Encyclopedia of Internet Technologies and Applications

Mario Freire
University of Beira Interior, Portugal

Manuela Pereira
University of Beira Interior, Portugal

Information Science REFERENCE

INFORMATION SCIENCE REFERENCE
Hershey • New York

Acquisitions Editor: Kristin Klinger
Development Editor: Kristin Roth
Senior Managing Editor: Jennifer Neidig
Managing Editor: Sara Reed
Copy Editor: Larissa Vinci and Mike Goldberg
Typesetter: Amanda Appicello and Jeffrey Ash
Cover Design: Lisa Tosheff
Printed at: Yurchak Printing Inc.

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue, Suite 200
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com/reference

and in the United Kingdom by
Information Science Reference (an imprint of IGI Global)
3 Henrietta Street
Covent Garden
London WC2E 8LU
Tel: 44 20 7240 0856
Fax: 44 20 7379 0609
Web site: http://www.eurospanonline.com

Library of Congress Cataloging-in-Publication Data

Encyclopedia of Internet technologies and applications / Mario Freire and Manuela Pereira, editors.
p. cm.
Summary: "This book is the single source for information on the world's greatest network, and provides a wealth of information for the average Internet consumer, as well as for experts in the field of networking and Internet technologies. It provides the most thorough examination of Internet technologies and applications for researchers in a variety of related fields"--Provided by publisher.
Includes bibliographical references and index.
ISBN 978-1-59140-993-9 (hardcover) -- ISBN 978-1-59140-994-6 (ebook)
1. Internet--Encyclopedias. I. Freire, Mário Marques, 1969- II. Pereira, Manuela.
TK5105.875.I57E476 2007
004.67'803--dc22
2007024949

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

Semantic Web Languages and Ontologies

Livia Predoiu
University of Mannheim, Germany

Anna V. Zhdanova
University of Surrey, UK

INTRODUCTION

On the current World Wide Web, most of the information is stored syntactically, i.e., only as data. The information that lies within the data can only be understood by humans and not automatically by computer programs. In order to overcome this issue, the idea of encoding the information not just syntactically but also with semantics has created a new notion of the Web called *Semantic Web*. This notion emerged together with developments of semi-structured languages like SGML and XML.

As the Web is a "many-to-many data-interchange medium," it works only if the exchange format provides interoperability at various levels (Decker et al., 2000). The first requirement is *universal expressive power,* which means that any type of data is expressible. The second requirement is *syntactic interoperability,* which means that data and information modelling languages have a wide implementation support. The third, and last, requirement is *semantic interoperability*, which means that mappings between terms within the data can be discovered automatically.

It has been agreed in research that the best way to represent semantics is to use formal ontologies. This agreement is due to the fact that ontologies can be seen as formal and structured representations of the concepts and relationships a domain exposes and thus of the data within as information. There exist different kinds of ontologies (e.g., lightweight and heavyweight) and different languages for describing them. Currently, XML-based W3C recommendations, such as RDF(S) and OWL, are the most widely-exploited ontology languages. However, new extensions and usages of existing ontology languages, as well as new ontology languages, continue to appear. The amount of domain-dependent ontologies also expands rapidly over time. Following the successes of RSS, FOAF, vCard, and Dublin Core, ontologies are now being constructed to describe practically every side of human life. Therefore, methods for ontology discovery and reuse gain further importance. Dynamicity of Semantic Web development makes it difficult for Internet application developers to be aware of existing initiatives and trends in ontologies and ontology languages. Choosing and adopting Semantic Web core components thus remains a challenge. To assist in addressing this challenge, we provide an overview of leading ontology languages and ontologies in this paper.

The paper is organized as follows. In the second section we present a background to the topic. State of the art and trends in the area of languages and ontologies on the Web are discussed in the third section. In the fourth section we identify challenges in this area and,finally, the fifth section concludes the article.

BACKGROUND

Semantic Web languages and ontologies are strongly related. While a Semantic Web Language can be used to specify an ontology, a Semantic Web language can also be seen as a an ontology with its modelling primitives being ontological entities. In this chapter we first describe the evolution of Semantic Web languages and then clarify the meaning of the term ontology.

HTML (Ragett et. al., 1999) was the promoter of the World Wide Web because it provided a standard for structuring documents such that browsers were able to display them in a uniform way. The first Web-based Ontology Language, SHOE, was based on HTML. The main disadvantage of HTML, namely being a very rigid and inflexible language by not allowing adaptation to different types of documents, was overcome by the standardization of XML (Bray et. al., 2004) and XML Schema.

