Ontologies for the Semantic Web

María Clara Vallés and Pablo R. Fillottrani

Departamento de Ciencias e Ingeniería de la Computación
Universidad Nacional del Sur
Bahía Blanca, Argentina
mcvalles@gmail.com
prf@cs.uns.edu.ar

Abstract

The Semantic Web attempts to reach a state in the future where everything on the Web will no longer be only machine-readable, but also machine-understandable. An ontology is an explicit specification of a conceptualization. A conceptualization is an abstract, simplified view of the world that we wish to represent for some purpose. Every knowledge base, knowledge-based system, or knowledge-level agent is committed to some conceptualization, explicitly or implicitly. By defining shared and common domain theories, ontologies help both people and machines to communicate concisely, supporting the exchange of semantics and not only syntax. Hence, the construction of domain-specific ontologies is crucial for the success and the proliferation of the Semantic Web.

1 Introduction

Since its beginning, the World Wide Web has played an important role in our everyday life, transforming the world towards a knowledge society. As a result, the way computers are used has diversified, gaining popularity and users. At present, the view of computers as an efficient way to access information on practically any subject, has gained special attention. Most of today's Web content is presented in a way that makes it suitable only for human consumption. In other words, information is expected to be consumed by individuals, not software programs.

Apart from hypertext links, which allow the possibility of linking a document to any other document, keyword-based search engines have turned into an essential tool for information management on today's web. However, the use of these tools in this fashion has some disadvantages, given the fact that it is the person who must browse documents, extract the information he or she is looking for, and discard the rest. The next step to take in order to solve this problem, would be the automatization of this process.

Unfortunately, there is a mayor inconvenience we must solve before we can achieve this task, which has to do with the fact that Web content must allow a computer program to sort out the meaning (semantics) of the information it browses. In other words, it is easy for a person to distinguish the information that is meaningful, because humans can "understand" the meaning of the Web content they read. However, in order for a software tool to interpret sentences and extract useful information for users, Web content should be represented in a form that is more machine-processable and which allows intelligent techniques to take advantage of these representations. We refer to this future Web plan, as the Semantic Web. Therefore, the Semantic Web attempts to reach a state in the future where everything on the Web will no longer be only machine-readable, but also machine-understandable.

It is important to understand that the Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, enabling computers and people to work in closer cooperation. The first steps towards incorporating the Semantic Web into the structure of the existing Web are already under way. In the near future, these developments will cause the dawning of significant new functionality as machines become much better able to process and "understand" the data that they merely display at present.

2 Knowledge Representation

For the semantic web to function, computers must have access to structured collections of information and sets of inference rules that they can use to conduct automated reasoning.

Semantic Web researches, accept that unanswerable questions represent a small price to pay to achieve versatility. The task of adding logic to the Web encompasses a series of complex decisions, given the fact that the logic must be strong enough to describe object properties, but not to powerful so as to avoid agents tricking themselves into considering paradoxes.

3 Technologies for developing the Semantic Web

Two important technologies for developing the Semantic Web are already in place: Extensible Markup Language (XML) and the Resource Description Framework (RDF).

3.1 Extensible Markup Language (XML)

One of the fundamental contributions towards the Semantic Web to date has been the development of XML.

XML provides an interoperable syntactical foundation upon which solutions to the larger issues of representing relationships and meaning can be built.

XML owns its name to the fact that it allows users to create their own tags that annotate Web pages or sections of text on a page. Scripts, or programs, can make use of these tags in sophisticated ways, but the script writer has to know what the page writer uses each tag for. Therefore, XML is purely sintactical, allowing users to add arbitrary structure to their documents but saying nothing about what the structures mean, i.e. its semantics. [2] There are many different XML schema languages, with different levels of expressivity. The most broadly supported schema language and the only one defined by the XML 1.0 specification itself is the document type definition (DTD). A DTD lists all the legal markup and specifies where and how it may be included in a document. DTDs are optional in XML.

