| |
|---|
| **Institution: University of Aberdeen** |
| **Unit of Assessment: 11 (Computer Science & Informatics)** |
| **Title of case study: Enabling Semantic Reasoning for Linked Data** |

## 1. Summary of the impact

Within this case study we present the TrOWL technology developed at the University of Aberdeen that enables more efficient and scalable exploitation of semantic data. TrOWL and its component algorithms – REL, Quill and the Aberdeen Profile Checker – have had non-academic impact in two key areas. With respect to *practitioners and professional services*, the technology has enabled the introduction of two important World Wide Web Consortium (W3C) standards: OWL2 and SPARQL 1.1. This has led to impact in the way that many companies work, across a range of sectors. Further, through partnership with specific companies, the use of TrOWL has changed the way they operate and the technical solutions they provide to clients. These collaborations have led to *economic* impacts in companies such as Oracle in "mitigat[ing] the losses of potential customers", and IBM in "using the TrOWL reasoning infrastructure in [their] Smarter Cities solutions".

## 2. Underpinning research

Recent years have seen rapid growth in the use of vocabularies defined in formal knowledge bases called ontologies to annotate data. Such semantic data can be represented as sets of subject-predicate-object triples. By employing logical reasoning facilities, new triples can be inferred from known triples. With over 30 billion triples of linked semantic data in online repositories and over 1.7 trillion triples of fast growing online semantic data (contributed by, for example, Google and Facebook), efficient and scalable reasoning over semantic data is one of the most pressing research problems in the field of the semantic web. The challenge is to balance the trade-off between representational expressiveness (the need to describe web data) and the efficiency and scalability of reasoning. To answer the challenge, Jeff Z. Pan and his research team have developed advanced reasoning techniques, such as approximation, reasoning parallelisation, and reasoning on streaming data. They have used this expertise to help shape and establish the introduction of key industrial standards OWL 2 [R6] and SPARQL 1.1. Further, Pan and his team have evaluated the usefulness of these techniques in real world scenarios, and have worked closely with a range of commercial partners to enable the exploitation of these technologies through to products. Jeff Pan joined the unit as a lecturer in October 2005, promoted to Senior Lecturer in 2011 and Reader in 2013, and has been a full-time member of academic staff throughout this period.

The key result of the research is TrOWL – a tractable reasoning and querying infrastructure, which is available for public download from http://trowl.eu. TrOWL is the most well known approximate ontology reasoning system that advocates and takes advantage of the two-layer architecture of the OWL standard (second version), by approximating ontologies in OWL 2 DL (in the expressive layer) into ontologies in the tractable profiles (in the tractable layer). TrOWL provides quality-preserving, approximate reasoning techniques to improve the efficiency and scalability of OWL 2 ontology reasoning, and has been developed over several years. During 2005-2007 (through the EPSRC-funded *Advanced Knowledge Technologies* project [GR/N15764/01] and EU-funded *Knowledge Web* [ST-2004-507842]), ARK researchers developed the first semantic approximation approach [R1], which produces optimal approximations (of input ontologies) that database-style queries cannot distinguish from input ontologies themselves. This is the Quill reasoner and query engine, implementing semantic approximation. During the second phase (Feb 2008-Apr 2011), the current TrOWL system was developed. This integrates the Quill reasoner and querying engine [R1], with REL and the Aberdeen Profile Checker. REL is a reasoner that implements real-time and high-recall syntactic approximation [R2] for the expressive OWL 2 DL. The Aberdeen Profile Checker is used as part of the TrOWL system to determine whether an input ontology complies with standard OWL2 'Profiles' (fragments of the SROIQ Description Logic defined by placing restrictions on the axioms employed for inference). This profile checker enables different reasoners to be employed, given the input ontology, enhancing the efficiency of the reasoning [R3]. Throughout this period the researchers collaborated closely with SAP and BOC Information Systems to evaluate the performance of TrOWL in solving challenging real-world problems. This included understanding how the underpinning research can be used to improve reasoning services in SAP process engineering [R4]. The TrOWL system [R3] underwent further extensive evaluation

in the *MOST* project [R5] – a project aimed at improving software engineering by leveraging ontology and reasoning technologies.

## 3. References to the research

[R1]   J Z Pan & E Thomas. Approximating OWL-DL Ontologies. In *Proc. of the 22nd AAAI Conference*. 1434-1439. 2007. http://homepages.abdn.ac.uk/jeff.z.pan/pages/PaTh07.pdf

** This paper presents the semantic approximation technology, Quill, that underpins the query answering service of TrOWL.  *Best indicates quality of underpinning research*.

[R2]   Y Ren, J Z Pan & Y Zhao. Soundness Preserving Approximation for TBox Reasoning. In *Proc. of the 24th AAAI Conference*. 351-356. 2010. [Pan3 in the REF2 for this unit.]