However, syntactic interoperability is not enough for the purpose of automatic processing of queries posed to the huge amount of data available on the Internet. In order to enable information sharing, information integration, and retrieval of semantically similar documents available on the Web, languages offering means to model semantics are required. XML is not appropriate for semantic interoperability because it "just describes grammars" and therefore it is not able "to recognize a semantic unit from a particular domain because XML aims at document structure and imposes no common interpretation of the data contained in the document" (Decker et. al, 2000).

A first step to syntactic and semantic interoperability has been reached by means of the standardization of RDF/RDF schema. RDF has been attached to an XML syntax but it has a different and much richer data model. RDF is designed to provide "a basic object-attribute-value data model." By this means, chaining (i.e., any object can play the role of a value) and reification (any RDF statement can be the object or value of a triple) can be modelled. However, "this intentional semantics is described only informally in the standard" and "apart from that, RDF does not make any data-modelling commitments" (Decker et al., 2000). But RDF Schema introduces "some simple ontological concepts" (McBride 2004) and provides means for defining more complex and formal Ontology Languages on top of it.

The first full-fledged ontology languages to be defined on top of RDF/RDF Schema and standardized by the W3C were OIL and DAML+OIL. They correspond to some Description Logics variants and thus provide well-known algorithms for automatic reasoning and, therefore, the automatic handling of information by software tools.

The term ontology originates from philosophy where it denotes "*the theory or study of being as such, i.e., the basic characteristics of all reality.*"[1] In computer science, different views of the meaning of the term "ontology" exist (Guarino & Giaretta, 1995). Studer et al. (1998) have merged different definitions and set up the following one which has been adopted by the Semantic Web community:

An ontology is a formal, explicit specification of a shared conceptualization. Conceptualization refers to an abstract model of some phenomenon in the world by having identified the relevant concepts of that phenomenon. Explicit means that the type of concepts used, and the constraints on their use are explicitly defined. Formal refers to the fact that the ontology should be machine-readable. Shared reflects the notion that an ontology captures consensual knowledge, that is, it is not private of some individual, but accepted by a group.

Ontologies can be categorized according to different criteria. Guarino (1998) classifies ontologies according to their usage and application field. He distinguishes between top-level, domain, task, and application ontologies. Lassila and McGuiness (2002) classify ontologies according to their expressivity and internal structure. They mainly distinguish between *lightweight ontologies* and *heavyweight ontologies*. For heavyweight ontologies, a proof theory can be defined and thus automatic reasoning is possible. Automatic reasoning means that the computer can deduce new information that is encoded implicitly in the ontology. Automatic deduction of information implicitly encoded in the ontology leads to powerful means of querying these so-called knowledge bases. Ontologies are also used for Information Integration purposes as the semantics of different, seemingly disconnected, knowledge bases or ontologies as specified by means of the ontologies.

LANGUAGES AND ONTOLOGIES ON THE WEB

In this section we describe state of the art in Semantic Web languages and typical widespread ontologies lying in the core area of current Web applications. Furthermore, we identify domains where the Semantic Web is present factually, though not formally, or not standardized, and outline further areas of use for ontology languages.

Semantic Web Languages on the Web

RDF(S)

RDF (Lassila & Swick, 1999; Manola & Miller, 2004) became a W3C recommendation in 1999. It is a general-purpose language for representing resources on the Web in terms of named properties and values (McBride, 2004). With RDF, it is not possible to define the relationships between properties and resources. For this purpose, RDF Schema (Brickley & Guha, 2004) has been specified. It became a W3C recommendation in 2004 and is basically an extension of RDF. More specifically, it is a formal description language for eligible RDF expressions. In particular, a schema defines the kinds of properties available for resources (e.g., title, author, subject, size, colour, etc.) and the kind of resource classes being described (e.g., books, Web pages, people, companies, etc.). RDF Schema is a simple ontology and a simple ontology definition language. RDF and RDF Schema are usually denoted RDF(S).

RDF(S) is based on some syntactical principles of XML (e.g., URIs) and has been equipped with an XML syntax as well. The most basic Semantic Web language which provides the syntactical basis for all other Semantic Web languages is RDF(S). RDF(S) is not provided completely with a formal logical semantics, and thus reasoning is only possible in part.

Topic Maps

Topic maps are a data modelling language that became an ISO standard (ISO/IEC 13250) in 2000. A Topic Map offers a means to create an index of information which resides *outside* of that information. It describes the information in documents and databases by linking into them using URIs. A topic map consists of topics, associations (relationships between topics), and occurrences (information resources relevant to a topic). Topics and occurrences can be typed. Types in topic maps are themselves topics and thus there is no real difference between a topic and a type.

