

OntoFox and its Application in the Development of the Brucellosis Ontology

Zuoshuang Xiang, Yu Lin, Yongqun He

University of Michigan, Ann Arbor, USA

Abstract. OntoFox (<http://ontofox.hegroup.org/>) is a web-based system to support ontology reuse. OntoFox allows users to input terms, fetch selected properties, annotations, and certain classes of related terms from source ontologies, and save the results using the RDF/XML serialization of the OWL. Currently >80 existing ontologies are available in OntoFox for direct term extraction and importing. OntoFox is an efficient method that promotes ontology sharing and interoperability. As a demonstration, OntoFox has recently been applied for the development of Brucellosis Ontology (BO).

1 Introduction

To avoid duplication of effort during ontology development, it is advised to import pre-existing ontology terms and knowledge into a new ontology if possible. The Web Ontology Language (OWL; <http://www.w3.org/TR/owl-syntax/>) provides a mechanism to import ontologies. This approach leads to import of the whole ontology. However, importing a whole ontology may not be practical and needed, especially when the source ontology is very large and the majority of the terms in the source ontology are not relevant to the new ontology.

To address flexibility of ontology reuse, we have developed OntoFox (<http://ontofox.hegroup.org/>) [1], a web-based application for retrieving ontology term information and importing it into a target ontology. OntoFox follows and expands the principle of MIREOT (Minimum information to reference an external ontology term) [2]. Inspired by existing ontology modularization techniques, OntoFox also develops a new SPARQL-based ontology term extraction algorithm that extracts terms related to a given set of signature terms. In addition, OntoFox provides an option to extract all terms and annotations in an ontology hierarchy rooted at a specified ontology term [1].

As a demonstration, here we show how OntoFox is used to facilitate the development of the Brucellosis Ontology (BO) [3]. BO is a new community-based ontology in the domain

of brucellosis, a zoonotic disease caused by infection of an intracellular Gram-negative bacterium *Brucella* in human and animals. BO is an extension ontology of the Infectious Disease Ontology (IDO) [4]. The development of BO requires a large number of terms from existing ontologies.

2 Features and Usage

The OntoFox input data is processed using PHP, and the processed input data is used in SPARQL queries against an RDF triple store, e.g., the Neurocommons SPARQL endpoint (<http://neurocommons.org>). The returned results are then parsed and converted into an OntoFox output file.

OntoFox provides a user-friendly web form for data input. Alternatively, a user can generate an OntoFox input text file and upload it to OntoFox. A detailed description of OntoFox input file format is available through the OntoFox online tutorial: <http://ontofox.hegroup.org/tutorial>). Briefly, the steps using OntoFox web form to generate OntoFox input include:

1. **Select source ontology.** A dropdown menu can be used to select one ontology out of the available ontologies in OntoFox. If your favorable ontology is not available in OntoFox, a SPARQL endpoint and a graph URI points to the ontology can be specified instead.

