AberOWL: an ontology portal with OWL EL reasoning

Luke Slater [1]*, Georgios V Gkoutos[2], Paul N Schofield[3], Robert Hoehndorf[1]

[1] Computational Bioscience Research Center, King Abdullah University of Science and Technology, 4700 KAUST, 23955-6900, Thuwal, Saudi Arabia
[2] Department of Computer Science, Aberystwyth University, Aberystwyth, SY23 3DB, Wales, United Kingdom
[3] Department of Physiology, Development and Neuroscience, University of Cambridge, Downing Street, CB2 3EG, England, United Kingdom

ABSTRACT

The field of biological and biomedical science quickly generate large quantities of data and knowledge; often, domain knowledge is formalised using ontologies expressed in the Web Ontology Language (OWL). Ontology repositories such as Bioportal and Ontobee have been an important infrastructural component for managing ontologies, specifically to search, browse and download ontologies over the Web. We present the AberOWL system, a novel ontology repository that allows access to multiple ontologies through automated reasoning, utilizing parts of the OWL of the ontologies alongside a web interface and web services. AberOWL contains over 300 ontologies and integrates reasoning over ontologies with access to literature and SPARQL endpoints.

1 INTRODUCTION

Major ontology repositories such as BioPortal (Noy *et al.*, 2009), OntoBee (Xiang *et al.*, 2011), and the Ontology Lookup Service (Cote *et al.*, 2006), have existed for a number of years, and currently contain several hundred ontologies. They allow ontology creators to upload, manage and release their work to the wider community. For the end-user, they provide a web front-end for browsing, comparing, visualising, downloading, searching, and otherwise processing ontologies.

One feature that the existing ontology repositories lack is to utilize automated reasoning over ontologies to enrich the set of services they provide. Automated reasoning over the axioms in the ontologies enables the use of deductive inference when processing an ontology, which can improve the utility of many services provided by an ontology repository:

- *Verify Consistency and coherence* – The releases and updates of ontologies can be automatically classified by the repository, checking them for logical consistency and existence of unsatisfiable classes, and informing the ontology maintainers and the users of the result.
- *Semantic Query* – Searching can make use of inferred knowledge ‘created’ during the reasoning process. Furthermore, the results can be integrated in retrieval of text documents or data characterized with ontologies accessible through through public SPARQL endpoints (Jupp *et al.*, 2014; The Uniprot Consortium, 2007; Belleau *et al.*, 2008; Williams *et al.*, 2012).
- *Versioning* – Ontology versions can be compared semantically based on the inferred knowledge.
- *Reasoning API* – Applications building on the platform have access to classified data, and may build applications requiring ontology semantics and inferred knowledge without having to classify the ontology on the client-side.

However, enabling automated reasoning over multiple ontologies is a challenging task since automated reasoning can be highly complex and costly in terms of time and memory consumption (Tobies, 2000). In particular, ontologies formulated in the Web Ontology Language (OWL) (Grau *et al.*, 2008) can utilise statements based on highly expressive description logics (Horrocks *et al.*, 2000), and therefore queries that utilise automated reasoning cannot, in general, be guaranteed to finish in a reasonable amount of time.

Previous approaches to reasoning over a large set of ontologies have often involved working with existing collections of ontologies, usually from one of the large repositories such as Bioportal (Del Vescovo *et al.*, 2011; Sazonau *et al.*, 2013). Several approaches have employed RDFS reasoning (Patel-Schneider *et al.*, 2004) for answering queries over Bioportal’s set of ontologies through a SPARQL interface (Salvadores *et al.*, 2012, 2013). However, RDFS semantics is different from the semantics of OWL in which most of the ontologies are formalized. Alternatively, systems such as OntoQuery (Tudose *et al.*, 2013) provide access to ontologies through automated reasoning but limit the number of ontologies.

