THE SEMANTIC WEB

Methods, Applications and Future Trends

Rudi Studer, Sudhir Agarwal, Raphael Volz
Institute of Applied Informatics and Formal Description Methods (AIFB),
University of Karlsruhe,
D-76128 Karlsruhe, Germany.
{studer,agarwal,volz}@aifb.uni-karlsruhe.de

Abstract The paper presents the vision of the Semantic Web and describes ontologies and associated metadata as the building blocks of the Semantic Web. Current research topics and promising application areas are discussed as well.

Keywords: Semantic Web, ontologies, metadata.

1. Introduction

The development of the World Wide Web is a great success story with respect to the number of users and the amount of information that is nowadays offered by the WWW. However, most of the information that is available has to be interpreted by humans; machine support is rather limited. In order to get rid of that limitation, Tim Berners-Lee, the inventor of the WWW, coined the vision of the Semantic Web: to make the contents of the WWW accessible and interpretable by machines [Berners-Lee et al., 2001; Fensel et al., 2002; Davies et al., 2002].

Today it is almost impossible to integrate information that is spread over several Web or intranet pages. Consider, e. g., the query for a data mining expert in a company intranet, where the only explicit information stored are the relationships between people and the projects they work in on the one hand, and between projects and the topics they address on the other hand. In that case, a skills management system should be able to combine the information on the employees' home pages with the information on the projects' home pages in order to find the respective expert. To realize such scenarios, metadata have to be interpreted and appropriately combined by machines [Lau and Sure, 2002].

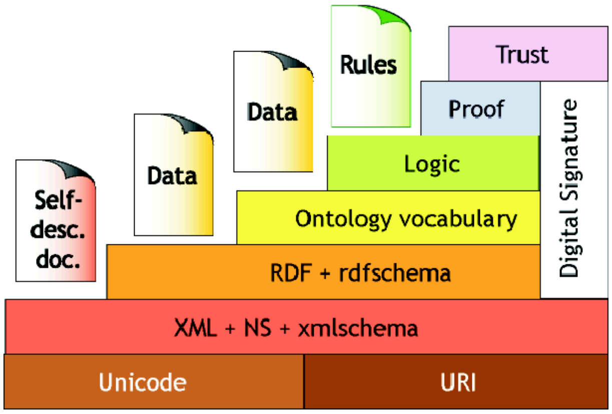

Figure 1. The layers of the Semantic Web.

The process of building the Semantic Web is still in genesis, but first standards, e.g. for the underlying data model and an ontology language already appeared. However, those structures are now to be filled with life in applications. In order to make this task feasible, one should start with the simpler tasks first. The following steps show the direction where the Semantic Web is heading:

1 providing a common syntax for machine understandable statements,

2 establishing common vocabularies,

3 agreeing on a logical language,

4 using the language for exchanging proofs.

Berners-Lee suggested a layer structure for the Semantic Web (cf. Figure 1) . This structure reflects the steps listed above. It follows the understanding that each step alone will already provide added value, so that the Semantic Web can be realized in an incremental way.

2. Layers of the Semantic Web

Unicode/Unified Resource Identifiers, XML, RDF, ontologies, logic, proof, and trust are suggested by Berners-Lee[1]: and discussed in detail for instance in [Patel-Schneider and Fensel, 2002] which also addresses open issues.

[1] see `http://www.w3.org/DesignIssues/Semantic.html`

On the first two layers, a common syntax is provided. *Uniform resource identifiers (URIs)* provide a standard way to refer to entities,[2] while *Unicode* is a standard for exchanging symbols. The *Extensible Markup Language (XML)* fixes a notation for describing labeled trees, and XML Schema allows to define grammars for valid XML documents. XML documents may refer to different *namespaces* to disambiguate between equally named tags. The formalizations on these two layers are nowadays widely accepted, and the number of XML documents is increasing rapidly.

The *Resource Description Framework (RDF)* can be seen as the first layer which is part of the Semantic Web. According to the W3C recommendation [W3C, 1999], RDF "is a foundation for processing metadata; it provides interoperability between applications that exchange machine-understandable information on the Web." RDF descriptions consist of three types of entities: resources, properties, and statements. Resources may be web pages, parts or collections of web pages, or any (real-world) objects which are not directly part of the WWW.

