Semantic Web 2 (2011) 67–70
DOI 10.3233/SW-2011-0048
IOS Press

Editorial

Taking flight with OWL2

Michel Dumontier
Department of Biology, School of Computer Science, Carleton University, 1125 Colonel By Drive, Ottawa, Canada K1S5B6
E-mail: michel_dumontier@carleton.ca

Abstract. The release of OWL2, the latest incarnation of the Web Ontology Language, in the fall of 2009 delivered a new set of features drawn from user communities, researchers, and developers. These features build on and extend the original version in a way that is not only more expressive, but also addresses the burden of implementing the full language through tractable subsets for more efficient reasoning. Increased understanding and a focus on scalability has also moved OWL from toy applications towards large scale knowledge bases for data verification, question answering and knowledge discovery.

Keywords: Semantic Web, OWL, RDF, triples, ontology, application

What's new with OWL2

OWL2 [24] introduced a number of constructs that increased the overall expressivity of ontologies in not just what one could craft in terms of class expressions, but more significantly, what one could say about relations. Of note in terms of class expressions, qualified cardinality restrictions (QCR) allowed users to express the number and kind of objects in a relation with the subject. Prior to this, users worked around it by introducing numerous domain-tainted relations i.e. 'has wheel part' with the caveat that users had to know how to use this specific relation rather than a more general 'has part'. Thus, QCRs enable powerful reuse of the same relation in different class expressions, and supports the idea of semantic interoperability through a shared set of domain-independent relations.

OWL2 also included a number of features to strengthen the semantics of relations, whether to say that they were disjoint (individuals paired by one relation must not share relations with which it is disjoint), or to indicate that relations are reflexive, irreflexive or asymmetric. One significant addition is that of so-called role chains, where a relation is inferred from the composition of two or more relations. Thus, it becomes possible to express that if an individual a has a parent *p* and that parent *p* has a sibling *s* and sibling *s* has a child *c* then *a* is the first cousin of c, and vice versa; see Robert Steven's blog entry [21].

Finally, the most notable development was the introduction of syntactic subsets of OWL which have more attractive computational properties – principally that they are tractable and offer polynomial time reasoning, instead of the worst case 2NEXPTIME-completeness.

OWL: Early struggles to catch wind

In the late 1990's and early 2000's, interoperability pretty much meant using XML as common syntax and XML Schema to define the document structure, principally by constraining the number, position and type of elements. For better or for worse, the main driver for XML was data exchange, and the semantics were in the documentation. People started to look at OWL to compose and share controlled vocabularies, but they certainly didn't care for OWL semantics. The most significant challenge often involved a lack of or deep misunderstanding of the lack of Unique Name assumption (i.e. different names don't imply different individuals) as well as the Open World Assumption (i.e. do not assume something is false – it needs to be explicitly stated). One of the early adopters to face this problem was BioPAX, an effort to exchange pathway data, which wanted more

in terms of constraints, rather than anything having to do with automated reasoning [17]. Moreover, the wider Semantic Web community has time and time again expressed a certain exasperation with OWL – many complaints stemming from a belief that the language is simply too complex for their applications. While it may very well be the case that OWL is apparent overkill for simple database, data exchange or web applications, there are many indicators that the problem lays less with applicability than with an adequate understanding of the technology. Of course, the right information hasn't been readily consumable for end-users – most of the necessary information on how to use OWL effectively has been locked up in theoretical contributions or could only be garnered by attending conferences (ISWC, ESWC) or workshops devoted to OWL (e.g. OWL: Experiences and Directions). The OWL2 primer [8], a practical guide to introduce users to both the syntax and semantics of OWL2, makes enormous strides to provide this basic level of education for a broader audience. New books are also providing some insight as to the meaning and use of OWL constructs [1,9,20,23].

