A Revised Architecture for Semantic Web Reasoning

Peter F. Patel-Schneider

Bell Labs Research, Lucent Technologies
pfps@research.bell-labs.com

Abstract. The current architecture for the Semantic Web, with its emphasis on RDF syntactic and semantic compatability, has severe problems when expressive Semantic Web languages are incorporated. An architecture less tied to RDF is proposed. In this architecture different Semantic Web languages can have different syntaxes but must use the same models. This revised architecture provides significant advantages over the currrent Semantic Web architecture while still remaining true to the vision of the Semantic Web.

1 Introduction

Because the aim of the Semantic Web is to make information on the Web more processable by computers, reasoning must be a vital part of the Semantic Web. Initial accounts of the Semantic Web, in particular, the initial versions of The Resource Description Framework (RDF) [Ora Lassila and Ralph R. Swick, 1999] and its schema extension, the RDF Schema Specification (RDFS) [Dan Brinkley and R. V. Guha, 2000], were without a formal semantic account, and thus did not support reasoning.

The current RDF recommendations include both a formal syntax [Dave Beckett, 2004] and semantics [Hayes, 2004] for RDF and RDFS [Dan Brinkley and R. V. Guha, 2004]. The W3C Web Ontology Language (OWL) [Dean *et al.*, 2004] has been given a semantics [Patel-Schneider *et al.*, 2004] that fits well on top of RDF. This thus appears to put the Semantic Web on a firm semantic foundation.

However, making OWL an extension of RDF was not without problems. This shows up in two versions of OWL: OWL DL, which has a different, mostly-compatible semantics from RDF and only extends part of RDF; and OWL Full, which has full compatability with RDF, but does not enjoy the computational benefits of OWL DL.

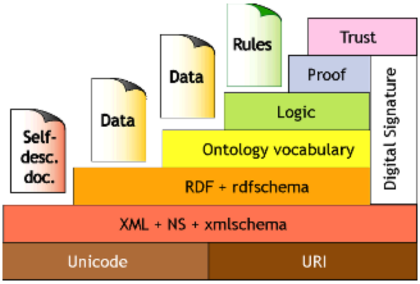
Extensions beyond OWL to first-order logic are even more problematic. In fact, an extension of RDF to incorporate all of first-order logic gives rise to paradoxes [Patel-Schneider, 2005], because a truth predicate is needed to encode first-order logic in the RDF syntax.

The problem is that RDF is not suitable as a basis for both the syntax *and* semantics of the Semantic Web. Some way of escaping from the RDF-provided meaning of the RDF syntax is needed for expressive Semantic Web languages.

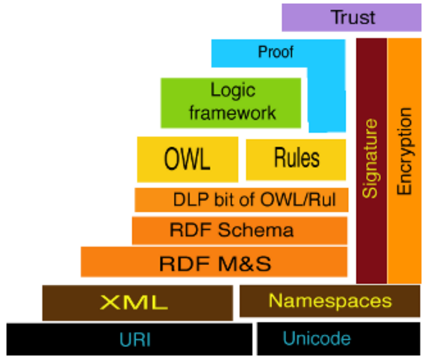
2 The Current Semantic Web Architecture

The current architecture of the Semantic Web is based on the well-known Semantic Web stack described by Tim Berners-Lee in 2000. (See the left-hand side of Figure 1.) This

F. Fages and S. Soliman (Eds.): PPSWR 2005, LNCS 3703, pp. 32–36, 2005.



http://www.w3.org/2000/Talks/1206-xml2k-tbl/



http://www.w3.org/2005/Talks/0511-keynote-tbl/

Fig. 1. Initial and New Semantic Web architecture

is only a high-level picture of the Semantic Web, and thus leaves out a lot of details. The general impact of this picture, particularly as it has been interpreted during the development of RDF, RDFS, and OWL, is that RDF forms the basis of the Semantic Web, both for syntax and semantics.