3.2 Resource Description Framework (RDF)

The Resource Description Framework (RDF), developed under the auspices of the World Wide Web Consortium (W3C), is an infrastructure that enables the encoding, exchange, and reuse of structured metadata. This infrastructure enables metadata interoperability through the design of mechanisms that support common conventions of semantics, syntax, and structure. RDF does not stipulate semantics for each resource description community, but rather provides the ability for these communities to define metadata elements as needed. RDF uses XML as a common syntax for the exchange and processing of metadata. By exploiting the features of XML, RDF imposes structure that provides for the unambiguous expression of semantics and, as such, enables consistent encoding, exchange, and machine-processing of standardized metadata.

3.2.1 The RDF Data Model

RDF provides a model for describing resources. Resources have properties (attributes or characteristics). RDF defines a resource as any object that is uniquely identifiable by an Uniform Resource Identifier (URI). The properties associated with resources are identified by property-types, and property-types have corresponding values. Property-types express the relationships of values associated with resources. A collection of these properties that refers to the same resource is called a description. [4]

3.2.2 The RDF Schema

RDF Schemas (RDF-S) are used to declare vocabularies, the sets of semantics property-types defined by a particular community. RDF schemas define the valid properties in a given RDF description, as well as any characteristics or restrictions of the property-type values themselves. The XML namespace mechanism serves to identify RDF Schemas. [4]

4 Ontologies

An ontology is an explicit specification of a conceptualization. A conceptualization is an abstract, simplified view of the world that we wish to represent for some purpose. Every knowledge base, knowledge-based system, or knowledge-level agent is committed to some conceptualization, explicitly or implicitly.

In recent years the development of ontologies has become common on the World-Wide Web. Many disciplines now develop standardized ontologies that domain experts can use to share and annotate information in their fields, and which can be used for reasoning about the objects within a particular domain. An ontology defines a common vocabulary for researchers who need to share information in a domain. It includes machine-interpretable definitions of basic concepts in the domain and relations among them.[5]

Some of the reasons for developing ontologies are:

- To share common understanding of the structure of information among people or software agents
- To enable reuse of domain knowledge
- To make domain assumptions explicit
- To separate domain knowledge from the operational knowledge
- To analyze domain knowledge

Often an ontology of the domain is not a goal in itself. Developing an ontology is akin to defining a set of data and their structure for other programs to use. Problem-solving methods, domain-independent applications, and software agents use ontologies and knowledge bases built from ontologies as data. By defining shared and common domain theories, ontologies help both people and machines to communicate concisely, supporting the exchange of semantics and not only syntax. Hence, the construction of domain-specific ontologies is crucial for the success and the proliferation of the Semantic Web.[3]

4.1 Web Ontology Language - OWL

An ontology language is a formal language used to encode the ontology. There are a number of such languages for ontologies, one of them being OWL. OWL is intended to be used when the information contained in documents needs to be processed by applications, as opposed to situations where the content only needs to be presented to humans. OWL can be used to explicitly represent the meaning of terms in vocabularies and the relationships between those terms. OWL has more facilities for expressing meaning and semantics than XML, RDF, and RDF-S, and thus OWL goes beyond these languages in its ability to represent machine interpretable content on the Web. [6]

OWL has been designed to meet the need for a Web Ontology Language, and is part of the growing stack of W3C recommendations related to the Semantic Web:

- XML provides a surface syntax for structured documents, but imposes no semantic constraints on the meaning of these documents.
- XML Schema is a language for restricting the structure of XML documents and also extends XML with data types.
- RDF is a datamodel for objects ("resources") and relations between them, provides a simple semantics for this datamodel, and these datamodels can be represented in an XML syntax.
- RDF Schema is a vocabulary for describing properties and classes of RDF resources, with a semantics for generalization-hierarchies of such properties and classes.
- OWL adds more vocabulary for describing properties and classes. OWL provides three increasingly expressive sublanguages designed for use by specific communities of implementers and users.
- OWL Lite supports those users primarily needing a classification hierarchy and simple constraints.
- OWL DL supports those users who want the maximum expressiveness while retaining computational completeness (all conclusions are guaranteed to be computable) and decidability (all computations will finish in finite time). OWL DL includes all OWL language constructs, but they can be used only under certain restrictions.
- OWL Full is meant for users who want maximum expressiveness and the syntactic freedom of RDF with no computational guarantees. It is unlikely that any reasoning software will be able to support complete reasoning for every feature of OWL Full.