Data provisioning: Linked data and RDF

For better or for worse, the Semantic Web effort is *currently* about exposing and linking data using URIs – under the moniker of Linked Data – and acts as a first step in getting people to share their data by any means possible. However, when data are represented using the Resource Description Framework (RDF), there is often little or no commitment to RDF semantics. Instead, the data is generally considered as a set of nodes and edges which extend from one dataset into a growing web of linked data. The Bio2RDF project [4], which provides billions of statements from scores of life science databases also only makes a lightweight commitment to simple RDF semantics and there is little return from reasoning over the knowledge base. Certain vocabularies augment the semantics of their relations with some features from OWL, but this is often limited to transitive or functional / inverse functional object properties. But since SPARQL based queries over RDF-based linked data seem to satisfy most users, why bother with more?

The trouble with triples

Together, RDF/RDFS and SPARQL 1.1 offer a compelling set of technologies to address a major data interoperability problem – that of having a common syntax with lightweight semantics along with a powerful query language. Even querying data from multiple sources using transitive relations and a simple hierarchy can vastly improve the query experience – something that is offered by SKOS when querying terminologies. So why do we need more? Well, if you ask anybody about the quality of the semantically annotated data, most remark that the linked data web is a scary place to explore because every dataset commits to a different i) conceptualization and ii) formalization – that is to say – what the data represents and how it is represented varies so dramatically that when one takes a close look there are intrinsic errors in a single dataset or in the combination of datasets and that even datasets containing the same kind of data requires different queries. A good example of the trouble with triples is that one might see a set of triples (using their human readable labels instead of URIs) such as:

`tailless mouse' `species' `mouse'
`tailless mouse' `lacks part' `tail'

For which the intent is to express that every instance of a tailless mouse is a mouse for which there is never an instance of a tail as a part. However, the RDF representation above states that a tailless mouse holds the relation 'species' with mouse as opposed to indicate that every instance of a tailless mouse is an instance of a mouse. It also states that a tailless mouse holds the relation 'lacks part' with a tail instead of stating that for every instance of a tailless mouse there is never an instance of tail as a part. While linked data is using RDF as a vehicle to publish data, much work remains to accurately formalize the knowledge such that the meaning is preserved.

Ontological commitment: The path to semantic redemption

So how do we solve the semantic interoperability problem? To a large part it requires that data (in the form of triples or more generally n-tuples) be faithfully represented so that statements are represented as accurately as possible so that the interpretation is both correct and unambiguous. A proper formalization of the above example requires a more expressive language like OWL and a commitment to the meaning of the relations, particularly those that hold among all instances of a given class. Thus, a more

accurate formalization in OWL (Manchester OWL syntax [14,15]) is:

> `tailless mouse' EquivalentTo `mouse' and not (`has part' some `tail')

The interesting thing here is that the formalization of knowledge is less about triples, and more about axioms that create a truth value for a statement which can be checked by a reasoner. This doesn't mean that we have to commit to a single interpretation – in fact we often want a set of possibilities to be included – it just means that when you examine the meaning of the statement, we can derive all the meanings of that statement (including when there is an element of vagueness). In recent work, Hoehndorf et al. [11] demonstrated the formalization of statements made in OBO, a language to compose biological ontologies, into OWL, such that the relations can be expanded into expressions that can be automatically reasoned about by OWL reasoners. In this way, even ontologies using different relations could be made semantically interoperable by committing relations with ambiguous semantics into a coherent representation.

Real world applications of OWL

As part of the 2010 launch of the Semantic Web Journal, we put out a call for papers on the real world applications of OWL [22], with the hope to get reports of how using OWL helped solve a problem that would otherwise be challenging. Out of eight submissions, we have so far only accepted one – "FactForge: A fast track to the Web of data" [5] by Barry Bishop and colleagues from Ontotext AD – whose contribution was to define a reasonable view (RAV) over a subset of linked data comprising of 282M entities and 100k relations using BigOWLIM (now OWLIM cluster) with a reduced ruleset corresponding to OWL Horst, a subset of OWL RL. With the exception of disjoint class axioms, the authors note that since their linked data contained little else in terms of OWL that would trigger rules for consistency checking. Once the authors repaired inconsistencies, reasoning added an additional 881M statements to the 1.1B assertions, thereby making it one of the largest examples of reasoning over real data (as opposed to standard reasoning datasets or large scale synthetic data).

