Extending Datatype Support in Web Ontology Reasoning

Jeff Z. Pan and Ian Horrocks
Department of Computer Science, University of Manchester, UK
{pan,horrocks}@cs.man.ac.uk

The Semantic Web is a vision of the next generation Web, which aims at machine understandability. "Semantic" markup will be added to Web resources, specifying their meanings so as to make them more accessible to software agents. Markups will use ontologies for shared understanding within certain domain. DAML+OIL [van Harmelen *et al.*, 2001] is a Web ontology language, which is compatible with existing Web standards, i.e. RDF. It is formally specified and have adequate expressive power—on the one hand, it is much more expressive than RDF, while on the other hand, it is believed to be still decidable.

DAML+OIL is in fact a Description Logic (DL). Significant efforts have already been devoted to the investigation of suitable DLs to provide reasoning support for DAML+OIL—in particular, [Horrocks and Sattler, 2001] have presented the $\mathcal{SHOQ}(\mathbf{D})$ DL, which is believed to be a good candidate, along with a sound and complete algorithm for deciding concept satisfiability, a basic reasoning service for DLs and ontologies.

A key feature of $\mathcal{SHOQ}(\mathbf{D})$ is that, like DAML+OIL, it supports *datatypes*(e.g., string, integer) as well as the usual abstract concepts (e.g., animal, plant). $\mathcal{SHOQ}(\mathbf{D})$, however, only supports unary datatype predicates. While this is enough for the *current version* of the DAML+OIL language, it may not be adequate for some Semantic Web applications and for possible future extensions of DAML+OIL.

E.g., ontologies used in e-commerce may need to classify different customers according to the numbers of their friends' email addresses they provide, and to reason that a customer who provides at least 10 *friends' email addresses*, and at least 5 of them are from UK, at least 5 of them have the same domain as the *customer's email address*, e.g. hotmail.com, is a kind of customers who are entitled to have 5% cash back during the promotion. Here "*friend's email address*" and "*customer's email address*" have concrete values (string). "From UK (*friends' email address*)" is a unary datatype predicate, and "the same domain as (*friends' email address*,*customer's email address*)" is a binary predicate. As shown above, unary predicates are *not enough*, while n-ary predicates, as well as number restrictions with datatypes, are often necessary in (Web) ontology applications.

[Baader and Hanschke, 1991] extended the well known $\mathcal{ALC}$ DL with concrete domain. Though $\mathcal{ALC}(\mathcal{D})$ is proved decidable, $\mathcal{ALC}(\mathcal{D})$ with general TBox is found undecidable. In order to extend *expressive* DLs with concrete domains, [Horrocks and Sattler, 2001] proposed a simplified approach on concrete domain and gave the $\mathcal{SHOQ}(\mathbf{D})$ DL. [Pan, 2001] investigated the simplifying constraints introduced in [Horrocks and Sattler, 2001]. The main difference between the two approaches is that the latter one uses concrete roles, instead of features.

We extend the $\mathcal{SHOQ}(\mathbf{D})$ DL with n-ary datatype predicates and number restrictions on datatypes, to give the $\mathcal{SHOQ}(\mathbf{D_n})$ DL. The kind of customers described in the above example can be defined as a $\mathcal{SHOQ}(\mathbf{D_n})$-concept

$$customer \sqcap \geqslant 10 friends-email \sqcap \geqslant 5 friends-email. \texttt{from-uk} \sqcap \geqslant 5 friends-email, email.\texttt{same-domain}$$

where *friends-email* and *email* are concrete roles, `from-uk` is unary datatype predicate and `same-domain` is binary datatype predicate.

We prove that the tableau algorithm we give is a *decision procedure* for concept satisfiability and subsumption of the $\mathcal{SHOQ}(\mathbf{D_n})$-concept w.r.t. terminologies. With its support for both nominals and n-ary datatypes, $\mathcal{SHOQ}(\mathbf{D_n})$ is well suited to provide reasoning support for ontology language in general, and Semantic Web ontology language in particular, e.g. the DAML+OIL language and suggests that *future version* of DAML+OIL can have n-ary datatype predicates and qualified number restrictions with datatypes. As future work, an optimised implementation based on the FaCT system is planned, and will be used to test empirical performance.

References

[van Harmelen *et al.*, 2001] Frank van Harmelen and Peter F. Patel-Schneider and Ian Horrocks. *A Model-Theoretic Semantics of DAML+OIL(March 2001)*.

[Horrocks and Sattler, 2001] Ian Horrocks and Ulrike Sattler. Ontology Reasoning for the Semantic Web. In *Proc. of the 17th Int. Joint Conf. on Artificial Intelligence (IJCAI'01)*,2001.

[Baader and Hanschke, 1991] F. Baader and P. Hanschke. A Scheme for Integrating Concrete Domains into Concept Languages. In *Proceedings of IJCAI-91*, 1991.

[Pan, 2001] Jeff Z. Pan. Web Ontology Reasoning in the SHOQ(Dn) Description Logic. In *Proceedings of the Methods for Modalities 2 (M4M-2)*,ILLC, University of Amsterdam, 2001.