

HINO: BFO-aligned ontology representation of human molecular interactions and pathways

Zoushuang Xiang, Yongqun He

Department of Microbiology and Immunology, Unit for Laboratory Animal Medicine, and Center for Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI 48109, USA

ABSTRACT

Reactome is a manually-curated, peer-reviewed knowledgebase of human pathways and processes. As a standard pathway data exchange format, BioPAX represents individual interactors (e.g., a protein), interactions, and pathways as instances instead of classes. This study aims to represent various human interaction pathways and networks as classes via a formal ontology aligned with the Basic Formal Ontology (BFO). Towards this goal, the Human Interaction Network Ontology (HINO) was generated by extending the BFO-aligned Interaction Network Ontology (INO). All human pathways and associated processes and interactors listed in the Reactome database were first converted to ontology classes by aligning them within the framework of the INO ontology. Related terms and associated relations and hierarchies from external ontologies (e.g., CHEBI and GO) were retrieved and imported into HINO. Resulting HINO OWL file was transferred to a RDF triple store. Detailed information about individual HINO ontology terms can be displayed using the linked ontology data server Ontobee. The RDF triples stored in the RDF triple store are queryable through a SPARQL program. Such an ontology system supports advanced pathway data integration and applications.

INTRODUCTION

Hundreds of biological interaction and network pathways exist. The pathway data in Reactome are stored in a relational database and can be downloaded in BioPAX, an OWL-based biological pathway data exchange format. The BioPAX ontology format represents basic high level pathways and associated interactions and entities participating in the interactions and pathways. Individual Reactome interactions and pathways are then represented as instances of BioPAX ontology classes. While the approach of listing individual pathways as instances in OWL is reasonable for data exchange, different levels of pathways (e.g., apoptosis pathways), reactions (e.g., phosphorylation), and entities (e.g., protein) are universals (or classes) instead of particulars (or instances). Given appropriate conditions, these pathways always occur. Therefore, they are more appropriately represented as classes instead of instances in the scope of ontology. The Interaction Network Ontology (INO) is a Basic Formal Ontology (BFO)-aligned ontology that targets on the representation of interactions and networks. In this study, we report our development of the Human Interaction Network Ontology (HINO) as an extension of INO.

METHODS AND RESULTS

Our strategy is to use existing ontologies and ontology terms to represent entities in Reactome whenever possible.

To whom correspondence should be addressed: yongqunh@umich.edu

This strategy led to the generation of many ontology terms in INO particularly from the NCBI_Taxon, CHEBI, PRO, and GO. New ontology terms forming the hierarchies were then imported to HINO. INO serves as a core that provides many superclasses for terms in HINO.

Most terms in HINO come from the conversion of the Reactome human pathways from its BioPAX format. The Reactome BioPAX uses its own IDs to represent the GO terms. In HINO, the GO ontology term URIs and annotations are directly imported from GO. In addition, the original hierarchy of imported GO terms is also retrieved and imported. Similarly, HINO also imports many CHEBI terms and their relational hierarchy for representation of small molecules used in HINO. The use of original ontology term annotations and term hierarchy in HINO for representation of processes, small molecules, cellular components, and other related elements in human interactions and pathways allows clear visualization and analysis of specific human interaction pathways.

As of May 15, 2013, HINO includes 38,435 classes, 57 object properties, and 4 datatype properties. Hgroup RDF triple store (<http://sparql.hegroup.org/>) was used to store the RDF triple store. This triple store was generating using the Virtuoso RDF triple store management system. The Virtuoso server was used to read the HINO OWL file and automatically transfer the HINO OWL file to RDF triples. A HINO SPARQL program was generated in the INO website (<http://www.ino-ontology.org>). SPARQL based methods were used for further HINO data query and analysis. The HINO is openly available in Ontobee: <http://www.ontobee.org/browser/index.php?o=HINO>. It is also available in NCBO Bioportal. The source code is also freely available under the Apache License 2.0.

DISCUSSION

Different databases use different formats to represent the pathway contents. Our approach of representing pathway data using BFO-aligned ontology and incorporating as many reliable ontologies as possible provides a feasible way to eventually integrate different interaction pathways.

ACKNOWLEDGEMENTS

This project was supported by a NIH-NIAID grant (R01AI081062).