There exists SGML, XML, and RDF language support for Topic Maps. However, they are very simple and do not have a formal semantics and thus no sophisticated inference support. Nonetheless, because of their simplicity, they are often used in industry applications.

OWL

OWL (Dean et al., 2004) became a W3C recommendation in 2004. OWL is mainly based on OIL and DAML+OIL, which are obsolete Semantic Web languages and therefore not mentioned further here. OWL is equipped by an RDF syntax and includes three sub languages:

- *OWL-Lite* roughly consists of RDF(S) plus equality and 0/1-cardinality. It is intended for classification hierarchies and simple constraints. OWL-Lite corresponds semantically to the formal Description Logic *SHIF(D)* and cannot express the whole RDF vocabulary.
- *OWL-DL* contains the language constructs of OWL-Lite. OWL-DL corresponds semantically to the Description Logic *SHOIN*(D). Although strictly more expressive than OWL-Lite, it still provides computational completeness and decidability.
- *OWL Full* does not correspond to a formal logic anymore as it builds upon the complete RDF(S) vocabulary which also lacks a correspondence to a formal logic. The language incorporates maximum expressive power and syntactic freedom, but offers no computational guarantees.

Semantic Web Languages not Yet Standardized

In this subsection, we consider Semantic Web languages which have been submitted to the W3C and thus have communities promoting them. At least some of them can be expected to become W3C recommendations.

Examples of such languages are:

- **Languages based on the logic programming knowledge representation paradigm:** The trend to the aforementioned paradigm exists already since the year 2000 when the development of RuleML[2] was started. RuleML is a set of languages revolving around the Logic Programming paradigm and equipped with an RDF syntax. Other examples of Semantic Web Languages with Logic Programming semantics are WRL[3], a set of three-layered rule languages of increasing expressivity, and SWRL[4], a language which combines OWL and RuleML but is computationally intractable. Furthermore, a W3C working group[5] has been formed to establish standards for Semantic Web rule languages.
- **Semantic Web service modelling languages:** Semantic Web services will play an important role in the Semantic Web as they combine Web Services with semantics. Examples for Semantic Web Services Languages are WSML[6] and SWSL[7]. The languages serve for the specification of ontologies describing Semantic Web services. E.g., WSML is used to describe WSMO[8] and SWSL is used to describe SWSO[9].

The advantage of rule languages is that role chaining can be expressed and, thus, also the transitive hull of a relation can be computed. This is not possible in languages based on description logics. Furthermore, databases can be integrated much easier to the Web by means of rule languages. However, description logics have advantages as well, e.g., negation is based on the Open World Assumption (OWA) and thus seems to be more suitable for the Web in which Web sites providing information can become available or unavailable at any time.

Ontologies on the Web

vCard, FOAF, Dublin Core, and RSS

There are several examples of ontologies that became widely accepted and reused for the purpose of distributed data exchange and integration for semantic community portals. Very often these ontologies were organically grown and quickly found a large number of creative users, even though for a long time they were not endorsed by any of the popular standards committees. Two examples of the most often described domains are represented by ontologies describing a *person* and ontologies describing a *document*. Below, we provide typical examples of the person and document ontologies that gained a high degree of popularity.

Person ontologies:

1. **VCard**[10] is a schema to specify electronic business card profiles. Factually, vCard is a simple ontology to describe a person with 14 attributes, such as Family Name, Given Name, Street Address, Country, etc. The ontology provides a precise way to describe the instance data using RDF.

2. **FOAF** (Friend of a Friend) is a schema which is similar to VCard in a way that FOAF also is a wide-spread ontology to describe a person. FOAF schema provides 12 core attribute types that are similar to the attribute vCard provides: First Name, Last Name, E-mail address, etc., and the precise way to describe the instance data using RDF is also proposed by the FOAF-project.

Document/Web publication ontologies:

1. **Dublin Core**[11] stands for a vocabulary aimed to be used to semantically annotate Web resources and documents. The core vocabulary consists of 15 attributes to describe a document or a Web resource and contains parameters that express the primary characteristics of the documents, e.g., Title, Creator, Subject, Description, Language, etc.
2. **RSS**[12] is variably used as a name by itself and as an acronym for “RDF Site Summary,” “Rich Site Summary,” or “Really Simple Syndication.” The RSS ontology specifies the model, syntax, and syndication feed format and consists of four concepts: “channel,” “image,” “item,” and “textinput,” each of them having some attributes like “title,” “name,” and “description.”

Internet Ontologies not Formalized Semantically

In this subsection, we consider popular schemata which are most typical for the Web, and are used for representation of information commonly found on the Web. Currently there exist a large number of schemata that are shared among people and applications and could be ontologies, but they do not employ common Semantic Web languages.