All Semantic Web documents thus should have the syntax of RDF, and this syntax should be read as encoding RDF triples [Graham Klyne and Jeremy Carroll, 2004], which form the abstract syntax of RDF. Further, the meaning of these triples should include their RDF model-theoretic meaning [Hayes, 2004], that is, all triples can be thought of a atomic facts.

There are other aspects of the Semantic Web Architecture. These include the use of URI references as identifiers, XML Schema datatypes as datatypes, and the use of model-theoretic entailment as the primary semantic relationship. As well, the Semantic Web was envisioned as a stack of languages, each building directly and completely on the lower languages. Thus the ontology layer built on the RDF layer, and the logic layer built on the ontology layer.

Recent accounts of the Semantic Web architecture (the right-hand side of Figure 1) have split the single stack into two side-by-side extensions of RDF for ontologies and rules. However, this does not change the fundamental role of RDF in the Semantic Web architecture.

3 Problems with the Current Architecture

Unfortunately, RDF is just not adequate for this fundamental place in the Semantic Web architecture. The problems are two-fold, both syntactic and semantic.

On the syntactic side, all RDF has to offer for syntax is triples. It is true that triples are indeed adequate for encoding all sorts of ground information, and thus appear form an adequate syntactic basis. However, encoding complex syntactic information in triples is painful, as shown by the difficult official OWL syntax. For example, the triple encoding of the simple Description Logic construct $(\forall r.C) \sqcap (\geqslant 3\, r)$ is

_:c rdf:type owl:Class . _:c owl:intersectionOf _:l1 . _:l1 rdf:type rdf:List . _:l1 rdf:first _:c1 . _:l1 rdf:rest _:l2 . _:l2 rdf:type rdf:List . _:l2 rdf:first _:c2 . _:l2 rdf:rest rdf:nil . _:c1 rdf:type owl:Restriction . _:c1 owl:onProperty r . _:c1 owl:allValuesFrom C . _:c2 rdf:type owl:Restriction . _:c2 owl:onProperty r . _:c2 owl:minCardinality 3 .

Not only is this ugly and verbose, there is the problem of what to do with various sorts of malformed bits of syntax, such as constructs with missing or extra triples.

The semantic side is even more problematic. The problem arises because each triple in RDF is a fact. This means that all the triples needed to encode syntax are facts, and these facts must be true before they can be inferred, independent of any other meaning that these syntactic facts encode. This requires complex machinery to require that the these facts are true when necessary, and this complex machinery can cause semantic paradoxes.

The semantic paradoxes were avoided in OWL by not requiring that self-referential syntactic structures must be inferrrable. This means that certain kinds of inferences that one might want are not inferrable in OWL, but does mean that OWL has a non-paradoxical semantics. However, this kind of solution is not available for first-order logic with equality, as equality can be used as a substitute for self-reference [Patel-Schneider, 2005].

4 A Revised Architecture for the Semantic Web

The problem is the use of triples as facts that encode syntax. So retaining the use of triples to encode syntax but making them not be facts will indeed eliminate the problem. However, this doesn't make triples any nicer for encoding syntax. As well, the syntax-encoding triples don't mean what they mean in RDF, so treating them as RDF doesn't make sense. This means that RDF tools will not perform correctly on these triples.

A different way to go is to allow different syntaxes for different Semantic Web languages. To fit into the Semantic Web vision, these syntaxes should use IRI references as identifiers and XML Schema datatypes. To fit into the World Wide Web these syntaxes should be XML dialects.

What then unifies the Semantic Web is a common semantic framework. Separate syntaxes for OWL, rule languages like SWRL [Horrocks *et al.*, 2005], and first-order logic including variants like Simple Common Logic (http://www.w3.org/2004/12/rules-ws/paper/103/) can be specified and their semantics given as extra conditions added to the RDF model theory. The new syntaxes can be made full partners in the Semantic Web by giving them MIME types. This does require the use of multiple parsers, but writing parsers is quite easy. To keep the single stack view of the Semantic Web only requires that systems for the higher levels also parse the syntax for the lower levels.

