The Semantic Web Explained

The Semantic Web is a new area of research and development in the field of computer science, aimed at making it easier for computers to process the huge amount of information on the web, and indeed other large databases, by enabling them not only to read but also to understand the information.

Based on successful courses taught by the authors, and liberally sprinkled with examples and exercises, this comprehensive textbook describes not only the theoretical issues underlying the Semantic Web, but also algorithms, optimisation ideas and implementation details. The book will therefore be valuable to practitioners as well as students, indeed to anyone who is interested in Internet technology, knowledge engineering or description logics.

Supplementary materials available online include the source code of program examples and solutions to selected exercises.

Péter Szeredi is a Professor at the Budapest University of Technology and Economics (BUTE). In the mid 1970s he authored the first Hungarian Prolog interpreter and led the development of the MProlog system, a pioneering Hungarian software product sold worldwide in the 1980s. His main fields of interest include: declarative programming (logic programming, constraint programming, the Prolog language); semantic technologies (Semantic Web, semantic integration); exploiting parallelism and programming languages supporting parallelism; as well as the implementation of programming languages. He is the author and co-author of about 90 peer-reviewed publications, including 14 books or book chapters and eight journal publications. In recognition of his pioneering work in the field, he was honoured by the Association of Logic Programming as one of the 15 Founders of Logic Programming.

Gergely Lukácsy is a Software Architect at Cisco Systems, Ireland. He received his Ph.D. in Computer Science from the Budapest University of Technology, Hungary, in June 2008. The main contribution in his Ph.D. dissertation was a novel description logic reasoning approach that works effectively in cases when there are huge amounts of data. This result is particularly important for the Semantic Web area, where it is essential to have systems that can reason efficiently over large data sets. Dr Lukácsy has published more than 25 articles in journals, books, refereed conference proceedings and workshops. He has been working on various research projects involving semantic technologies and logic programming since 2000. His main results include the DLog reasoner, which uses logic programming techniques for efficient reasoning, including program analysis and transformation techniques and parallelisation. He has also been involved in the design and development of logic-based information integration systems that use knowledge representation, constraint logic programming and reasoning techniques.

Tamás Benkő is a Research Associate at the Digital Enterprise Research Institute, Ireland. He is involved in the development of the Semantic Web indexing project Sindice. His main fields of interest include declarative programming (logic programming, constraint programming, the Prolog language) and semantic technologies (Semantic Web, semantic integration). He is the author or co-author of several related conference papers and books.
The Semantic Web Explained
# Contents

**Introduction**

<table>
<thead>
<tr>
<th>Part I</th>
<th>The Semantic Web</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>The World Wide Web today</td>
<td>9</td>
</tr>
<tr>
<td>1.1.</td>
<td>The architecture of the web</td>
<td>9</td>
</tr>
<tr>
<td>1.2.</td>
<td>Conventional web search engines</td>
<td>21</td>
</tr>
<tr>
<td>1.3.</td>
<td>Problems with searching the web</td>
<td>37</td>
</tr>
<tr>
<td>1.4.</td>
<td>Possible solutions</td>
<td>43</td>
</tr>
<tr>
<td>1.5.</td>
<td>Summary</td>
<td>48</td>
</tr>
<tr>
<td>2.</td>
<td>The Semantic Web and the RDF language</td>
<td>52</td>
</tr>
<tr>
<td>2.1.</td>
<td>Introduction</td>
<td>52</td>
</tr>
<tr>
<td>2.2.</td>
<td>The XML language</td>
<td>55</td>
</tr>
<tr>
<td>2.3.</td>
<td>The basics of the RDF language</td>
<td>65</td>
</tr>
<tr>
<td>2.4.</td>
<td>The XML syntax for RDF</td>
<td>73</td>
</tr>
<tr>
<td>2.5.</td>
<td>Special RDF constructs</td>
<td>84</td>
</tr>
<tr>
<td>2.6.</td>
<td>Representing background knowledge: RDF schema</td>
<td>98</td>
</tr>
<tr>
<td>2.7.</td>
<td>Examples of RDF use</td>
<td>110</td>
</tr>
<tr>
<td>2.8.</td>
<td>Summary</td>
<td>118</td>
</tr>
<tr>
<td>3.</td>
<td>Managing and querying RDF sources</td>
<td>122</td>
</tr>
<tr>
<td>3.1.</td>
<td>RDF descriptions on the web</td>
<td>122</td>
</tr>
<tr>
<td>3.2.</td>
<td>Parsing and processing metadata</td>
<td>127</td>
</tr>
<tr>
<td>3.3.</td>
<td>Querying RDF sources</td>
<td>130</td>
</tr>
<tr>
<td>3.4.</td>
<td>Reasoning during the execution of queries</td>
<td>156</td>
</tr>
<tr>
<td>3.5.</td>
<td>Optimising RDF queries</td>
<td>163</td>
</tr>
<tr>
<td>3.6.</td>
<td>Summary</td>
<td>168</td>
</tr>
</tbody>
</table>
Part II Ontologies and logics 173

4. Description Logic 175
  4.1. Introduction 175
  4.2. An informal introduction to the DL formalism 178
  4.3. The $AC$ DL language 183
  4.4. The $AC$ language family 189
  4.5. The $SHIQ$ language family 197
  4.6. Terminological reasoning tasks 206
  4.7. Eliminating terminological axioms 212
  4.8. ABoxes 217
  4.9. Mapping DL languages to first-order logic 227
  4.10. Advanced DL languages 231
  4.11. Summary 243

5. Reasoning on simple DL languages 246
  5.1. Reasoning in the $AC$ language 246
  5.2. The $ALCN$ tableau algorithm – an introduction 254
  5.3. The $ALCN$ tableau algorithm for empty TBoxes 267
  5.4. The $ALCN$ tableau algorithm for non-empty TBoxes 281
  5.5. Handling ABoxes in the $ALCN$ tableau algorithm 291
  5.6. Summary 299

6. Implementing a simple DL reasoning engine 302
  6.1. Introduction 302
  6.2. Executing the reasoning engine 303
  6.3. Data structures 304
  6.4. Transforming concepts to negation normal form 308
  6.5. The entry point of the tableau algorithm 311
  6.6. Tableau transformation rules 314
  6.7. Blocking 337
  6.8. Auxiliary functions 339
  6.9. Challenges 344
  6.10. Summary 345

7. The $SHIQ$ tableau algorithm 346
  7.1. An outline of the $SHIQ$ tableau algorithm 346
  7.2. Transitive roles; the language $S$ 348
  7.3. Role hierarchies – extension $H$ 352
  7.4. Inverse roles – extension $I$ 356
  7.5. Functional restrictions – extension $F$ 362
  7.6. Qualified number restrictions – extension $Q$ 374
  7.7. The complete $SHIQ$ tableau algorithm 375
  7.8. The $SHIQ$ tableau algorithm for ABoxes 385
CONTENTS

7.9. Optimisation of the tableau algorithm 387
7.10. Summary 404

Part III Ontologies and the Semantic Web 405

8. The Web Ontology Language 407
  8.1. The language OWL – an introduction 407
  8.2. The ontology language OWL 1 412
  8.3. Revision of the OWL language: OWL 2 445
  8.4. Summary 450

References 451
Index 460