Knowledge Representation, Ontologies, and the Semantic Web*

Evimaria Terzi[1], Athena Vakali[1], and Mohand-Saïd Hacid[2]

[1] Informatics Dpt., Aristotle University,
54006 Thessaloniki, Greece
evimaria,avakali@csd.auth.gr
[2] Université Claude Bernard Lyon 1,
Laboratoire d'Ingénierie des Systèmes d'Information,
69622 Villeurbanne, France
mshacid@.bat710.univ-lyon1.fr

Abstract. A unified representation for web data and web resources, is absolutely necessary in nowdays large scale Internet data management systems. This representation will allow for the machines to meaningfully process the available information and provide semantically correct answers to imposed queries. Ontologies are expected to play an important role towards this direction of web technology which defines the so called, Semantic Web. The goal of this paper is to provide an overview of the Knowledge Representation (KR) techniques and languages that can be used as standards in the Semantic Web .

1 Introduction

The development of standards that guarantee interoperability at various levels, has been one of the major factors contributing in the successful establishment of the World Wide Web (*WWW*). As pointed out in [5], the current Web can be classified as the second generation Web. The first generation started with the "handwritten" HTML web pages. The second generation, made a step forward introducing machine - generated web pages or even active HTML pages. The common characteristic between the first two web generations is that they are both human - oriented. The third generation of the Web is what is widely known as the "Semantic Web" and its main difference from the previous two generations is that it aims to machine - readable information ([1]).

1.1 Applications of the Semantic Web

The Semantic Web will enable intelligent services (such as information brokers, search agents and information filters) to process information automatically and it

* This work is supported by the bilateral research collaboration programme Platon between Greece and France, 2002–2004 financed by the French and Greek Secretariat of Research and Technology, Ministry of Development.

X. Zhou, Y. Zhang, and M.E. Orlowska (Eds.): APWeb 2003, LNCS 2642, pp. 382–387, 2003.

is expected that the existing e-services (of limited functionality) will be replaced. A more detailed description of the applications and the areas of impact of the Semantic Web, that include search engines, intelligent agents and push systems, is given in [9].

The remainder of the paper is organized as follows: The next section identifies the role of Ontologies in the Semantic Web. Section 3 describes the core of support of semantics on the web, with reference on the XML and RDF standards, while in Section 4 the most popular Ontologies languages are outlined. Conclusions are given in Section 5.

2 Ontologies in the Semantic Web

For web-based systems in specific, the poor communication and lack of a common understanding (due to the different languages, the overlapping and/or mismatched concepts, the differences in structures and methods) may entail to poor knowledge and software interoperability. This problem of poor communication can be solved by the introduction of ontologies. Ontologies are expected to have a major impact in the development of the Semantic Web, and mainly in the information exchange process.

The definition of the ontology given in [11] can be adopted: "An ontology is a logical theory accounting for the intended meaning of a formal vocabulary, i.e., its ontological commitment to a particular conceptualization of the world. The intended models of a logical language using such a vocabulary are constraint by its ontological commitment (and the underlying conceptualization) by approximating these intended models." This definition captures the two most important and necessary features of an ontology which are the vocabulary and the corresponding definitions. The vocabulary includes the terms for classes and relations, while the definitions of these terms may be informal text, or may be specified using a formal language like predicate logic. The advantage of formal definitions is that these definitions allow a machine to perform much deeper reasoning, while the disadvantage is that these definitions are much more difficult to construct. Typically, a layered approach is adopted, in order to represent the knowledge on the web and in order to define the ontologies. At the lowest level, there is a need for a generic mechanism that will allow for expression of machine-readable semantics of data. The Schema layer lies on top of this layer whereas a formal *KR* language will be used as the third logical layer.

Ontology models and languages need to be established so that the potential of ontologies to be fully exploited. In order for this to be done, these languages and models should be based on existing mechanisms and schemata like RDF and XML. In this paper We mainly focus on the third layer and we present an overview of the already proposed languages which provide Web-based *KR* and knowledge sharing, mainly via ontologies. These languages are mostly based either on XML or RDF and their goal is satisfy the following three requirements [5]:

– *Global Expressive and Modeling Power:* which is ecessary to the existing high degree of heterogeneity.
– *Syntactic Interoperability:* High syntactic interoperability is achieved when the data expressed in a specific format can be easily read by many applications that use even already existing parsers. If such parsers do not exist and there is extra effort for the data to be read, then the interoperability is low.
– *Semantic Interoperability and Reasoning Support:* Semantic interoperability refers to the mapping between the unknown terms and known terms in the data.

