ENRICHMENT OF OBO ONTOLOGIES

Michael Bada, Ph.D. and Lawrence Hunter, Ph.D.
University of Colorado at Denver and Health Sciences Center, Aurora, CO, USA

An enormous effort has gone into the creation and maintenance of the Open Biomedical Ontologies (OBO) project, and this has been met with the ever-increasing use of the constituent ontologies by biological researchers. There are currently over 50 OBO ontologies, ranging over such domains as anatomy, behavior, phenotype, experiment, and sequence. The flagship OBO ontology, the Gene Ontology (GO), with its three subontologies detailing molecular functions, biological processes, and cellular components, in particular has experienced phenomenal growth in terms of numbers of terms and also of its extensive use by annotators to describe gene and gene-product entries in a number of prominent model-organism databases.

Although many of these ontologies are large, they are structurally quite simple, typically consisting of only a few relationship types apart from the fundamental is-a relationship that forms the backbone of an ontology. Furthermore, there are no links between terms from separate ontologies, even between terms from different GO subontologies. However, most concepts do have relationships with other concepts, as evidenced by their natural-language names and definitions, and computational agents cannot take advantage of these relationships if they are not formally represented. The work presented in this demo expands on previous efforts that have taken advantage of the compositionality of GO terms to produce more formal definitions of GO terms in the form of added relationships with other ontological terms.

In this demo will be shown a frame-based integration of the three GO subontologies, the Chemicals of Biological Interest ontology (ChEBI), and the Cell Type Ontology in which relationships between elements of the ontologies are modeled in a way that better captures the relational semantics between biological concepts represented by the terms, rather than between the terms themselves, than previous frame-based efforts. A methodology for creating suggested enriching assertions of the form [subject, relationship, object] by identifying patterns in GO terms, correlating these patterns and subpatterns with relationships, matching concepts to these patterns and subpatterns, and integrating these assertions into the ontologies will be described. Using this methodology, a large number of new, reliable assertions linking ontological terms using a variety of specific, hierarchically arranged relationship were created: A predicted assertion was made for 62% of GO terms that matched one of 31 patterns, and 97% of these predicted assertions were assessed to be valid; a further 429 assertions (corresponding to 6% of the matching terms) were manually created, resulting in an initial set of 4,497 assertions. Furthermore, this methodology programmatically integrates assertions into the base ontology such that each assertion is fully consistent with respect to higher (*i.e.*, more general) relevant class and slot levels.