

Chapter XLVI

Using Semantic Web Tools for Ontologies Construction

Gian Piero

University of Paris IV / Sorbonne, France

INTRODUCTION

The current state of Web technology – the “first generation” or “syntactic” Web – gives rise to well-known, serious problems when trying to accomplish, in a non-trivial way, essential 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” 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 of being able to access and retrieve information on the Web by “*understanding*” its proper semantic content (its meaning), and not simply by matching some keywords.

From a technical point of view, the Semantic Web vision is deeply rooted into an “ontological” approach, with some *proper characteristics* that differentiate it from the “classical” approach to the construction of ontologies based on a methodology of the frame” type (Chaudhri *et al.*, 1998) and on the use of tools in the “standard” Protégé style (Noy, Fergerson and Musen, 2000). We will describe these characteristics in the following Sections.

BACKGROUND INFORMATION

Berners Lee's Architectural Proposal for the Semantic Web

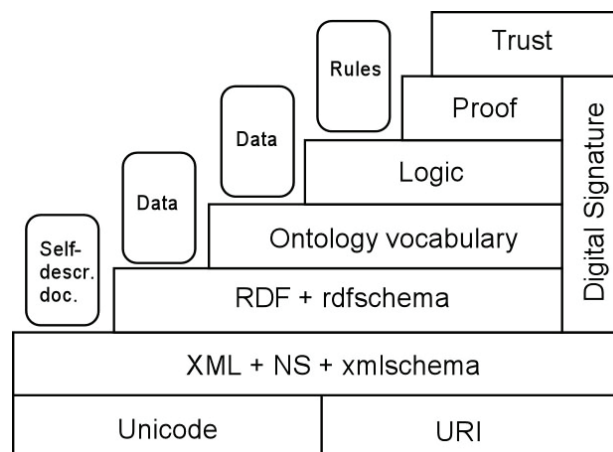
To support his vision, Berners-Lee has proposed in several papers an 'architecture' for the Semantic Web like that reproduced in Figure 1. Making abstraction now from all the discussions and criticisms that this proposal has brought up, see also the 'Conclusion', what is relevant for the topic of this article is the central position that 'ontologies' occupy in the architecture. We can note immediately that a first, important difference with respect to the 'classical' approach evoked above is that ontologies are no more considered 'in isolation': they are now supported by lower-level tools like XML and RDF and must also implement and additional logic level.

In the embedded architecture of Figure 1, 'Unicode' and 'URI' make up the basis of the hierarchy. 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. It is supported in many operating systems and all the modern browsers, and it enables a single

software product or a single website to be targeted across multiple platforms, languages and countries without re-engineering. URI (Uniform Resource Identifier) represents a generalization of the well-known URL (Uniform Resource Locator), which 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 an RDF statement. 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'.

XML (eXtensible Markup Language), see (Bray *et al.*, 2004), has been created to overcome some difficulties proper to HTML (Hypertext Markup Language), developed in 1989 by Tim Berners-Lee as a means for sharing information from any locations. An HTML file is a text file characterized by the presence of a small set of 'tags' – like <Head>, <Body>, <Input>, <Applet>, etc. – that instruct the Web browsers how to display a given Web page. HTML is, then, a

Figure 1. Semantic Web architecture according to Tim Berners-Lee



14 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/using-semantic-web-tools-ontologies/20727

Related Content

Data Warehouses

Antonio Badia (2005). *Encyclopedia of Database Technologies and Applications* (pp. 121-126).
www.irma-international.org/chapter/data-warehouses/11133

Expert Detection and Recommendation Model With User-Generated Tags in Collaborative Tagging Systems

Mengmeng Shen, Jun Wang, Ou Liu and Haiying Wang (2020). *Journal of Database Management* (pp. 24-45).
www.irma-international.org/article/expert-detection-and-recommendation-model-with-user-generated-tags-in-collaborative-tagging-systems/266403

SWIFT: A Distributed Real Time Commit Protocol

Udai Shanker, Manoj Misra and Anil K. Sarje (2009). *Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends* (pp. 737-743).
www.irma-international.org/chapter/swift-distributed-real-time-commit/20759

Modeling Temporal Dynamics for Business Systems

Gove N. Allen and Salvatore T. March (2003). *Journal of Database Management* (pp. 21-36).
www.irma-international.org/article/modeling-temporal-dynamics-business-systems/3297

INDUSTRY AND PRACTICE: How Clean is your Data?

Huw Price (1994). *Journal of Database Management* (pp. 36-42).
www.irma-international.org/article/industry-practice-clean-your-data/51131