Examples of such shared schemata include:

- RSS, this aforementioned format is more successful and widespread in its XML version rather than in the Semantic RDF version;
- WSDL[13], which is a current W3C candidate recommendation, and most other acknowledged Web service specification formalisms are based purely on XML technologies;
- EMMA[14] is a shared format facilitating interoperation and information exchange between fixed and mobile terminals supporting different modality. It is designed as an XML schema and the instance data are exchanged duly in the XML format.

Bringing these and new appearing schemata for information sharing to the use of a Semantic Web language will grant benefits. Such benefits would include:

- Improved interoperability between different knowledge domains (e.g., services and applications addressing the same topic)
- Integration of information arriving from different platforms (e.g., seamless switch of data stream from a mobile device to a fixed terminal, roaming)
- More accurate/expressive representation of the information entities (e.g., knowledge engineers have an opportunity to unambiguously define and share arbitrary datatypes—whereas XML schema formalism is not sufficient for this)

At the moment, transition from shared schemata to ontologies on the Web is hindered by the fact that currently XML has a larger community behind it than any Semantic Web language does. Thus, XML is used more widely due to the fact many developers share experience with this technology. Another barrier to transition to the Semantic Web is complexity of Semantic Web formalisms: Even the simplest Semantic languages are more complex than XML schema.

FUTURE TRENDS

In addition to the trend towards migration to the Semantic Web in knowledge representation, future trends in the area of Semantic Web languages and ontologies include addressing the following challenges.

Integrated Semantics for Monotonic and Nonmonotonic Negation

The languages that potentially seem to be prevalent in the Semantic Web will revolve around logic programming paradigms, and thus nonmonotonic negation under the CWA and description logic paradigms and monotonic negation under the OWA. These different kinds of negation need to be integrated semantically in order to enable a coherent and valid reasoning mechanism. Different preliminary ideas have been manifested, like scoped negation (Polleres et al., 2006) or Open Answer Set Programming (Heymans et al., 2005). However, these ideas are at a very preliminary stage and it still remains a big challenge to integrate monotonic and nonmonotonic negation properly into a Semantic Web language and provide tractable reasoning algorithms for such a language.

Representation of Context, Mappings, and Information Integration

The Internet forms a huge network of peers where each peer employs an independent set of ontologies, with each peer possibly using different knowledge representation formalisms. In this setting, the context or the mappings that connect the different ontologies needs to be represented in order to enable a coherent reasoning mechanism for the distributed information in the network. Approximation to a solution includes C-OWL (Bouquet et al., 2004) and community-driven ontology matching (Zhdanova & Shvaiko, 2006).

Convergence of Mobile and the Web: Adoption of Semantics

Mobile environments and the Web converge forming a shared communication sphere. Convergence of mobile and the Web causes the appearance of new settings to be supported, such as when the user utilizes mobile and fixed devices to interact with systems, and mobile applications become increasingly connected with the Internet. To ensure interoperation of mobile and Web applications, and tools in such a sphere, developers need to have a shared specification of objects belonging to the sphere and their roles. Certain ontologies have already been developed for the mobile communications area with employment of Semantic Web formalisms (Pfoser et al., 2002; Korpipää et al., 2004). However, widespread and global adoption of such ontologies remains a challenge.

CONCLUSION

State of the art and trends in Semantic Web languages and ontologies are presented in this article. Languages and ontologies in general are detailed, and the contributions of Semantic Web technologies to the Web have been discussed. Specific attention is paid to semantic languages recommended by W3C. Future challenges in this area have been outlined, including the development of an integrated semantics for monotonic and nonmonotonic negation, the representation of context and mappings, and the convergence of mobile and the Wweb.

REFERENCES

Bouquet, P., Giunchiglia, F., van Harmelen, F., Serafini, L., & Stuckenschmidt, H. (2004). Contextualizing ontologies. *Journal of Web Semantics, 1*(4), 325-343.

Bray, T., Paoli, J., Sperberg-McQueen, C.M., Maler, E., & Yergeau, F. (2004) Extensible markup language (XML) 1.0. W3C Recommendation, http://www.w3.org/TR/REC-xml/

de Bruijn, J., Lara, J. R., Polleres, A., & Fensel, D. (2005). OWL-DL vs. OWL flight: Conceptual modelling and reasoning for the semantic Web. In *Proceedings of the 14th International Conference on World Wide Web*.

Decker, S., Melnik, S., van Harmelen, F., Fensel, D., Klein, M., Broekstra, J., et al. (2000). The Semantic Web: The roles of XML and RDF. *IEEE Internet Computing, 4*(5), 63-67.