While FactForge (and OWLIM [6]) pushes the envelope in terms of reasoning with linked data, new work is coming out that demonstrates the value of large-scale generated OWL ontologies for consistency checking, question answering and knowledge discovery. Mungall and colleagues [18] turned towards axiomatic descriptions to check the correctness of the manual curation of the Gene Ontology and subsequently enhance its quality and coverage. In fact, the development of axiomatic descriptions are now becoming the norm for a number of ontology building efforts such as the cell cycle ontology [2], or in the integration of anatomy-phenotype ontologies [19]. Hoehndorf et al. [13] demonstrate disease gene discovery using an ontology containing more than 275,000 classes and 1M axioms, produced from the alignment of species-specific anatomy and phenotype ontologies against a common ontology to describe qualities. Similarly, formalization of RDF annotations embedded in XML-based biological models for simulation have demonstrated OWL's ability to uncover inconsistencies related to an abuse of SBML models and errors in curation [10]. Part of the solution for large scale reasoning also involves reducing the complexity of OWL-DL ontologies into OWL-EL for which there are now efficient reasoners. The El Vira software [12] can be used to convert the OWL files to the OWL EL subset and enable tractable automated reasoning over the combined ontologies using reasoners such as CB [16] or CEL [3]. A more complete comparison of OWL reasoners for OWL EL is now available [7].

Final thoughts

While early use of OWL pertained largely to the development of hand crafted ontologies for simple semantic annotation or to demonstrate some reasoning over a clever formalization, there is a growing sense that OWL2 with its computationally attractive OWL profiles now delivers in terms of building and reasoning about large scale OWL knowledge bases for what OWL is advertised for: consistency checking, question answering and knowledge discovery. Yet significant challenges remain towards having a sufficient understanding of how to conceptualize and formalize all the kinds of entities we wish to describe, and of particular interest are those that are already present in linked data. Certainly, much more work remains in terms of having a more coherent representation of knowledge on the Semantic Web along with the tools to execute powerful reasoning over it.

References

[1] D. Allemang and J. Hendler, *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*, Morgan Kaufmann, 2008.

[2] E. Antezana, M. Egana, W. Blonde, A. Illarramendi, I. Bilbao, B. De Baets, R. Stevens, V. Mironov, and M. Kuiper, The cell cycle ontology: An application ontology for the representation and integrated analysis of the cell cycle process, *Genome Biol*. **10**(5) (2009), R58.

[3] F. Baader, C. Lutz, and B. Suntisrivaraporn, CEL: A polynomial-time reasoner for life science ontologies, in: *Proc. of the 3rd Int. Joint Conf. on Automated Reasoning (IJCAR06)*, U. Furbach and N. Shankar, eds, Springer-Verlag, 2006, pp. 287–291.

[4] F. Belleau, M.A. Nolin, N. Tourigny, P. Rigault, and J. Morissette, Bio2RDF: Towards a mashup to build bioinformatics knowledge systems, *J. Biomed. Inform*. **41**(5) (2008), 706–716.

[5] B. Bishop, A. Kiryakov, D. Ognyanoff, I. Peikov, Z. Tashev, and R. Velkov, FactForge: A fast track to the web of data, *Semantic Web* **2**(2) (2011).

[6] B. Bishop, A. Kiryakov, D. Ognyanoff, I. Peikov, Z. Tashev, and R. Velkov, OWLIM: A family of scalable semantic repositories, *Semantic Web* **2**(1) (2011), 33–42.

