdam, the Netherlands (Chap. 10 and 11)
- Eduardo Mena, University of Zaragoza, Spain (Chap. 9)
- Christoph Schlieder, University of Bamberg, Germany (Chap. 8)
- Tim Verwaart, LEI Wageningen, the Netherlands (Chap. 7)
- Ubbo Visser, TZI, University of Bremen, Germany (Chap. 1, 2, 8 and 9)
- Thomas Voegelé, TZI, University of Bremen, Germany (Chap. 1, 2, 8 and 9)
- Holger Wache, TZI, University of Bremen, Germany (Chap. 2, 6 and 9)

Some of the work has been supported by the European Union under contracts IST-2001-33052 (WonderWeb) and IST-2001-34103 (SWAP). A significant part of the work has been carried out by the first author during his appointment at the Artificial Intelligence Group (Prof. Herzog) at the University of Bremen, Germany.

Amsterdam,  
January 2004

*Heiner Stuckenschmidt*  
*Frank van Harmelen*



---

# Contents

---

<b>Part I Information sharing and ontologies</b>	
<b>1</b>	<b>Semantic integration</b> . . . . . 3
1.1	Syntactic standards . . . . . 4
1.1.1	HTML: visualizing information . . . . . 4
1.1.2	XML: exchanging information . . . . . 5
1.1.3	RDF: a data model for meta-information . . . . . 6
1.1.4	The roles of XML and RDF . . . . . 8
1.2	The Problem of Heterogeneity . . . . . 10
1.2.1	Structural Conflicts . . . . . 10
1.2.2	Semantic Conflicts . . . . . 12
1.3	Handling information semantics . . . . . 14
1.3.1	Semantics from structure . . . . . 15
1.3.2	Semantics from text . . . . . 16
1.3.3	The need for explicit semantics . . . . . 17
1.4	Representing and comparing semantics . . . . . 19
1.4.1	Names and labels . . . . . 20
1.4.2	Term networks . . . . . 20
1.4.3	Concept lattices . . . . . 21
1.4.4	Features and constraints . . . . . 22
1.5	Conclusion . . . . . 23
<b>2</b>	<b>Ontology-based information sharing</b> . . . . . 25
2.1	Ontologies . . . . . 25
2.1.1	Shared vocabularies and conceptualizations . . . . . 26
2.1.2	Specification of context knowledge . . . . . 27
2.1.3	Beneficial applications . . . . . 29
2.2	Ontologies in information integration . . . . . 31
2.2.1	Content explication . . . . . 31
2.2.2	Additional roles of ontologies . . . . . 34
2.3	A framework for information sharing . . . . . 36

5.4.1	Generating ontology-based metadata	111
5.4.2	Using ontology-based metadata	112
5.5	Conclusions	114

---

## Part III Retrieval, integration and querying

---

<b>6</b>	<b>Retrieval and Integration</b>	119
6.1	Semantic integration	120
6.1.1	Ontology heterogeneity	120
6.1.2	Multiple systems and translatability	122
6.1.3	Approximate re-classification	123
6.2	Concept-based filtering	125
6.2.1	The idea of query-rewriting	126
6.2.2	Boolean concept expressions	127
6.2.3	Query re-writing	129
6.3	Processing complex queries	131
6.3.1	Queries as concepts	132
6.3.2	Query relaxation	134
6.4	Examples from a case study	137
6.4.1	Concept approximations	137
6.4.2	Query relaxation	138
6.5	Conclusions	140
<b>7</b>	<b>Sharing statistical information</b>	143
7.1	The nature of statistical information	144
7.1.1	Statistical metadata	145
7.1.2	A basic ontology of statistics	146
7.2	Modelling Statistics	150
7.2.1	Statistics as views	150
7.2.2	Connection with the domain	151
7.3	Translation to Semantic Web languages	155
7.3.1	Ontologies	155
7.3.2	Description of information	159
7.4	Retrieving statistical information	162
7.5	Conclusions	164
<b>8</b>	<b>Spatially-related information</b>	167
8.1	Spatial representation and reasoning	168
8.1.1	Levels of spatial abstraction	168
8.1.2	Reasoning about spatial relations	169
8.2	Ontologies and spatial relevance	170
8.2.1	Defining Spatial Relevance	171
8.2.2	Combined spatial and terminological matching	172
8.2.3	Limitations	174

<b>11 Evolution management</b> .....	231
11.1 Change detection and classification .....	232
11.1.1 Determining harmless changes .....	232
11.1.2 Characterizing changes .....	233
11.1.3 Update management .....	235
11.2 Application in a case study .....	236
11.2.1 The WonderWeb case study .....	236
11.2.2 Modularization in the case study .....	238
11.2.3 Updating the models .....	239
11.3 Conclusions .....	240

---

## Part V Conclusions

---

<b>12 Conclusions</b> .....	245
12.1 Lessons learned .....	245
12.2 Assumptions and Limitations .....	248
12.2.1 Shared Vocabularies .....	248
12.2.2 On demand translation .....	249
12.2.3 Modular Ontologies .....	250
12.3 Where are we now? .....	251
12.4 Is that all there is? .....	252
<b>A Proofs of theorems</b> .....	255
A.1 Theorem 6.6 .....	255
A.2 Theorem 6.11 .....	255
A.3 Theorem 6.14 .....	256
A.4 Theorem 10.9 .....	256
A.5 Theorem 10.11 .....	256
A.6 Lemma 11.1 .....	259
A.7 Theorem 11.2 .....	259
<b>References</b> .....	261
<b>Index</b> .....	275