The Semantic Web: Semantics for Data on the Web

Stefan Decker[#] and Vipul Kashyap[*]

[#]USC Information Sciences Institute, 4676 Admiralty Way, Marina Del Ray, CA 90292, USA
[*]LHNCBC, National Library of Medicine, 8600 Rockville Pike, Bethesda, MD 20894, USA
stefan@isi.edu, kashyap@lhc.nlm.nih.gov

Abstract

In our tutorial on Semantic Web (SW) technology, we explain the why, the various technology thrusts and the relationship to database technology. The motivation behind presenting this tutorial is discussed and the framework of the tutorial along with the various component technologies and research areas related to the Semantic Web is presented.

1. Motivation

The advent of the internet and the World Wide Web has established a ubiquitous communications and information infrastructure. However, information and services are still hard to find and integrate, resulting in higher costs for deploying relevant information and generating value from it. Several developments in the recent years and are fueling the SW effort: e-Science, e-business, and e-commerce, e-government. Some phenomena motivating the SW are (a) increased cost pressure and competition leading to interconnection of workflows and business processes, for sharing data and services; (b) portals aiming to unify access to information and services; and (c) need for collaboration and data sharing across multiple scientific communities (e.g., the GO (genomic data) and GEON (earth sciences) efforts). Various stakeholders need to work together to agree on interchange formats for data and services that adequately capture the semantics for current and future applications.

2. Framework

The broad dimensions of the SW effort and its multi-disciplinary underpinnings can be viewed from the *information* and *computational* aspects. We focus on the information aspect, organized across two dimensions: *semantic web content*, and *ontologies and schemata*.

2.1 Semantic Web Content

This refers to the myriad forms of data that will be presented on the SW along with metadata descriptions, exemplified by the equation ***SW content = data + metadata***. Data might be structured (e.g., relational), semi-structured (e.g., XML, RDF) and unstructured (e.g., raw text, multimedia). Metadata are primarily used to annotate data on the SW and may be specific to an information or application domain. The annotation process is fundamental to creating a SW infrastructure and may involve tools and semi-automatic techniques.

2.2 Ontologies and Schemata

This refers to the underlying vocabulary used to specify the semantics of the metadata descriptions. Collections of structured domain specific concepts may be used to create domain specific views on the underlying content. Schemata are a special case of metadata and the semantics of information they capture depends on the expressiveness of the meta-model. In the cases where the metadata is explicit (e.g., database/XML schemas) they may be mapped to other related metadata or ontological concepts.

3 Conclusions.

In the context of the framework discussed above, we will explore the multi-disciplinary aspects of SW research as enumerated in the table below. SW research provides an interesting framework for identifying synergies between various computer science disciplines and opens up new avenues for innovative cross-disciplinary research.

	Data	***Metadata***	***Ontologies***
DB/CM Systems	×	×	
KR Systems		×	×
Machine Learning		×	×
Stat. Clustering		×	×
Inform. Retrieval	×	×	
Nat. Lang. Proc.	×	×	×

Proceedings of the 29th VLDB Conference, Berlin, Germany, 2003