Automated Enrichment of BioOntologies

C. Pesquita [†], F. Couto, M. Silva

XLDB, Lasige, Department of Informatics, University of Lisbon, Portugal

[†] E-mail: *cpesquita@xldb.di.fc.ul.pt*

BioOntologies provide a commonly accepted scheme for the characterization of biological concepts that enable knowledge sharing and integration. Updating and maintaining a BioOntology requires experts and is very time-consuming given the amount of literature the experts have to analyze and the difficulty in reaching consensus about how to model that portion of biological knowledge.

We propose the development of automated processes for ontology enrichment that are capable not only of extracting domain knowledge but also of integrating it into the ontology. This will be tested on the Gene Ontology (GO), which is actually one of the most widely used and dynamic BioOntologies.

Given that most biological knowledge is stored in the unstructured text of scientific publications, text mining techniques will be employed to find novel GO terms and relationships. Also, we will explore the possibility of retrieving terms from other related ontologies through alignment techniques. The proposed work is composed of three steps: 1) identifying ontology areas that need to be enriched (i.e., areas in the GO that do not provide sufficient detail to annotators); 2) retrieving (from text, or via alignment with other ontologies) candidate terms to enrich that area of the ontology; 3) where needed, organize the candidate terms in a hierarchy according to GO's relationship types. With this work we hope to provide biomedical researchers with an extended version of GO that can be used "as is" or by GO developers as a starting point to enrich GO.