Institution: University of Manchester

Unit of Assessment: UoA 11 Computer Science and Informatics

Title of case study: OWL – an Ontology Language Standard with Sound Logical Underpinning

1. Summary of the impact

Ontologies are used to describe the meaning of terms in a domain. Manchester has had a leading role in the design of ontology languages, algorithms and tools. Through standardization, algorithm development and tool creation, we have significantly influenced the uptake of the Ontology Web Language (OWL) and Semantic Web Technologies by public service providers and industry. For example, the NCI thesaurus and SNOMED CT are medical terminologies in OWL; specialised semantic web companies such as Clark & Parsia, Racer Systems and TopQuadrant provide semantic technologies and services that build on OWL; and companies such as Oracle and B2i Healthcare include tool support for OWL.

2. Underpinning research

By the end of the 1990s, Description Logics (DLs) were well-understood logical formalisms, e.g., with respect to their computational complexity, model theory and reasoning services. They were used in niche applications, but of limited expressive power and with limited tool support – the latter due to the lack of a standardised syntax. Research carried out in Manchester dramatically improved the expressive power of and tool support for DLs, which led to the adoption of DLs as the logical underpinning of ontology languages, to their standardisation and integration with Semantic Web infrastructure in OWL and OWL 2, and to their usage in biohealth and eScience applications.

Key researchers from Manchester are:

- Sean Bechhofer (1993 – present: RA, Lecturer 2004, Senior Lecturer 2013)
- Ian Horrocks (1996 – 2007: PhD student, RA, Lecturer 1999, Senior Lecturer 2002, Professor 2003)
- Bijan Parsia (2006 – present: Lecturer, Senior Lecturer 2012)
- Uli Sattler (2003 – present: Senior Lecturer, Professor 2007)

Key research results that underpinned the impact were:

i. Motivated by application examples, the design and investigation of a range of extensions to existing DLs, establishing various (un)decidability and computational complexity results [2,3]. In particular, the design of the *SHIQ* family of DLs, which show good computational properties while being highly expressive: they support general concept inclusion, transitive roles and role hierarchies, inverse roles, cardinality restrictions and nominals. Each of these features was thought to be problematic yet desirable, and the research proved that their combination is of high computational complexity yet decidable.

ii. The design and analysis of reasoning algorithms for *SHIQ* DLs [2,3]. These mainly tableau-based algorithms provide sound and complete decision procedures for the basic DL reasoning problems, and depend on sophisticated blocking techniques for their termination. The implementation and optimisation of these algorithms in the *FaCT* reasoner provided evidence of their applicability in practice, overcoming the commonly held belief that such algorithms are impractical [1].

iii. Investigation of ontology engineering problems, in particular for modularity, entailment explanation and query answering, and development of logically sound yet practical solutions for these [4].

iv. Design and implementation of tools such as ontology-based editors, reasoners, APIs, and web applications that both showcase the benefits of OWL for various applications and demonstrate the practicality of the developed algorithms [5]. The resulting reasoners, editors, and APIs were highly influential: further research into optimisation in *FaCT++* showed that reasoners for complex DLs can indeed cope with large-scale ontologies; the ontology editor *OiLEd* was the first one that exposed ontology designers to a reasoner, and it heavily influenced Protégé and other editors; and the *OWL API* is the main API used to interact programmatically with an ontology and reasoners [6].

Our research involved a unique combination of investigations of the computational complexity of logics with informed language, algorithm design and tool development that changed the general understanding of what it means for an ontology language to be practical and expressive.

3. References to the research

Papers that describe this research have been published in international leading journals.

Key references:

[1] I. Horrocks. Using an Expressive Description Logic: FaCT or Fiction? In Proc. of the 6th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'98), pp. 636-647, 1998. [Google Scholar: 575 citations] KR Outstanding Paper Award.

[2] I. Horrocks, and U. Sattler. Decidability of SHIQ with complex role inclusion axioms. *Artificial Intelligence*, 160 (1-2), pp. 79-104. 2004. DOI: 10.1016/j.artint.2004.06.002 [Google Scholar: 97 citations].

