Position paper

Jeff Mischkinsky, Dmitry Lenkov, Dana Florescu, Ashok Malhotra
firstname.lastname@oracle.com
Oracle

Introduction

With the vast expansion of the World Wide Web during recent years the semantic integration of services and information sources on the Web has become a hot topic. The growing experience in adopting RDF and Ontologies for the classification and discovery of Web services shows their significant advantages. However, at the same time, the differences of XML (XML schema) and RDF data models create difficulties in the adoption of RDF for Web services that are already based on XML and XML schemas and their data models. Thus, the integration of these data models becomes a necessity for further advances of Web services technologies.

Web services and RDF

Web services are just a part of the evolution of component development technology. It has long been recognized that metadata is one of the most important aspects of component-based systems. Web services bring component technologies to the Web. Metadata is as important for Web services as it is for other component systems. Among different efforts to define metadata systems, W3C came up with a standard for managing metadata: RDF (Resource Description Framework). While the adoption of RDF has been steadily increasing, its adoption for Web services is rather slow.

Since the beginning Web services frameworks were developed using XML and XML schemas for representation of structured and semi-structured data. While RDF has been available for quite a few years, it did not gain any significant attention from Web services implementers and Web services standard developers until very recently. One of the explanations is that initially the main attention was given the interoperability on the syntactic level leaving semantics to be handled on the application level. Now there is a great deal of activity, much of which brings together core XML and RDF, and the Web Services that can built on these. The W3C has also devoted some related activities to Web Services. However, bringing together XML (and XML schemas) and RDF is not simple because data models used by them are different

Since this problem is common to Web services and information management (information sharing, distribution, and discovery), content management, etc., below we will be talking about semantic integration of Web services, information, etc. in general.

Semantic Integration

XML in general, has become an enormous success and is widely accepted as a standard means for the representation and communication of (semi-) structured data. The next generation of the Web, often characterized as the “Semantic Web”, will require certain level of automated semantic interpretation that can be done by machines. Thus, it

requires interoperability on the semantic level. The semantic interoperability will enable semantic integration of information and services on the Web

In this respect XML (with XML schemas) shows some limitations. The major limitation is that it just describes grammars for labeled (tagged) trees. No semantic information is captured by these grammars. It means that tags, attributes, and other language primitives are defined in an arbitrary way, assigning them semantics based on a conceptual domain model in author's mind. XML (with XML schemas) does not impose rules for such definitions and there are many ways how to denote semantically equivalent things. Thus it becomes hard to reconstruct the semantic meaning from an XML document.

The semantic meaning of XML documents, as well as documents using other syntaxes, can be expressed by the definitions of semantic metadata associated with a particular set of documents. RDF and RDFS provide standard means to describe both descriptive and semantic metadata. Semantic metadata characterizes the service or the content of the document. This kind of metadata can be used in machine processing to extract the semantics of data.

Two crucial RDF Schema constructions are subClassOf and subPropertyOf. RDF objects may be instances of one or more classes. The subClassOf construct from RDF Schema allows the specification of the hierarchical organization of such classes. The subPropertyOf construct does the same for properties. Furthermore, constraints on properties can be specified using domain and range constructs. Thus, RDF provides means to specify relationships between RDF named classes, in difference with XML providing named tags based syntax only.

However, RDF (with RDFS) is limited too. Besides the intended Object-Attribute-Value-semantics, no further data modeling commitments are made. In particular, no particular reserved terms for any further data modeling are defined. Just as XML Schema provides a vocabulary definition facility for XML, RDF Schema provides a similar facility for RDF. However, other standard processing techchniques like complex querying, updates, full text search and so on, are missing today from the RDF set of standards.

Ontology languages like OWL, built on top of RDF and RDFS, using its primitives, significantly extend the ability to define semantics of documents and information data sets. With these languages one can describe domain ontologies, by identifying hierarchies of concepts and relations together with axioms that can be used to derive new facts from existing ones. The ontology can thus be seen as a semantic interface for accessing services and information sources. This introduces a new, semantic-based generation of semantic integration architectures.

While the growing use of RDF and Ontologies, in particular RDF based, improves environments and tools for the semantic integration over the Web, it also introduces a significant problem. So far, approaches to application frameworks have been based on either RDF schemas or XML schemas. Despite multiple attempts to unify data models of two approaches, a purist and competitive attitude has prevailed among framework implementers and application developers.

Conclusion

The growing acceptance of W3C standard-based technologies for semantic integration over the Web using the potential offered by Web Services creates new opportunities for greater automation of distributed document processing and interactions. These opportunities should not be diminished by the fact that different sets of technologies are often competing for the same application domain, and sometimes inconsistently designed. It is absolutely necessary that W3C takes steps to guarantee the interroperability of the three sets of technologies: the XML-based ones (XML schema, XQuery, XSLT, etc.) , RDF-based ones (RDFS, SparQL, etc) and the Web Services (SOAP, WSDL, etc).

References

[1] Tim Berners-Lee and Mark Fischetti. *Weaving the we*b, chapter Machines and the Web, pages 177 – 198. Harper, San Francisco, 1999.