[7] K. Dentler, R. Cornet, A. ten Teije, and N. de Keizer, Comparison of reasoners for large ontologies in the OWL 2 EL profile, *Semantic Web* **2**(2) (2011).

[8] P. Hitzler, M. Krötzsch, B. Parsia, P.F. Patel-Schneider, and S. Rudolph. OWL 2 Web Ontology Language Primer. 2009 [cited 2011; available from: http://www.w3.org/TR/owl2-primer/].

[9] P. Hitzler, M. Krötzsch, and S. Rudolph, *Foundations of Semantic Web Technologies*, Chapman & Hall/CRC, 2009.

[10] R. Hoehndorf, M. Dumontier, J.H. Gennari, S. Wimalaratne, B.d. Bono, D. Cook, and G.V. Gkoutos, Integrating systems biology models and biomedical ontologies, *BMC Systems Biology* (2011), in press.

[11] R. Hoehndorf, M. Dumontier, A. Oellrich, D. Rebholz-Schuhmann, P.N. Schofield, and G.V. Gkoutos, Interoperability between Biomedical Ontologies through relation expansion, upper-level ontologies and automatic reasoning, *PLoS One* **6**(7) (2011), e22006.

[12] R. Hoehndorf, M. Dumontier, A. Oellrich, S. Wimalaratne, D. Rebholz-Schuhmann, P. Schofield, and G.V. Gkoutos, A common layer of interoperability for biomedical ontologies based on OWL EL, *Bioinformatics* **27**(7) (2011), 1001–1008.

[13] R. Hoehndorf, P.N. Schofield, and G.V. Gkoutos, PhenomeNET: A whole-phenome approach to disease gene discovery, *Nucleic Acids Res*. (2011).

[14] M. Horridge, N. Drummond, J. Goodwin, A. Rector, R. Stevens, and H.H. Wang, The Manchester OWL syntax, in: *OWL Experiences and Directions*, Athens, Georgia, 2006.

[15] M. Horridge and P.F. Patel-Schneider, OWL 2 Web Ontology Language: Manchester Syntax. 2009. Available from: http://www.w3.org/TR/owl2-manchester-syntax/.

[16] Y. Kazakov, Consequence-driven reasoning for horn SHIQ ontologies, in: *Proc. of the 21st International Conference on Artificial Intelligence (IJCAI 2009)*, H. Kitano, ed., Morgan Kaufmann Publishers Inc., 2009, pp. 2040–2045.

[17] J.S. Luciano and R.D. Stevens, OWL: PAX of mind or the AX? Experiences of using OWL in the development of BioPAX, in: *OWL: Experiences and Directions*, Washington, DC, 2008.

[18] C.J. Mungall, M. Bada, T.Z. Berardini, J. Deegan, A. Ireland, M.A. Harris, D.P. Hill, and J. Lomax, Cross-product extensions of the gene ontology, J. Biomed. Inform. **44**(1) (2011), pp. 80–86.

[19] C.J. Mungall, G.V. Gkoutos, C.L. Smith, M.A. Haendel, S.E. Lewis, and M. Ashburner, Integrating phenotype ontologies across multiple species, *Genome Biol.* **11**(1) (2010), R2.

[20] P.N. Robinson, *Introduction to Bio-Ontologies*, Chapman & Hall/CRC, 2011.

[21] R. Stevens, The Family History Knowledge Base, 2010 [cited 2011 July 30]. Available from: http://robertdavidstevens.wordpress.com/2010/05/04/the-family-history-knowledge-base/.

[22] Call for papers: Special Issue: Real-World Applications of OWL, *Semantic Web Journal*, 2011. Available from: http://www.semantic-web-journal.net/content/special-issue-real-world-applications-owl.

[23] Ontogenesis 2011. Available from: http://ontogenesis.knowledgeblog.org/.

[24] OWL 2 Web Ontology Language – Document Overview. 2009. Available from: http://www.w3.org/TR/owl2-overview/.