3 XML/RDF to Support Semantics on the Web

3.1 The XML Standard

XML [2] is intended as a markup language for arbitrary document structure,as opposed to HTML, which is a markup-language for a specific kind of hypertext documents. Regarding XML syntax it looks very familiar to this of HTML. An XML document consists of a properly nested set of open and closed tags, where each tag can have a number of attribute-value pairs. The important thing about XML is that the vocabulary of the tags as well as the way these tags can be combined is not predefined and it determined by the application and gives additional information about the text itself.

Although XML is characterized by high flexibility, there may be some problems that will occur when it comes to the machine processing issues of XML documents. In order for these problems to be solved, there is a need to define grammars for XML tags. This information is contained in a Document Type Definition (DTD) that specifies the valid elements, the contents of these elements and which attributes may modify an element. An XML document that is associated with a DTD and conforms to the rules that are defined by it is said to be valid. Notice though, that the DTD just provides the syntax of the XML document and gives no information about the semantics. The semantics of the DTDs and the XML Documents are just implicit and can only be perceived by humans that read the document's source code.

Recently, W3C has released an alternative to DTDs called XML Schema. XML Schemas provide greater flexibility in the definition of an XML application, even allowing the definition of complex data types, while they also use the same syntactic style as other XML Documents. However, XML Schema provides just an advanced XML grammar specification and data typing, and it still suffers from the semantic drawbacks of DTDs.

3.2 The RDF Schema

The real value and the basic conception of RDF [13] is its data model. More specifically, it defines a very simple data model of triples (subject, value, object). Every object O has an attribute A with value V. Such a triple corresponds to the relation that is commonly written as A(O,V). With this simple model objects and their properties may be presented. Coming back to the graph representation the RDF triples can also be represented by a labeled edge between two nodes: [O]-A→[V]. This notation allows the first and third elements of the triple to be combined in many ways.

Once we have the data model, there is a need to describe the characteristics of the objects being modeled. Just as XML Schema provides a vocabulary definition facility for XML, RDF Schema provides a similar facility for RDF. The main difference between RDF and XML Schemas is that in RDF case the schemas do not define a permissible syntax but instead classes, properties, and their interrelation: they operate directly at the data model level, rather than the syntax level. Scaled up to the Web RDF schemas are a key technology, as they will allow machines to make inferences about the data collected from the web.

3.3 XML – RDF and the Semantic Web Language Requirements

The Table 1 below shows whether XML and RDF described in the previous subsections can satisfy the basic three requirements imposed by the semantic web:

Table 1. XML/RDF and the Semantic Web Requirements

	Universal Expressive Power	**Syntactic** Inter-**operability**	**Semantic** Inter **-operability**
XML	Yes	Yes	Several Disadvantages
RDF	Yes	Yes	Several Advantages

4 Ontology Languages

There are two main approaches that have been followed towards the development of languages for ontologies representation. The first class of languages is based on First Order Logic (*FOL*), whereas the second class is based on ontology-representation. *FOL*-based languages have been adopted by organizations and cooperating partners whose main concern is not to make their knowledge publicly available over the Web. A relatively detailed description of the above ontology languages and their relationship with the previously analyzed web standards will follows.

4.1 SHOE – Simple HTML Ontology Extensions

SHOE is a *KR* language that allows ontologies to be designed and used directly to the World Wide Web (WWW). Based on the description provided in [8], SHOE's basic structure consists of ontologies, entities which define rules guiding what kind of assertions may be made and what kind of inferences may be drawn on ground assertions, and instances that make assertions to these rules. Regarding the *syntax* of SHOE, it has been primarily selected to fit seamlessly with that of HTML. Using similar syntax with HTML makes it easy for those that know HTML to learn SHOE as well. A slight variant of SHOE syntax also exists for compatibility with XML [14], which is a requirement for compatibility with today's applications since XML is an emerging standard for transmitting web documents. Two applications of SHOE developed in Univeristy of Maryland for research reasons are summarized in [8].

4.2 XOL – XML-Based Ontology Exchange Language

XOL is a language for ontology exchange. It is "designed to provide a format for exchanging ontology definitions among a set of interested parties. The ontology definitions that XOL is designed to encode include both schema information (meta-data), such as class definitions from object databases, as well as non-schema information (ground facts), such as object definitions from object databases. The syntax of XOL is based on XML that is a widely adopted web standard with high expressive power and syntactic interoperability, which are the basic requirements for the syntax of a language that can be used in the context of the Semantic Web. The semantics of XOL are based on OKBC-Lite, which is a simplified form of the knowledge model for the OKBC (Open Knowledge Base Connectivity.) [3], that is an API (application program interface) for accessing frame knowledge representation systems. The XOL was designed in response to a study of ontology languages performed by the Bio-Ontology Core Group. The study had found that no existing ontology-exchange language satisfied the requirements of the bio-informatics community.

