Aber-OWL: a framework for ontology-based data access in biology

Robert Hoehndorf[1], Luke Slater[2], Paul N Schofied[3], and Georgios V Gkoutos[2]

[1] Computational Bioscience Research Center, King Abdullah University of Science and Technology, Thuwal 23955-6900, Kingdom of Saudi Arabia
`robert.hoehndorf@kaust.edu.sa`,
[2] Department of Computer Science, Aberystwyth University, Aberystwyth, SY23 3DB, UK,
`{lus11,geg18}@aber.ac.uk`,
[3] Department of Physiology, Development & Neuroscience, University of Cambridge, Downing Street, CB2 3EG, UK,
`pns12@hermes.cam.ac.uk`

Abstract. Many ontologies have been developed in biology and these ontologies increasingly contain large volumes of formalized knowledge commonly expressed in the Web Ontology Language (OWL). Computational access to the knowledge contained within these ontologies relies on the use of automated reasoning. We have developed the Aber-OWL infrastructure that provides reasoning services for bio-ontologies. Aber-OWL consists of an ontology repository, a set of web services and web interfaces that enable ontology-based semantic access to biological data and literature. Aber-OWL is freely available at http://aber-owl.net. Aber-OWL provides a framework for automatically accessing information that is annotated with ontologies or contains terms used to label classes in ontologies. When using Aber-OWL, access to ontologies and data annotated with them is not merely based on class names or identifiers but rather on the knowledge the ontologies contain and the inferences that can be drawn from it.

Keywords: ontology-based data access, ontology repository, semantic query

While ontology repositories, such as BioPortal (8) and the Ontology Lookup Service (OLS) (2), provide web services and interfaces to access ontologies, including their metadata such as author names and licensing, the list of classes and asserted structure, they do not enable computational access to the semantic content of the ontologies and the inferences that can be drawn from them. Access to the semantic content of ontologies usually requires further inferences to reveal the consequences of statements (axioms) asserted in an ontology; these consequences may be automatically derived using an automated reasoner. To the best of our knowledge, no reasoning infrastructure that supports semantically enabled access to biological and biomedical ontologies currently exists.

Here, we present Aber-OWL, a reasoning infrastructure over ontologies consisting of an ontology repository, web services that facilitate semantic queries over ontologies specified by a user or contained in Aber-OWL's repository, and a user interface. The Aber-OWL infrastructure can not only enable access to knowledge contained in ontologies, but crucially can also be used for semantic queries over data annotated with ontologies, including the large volumes of data that are increasingly becoming available through public SPARQL endpoints (6). Allowing access to data through an ontology is known as the "ontology-based data access" paradigm (1; 7), and can exploit formal information contained in ontologies to identify possible inconsistencies and incoherent descriptions (4), enrich possibly incomplete data with background knowledge so as to obtain more complete answers to a query (e.g., if a data item referring to an organism has been characterized with findings of pulmonary stenosis, overriding aorta, ventricular septal defect, and right ventricular hypertrophy, and the ontology – or the set of ontologies it imports – contains enough information to allow, based on these four findings, the inference of a Tetralogy of Fallot condition, then the data item can be returned when querying for Tetralogy of Fallot even in the absence of it being explicitly declared in database) (5; 1), enrich the data schema used to query data sources with additional information (e.g., by using a class in a query that is an inferred super-class of one or more classes that are used to annotate data items, but the class itself is never used to characterize data) (1), and provide a uniform view over multiple data sources with possibly heterogeneous, multi-modal data (1; 7).

To demonstrate how Aber-OWL can be used for ontology-based access to data, we provide a service that performs a semantic search over Pubmed and Pubmed Central articles using the results of an Aber-OWL query, and a service that performs SPARQL query extension so that the results of Aber-OWL queries can be used to retrieve data accessible through public SPARQL endpoints. In Aber-OWL, following the ontology-based data access paradigm (7; 1), we specify the features of the relevant information on the ontology- and knowledge level (3), and retrieve named classes in ontologies satisfying these condition using an automated reasoner, i.e., a software program that can identify whether a class in an ontology satisfies certain conditions based on the axioms specified in an ontology. Subsequently, we embed the resulting information in database, Linked Data or literature queries.

Aber-OWL can be accessed at http://aber-owl.net. The Aber-OWL software is freely available at https://github.com/reality/SparqOWL can be installed locally by users who wish to provide semantic access to their own ontologies and support the use of their ontologies in semantic queries.

Bibliography

[1] Bienvenu, M., ten Cate, B., Lutz, C., Wolter, F.: Ontology-based data access: a study through disjunctive datalog, csp, and mmsnp. In: PODS. pp. 213–224 (2013)

[2] Cote, R., Jones, P., Apweiler, R., Hermjakob, H.: The ontology lookup service, a lightweight cross-platform tool for controlled vocabulary queries. BMC Bioinformatics 7(1), 97+ (2006), http://dx.doi.org/10.1186/1471-2105-7-97

[3] Guarino, N.: The ontological level. In: Casati, R., Smith, B., White, G. (eds.) Philosophy and the Cognitive Sciences, pp. 443–456. Hölder-Pichler-Tempsky, Vienna (1994)

[4] Hoehndorf, R., Dumontier, M., Oellrich, A., Rebholz-Schuhmann, D., Schofield, P.N., Gkoutos, G.V.: Interoperability between biomedical ontologies through relation expansion, upper-level ontologies and automatic reasoning. PLOS ONE 6(7), e22006 (July 2011)

[5] Hoehndorf, R., Schofield, P.N., Gkoutos, G.V.: Phenomenet: a whole-phenome approach to disease gene discovery. Nucleic Acids Res 39(18), e119 (2011), dOI:10.1093/nar/gkr538

[6] 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., Jenkinson, A.M.: The EBI RDF platform: linked open data for the life sciences. Bioinformatics 30(9), 1338–1339 (2014), http://bioinformatics.oxfordjournals.org/content/30/9/1338.abstract

[7] Kontchakov, R., Lutz, C., Toman, D., Wolter, F., Zakharyaschev, M.: The combined approach to ontology-based data access. In: IJCAI. pp. 2656–2661 (2011)

[8] 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., Musen, M.A.: Bioportal: ontologies and integrated data resources at the click of a mouse. Nucleic acids research 37(Web Server issue), W170–173 (July 2009), http://dx.doi.org/10.1093/nar/gkp440