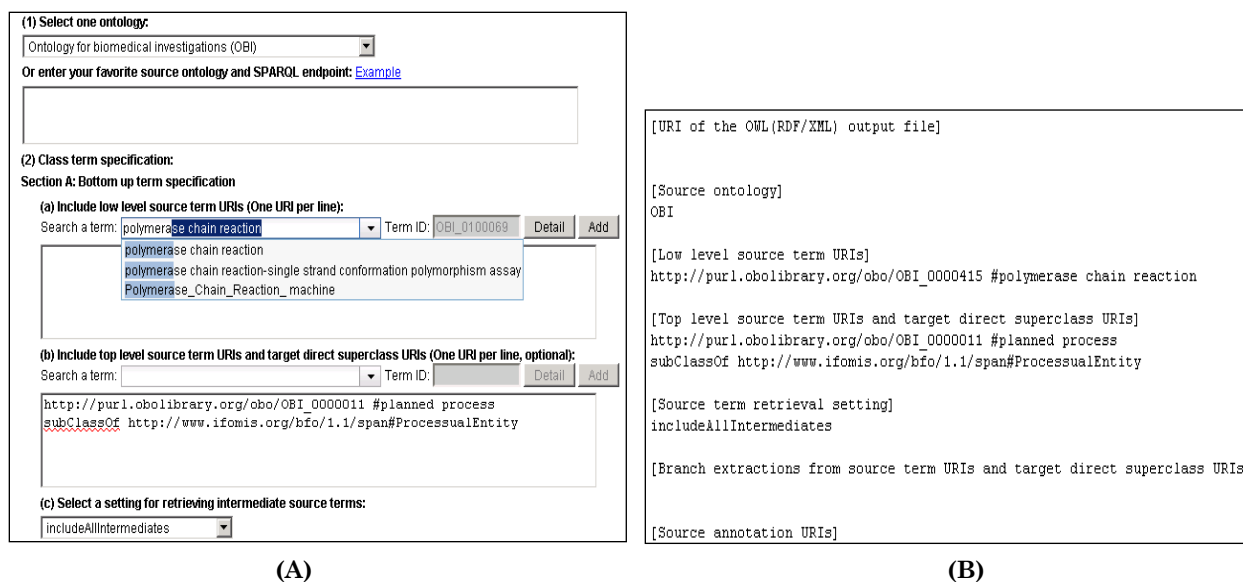


Figure 1. OntoFox retrieval of ontology information of 'polymerase chain reaction' from OBI.

(A) A screenshot of how to provide OntoFox input data;

(B) OntoFox input text file that matches the input in (A) and can be automatically generated by the server.

2. **Class term specification.** This can be done in two formats:

a) Bottom up term specification. To implement this, three sets of information are needed: Low level source term URIs, top level source term URIs and target direct superclass URIs, and setting for retrieving intermediate source terms [1].

b) Top-down branch term specification. This feature is designed to extract all terms of an ontology hierarchy branch under a specific ontology term.

3. **Annotation/Axiom Specification.** This allows a user to specify which annotations and axioms to be included in information retrieval.

4. **Specification of the URI of the final OWL (RDF/XML) output file.** The specified URI will be automatically added to the OntoFox output file. If the target ontology includes the same URI information, no additional edition is required for the import of the OntoFox output results into the target ontology.

As an example, we show how to extract the term 'polymerase chain reaction' from OBI into BO (Fig. 1). BO requires the term 'polymerase chain reaction', its relevant

restrictions, and its direct superclass term and top class terms from OBI. After a user types the first few characters of a term, OntoFox automatically provides a list of matching terms for selection, which facilitates the term inputting process (Fig. 1A). After all requirements are specified, an OntoFox input text file will be generated by OntoFox (Fig. 1B).

Under the platform of the Protégé ontology editor (<http://protege.stanford.edu/>), an OntoFox output OWL file can be directly visualized and imported in the target ontology (e.g., BO) using the OWL import function. This approach allows efficient import of minimum information of external ontology terms into target ontology.

In the past year, OntoFox has been used over 2,000 times by more than 700 unique users from 51 countries. We have received many requests and questions during the time. Our software demonstration will detail how the newly developed OntoFox program works and answer some common questions raised by OntoFox users.

3 Use Case: Application of OntoFox in BO Development

Currently BO imports 232 individual terms from eight existing ontologies using OntoFox.

These eight ontologies include: Chemical Entities of Biological Interest (ChEBI), Gene Ontology (GO), Information Artifact Ontology (IAO), NCBI_Taxon, OBI, Ontology for General Medical Science (OGMS), Protein Ontology, and Vaccine Ontology. All OntoFox input files for these ontologies exist on the BO Sourceforge website:

<http://bo-ontology.svn.sourceforge.net/viewvc/bo-ontology/trunk/src/ontology/imports/>.

It is noted that the input files for GO, IAO, and OGMS are not listed in this website. It is because those terms partially imported to BO come from IDO, which is fully imported to BO as its immediate upper level ontology. The ontology term 'polymerase chain reaction' (Fig. 1) is one of the OBI terms imported to BO using OntoFox. The OntoFox output OWL files are located in the parent directory of the above folder.

It is extremely difficult to partially import these terms to BO without such a tool as OntoFox. To our knowledge, the only other tool for such task is MIREOT [2]. However, the usage of MIREOT requires command line programming and does not include a user-friendly interface.

4 Summary

OntoFox is a new web server that automatically extracts external ontology terms and their annotations for efficient ontology development. OntoFox is easy to use without a prior knowledge of SPARQL query and command line programming. OntoFox provides an efficient approach to facilitate ontology sharing and interoperability.

Acknowledgements

This research is supported by a NIH R01 grant (1R01AI081062).

References

1. Xiang Z, Courtot M, Brinkman RR, Ruttenberg A, He Y. OntoFox: web-based support for ontology reuse. *BMC Research Notes*. 2010, 3:175.
2. Courtot M, Gibson F, Lister AL, Malone J, Schober D, Brinkman RR, Ruttenberg A. 2011. MIREOT: the Minimum Information to Reference an External Ontology Term. *Applied Ontology*. 6 (1): 23-33.
3. Brucellosis Ontology (BO). Available at: <http://sourceforge.net/projects/bo-ontology>.
4. Infectious Disease Ontology (IDO). Available at: <http://infectiousdiseaseontology.org>.