The AberOWL (Hoehndorf *et al.*, 2015) system is a novel ontology repository which allows access to multiple ontologies through automated reasoning, utilising the OWL semantics of the ontologies. AberOWL mitigates the complexity challenge by using a reasoner which supports only a subset of OWL (i.e., the OWL EL profile (Motik *et al.*, 2009)), ignoring ontology axioms and queries that do not fall within this subset. This enables the provision of polynomial-time reasoning, which is sufficiently fast for many practical uses even when applied to large ontologies.

In the demonstration, we will show AberOWL and its functionality, highlighting the differences between our system over traditional ontology repositories that do not utilize automated reasoning.

2 ABEROWL

AberOWL consists of an ontology repository, web services which facilitate semantic queries over specific ontologies or the entire set of ontologies contained in the repository, and a user interface. It aims to provide a full framework for interacting with ontologies through an automated reasoner.

2.1 AberOWL Server

The core of AberOWL is a server which handles the loading, classification of and interaction with ontologies, exposing its

---

*To whom correspondence should be addressed: luke.slater@kaust.edu.sa functionality through a JSON REST API. Specifically, the AberOWL server provides the possibility to query one or all ontologies using the Manchester OWL Syntax. It also supports classifying new ontologies and new versions of existing ontologies, immediately making these available through the API.

Additionally, the AberOWL server provides a searchable index of all classes available in its set of ontologies, including class labels, synonyms, description and other associated information. It utilises an Apache Lucene (Lucene, 2005) index to provide quick searching for classes over the entire set of ontologies. Search over the Lucene index is currently used to facilitate autocompletion in the web interface and provide information about classes in the ontology.

2.2 AberOWL Sync

Another component of AberOWL scans existing ontology repositories, such as BioPortal, and a set of URLs, for new ontologies and new versions of existing ontologies. When a new ontology or a new version of an ontology is found, a request is sent to the AberOWL server to load the new ontology or to retrieve and classify a new version of an existing ontology.

2.3 AberOWL Repository

The AberOWL repository is a web frontend to AberOWL which interacts with the AberOWL server to provide a user interface for querying ontologies, alongside features found in other ontology repositories such as ontology upload, browsing of ontologies, and downloading ontologies and their versions. New ontologies and new versions of ontologies can be uploaded, after which they are immediately classified by the AberOWL server and made available for querying.

The prime goal of the AberOWL repository is to have a user-interface in which all structural information about ontologies is derived in real time from the AberOWL reasoning server. Currently, the interface can be used to upload and classify new ontologies, browse and query ontologies, visualize ontology structure and retrieve information on ontology classes.

3 CONCLUSION AND FUTURE WORK

In the future, we plan to improve AberOWL's potential to retrieve a single class based on search of the class' metadata and axioms. Furthermore, we intend to provide the possibility for the visual comparison of ontology versions, utilizing automated reasoning to show differences in the inferences that can be drawn from both ontologies.

A second area of research is to increase AberOWL's integration with biomedical data. In particular, we aim to more closely integrate AberOWL with publicly available SPARQL endpoints to enable access to these SPARQL endpoints through automated reasoning over ontologies.

ACKNOWLEDGEMENTS

REFERENCES

Belleau, F., Nolin, M., Tourigny, N., Rigault, P., and Morissette, J. (2008). Bio2RDF: Towards a mashup to build bioinformatics knowledge systems. *Journal of Biomedical Informatics*, **41**(5), 706–716.

Cote, R., Jones, P., Apweiler, R., and Hermjakob, H. (2006). The ontology lookup service, a lightweight cross-platform tool for controlled vocabulary queries. *BMC Bioinformatics*, **7**(1), 97+.

Del Vescovo, C., Gessler, D. D., Klinov, P., Parsia, B., Sattler, U., Schneider, T., and Winget, A. (2011). Decomposition and modular structure of bioportal ontologies. In *The Semantic Web–ISWC 2011*, pages 130–145. Springer.

Grau, B., Horrocks, I., Motik, B., Parsia, B., Patelschneider, P., and Sattler, U. (2008). OWL 2: The next step for OWL. *Web Semantics: Science, Services and Agents on the World Wide Web*, **6**(4), 309–322.

