Chapter 3.15
RDF and OWL for Knowledge Management

Gian Piero Zarri
University Paris-EstFrance

INTRODUCTION

As Web-based content becomes an increasingly important knowledge management resource, Web-based technologies are developing to help harness that resource in a more effective way.

The current state of these Web-based technology – the ‘first generation’ or ‘syntactic’ Web – gives rise to well known, serious problems when trying to accomplish in a non-trivial way essential management tasks like indexing, searching, extracting, maintaining and generating information. These tasks would, in fact, require some sort of ‘deep understanding’ of the information dealt with: in a ‘syntactic’ Web context, on the contrary, computers are only used as tools for posting and rendering information by brute force. Faced with this situation, Tim Berners-Lee first proposed a sort of ‘Semantic Web’ (SW) where the access to information is based mainly on the processing of the semantic properties of this information: “… the Semantic Web is an extension of the current Web in which information is given well-defined meaning (emphasis added), better enabling computers and people to work in co-operation” (Berners-Lee et al., 2001: 35). The Semantic Web’s challenge consists then in being able to manage information on the Web by ‘understanding’ its proper semantic content (its meaning), and not simply by matching some keywords.
GENERAL BACKGROUND

An updated representation of the Semantic Web’s architecture originally proposed by Berners-Lee—the well-known SW ‘cake’ see, e.g., http://www.w3.org/2007/03/layerCake.png—is reproduced in Figure 1.

‘URIs’ and ‘IRIs’ make up the basis of this hierarchy. URI (Uniform Resource Identifier) represents a generalization of the well-known URL (Uniform Resource Locator) that is used to identify a ‘Web resource’ (e.g., a particular page) by denoting its primary access mechanism (essentially, its ‘location’ on the network). URI has been created to allow recording information about all those ‘notions’ that, unlike Web pages, do not have network locations or URLs, but that need to be referred to in a statement formulated in some SW language. These notions include network-accessible things, such as an electronic document or an image, things that are not network-accessible, such as human beings, corporations, and bound books in a library, or abstract concepts like the concept of a ‘creator’. To be able to identify resources making use of languages different from English, IRIs (International resource Identifiers) have then been introduced as a complement to URIs. An IRI is a sequence of characters from the Universal character set, see Unicode/ISO 10646. The Unicode Standard provides a unique numerical code for every character that can be found in documents produced according to any possible language, no matter what is the hardware and software used to deal with such documents.

XML (eXtensible Markup Language), see (Bray et al., 2004), has been created to overcome some difficulties proper to HTML (Hypertext Markup Language). An HTML file is a text file characterized by the presence of a small set of ‘tags’—like <Head>, <Body>, <Input>, <Applet>, <Font> etc.—that instruct the Web browsers how to display a given Web page. HTML is, then, a ‘presentation-oriented’ markup tool. In spite of its evident utility, HTML suffers from a number of limitations, from its lack of efficiency in handling the complex client/server communication of today applications to (mainly) the impossibility of defining new tags to customize exactly the user’s needs.

Figure 1. Updated view of the Semantic Web architecture according to Tim Berners-Lee