In RDF, resources are always addressed by URIs. Properties are specific attributes, characteristics, or relations describing resources. A resource together with a property having a value for that resource form an RDF statement. A value is either a literal, a resource, or another statement. Statements can thus be considered as object–attribute–value triples.

The middle part of Figure 2 shows an example of RDF statements. Two persons (i. e., their Web pages) are represented as resources 'URI-GST' and 'URI-AHO'. The statement on the lower right consists of the resource 'URI-AHO' and the property 'cooperates with' with the value 'URI-GST' (which again is a resource). The resource 'URI-SWMining' has as value for the property 'title' the literal 'Semantic Web Mining'. Such statements may be attached to web pages by annotation tools, as e.g. OntoMat Annotizer [Handschuh et al., 2002].

The data model underlying RDF is basically a directed, labelled pseudograph. RDF Schema defines a simple modeling language on top of RDF which includes classes, is-a relationships between classes and between properties, and domain/range restrictions for properties. RDF and RDF Schema are encoded in XML syntax, but they do not employ the tree semantics of XML.

[2] *URL (uniform resource locator)* refers to a locatable URI, e.g. an `http://...` address. It is often used as a synonym, although strictly speaking URLs are a subclass of URIs, see `http://www.w3.org/Addressing`.

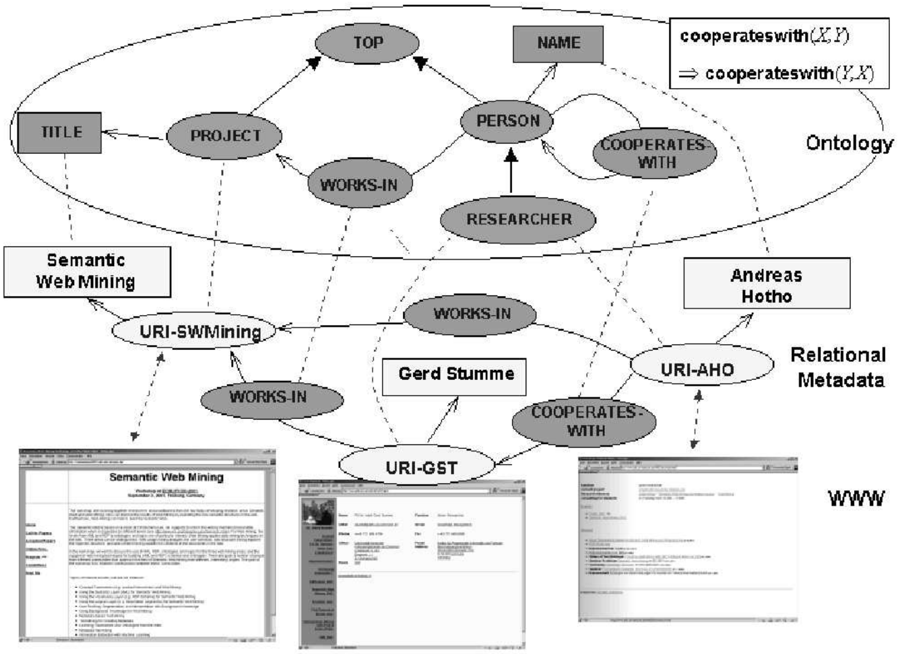

Figure 2. The relation between the WWW, relational metadata, and ontologies.

The next layer is the *ontology vocabulary.* Following [Gruber, 1993], an ontology is "an explicit formalization of a shared understanding of a conceptualization". This high-level definition is realized differently by various research communities and thereby in ontology representation languages. However, most of these languages have a certain understanding in common, as most of them include a set of concepts, a hierarchy on them, and relations between concepts. Some of them also include axioms in some specific logic. We will discuss the most prominent approaches in more detail in the next section.

Logic is the next layer according to Berners-Lee, cf. Figure 1. However, nowadays research considers usually the ontology and the logic levels together, as ontologies are already based on logic and should allow for logical axioms [Kifer et al., 1995]. By applying logical deduction, one can infer new knowledge from the information which is stated explicitly. For instance, the axiom saying that the 'cooperates with'-relation is symmetric (cf. Figure 2) allows to logically infer that the person addressed by 'URI-AHO' is cooperating with the person addressed by 'URI-GST' although only the person "GST" specifies his cooperation with the person "AHO". The kind of inference that is possible depends heavily on the logics chosen.

