Semantic Web, RDF, and Portals

Ah Lian Kor
Leeds Metropolitan University, UK

Graham Orange
Leeds Metropolitan University, UK

INTRODUCTION

In existing literature, Semantic Web portals (SWPs) are sometimes known as semantic portals or semantically enhanced portals. It is the next generation Web portal which publishes contents and information readable both by machines and humans. A SWP has all the generic functionalities of a Web portal but is developed using semantic Web technologies. However, it has several enhanced capabilities such as semantics-based search, browse, navigation, automation processes, extraction, and integration of information (Lausen, Stollberg, Hernandez, Ding, Han & Fensel, 2004; Perry & Stiles, 2004). To date the only available resources on SWPs are isolated published Web resources and research or working papers. There is a need to pool these resources together in a coherent way so as to provide the readers a comprehensive idea of what SWPs are, and how they could be built, and these will be supported by some appropriate examples. Additionally, this article will provide useful Web links for more extensive as well as intensive reading on the subject.

The SWP is an amalgam of the three following components: semantic Web, Web services, and Web portal. In this article, we will only discuss the architecture of the semantic Web, the RDF (resource description framework) language, and syntax used for representing information in the Web. The discussion on ontology Web languages, semantic query, features of a Web portal, and Web services can be found in the article "Ontology, Web Services and Semantic Web Portals" of this encyclopedia.

SEMANTIC WEB

SW Architecture

The Semantic Web provides a common framework for data sharing and reuse across applications, businesses, and communities. The semantic Web technologies in the semantic Web architecture (Berners-Lee, 2005a) are depicted in Figure 1. This architecture is an extension of the widely quoted semantic

Figure 1. The Semantic Web architecture (Berners-Lee, 2005a)

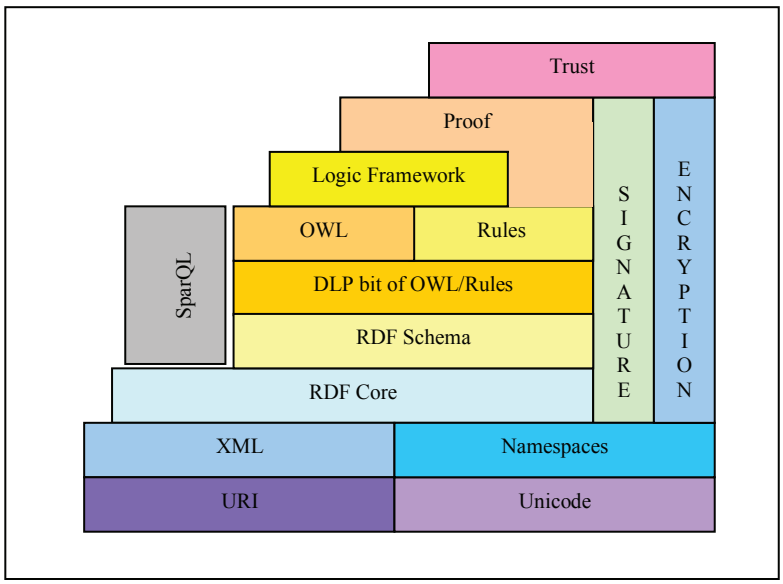
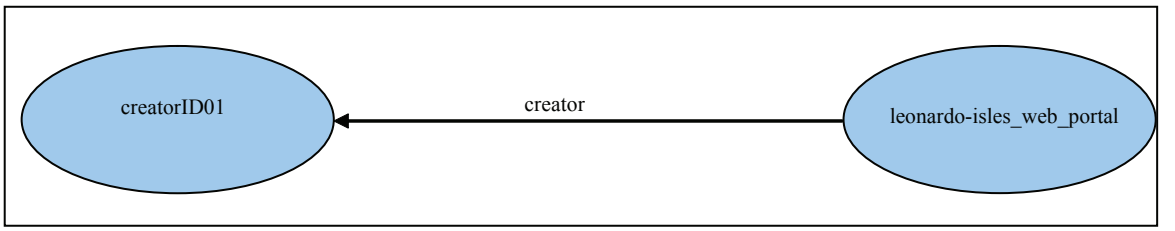


Figure 2. A RDF graph representation of a statement



Web "layered cake" model (Berners-Lee, Hendler & Lassila, 2001) which begins with simple mechanisms for naming, identifying and locating resources (URIs) at the lowest layer, and rising through layers of increasing sophistication to the highest, the *Trust* (security) layer (JISC, 2005).

A *uniform resource identifier* (URI) is an identifier which consists of short strings of characters that represent names or addresses of Web resources such as documents, images, files, services, or electronic mailboxes. According to the URI Planning Interest Group (2001), some examples of URIs are: *uniform resource names* (URNs), *uniform resource citations* (URC), or *uniform resource locators* (URLs). URIs can be used to refer to objects that are accessible through the WWW (e.g., Web resources – URLs which begin with *http:*), objects that are not accessible through the WWW (e.g., books in the library with URNs such as urn:isbn: 072142144X), or abstract concepts (e.g., the creator of a Web resource).

Extended Markup Language (XML) is a Web technology which adds style to Web documents and services. It is a tool for describing data while HTML controls the displaying and formatting of the data. The structure, contents, and semantics represented in an XML document are defined by the XML Schema Definition Language which is also used to express shared vocabularies. An XML namespace (XMLNS) is a collection of names used in XML documents, which has a unique URI.

In 2004, the World Wide Web Consortium declared Semantic Web languages resource description framework (RDF), Web Ontology Language (OWL), and SPARQL official W3C recommendations. Information is represented and exchanged between applications through the Web using RDF where RDF specifications are built on XML and URIs on technologies. XML provides the syntax and plays a pivotal role in data manipulation and transmission on the Web or across incompatible systems. On the other hand, OWL exploits the use of ontologies for publishing, sharing, and reusing information. It also supports semantic-based query, use of software agents, and knowledge management. OWL also uses URIs for naming purposes and it is built on RDF and RDF schema (RDF-S). SPARQL is a W3C standard for RDF query language which is similar to SQL, a query language for a relational database system.

The *digital signature* component is for detecting alterations in Web documents (Koivunen & Miller, 2001). The three top layers—*Logic*, *Proof*, and *Trust*—are still in their embryonic stage. The *Logic* layer enables the writing of inference rules while the *Proof* layer executes the rules to test the truth of statements, and, together with the *Trust* layer mechanism for applications (e.g., transactions involving privacy in e-commerce), evaluate the trustworthiness of a given proof (Koivunen & Miller, 2001).

RDF

According to W3C (Manola & Miller, 2004), RDF is a Web language that can represent information about a Web resource (e.g., author, title, creation date, etc.). However, if the Web resource concept is generalized, then it means that RDF can represent information about anything that is identified by URIs even though they cannot be retrieved directly. Additionally Web resources are described in terms of properties and proties values. The XML language used to write RDF documents is known as RDF/XML.

A RDF triple contains three components, namely, a subject, predicate, and object. A RDF data model can be represented by the following triple ‹subject, predicate, object›. An example of a statement is: the creator of a "leonardo-isles_Web_portal" (a resource) is "creatorID01" (ID of one of the project partners). The "leonardo-isles_Web_portal" is a subject (resource), "creator," a predicate (property of resource), and "creatorID01," an object (value of property). This statement can be represented by a simple RDF graph (Figure 2) which has two nodes and an arc identified by a URI. However, only the node for the object may be a literal (string or integer) or blank.

Tim Berners-Lee (2005b) uses Notation 3 or N3 to represent a RDF statement or, in other words, express RDF