This revised architecture does not really require extra work in writing non-parsing tools, as one might expect. User-interface, reasoning, and other tools for the Semantic Web are of necessity already tailored for each level of the Semantic Web stack. For example, a tool for building RDF knowledge bases is not useful for ontology building in OWL, even though OWL can be written as RDF triples. Separate ontology-building tools like Protege or OilEd are needed for this purpose. Such tools are likely to be easier to write in the revised architecture, as they will not have to worry about the kinds of malformed syntax that are possible when encoding complex syntactical constructs in RDF triples.

Certain reasoning tasks can also be easier in this revised syntax. It has recently (private communication, but should be public shortly) been shown that a source of undecidability in OWL Full is the RDF ability to manipulate the RDF properties that encode OWL syntax.

This multi-syntax architecture has other potential benefits. It is possible to use content negotiation to allow less-capable systems to access approximations of complex information. For example, if an RDFS-only system asks for an OWL ontology it can be given the classification taxonomy of that ontology, as computed by an OWL reasoner, in RDFS form. This gives the RDFS system much better information that it would have if it was simply given the encoding of the syntax of the OWL ontology.

5 Future Directions

It would also be possible to generalize the revised Semantic Web architecture in several ways. One could lift the requirement that systems handle lower levels of the stack, turning the Semantic Web stack into a a collection of languages with a common semantic framework. One could also loosen the requirement of a common semantic framework into simply some sort of semantic compatability. Both these generalizations involve considerable work—in the first to determine whether or not the result would balkanize the Semantic Web and thus reduce its viability, in the second to determine what sorts of semantic compatability are desirable or required.

References

[Dan Brinkley and R. V. Guha, 2000] Resource description framework (RDF) schema specification 1.0. W3C Candidate Recommendation, http://www.w3.org/TR/2000/CR-rdf-schema-20000327, March 2000.

[Dan Brinkley and R. V. Guha, 2004] RDF vocabulary description language 1.0: RDF schema. W3C Recommendation, http://www.w3.org/TR/rdf-schema, 2004.

[Dave Beckett, 2004] RDF/XML syntax specification (revised). W3C Recommendation, http://www.w3.org/TR/rdf-syntax-grammar/, 2004.

[Dean *et al.*, 2004] Mike Dean, Guus Schreiber, Sean Bechhofer, Frank van Harmelen, Jim Hendler, Ian Horrocks, Deborah L. McGuinness, Peter F. Patel-Schneider, and Lynn Andrea Stein. OWL web ontology language: Reference. W3C Recommendation, http://www.w3.org/TR/owl-ref/, 2004.

[Graham Klyne and Jeremy Carroll, 2004] Resource description framework (RDF): Concepts and abstract syntax. W3C Recommendation, http://www.w3.org/TR/rdf-schema, February 2004.

[Hayes, 2004] Patrick Hayes. RDF semantics. W3C Recommendation, http://www.w3.org/TR/rdf-mt/, 2004.

[Horrocks *et al.*, 2005] Ian Horrocks, Peter F. Patel-Schneider, Sean Bechhofer, and Dmitry Tsarkov. OWL Rules: A proposal and prototype implementation. *Journal of Web Semantics*, 2005.

[Ora Lassila and Ralph R. Swick, 1999] Resource description framework (RDF): Model and syntax specification. W3C Recommendation, 22 February 1999, http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/, February 1999.

[Patel-Schneider *et al.*, 2004] Peter F. Patel-Schneider, Patrick Hayes, and Ian Horrocks. OWL web ontology language: Semantics and abstract syntax. W3C Recommendation, http://www.w3.org/TR/owl-semantics/, 2004.

[Patel-Schneider, 2005] Peter F. Patel-Schneider. Building the semantic web tower from rdf straw. In *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence*. International Joint Committee on Artificial Intelligence, August 2005.