** This paper presents the syntactic approximation technology, REL, that underpins the ontology classification service of TrOWL.  *Best indicates quality of underpinning research.*

[R3]   E Thomas, J Z Pan & Y Ren. TrOWL: Tractable OWL 2 Reasoning Infrastructure. In *Proc. of the 7th Extended Semantic Web Conference*. 431-435. 2010. http://dx.doi.org/10.1007/978-3-642-13489-0_38

** This paper describes the core components of the TrOWL system: REL, Quill and the Aberdeen Profile Checker.  *Best indicates quality of underpinning research.*

[R4]   Y Ren, G Groener, J Lemcke, T Rahmani, A Friesen, Y Zhao, J Z. Pan & S Staab. Validating Process Refinement with Ontologies. In *Proc. of Int. Workshop on Description Logics - DL2009*. http://www.cs.ox.ac.uk/DL2009/proceedings/oral/Ren_Groener_Lemcke_Rahmani_Friesen_Zhao_Pan_Staab.pdf

** This paper presents initial collaboration results with SAP on process refinement validation.

[R5]   J Z Pan, S Staab, U Aßmann, J Ebert & Y Zhao (Eds). Ontology-Driven Software Development. Springer, 2013. ISBN: 978-3-642-31225-0 (Print) 978-3-642-31226-7 (Online). http://dx.doi.org/10.1007/978-3-642-31226-7

** This book compiles the results of the MOST project, including how to use TrOWL to support ontology driven software engineering.

[R6]   W3C Web Ontology Language OWL 2 Recommendation: http://www.w3.org/TR/2009/REC-owl2-overview-20091027/

## 4. Details of the impact

The TrOWL system, incorporating the Quill and REL reasoners and the Aberdeen Profile Checker, has had important *impacts on practitioners and professional services* by enabling the approval of the World Wide Web Consortium (W3C) standards *OWL 2* and *SPARQL 1.1*. Jeff Pan was a founding member of the OWL 2 working group (the developers of the OWL 2 standard) in 2007, and continued to be an active participant of this group through to it being established as a standard (2009). (Note that minor revisions have subsequently been made to ensure consistency across W3C standards.) Further, Jeff Pan acted as reviewer of the SPARQL 1.1 standard. Within the W3C, standards are referred to as 'Recommendations'. Before this, they go through the 'Candidate Recommendation' phase. According to W3C, "[t]he goal of the Candidate Recommendation (CR) phase is to demonstrate the existence of multiple interoperable and practically useful […] systems." For a Candidate Recommendation (CR) to become a W3C Recommendation (standard), there must be multiple, independent implementations that pass specified test cases. This is a time-limited phase (45 days in the case of OWL 2) and requires prompt input from the community. The REL and Quill components of TrOWL were among the first systems to support the 'Profiles' of OWL 2 [S6]. These tests include both reasoning (in which REL and Quill have contributed to a successful candidate recommendation phase) and ontology profile checking (in which the Aberdeen Profile Checker was the primary contributor). The Aberdeen Profile Checker is one of only two syntax constraints implementations for OWL 2. REL is one of the 3 implementations passing the OWL 2 EL object property chain test. Quill is one of the 3 implementations passing the OWL 2 QL import test. TrOWL is also one of the first systems to support the entailment regime in

the new semantic web query language standard SPARQL 1.1 [S7], which enabled this important reasoning feature to be approved in the standard. In the 6 SPARQL DL entailment tests, TrOWL was 1 of the 3 implementations to pass. The TrOWL system thus significantly contributed to both the OWL 2 and SPARQL 1.1 Candidate Recommendations becoming Recommendations (OWL2: 27 October 2009 [R6,S6]; and SPARQL 1.1: 21 March 2013 [S7]). These standards are widely used by government (e.g., data.gov), scientific communities (e.g., the bio-medical community), companies (e.g., BBC & BestBuy) and community-driven efforts (e.g., DBpedia).

Jeff Pan has further developed *impacts on practitioners and professional services* through tutorials and invited talks at industry-facing events. These include the AAAI 2010 tutorial on *Large-Scale Ontology Reasoning and Querying*. The leading industry conference in the area is Semantic Technology & Business (SemTechBiz) including *Semantic Web Enabled Software Engineering — State of the Art* in the 'Developing Semantic Software' track (SemTechBiz 2011, London). Jeff Pan and Oracle jointly presented *OWL-DBC: Integrating TrOWL Reasoner with Oracle Spatial and Graph* in the 'Product/Service Available Now' track (SemTechBiz 2013, San Francisco).