Ding, L., Zhou, L., Finin, T., & Joshi, A. (2005). How the semantic Web is being used: An analysis of FOAF documents. In *Proceedings of the 38th International Conference on System Sciences.*

Guarino, N. (1998). Formal ontology in information systems. In N. Guarino (Ed.), *First international conference on formal ontology in information systems.*

Guarino, N., & Giaretta, P. (1995). Ontologies and knowledge bases: Towards a terminological clarification. In N. Mars (Ed.), *Towards very large knowledge bases* (pp. 25-32).

Heymans, S., Van Nieuwenborgh, D., & Vermeir, D. (2005). Guarded open answer set programming. In *Proceedings of the 8th International Conference on Logic Programming and Non Monotonic Reasoning*.

Korpipää, P., Häkkilä, J., Kela, J., Ronkainen, S., & Känsälä I. (2004). Utilising context ontology in mobile device application personalization. *ACM International Conference Proceeding Series*, Vol. 83 (pp. 133-140).

Lassila, O., & McGuiness, D. (2002). The role of frame-based representation on the Semantic Web. *Electronic Transactions on Artificial Intelligence*.

Lassila, O., & Swick, R. (1999). *Resource description framework (RDF) model and syntax specification*. W3C Recommendation. http://www.w3.org/TR/REC-rdf-syntax/

Manola, F., & Miller, E. (2004). RDF Primer. W3C Recommendation. http://www.w3.org/TR/rdf-primer/

Pfoser, D., Pitoura, E., & Tryfona, N. (2002). Metadata modeling in a global computing environment. In *The 10th ACM International Symposium on Advances in Geographic Information Systems*.

Polleres, A., Feier, C., & Harth, A. (2006). Rules with contextually scoped negation. In *Proceedings of the 3rd European Semantic Web Conference*. Springer-Verlag. LNCS 4011.

Ragett, D., le Hors, A., & Jacobs, I. (1999). HTML 4.01 Specification. W3C Recommendation. http://www.w3.org/TR/REC-html40/

Studer, R., Benjamins, V.R., & Fensel, D. (1998). Knowledge engineering: Principles and methods. *IEEE Transactions on Data and Knowledge Engineering 25*(1-2),161-197.

Zhdanova, A. V., & Shvaiko, P. (2006). Community-driven ontology matching. In *Proceedings of the 3rd European Semantic Web Conference*. Springer-Verlag. LNCS 4011.

KEY TERMS

Closed World Assumption (CWA): The presumption that what is not currently known to be true is false. This assumption introduces nonmonotonicity if the world is not closed, i.e., new information can be introduced or deleted. The opposite of the CWA is the OWA, stating that lack of knowledge does not imply falsity.

Heavyweight Ontology: An ontology of different, but rather higher expressivity, which bases on a formal logic. With such an ontology, it is possible to perform formal reasoning.

Lightweight Ontology: An ontology which corresponds rather to a vocabulary and usually does not base on a formal logic.

Ontology: A data model that represents the objects, sets of similar objects (i.e., classes), and their interrelations within a domain of discourse.

Ontology Language: A formal language used to encode an ontology.

Open World Assumption (OWA): The presumption that what is not stated true is unknown and thus cannot be assumed to be either true or false. The OWA is considered being implicit in the Web as new Web sites can be connected and disconnected to the Web at any time.

Semantic Interoperability: Applications can understand the meaning of representations and thus can setup automatically mappings between different representations by content analysis.

Semantic Web: The next evolutionary step of the World Wide Web. It bases on language standards that provide not only universal expressive power and syntactic interoperability, but also semantic interoperability.

Syntactic Interoperability: Applications can take advantage of parsers and APIs providing syntactical manipulation facilities. If a language is standardized, it is used actively and required parsers and APIs are implemented.

ENDNOTES

[1] This definition has been taken from the Encyclopaedia Britannica.
[2] http://www.ruleml.org
[3] http://www.w3.org/Submission/WRL/
[4] http://www.w3.org/Submission/SWRL/
[5] http://www.w3.org/2005/rules/wg
[6] http://www.w3.org/Submission/WSML/
[7] http://www.w3.org/Submission/SWSF-SWSL/
[8] http://www.w3.org/Submission/WSMO/
[9] http://www.w3.org/Submission/SWSF-SWSO/
[10] http://www.w3.org/TR/vcard-rdf
[11] http://dublincore.org
[12] http://web.resource.org/rss/1.0
[13] http://www.w3.org/TR/wsdl20/
[14] http://www.w3c.org/TR/emma/