[3] I. Horrocks and U. Sattler. A Tableaux Decision Procedure for SHOIQ. *Journal of Automated Reasoning*, Springer Verlag, 39(3), 245-429, 2007. DOI: 10.1007/s10817-007-9079-9 [Google Scholar: 419 citations].

Other references:

[4] B. Cuenca Grau, I. Horrocks, Y. Kazakov, and U. Sattler. Modular Reuse of Ontologies: Theory and Practice. *Journal of Artificial Intelligence Research (JAIR*), Vol. 31, pp. 273-318, 2008. DOI: /10.1613/jair.2375. [Google Scholar: 215 citations].

[5] BC Grau, I Horrocks, B Motik, B Parsia, P Patel-Schneider, U Sattler, OWL 2: The next step for OWL, *J. Web Semantics* 6 (4), 309-322, 2008. DOI: 10.1016/j.websem.2008.05.001 [Google Scholar: 351 citations].

[6] M. Horridge, S. Bechhofer. The OWL API: A Java API for OWL ontologies. *Semantic Web* 2(1): 11-21, 2011. DOI:10.3233/SW-2011-0025. [Google Scholar: 70 citations].

4. Details of the impact

Context

There were promising, prototypical implementations of DL reasoners that were respected in the knowledge representation community, but found little adoption: the logics were inexpressive, there was limited tool support, no methodologies, no editors and no standardized syntax.

Pathways to Impact

The research that underpins the impact was reported in high profile publications that were widely read and cited. Associated with these publications, several software systems were released that enabled experimentation with and application of the research results. For example, the DL reasoners *FaCT* and *FACT++* each received tens of thousands of downloads, as did OiLEd, the first ontology editor that was tightly integrated with a DL reasoner, thereby providing a tool to showcase the feasibility and potential benefits of supporting ontology engineers through DL reasoning. These tools, together with the *OWL API*, led to a wide user base and support by other tool developers, which helped to establish the utility and maturity of DLs for a range of applications, including in the web. This, in turn, enabled Horrocks and others to convince the Semantic Web community to adopt DLs as the logical foundations of Semantic Web Ontology languages.

Reach and Significance of the Impact

This section describes impacts that have resulted from the research. Although some of these are economic, there have also been important impacts on practitioners and service providers through the introduction of influential *de jure* standards and changes to best practice, in particular for the design and development of widely deployed biomedical terminologies.

Development of Standards in the W3C. The widespread adoption of the research results has been made possible by the development of World Wide Web Consortium (W3C) standards for ontology languages and associated interfaces. These standards are designed by *working groups* that have members from outside academia and require serious implementation and usage efforts to be established. Standards that build directly on Manchester research, and in which Manchester authors collaborated with other academics and industrial partners include:

- OWL (2004): the first W3C ontology language standard, which builds on Manchester research as described in [5]; contributors: Horrocks and Bechhofer.
- OWL 2 (2009): a revision of OWL, heavily driven by the OWL Experiences and Directions group founded by Horrocks, Parsia, *et al*. and by research on qualified number restrictions, key constraints, and rich property axioms by Horrocks, Parsia, and Sattler; contributors: Horrocks, Parsia and Sattler.
- SKOS (2009): a standard language for knowledge organisation systems based on OWL; contributors: Bechhofer.
- SPARQL (2008) and SPARQL 1.1 (2010-12): a query language for RDF and OWL, influenced by work on OWL query answering by Horrocks and Sattler; contributors: Parsia.

These *de jure* standards have been widely deployed throughout the REF period, as described below, thus impacting on practitioners who both develop and apply semantic web technologies.

Uptake of Reasoning Algorithms and Tools. The standards have enabled the development of products and tools that in turn have facilitated the widespread application of semantic web techniques. In this section we focus principally on two examples:

- *Pellet*: Pellet, a commercial reasoning system that supports OWL 2, is marketed by Clark & Parsia, and includes algorithms developed at Manchester [A]. Pellet has been used by organisations including NASA, US Army, US Banking Institutions, NATO, NCI, Ordnance Survey and iPlant Collaborative [A]. In addition, Pellet has been integrated with and is used in Oracle 11g [A].
- *Protégé*: Protégé is the most widely used computer system for engineering ontologies, with 225,000 registered users, and more than 17,000 members of email discussion groups [B]. Protégé supports OWL 2, the user interface for the current version of Protégé was designed and implemented at Manchester, and Protégé implements the Manchester OWL API [B].