Each of these sublanguages is an extension of its simpler predecessor, both in what can be legally expressed and in what can be validly concluded. Ontology developers adopting OWL should consider which sublanguage best suits their needs. OWL Full can be viewed as an extension of RDF, while OWL Lite and OWL DL can be viewed as extensions of a restricted view of RDF. Every OWL (Lite, DL, Full) document is an RDF document, and every RDF document is an OWL Full document, but only some RDF documents will be a legal OWL Lite or OWL DL document.

5 Agents

The real power of the Semantic Web will be realized when people create many programs that collect Web content from diverse sources, process the information and exchange the results with other programs. Ontologies can be imagined as operating one level above RDF. Agents operate one level above ontologies, they examine different ontologies to find new relations among terms and data in them. The effectiveness of such software agents will increase exponentially as more machine-readable Web content and automated services (including other agents) become available. The Semantic Web promotes this synergy: even agents that were not expressly designed to work together can transfer data among themselves when the data comes with semantics.[1]

An important facet of agents functioning will be the exchange of "proofs" written in the Semantic Web's unifying language (the language that expresses logical inferences made using rules and information such as those specified by ontologies). Another vital feature will be digital signatures, which are encrypted blocks of data that computers and agents can use to verify that the attached information has been provided by a specific trusted source. Agents should be skeptical of assertions that they read on the Semantic Web until they have checked the sources of information.

In the Semantic Web, the consumer and producer agents can reach a shared understanding by exchanging ontologies, which provide the vocabulary needed for discussion. Agents can even "bootstrap" new reasoning capabilities when they discover new ontologies. Semantics also makes it easier to take advantage of a service that only partially matches a request.

A typical process will involve the creation of a "value chain" in which subassemblies of information are passed from one agent to another, each one "adding value" to construct the final product requested by the end user. To create complicated value chains automatically on demand, some agents will exploit artificial-intelligence technologies in addition to the Semantic Web. But the Semantic Web will provide the foundations and the framework to make such technologies more feasible.

In the next step, the Semantic Web will break out of the virtual realm and extend into our physical world. URIs can point to anything, including physical entities, which means we can use the RDF language to describe devices such as cell phones and TVs. Such devices can advertise their functionality, what they can do and how they are controlled, much like software agents. Such a semantic approach opens up a world of exciting possibilities.

6 Conclusion

The potential implications of widespread adoption of semantic web technologies, promises a knowledge revolution. If properly designed, it would not only become a tool for conducting individual tasks, but it would also assist the evolution of human knowledge as a whole. Once the web has been sufficiently populated with rich metadata, searching on the web will become easier as search engines have more information available, and thus searching can be more focused. As more groups develop ontologies, Semantic Web tools allow them to link their schemes and translate their terms, gradually expanding the number of people and communities whose Web Software can understand one another automatically. The web of today, the vast unstructured mass of information, may be transformed into something more manageable - and thus something far more useful, allowing agents and users to work and learn together.

References

[1] Tim Berners-Lee, James Hendler, and Ora Lassila. The semantic web. *Scientific American*, 2001.

[2] Elliotte Rusty Harold and W. Scott Means. *XML in a Nutshell*. O'Reilly, June 2002.

[3] Alexander Maedche and Steffen Staab. Ontology learning for the semantic web. *IEEE Intelligent Systems*, 16(2), March 2001.

[4] Eric Miller. An introduction to the resource description framework. *D-Lib Magazine*, May 1998.

[5] Natalya F. Noy and Deborah L. McGuinness. Ontology development 101: A guide to creating your first ontology, September 2005.

[6] W3C Recommendation. Owl web ontology language, February 2004.