The Web Ontology Language (OWL) and Its Applications

Jorge Cardoso
University of Coimbra, Portugal & KSRI/Karlsruhe Institute of Technology, Germany

Alexandre Miguel Pinto
University of Coimbra, Portugal

INTRODUCTION

Initially, the WWW was primarily composed of documents written in HTML (hyper text markup language), a language that is useful for visual presentation. HTML is a set of "markup" symbols contained in a Web page intended for display on a Web browser. At a later stage the distinct aspects of web page presentation style became separated from the contents, the former being specified in CSS (cascading style sheets). Still, most of the information on the Web is designed only for human consumption. Humans can read Web pages and understand them, but their inherent meaning is not shown in a way that allows their interpretation by computers (Cardoso & Sheth, 2006).

Since this setting does not allow computers to understand the meaning of Web pages (Cardoso, 2007), the W3C (World Wide Web Consortium) started to work on the concept of Semantic Web (SW) with the objective of developing approaches and solutions for data integration and interoperability purpose. The goal was to develop ways to allow computers to understand Web information.

One of the cornerstones of the SW is the Web Ontology Language (OWL). OWL can be used by applications that need to understand and to reason over the meaning of information instead of just parsing data for display purposes. The aim of this article is to present OWL, which can be used to develop Semantic Web applications that understand information and data on the Web. This language was proposed by the W3C and was designed for publishing, sharing data, and automating data understood by computers using ontologies. To fully comprehend OWL we need first to study its origin, its logic (semantic) foundations, and the basic blocks of the language. Therefore, we start by briefly introducing XML (eXtensible Markup Language), RDF (Resource Description Framework), and RDF Schema (RDFS). These concepts are important since one of the main syntaxes OWL takes is the XML format, and also because conceptually OWL is an extension of RDF and RDFS.

BACKGROUND

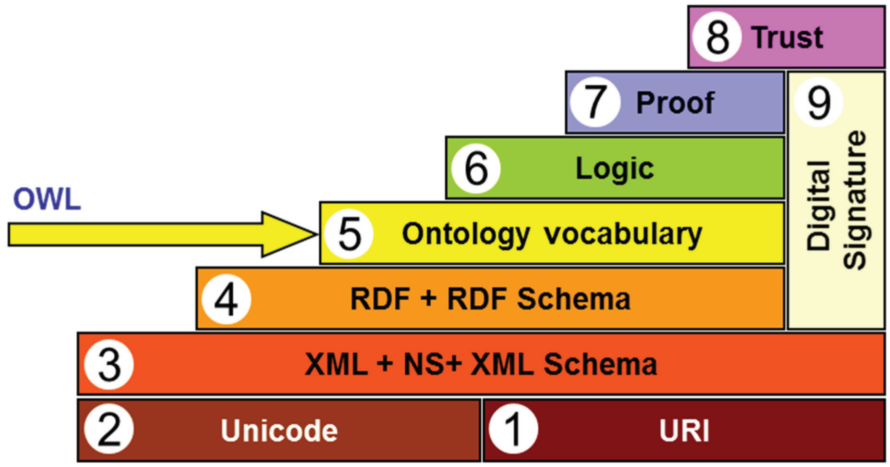
The Semantic Web Stack

The Semantic Web identifies a set of technologies and standards that form the basic building blocks of an infrastructure that supports the vision of the meaningful Web. Figure 1 illustrates the different parts of the SW architecture. It starts with the foundation of URI (universal resource identifier) and Unicode. URI is a formatted string that serves as a means of identifying abstract or physical resources. For example, http://eden.dei.uc.pt/~jcardoso/ identifies the location from where a Web page can be retrieved and urn:isbn:3-540-24328-3 identifies a book using its ISBN. Unicode provides a unique number for every character, independent of the underlying platform, program, or language.

Directly above URI and Unicode we find the syntactic interoperability layer in the form of XML, which in turn underlies RDF and RDFS. Web ontology languages are built on top of RDF and RDFS. The last three layers are logic, proof, and trust, of which the top one has not been significantly explored yet. Some of the layers rely on the digital signature component to ensure security.

DOI: 10.4018/978-1-4666-5888-2.ch755

Figure 1. Semantic Web layered architecture (Berners-Lee et al., 2001)



Next we briefly describe the most basic layers (XML, RDF, and RDFS). While the notions presented have been simplified, they give a reasonable conceptualization of some of the simplest components of the SW.

XML

The extensible markup language (XML) (XML, 2007) was originally pictured as a language for defining new document formats for the WWW. An important feature of this language is the separation of content from presentation, which makes it easier to select and/or reformat data. XML is a text-based format that provides mechanisms for describing document structures using markup tags (words surrounded by '<' and '>'). Both HTML and XML representations use tags such as <h1> or <name>, and information between those tags, referred to as the content of the tag. However, there are significant differences between HTML and XML. XML is case sensitive while HTML is not; and HTML has predefined elements and attributes whose behavior is well specified, while XML does not. Instead, users can create their own XML vocabularies that are specific to their application or business needs.

The following example shows an XML document identifying a 'Contact' resource. The document includes various metadata markup tags, such as <first_name>, <last_name>, and <e-mail>, which provides various details about a contact.

```
<Contact contact_id="1234">
     <first_name> Jorge </first_name>
     <last_name> Cardoso </last_name>
     <organization> University of Co-
imbra </organization>
     <email> jcardoso@dei.uc.pt </
email>
     <phone> +351 239 790 051 </
phone>
</Contact>
```

While XML has gained awareness, XML is simply a way of standardizing data formats. But from the point of view of semantic interoperability, XML has restrictions. E.g., there is no way to recognize the semantics of a particular domain because XML aims at a document structure and enforces no common interpretation of the data. Although XML is simply a data-format standard, it is part of a set of technologies that constitute the foundations of the SW.

RDF

The Resource Description Framework (RDF) (RDF, 2002) was developed by the W3C to provide a common way to describe information so it could be read and understood by computer applications. RDF was designed using XML as the underlying syntax. It provides a model for describing resources on the Web. A resource is an element (document, Web page, printer, user, etc.) on the Web that is uniquely identifiable