4.3 OIL – Ontology Interchange Language

OIL [4,6,7] proposes a "joint standard for integrating ontologies with existing and arising web standards". OIL is a Web-based representation and inference layer for ontologies, which combines the widely used modeling primitives from frame-based languages with the formal semantics and reasoning services provided by description logics. OIL's *syntax* is layered on RDF. One of the design goals of OIL was to maximize integration with RDF applications. Thus most RDF Schemas are valid OIL ontologies and most of OIL ontologies can be partially understood by RDF processors. However, unlike RDF, OIL has very well defined *semantics*.

There are mainly multiple layers of OIL, where each subsequent layer adds functionality to the previous one. Core OIL is basically RDFS without reification. Standard OIL adds a number of description logic primitives to the Core OIL, and is the base of most of the work done with OIL today. Instance OIL adds the capability to model instances essentially using their description in RDF. Finally, Heavy OIL is an undefined layer that will include future extensions to the language. This layered approach allows applications to use pre-defined subsets of the language to manage complexity.

Nowadays, DARPA Agent Markup Language + Ontology Interchange Language (DAML+OIL) [10], [12] is a new language created with the best features of SHOE, DAML, OIL and several other markup approaches. It is considered to be the most advanced web ontology language, and it is expected to provide the basis for future web standards for ontologies.

5 Conclusions

Ontologies are expected to play an important role in the context of Semantic Web. This paper identifies on the major role of the Ontologies in the Semantic web, identifies the currently used standards and describes languages that have been proposed in order to define and describe ontologies on the web.

References

1. T. Berners – Lee: "Weaving the Web". Harper, San Francisco 1999.
2. T. Bray, J. Paoli and C. Sperberg-McQueen: " Extensible Markup Language (XML)". W3C (World Wide Web Consortium), February 1998. Available at http://www.w3.org/TR/1998/REC-xml-19980210.html.
3. V.K. Chaudhri, A. Farquhar, R. Fikes, P.D. Karp, J.P. Rice: "Open Knowldege Base Connectivity 2.0.2".http://WWW-KSL-SVC.stanford.edu:5915/doc/release/okbc/okbc-spec/okbc-2-0-2.pdf, 1998.
4. S. Decker, D. Fensel, F. Harmelen, I. Horrocks, S. Melnik, M. Klein and J. Broekstra: " Knowledge Representation on the Web". In Proceedings of the 2000 International Workshop on Description Logics (DL2000). Aachen, Germany, August 2000.
5. S. Decker et al: " The Semantic Web – on the respective Roles of XML and RDF.
6. D. Fensel et al.: "OIL in a nutshell". In Knowledge Acquisition, Modeling, and Management, Proceedings of the European Knowledge Acquisition Conference (EKAW-2000), R. Dieng et al. (eds.), Lecture Notes in Artificial Intelligence, LNAI, Springer-Verlag, October 2000.
7. D. Fensel et al.: "OIL: An ontology infrastructure for the Semantic Web.". In IEEE Intelligent Systems, 16(2):38–45, 2001.
8. J. Heflin, J. Hendler and S. Luke: "SHOE: A Knowledge Representation Language for Internet Applications". Technical Report CS-TR-4078 (UMIACS TR-99-71). 1999.
9. J. Heflin: " Towards the Semantic Web: Knowledge Representation in a Dynamic, Distributed Environment". PhD Thesis, University of Maryland, Computer Sciences Dept. 2001.
10. J. Hendler and D. McGuinness: " The DARPA agent markup language". IEEE Intelligent Systems, 15(6):72–73, November/December 2000.
11. N. Guarino: "Formal Ontology and Information Systems". In Proceedings of Formal Ontology and Information Systems, Trento, Italy, June 1998. IOS Press.
12. Joint US/EU Ad Hoc Agent Markup Language Committee. DAML+OIL, 2001.
13. "Resource Description Framework (RDF) Schema Specification 1.0", W3C Recommendation, March 2001. Available at http://www.w3.org/RDF/
14. eXtensible MArkup Language (XML) 1.0, Available at: http://www.w3.org/XML/