Manchester techniques and tools, in conjunction with the W3C Standards, inform other commercial platforms. For example: the commercial reasoner RacerPro incorporates algorithms designed in Manchester and pioneered in FaCT, and implements the Manchester OWL API [C]; and the KnowledgeServer semantic infrastructure of derivo GmbH implements OWL and the OWL API [D].

OWL Tools and Ontologies. Today, thousands of OWL ontologies are available on the Web. The National Centre for Biomedical Ontologies BioPortal ontology repository [E] contains 365 OWL ontologies, which are used in all areas of biomedical activity. The following three ontologies are amongst the most well established and widely used ones and are built using OWL and OWL 2:

- SNOMED CT [F], from the International Health Terminology Standards Development Organisation, is the prime medical thesaurus and is used worldwide in a variety of healthcare applications, e.g., in NHS Connecting for Health. The company B2i Healthcare provides specialist support around OWL and SNOMED CT, in particular *Snow OWL* [G].
- National Cancer Institute (NCI) Thesaurus [H] is a key biomedical research vocabulary used in OWL that uses Pellet for classification [A].
- The 11th version of the International Statistical Classification of Diseases and Related Health Problems (ICD-11) is currently being developed in OWL under WHO leadership [I]. ICD is used to classify diseases and other health problems, e.g., in death certificates and health records. These records also provide the basis for the compilation of national mortality and morbidity statistics by WHO Member States, and are used for reimbursement and resource allocation decision-making by governments.

The W3C maintains a list of use-cases of semantic web standards [J] that detail the users that have built on the standards to which Manchester contributed. These include: the Norwegian National Broadcaster, Cleveland Clinic, Ordnance Survey, IBM, National Archives of Korea, Food and Agriculture Organization of the United Nations, BBC, Chevron, Renault, Agfa Healthcare and Vodaphone.

5. Sources to corroborate the impact

Supporting material is available from the university for the corroborating sources below.

[A] Letter from CEO of Clark & Parsia (http://clarkparsia.com/). Confirms use of Manchester research in Pellet, and provides details of its use.

[B] Letter from Professor of Medicine (Biomedical Informatics), Stanford Centre for Biomedical Informatics Research. Confirms the role of Manchester research in Protégé, and the scale/nature of the user community.

[C] Racer Systems GmbH & Co: RacerPro User's Guide, Version 2.0, October 2012 (http://www.racer-systems.com), also (http://www.racer-systems.com/technology/references.phtml), on 29th August 2013. Confirms the influence of Manchester research on RacerPro.

[D] derivo GMBH (http://www.derivo.de/), also (http://www.derivo.de/ressourcen/owllink.html), on 29th August 2013. Confirms the use of the Manchester OWL API.

[E] BioPortal: (http://bioportal.bioontology.org/), on 29th August 2013. Provides information on the number and scale of biological ontologies.

[F] SNOMED CT Technical Implementation Guide, International Health Terminology Standards Organisation (www.snomed.org/tig.pdf). Confirms role of OWL in SNOMED CT.

[G] B2i Healthcare Ltd (http://www.b2international.com/), on 29th August 2013. Demonstrating a commercial use of OWL.

[H] NCI Thesaurus (http://ncicb.nci.nih.gov/download/evsportal.jsp), on 29th August 2013. Confirms role of OWL in NCI Thesaurus.

[I] Tudorache T, Falconer S, Nyulas C, Storey MA, Ustün TB, Musen MA. Supporting the Collaborative Authoring of ICD-11 with WebProtégé. *AMIA Annu Symp Proc.* 2010 Nov 13; 2010:802-6, http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3041458/. Confirms that ICD-11 uses OWL.

[J] W3C list of use cases: http://www.w3.org/2001/sw/sweo/public/UseCases/, on 29th August 2013.