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

Hoehndorf, R., Slater, L., Schofield, P. N., and Gkoutos, G. V. (2015). Aber-owl: a framework for ontology-based data access in biology. *BMC Bioinformatics*.

Horrocks, I., Sattler, U., and Tobies, S. (2000). Practical reasoning for very expressive description logics. *Logic Journal of the IGPL*, **8**(3), 239–264.

Jupp, S., Malone, J., Bolleman, J., Brandizi, M., Davies, M., Garcia, L., Gaulton, A., Gehant, S., Laibe, C., Redaschi, N., Wimalaratne, S. M., Martin, M., Le Novre, N., Parkinson, H., Birney, E., and Jenkinson, A. M. (2014). The EBI RDF platform: linked open data for the life sciences. *Bioinformatics*, **30**(9), 1338–1339.

Lucene, A. (2005). A high-performance, full-featured text search engine library. *URL: http://lucene. apache. org*.

Manola, F. and Miller, E., editors (2004). *RDF Primer*. W3C Recommendation. World Wide Web Consortium.

Motik, B., Grau, B. C., Horrocks, I., Wu, Z., Fokoue, A., and Lutz, C. (2009). Owl 2 web ontology language: Profiles. Recommendation, World Wide Web Consortium (W3C).

Noy, N. F., Shah, N. H., Whetzel, P. L., Dai, B., Dorf, M., Griffith, N., Jonquet, C., Rubin, D. L., Storey, M.-A. A., Chute, C. G., and Musen, M. A. (2009). Bioportal: ontologies and integrated data resources at the click of a mouse. *Nucleic acids research*, **37**(Web Server issue), W170–173.

Patel-Schneider, P. F., Hayes, P., and Horrocks, I. (2004). Owl web ontology language semantics and abstract syntax section 5. rdf-compatible model-theoretic semantics. Technical report, W3C.

Salvadores, M., Horridge, M., Alexander, P. R., Fergerson, R. W., Musen, M. A., and Noy, N. F. (2012). Using sparql to query bioportal ontologies and metadata. In *The Semantic Web–ISWC 2012*, pages 180–195. Springer.

Salvadores, M., Alexander, P. R., Musen, M. A., and Noy, N. F. (2013). Bioportal as a dataset of linked biomedical ontologies and terminologies in rdf. *Semantic web*, **4**(3), 277–284.

Sazonau, V., Sattler, U., and Brown, G. (2013). Predicting performance of owl reasoners: Locally or globally? Technical report, Technical report, School of Computer Science, University of Manchester.

Seaborne, A. and Prud'hommeaux, E. (2008). SPARQL query language for RDF. W3C recommendation, W3C. http://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/.

The Uniprot Consortium (2007). The universal protein resource (uniprot). *Nucleic Acids Res*, **35**(Database issue).

Tobies, S. (2000). The complexity of reasoning with cardinality restrictions and nominals in expressive description logics. *J. Artif. Int. Res.*, **12**(1), 199–217.

Tudose, I., Hastings, J., Muthukrishnan, V., Owen, G., Turner, S., Dekker, A., Kale, N., Ennis, M., and Steinbeck, C. (2013). Ontoquery: easy-to-use web-based owl querying. *Bioinformatics*, **29**(22), 2955–2957.

Williams, A. J., Harland, L., Groth, P., Pettifer, S., Chichester, C., Willighagen, E. L., Evelo, C. T., Blomberg, N., Ecker, G., Goble, C., and Mons, B. (2012). Open phacts: semantic interoperability for drug discovery. *Drug Discovery Today*, **17**(2122), 1188 – 1198.

Xiang, Z., Mungall, C. J., Ruttenberg, A., and He, Y. (2011). Ontobee: A linked data server and browser for ontology terms. In *Proceedings of International Conference on Biomedical Ontology*, pages 279–281.