Proof and *trust* are the remaining layers. They follow the understanding that it is important to be able to check the validity of statements made in the (Semantic) Web. Therefore the creators of statements should be able to provide a proof which is verifiable by a machine. At this level, it is not required that the machine of the reader of the statements finds the proof itself, it 'just' has to check the proof provided by the creator. These two layers are rarely tackled in today's research. Therefore we will focus our interest on the XML, RDF, ontology and logic layers in the remainder of this paper.

3. Ontologies: Languages and Tools

A priori, many formal knowledge representation mechanisms[3] may play the role of a Semantic Web language.

Description Logics (DL) are currently the most popular framework for defining a Web ontology language. DLs are subsets of first order logic which aim at being as expressive as possible while still being decidable. The description logic $\mathcal{SHOQ}(D)$ provides the basis for the W3C proposal for a Web ontology language, i.e., OWL [Dean and Schreiber, 2003]. OWL is the result of joining the efforts of two research programs: The DARPA Agent Markup Language DAML+OIL[4] was created as part of the DAML research programme, OIL (Ontology Inference Layer) [Fensel et al., 2000] was the result of the Semantic Web research programme of the European Union.

Several tools are in use for the creation and maintenance of ontologies and metadata, as well as for reasoning within them. *Ontoedit* [Sure et al., 2002] is an ontology editor which is connected to *Ontobroker* [Decker et al., 1999], an inference engine for F–Logic [Kifer et al., 1995]. Ontobroker provides means for semantic based query handling over distributed resources. F–Logic has also influenced the development of Triple [Sintek and Decker, 2002], an inference engine which allows to model features of UML, Topic Maps, or RDF Schema.

FaCT[5] provides inference services for the Description Logic $\mathcal{SHIQ}$. The reasoning implemented by the FaCT inference engine may also be used in the OilEd [Bechhofer et al., 2001] ontology editor.

The Karlsruhe Ontology Framework KAON [Bozsak et al., 2002], [Maedche et al., 2003] is a novel open-source infrastructe that takes a

[3] See [Studer et al., 1998] for a general discussion.
[4] `http://www.daml.org`
[5] `http://www.cs.man.ac.uk/~horrocks/FaCT`

holistic approach to ontology management and is targeted for business applications. It includes a comprehensive tool suite allowing easy ontology creation and management, as well as advanced components for building ontology-based applications supporting latest Web standards.

4. Research Areas Related to the Semantic Web

There are many research areas related to the Semantic Web, like multi-agent systems or human-computer interfaces. Because of space restrictions we only discuss databases and Web Mining.

Databases The database community has a rather long history in addressing semantic aspects, most notably in the context of semantic data models that were a prominent research topic in the 1980s [Hull and King, 1987]. Semantics played and still play an important role in topics like view management, schema transformation and integration, and query processing, just to mention a few of them. However, the Semantic Web raises new challenges that were not present in the more classical database scenarios [Sheth and Meersman, 2002]: the collection of metadata that will be created as part of the Semantic Web can be seen as a massive new distributed database whose size can be of the same order of magnitude as the data itself and whose complexity may be even higher. Therefore, methods for achieving scalability and robustness in the Semantic Web have to be developed.

The specific characteristics of the data models used in the Semantic Web pose also new research questions, e.g. for view management. In [Volz et al., 2003] a first approach has been developed to provide a view mechanism for an RDF-based data model. Handling properties as first class citizens and being able to cope with non-strict typing are some of the specific characteristics that have to be addressed by a view mechanism for the Semantic Web.

Semantic Web Mining The novel research area of Semantic Web Mining aims at combining the two fast-developing research areas Semantic Web and Web Mining. Web mining is the application of data mining techniques to the content, structure, and usage of Web resources. This can help to discover global as well as local structure within and between Web pages. This means that Web mining is an invaluable help in the transformation from human understandable content to machine understandable semantics. Three areas of Web mining are commonly distinguished: content mining, structure mining, and usage mining [Berendt et al., 2002].

As the Semantic Web enhances the first generation of the WWW with formal semantics, it offers a good basis to enrich Web Mining: The types of (hyper)links are now described explicitly, allowing the knowledge engineer to gain deeper insights in Web structure mining; and the contents of the pages come along with a formal semantics, allowing her to apply mining techniques which require more structured input. On the other hand, Web Mining can help setting up the Semantic Web. It can help to learn structures for knowledge organization (e. g., ontologies) from the Web and to provide the population of such knowledge structures.

