

Ontodog: A Web-based Ontology Community View Generator

Jie Zheng^{1*}, Zuoshuang Xiang^{2*}, Christian J. Stoeckert Jr¹, Yongqun He²

¹ Department of Genetics, University of Pennsylvania, Philadelphia, PA 19104, USA

² Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI 48109, USA

ABSTRACT

Reference ontologies are often very large and complex. When applied to a specific application, generally a subset of one reference ontology is needed. Moreover, the labels of ontology terms that were given in the perspective of ontology developers might not be preferred labels to the end users. Therefore, it is desirable to have a community view of a reference ontology that is a subset of the ontology including the terms needed for a particular application or community with user-preferred labels. Ontodog is a web-based system to support generation of ontology community views. Ontodog (<http://ontodog.hegroup.org/>) allows users to provide terms of interest in a source ontology and customized annotation information, such as user-preferred label. With these inputs, Ontodog can extract a subset of the source ontology containing all the terms of interest and generate user specified annotations in RDF/XML format (i.e., OWL files) which can be used to build an ontology community view. Currently over 100 ontologies including all OBO Foundry ontologies are available in Ontodog to generate views for a specific application or community. We demonstrate the application of Ontodog in generating ontology community views using the Ontology for Biomedical Investigations (OBI) as the source ontology.

1 INTRODUCTION

Biomedical ontologies have been widely used in various applications that facilitate biomedical data integration and sharing. With large reference ontologies, use of the entire ontology can impact the efficiency of an application (e.g. reasoning, semantic similarity computation) when only a portion of the ontology is needed. Reference ontologies may be developed collaboratively by various communities aiming to annotate data consistently regardless of specific research areas or technologies. For example, the Ontology for Biomedical Investigations (OBI) has been developed by more than 20 communities^[1]. The term labels were agreed upon by consensus and generally chosen for ontological clarity. Consequently, they may not be the preferred labels for a particular community or application users. For ontology users, it is desirable to have their own preferred labels in a customized view. We developed Ontodog, a web-based tool, to address these needs and generate ontology community views automatically without need for programming skills and requiring minimal ontology knowledge.

2 FEATURES AND USAGE

The Ontodog processes input data using PHP, issues SPARQL queries against an RDF triple store, e.g., the default SPARQL endpoint hosted by the He group (<http://sparql.hegroup.org/sparql>), to validate whether terms exist in the source ontology or retrieve terms from the

source ontology, and reformat the output files using OWL-API. Then the RDF/XML format output files are provided to the users for download.

Ontodog provides a user-friendly web form for data inputs. It contains two sections. The first section collects source ontology information and tagged terms of interest. The tagged terms of interest and user-specified annotations for the selected terms are provided either in a tab-delimited or Excel format file. The second section has three parts corresponding to three ontology output files. Output file 1 includes all tagged terms annotated with a user-specified annotation property to indicate they are community subset. Output file 2 has customized annotations for the terms, generally used for adding user-specified labels. Output file 3 is the subset of the source ontology including all terms tagged in the input term file and related terms/axioms to support proper reasoning. OntoFox SPARQL related term retrieval approach^[2] is adopted for ontology subset extraction.

Ontodog allows users to generate all or any combination of the ontology output files described above based on their needs. The OWL import mechanism is used to build different ontology community views using Ontodog output files.

3 USE CASE: APPLICATION OF ONTODOG IN OBI COMMUNITY VIEWS GENERATION

The Functional Genomics Data (FGED) Society (<http://www.fged.org/>) supported the migration of the MGED Ontology to OBI for functional genomics data annotation. Criteria for this move included a simplified set of terms with (molecular) biologist-community friendly labels. Ontodog was used to generate the OBI FGED view (<http://biportal.bioontology.org/ontologies/1123>) for this purpose. 2279 classes identified as relevant for this community were extracted from OBI (3501 classes).

ACKNOWLEDGEMENTS

This research is supported by NIH grants R01AI081062, P41HG003619, and R01GM93132.

REFERENCES

1. Brinkman RR, *et al* (2010) Modeling biomedical experimental processes with OBI. *J. Biomed. Semantics*. **1**(Suppl. 1), S7.
2. Xiang Z, *et al* (2010) OntoFox: web-based support for ontology reuse. *BMC Res Notes*. **3**:175.

* Contribute equivalently