A key means to following through on this impact is via direct collaboration with leading organisations, to enable practitioners to use TrOWL technology in the conduct of their work:

- Oracle have been collaborating with ARK researchers to develop OWL-DBC, an integration solution of TrOWL and Oracle products [S8,S9]. The integration of TrOWL is recommended by Oracle to provide efficient reasoning support to the Resource Description Framework (RDF) Semantic Graph of their latest release of mainline database system Oracle DB, which is widely used to provide enterprise-level database management solutions. Oracle have found that *[text removed for publication]*. In the white paper of OWL-DBC, we also defined the best practices for dealing with knowledge bases with different properties by leveraging the integration of Oracle DB and TrOWL" [S1]. This leading company is, therefore, promoting the use of results of our underpinning research;

- BOC Information Systems were "[i]nspired by the promising (MOST) project end results [and] initiated a follow-up project to further delve into investigating possibilities of integrating the semantic technology into its own software platform" [S2]. As a result they "found out that TrOWL matched the specific requirements in [their] problem domain the best among other peer systems in terms of performance in resolving complex reasoning tasks" [S2]. Details of the technical solution are not published, but the key idea is to leverage the efficient and high-recall syntactic approximate reasoning service of TrOWL (REL) to automatically detect inconsistencies between software models and their meta-model specifications during model-driven software engineering;

- IBM has "integrat[ed] TrOWL with [their] diagnosis and predictive reasoning models" to "reach scalable performance in [their] real-world urban context of large, heterogeneous and stream data". As a result of this, IBM are "using the TrOWL reasoning infrastructure in [their] Smarter Cities solutions, such as those [they] use in testing real and live stream data from Dublin City in Ireland" [S3]. The level of collaboration (on-going since 2012) and scope of this impact is made clear by the statement from IBM that "[o]ur joint effort is part of a bigger picture of building distributed information infrastructures for Smarter Cities that support individuals, groups and organizations alike" [S3];

- SAP exploits TrOWL to validate and explain the refinement relations between business process models at different levels of abstraction. The initial solution was published in [R4] and the revised and detailed solution was published in Chapter 10 of [R5]. As stated by SAP "[w]ith the use of TrOWL, automatic refinement validation of business processes can be performed significantly faster than with alternative ontological technologies" [S4];

- Semantic Arts is testing TrOWL to reason with its upper ontology GIST as well as its client enterprise ontologies. The key challenge that they are facing is the efficiency of ontology reasoning, which is addressed well by the approximate reasoning service from TrOWL. One senior ontology consultant said "[I] tried TrOWL on one client ontology (1300 classes, 500 properties) that takes several minutes on FaCT++ [another reasoner] and it was sub-second with TrOWL (972ms, to be exact), so I was blown away" [S5].

A number of these impacts on practitioners have led to TrOWL technologies being adopted in stakeholder's business operations and products, leading to significant *economic impacts*. With respect to Oracle products, they state that "[w]e (Oracle) […] are very happy to see the first release of OWL-DBC (on 4th Feb, 2013), which combines the best of Oracle's Spatial and Graph with the powerful reasoning engine TrOWL". An example of how this follows through to benefits for Oracle's customers is that this integrated solution ***"[text removed for publication]"*** [S1]. It is worth noting that Oracle has been the leading life sciences software vendor since 2010, according to IDC Health Insights, an advisory services and market research firm focused on the healthcare and life science market segments. The fact that IBM is "using the TrOWL reasoning infrastructure in [their] Smarter Cities solutions" to "reach scalable performance" [S3] has economic impact for this global business. According to Navigant Research's Leadership Report on Smart Cities (published 3rd Quarter of 2013) [S10] "the global market will grow from $6.1 billion annually in 2012 to more than $20 billion in 2020". IBM has a "leadership position" in the Smart City solutions market [S10], and TrOWL has been used in IBM's solutions for 6 months during the REF period.

## 5. Sources to corroborate the impact

[S1]   Architect, Oracle – provides a statement to corroborate the impact of TroWL and OWL-DBC, which combines Oracle's Spatial and Graph with TrOWL.

[S2]   Managing Director, BOC Information Systems – confirms the impact of the TrOWL project on their practices in resolving complex reasoning tasks.

[S3]   Lead Investigator in Large-Scale Reasoning Systems, IBM – confirms the use of TrOWL in IBM's Smart Cities solutions to build distributed information infrastructures.

[S4]   Researcher, SAP – confirms collaboration through MOST, and how TroWL provides a superior method to support validation of business processes refinement.

[S5]   Senior Ontology Consultant, Semantic Arts – corroborates claims in relation to the testing of TrOWL on client ontologies.

[S6]   W3C OWL 2 implementations and test results:
http://www.w3.org/2007/OWL/wiki/Test_Suite_Status – corroborates use of REL, Quill and the Aberdeen Profile Checker in testing phase for OWL2.

[S7]   W3C SPARQL 1.1 implementations and test results:
http://www.w3.org/2009/sparql/implementations/ – corroborates use of TrOWL in testing phase for SPARQL 1.1.

[S8]   Oracle Database version 12c, RDF Semantic Graph - Technical Information:
http://www.oracle.com/technetwork/database-options/spatialandgraph/documentation/rdfsem-techinfo-1916685.html

[S9]   Presentation by Xavier Lopez, Director of Product Management, Oracle in 2012 noting integration of TrOWL as a new inference function in Oracle DB release 12c:
http://download.oracle.com/otndocs/tech/semantic_web/pdf/semtech_datamining_v8.pdf

[S10] Navigant Research's Leadership Report on Smart Cities:
http://www.navigantresearch.com/wp-assets/uploads/2013/07/LB-SCITS-13-Executive-Summary.pdf