5. Application Areas

Different application areas benefit from the Semantic Web. We briefly present some areas which are currently under development.

Web Services Semantic Web enabled Web Services are one of the heavily discussed topics within the Semantic Web community. Fensel et al. define in [Fensel and Bussler, 2002] a Web Service Modeling Framework which provides a conceptual model for describing and developing web services. There, Semantic Web technology is used for service description and service discovery.

An application scenario for the configuration of Web services is presented in [Felfernig et al., 2002]. This approach is ontology based and aims at overcoming the huge effort needed for enabling B2B solutions such as eProcurement. Paolucci et al. [Paolucci et al., 2002] and Burstein et al. [Burstein et al., 2002] use DAML-S for defining an ontology for Web Services. The formers show the advantage of a declarative description for supporting the matching process by exploiting ontologies. The latter presents different aspects of web service ontologies and focus on the service grounding. The objective the authors aim at in [Agarwal et al., 2003] is that a human user can seamlessly surf the existing World Wide Web and the emerging web services and that he can easily compose and invoke web services on the fly without being a software engineer. They present a framework, which trades off between having a reasonably easy to use interface for web services and the complexity of web service workflows.

E-Learning Sharing knowledge is the main idea of education. With the growing amount of educational material in the WWW, this idea gets a new dimension and generates new technical challenges. Metadata schemes for the exchange of educational Web resources have been in use for a number of years. These metadata schemes, for example LOM

(Learning Objects Metadata)[6], usually extend the Dublin Core standard[7]. However, these standards lack a precise machine-interpretable semantics to describe the content of the learning objects. In [Schmitz et al., 2002] an approach for accessing and browsing distributed learning repositories is described that exploits ontologies and associated relational metadata.

Peer-To-Peer networks Peer-To-Peer networks can be seen as distributed repositories. In order to retrieve data from another peer, it is of paramount importance to know what the other peer provides. While some Peer-to-Peer scenarios (e. g., music networks such as Gnutella) work well with limited amount of metadata, most applications need more semantic information in order to find relevant peers efficiently. The Edutella framework, for instance, was established to support the exchange of RDF based repositories between peers. [Nejdl et al., 2002] describe use case scenarios for annotation and replication and propose the Modification Exchange Language MEL for such distributed RDF repositories. [Nejdl et al., 2003] discuss the RDF-based P2P networks as a specific example of a new type of P2P networks, schema-based P2P network, and describe the use of super-peer based topologies for these networks. In [Hatala and Richards, 2002], requirements and corresponding solutions for easy interaction between content providers and consumers are described.

Knowledge Management and Knowledge Portals New trends in knowledge management pave the way from the more document oriented view on knowledge management to a more knowledge item oriented view [Staab et al., 2001]. Such approaches again rely on Semantic Web methods, especially ontologies and metadata. In [Davies et al., 2002] a collection of methods and tools is described that exploit Semantic Web techniques for applications like skills management or for supporting virtual enterprises.

Knowledge portals provide views onto domain-specific information on the World Wide Web for facilitating their users to find relevant information. The extensive maintenance needed for keeping a portal up to date can be simplified by using an ontology as conceptual backbone for acquiring, maintaining, and providing information. SEAL [Maedche et al., 2002] is a comprehensive architecture for a semantic portal offering a broad range of tools for improving the cost/benefit ratio of semantic

[6] see `http://ltsc.ieee.org/wg12/`

[7] `http://dublincore.org`

portals. SEAL-II [Hotho et al., 2001] tackles the soft spot between unstructured knowledge and richly structured knowledge. To achieve that objective, SEAL-II includes a broad collection of techniques to instantiate knowledge and to browse through collections of documents and knowledge elements.

6. Conclusion

The development of the Semantic Web is a fast moving process that raises a lot of research challenges that have to be addressed in an interdisciplinary way. Furthermore, it paves the way to new types of applications that exploit the semantic basis of this new generation of the WWW. These aspects make the Semantic Web so fascinating.

Acknowledgments

We thank all our colleagues at the Institute AIFB and at the FZI Research Center for Information Technologies for a lot of fruitful discussions. Part of this research is funded by EU in the IST project WonderWeb (project no. 2001-33052), by bmb+f in the SemIPort project (project no. 08C5939) and by bmb+f in the SESAM project.

References

Agarwal, Sudhir, Handschuh, Siegfried, and Staab, Steffen (2003). Surfing the service web. In *ISWC2003: 2nd International Semantic Web Conference 2003, Sanibel Island, Florida, USA*.

Bechhofer, S., Horrocks, I., Goble, C., and Stevens, R. (2001). OilEd: A reasonable ontology editor for the Semantic Web. In *Proceedings of KI2001, Joint German/Austrian conference on Artificial Intelligence*, volume 2174 of *LNAI*, pages 396–408. Springer.

Berendt, B., Hotho, A., and Stumme, G. (2002). Towards semantic web mining. In *ISWC2002: Ist International Semantic Web Conference, Sardinia, Italy*, Lecture Notes in Computer Science, pages 264–278. Springer.

Berners-Lee, T., Hendler, J., and Lassila, O. (2001). The semantic web. *Scientific American, online at http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21.*

Bozsak, E., Ehrig, M., Handschuh, S., Hotho, A., Maedche, A., Motik, B., Oberle, D., Schmitz, C., Staab, S., Stojanovic, L., Stojanovic, N., Studer, R., Stumme, G., Sure, Y., Tane, J., Volz, R., and Zacharias, V. (2002). KAON - Towards a Large Scale Semantic Web. In Bauknecht, K., Tjoa, A. Min, and Quirchmayr, G., editors, *E-Commerce and Web Technologies, Third International Conference, EC-Web 2002, Aix-en-Provence, France, September 2-6, 2002, Proceedings*, volume 2455 of *Lecture Notes in Computer Science*, pages 304–313. Springer.

Burstein, M., Hobbs, J., Lassila, O., Martin, D., McDermott, D., McIlraith, S., Narayanan, S., Paolucci, M., Payne, T., and Sycara, K. (2002). DAML-S: Web Service Description for the Semantic Web. In *ISWC2002: Ist International Semantic Web*

Conference, Sardinia, Italy, Lecture Notes in Computer Science, pages 348–363. Springer.

Davies, John, Fensel, Dieter, and van Harmelen, Frank, editors (2002). *Towards The Semantic Web: Ontology-Driven Knowledge Management*. Wiley.

Dean, M. and Schreiber, G. (2003). Web ontology language (OWL) reference version 1.0. Internet: http://www.w3.org/TR/owl-ref/.

Decker, S., Erdmann, M., Fensel, D., and Studer, R. (1999). Ontobroker: Ontology based Access to Distributed and Semi-structured Information. In et al., R. Meersman, editor, *Database Semantics: Semantic Issues in Multimedia Systems.*, pages 351–369. Kluwer Academic.

Felfernig, A., Friedrich, G., Jannach, D., and Zanker, M. (2002). Semantic Configuration of Web Services in the CAWICOMS Project. In *ISWC2002: Ist International Semantic Web Conference, Sardinia, Italy*, Lecture Notes in Computer Science, pages 192–205. Springer.

Fensel, D. and Bussler, C. (2002). The Web Service Modeling Framework WSMF. *Electronic Commerce Research and Applications*, 1(2):113–137.

Fensel, D., Horrocks, I., van Harmelen, F., Decker, S., Erdmann, M., and Klein, M. (2000). Oil in a nutshell. In *Knowledge Acquisition, Modeling, and Management, Proceedings of the European Knowledge Acquisition Conference (EKAW-2000)*, volume 1937 of *LNAI*, page 1 ff. Springer.

Fensel, Dieter, Hendler, James, Lieberman, Henry, and Wahlster, Wolfgang, editors (2002). *Spinning the Semantic Web*. MIT Press.

Gruber, T. R. (1993). Towards Principles for the Design of Ontologies Used for Knowledge Sharing. In Guarino, N. and Poli, R., editors, *Formal Ontology in Conceptual Analysis and Knowledge Representation*, Deventer, The Netherlands. Kluwer Academic Publishers.

Handschuh, S., Staab, S., and Ciravegna, F. (2002). S-CREAM — semi-automatic CREAtion of Metadata. In *13th International Conference on Knowledge Engineering and Knowledge Management (EKAW02)*, volume 2473 of *LNAI*, page 358 ff.

Hatala, M. and Richards, G. (2002). Global vs. Community Metadata Standards: Empowering Users for Knowledge Exchange. In *ISWC2002: Ist International Semantic Web Conference, Sardinia, Italy*, Lecture Notes in Computer Science, pages 292–306. Springer.

Hotho, A., Maedche, A., Staab, S., and Studer, R. (2001). SEAL-II — the soft spot between richly structured and unstructured knowledge. *Journal of Universal Computer Science (J.UCS)*, 7(7):566–590.

Hull, R. and King, R. (1987). Semantic database modeling: survey, applications, and research issues. *ACM Computing Surveys (CSUR)*, 19(3):201–260.

Kifer, M., Lausen, G., and Wu, J. (1995). Logical foundations of object-oriented and frame-based languages. *Journal of the ACM*, 42:741–843.

Lau, T. and Sure, Y. (2002). Introducing ontology-based skills management at a large insurance company. In *Modellierung 2002: Modellierung in der Praxis - Modellierung fuer die Praxis*, Tutzing.

Maedche, A., Motik, B., Stojanovic, L., Studer, R., and Volz, R. (2003). An infrastructure for searching, reusing and evolving distributed ontologies. In *WWW2003: The Twelfth International World Wide Web Conference, Budapest, Hungary*.

Maedche, A., Staab, S., Studer, R., Sure, Y., and Volz, R. (2002). SEAL – Tying up information integration and web site management by ontologies. *IEEE-CS Data Engineering Bulletin, Special Issue on Organizing and Discovering the Semantic Web*.

Nejdl, W., Siberski, W., Simon, B., and Tane, J. (2002). Towards a Modification Exchange Language for Distributed RDF Repositories. In *ISWC2002: Ist International Semantic Web Conference, Sardinia, Italy*, Lecture Notes in Computer Science, pages 236–249. Springer.

Nejdl, W., Wolpers, M., Siberski, W., Schmitz, C., Schlosser, M., and Brunkhorst, I. (2003). Super-peer-based routing and clustering strategies for rdf-based peer-to-peer networks. In *WWW2003: The Twelfth International World Wide Web Conference, Budapest, Hungary*.

Paolucci, M., Kawamura, T., Payne, T.., and Sycara, K. (2002). Semantic matching of web services capabilities. In *ISWC2002: Ist International Semantic Web Conference, Sardinia, Italy*, Lecture Notes in Computer Science, pages 333–347. Springer.

Patel-Schneider, P. and Fensel, D. (2002). Layering the semantic web: Problems and directions. In *ISWC2002: Ist International Semantic Web Conference, Sardinia, Italy*, Lecture Notes in Computer Science, pages 16–29. Springer.

Schmitz, Ch., Staab, S., Stumme, G., Studer, R., and Tane, J. (2002). Accessing distributed learning repositories through a courseware watchdog. In *Proceedings of E-Learn 2002: World Conference on E-Learning in Corporate, Government, Healthcare and Higher Education*, Montreal, Canada.

Sheth, A. and Meersman, R. (2002). Amicalola report: Databases and information systems research challlenges and opportunities in semantic web and enterprises. *SIGMOD Record*, 31(4).

Sintek, M. and Decker, S. (2002). Triple — a query, inference, and transformation language for the semantic web. In *ISWC2002: Ist International Semantic Web Conference, Sardinia, Italy*, Lecture Notes in Computer Science, pages 364–378. Springer.

Staab, S., Studer, R., Schnurr, H.-P., and Sure, Y. (2001). Knowledge processes and ontologies. *IEEE Intelligent Systems*, 16(1):26–34.

Studer, R., Benjamins, V. Richard, and Fensel, D. (1998). Knowledge engineering: Principles and methods. *Data + Knowledge Engineering*, 25(1-2):161–197.

Sure, Y., Erdmann, M., Angele, J., Staab, S., Studer, R., and Wenke, D. (2002). OntoEdit: Collaborative ontology development for the semantic web. In *ISWC2002: Ist International Semantic Web Conference, Sardinia, Italy*, Lecture Notes in Computer Science, pages 221–235. Springer.

Volz, R., Oberle, D., and Studer, R. (2003). Views for light-weight web ontologies. In *Proceedings of ACM Symposium of Applied Computing (SAC)*, Melbourne, Florida, USA.

W3C (1999). *Resource Description Framework (RDF): Model and Syntax specification*. W3C recommendation, 22 February 1999, www.w3.org/TR/REC-rdf-syntax-19990222/.