

## AN ONTOLOGY GENERATION PLUGIN FOR OBO-Edit

**Wächter T. and M. Schroeder**

<sup>1</sup>Technische Universität Dresden Biotechnologisches Zentrum, Tatzberg 47-51, 01307 Dresden, Germany, Telephone: +49 (351) 463 40068

**Introduction:** Developing ontologies is a labor intensive process. Comfortable editors such as OBO-Edit, which is developed and maintained by the Gene Ontology Consortium, are a prerequisite for a good ontology design.

**Results:** To speed up the manual process we developed an OBO-Edit plug-in for semi-automated ontology generation. The ontology generation suggests terms, definitions and parent-child relationships based on text mining and natural language processing (NLP) techniques. State of the art NLP is used to rank the relevance of terms for the domain to be modeled. Previous studies showed that term generation can improve the completeness of an ontology by suggesting up to 89% good candidate terms in the top 50 ranked terms. Public resources such as Wikipedia, full text articles and web sites are incorporated to generate definitions, which follow the well defined structure "A is a B with property C" if available. Based on the generation of good definitions it is possible to suggest the likely parent of A, namely B.

**Methods:** In a three step procedure the plug-in supports the creation of new ontology terms from text. As textual resources free text or a query for PubMed abstracts can be submitted. In a first step the terminology mentioned in the text is retrieved and ranked according to its importance to the domain. Abbreviations and lexical variants are recognized and terms similar to existing OBO terms are indicated as displayed. The list of candidate terms can be searched and filtered with regular expression patterns to focus on specific lexical aspects of interest. In a second step definitions for terms are generated and presented to the curator. The defined candidate terms, which are enriched with synonyms are in the final step inserted into the ontology. We addressed the difficulty of finding the correct position in a tree structure by using all information available to the plug-in. All potential parent terms e.g. all terms of the Gene Ontology, are displayed as list and are ranked higher, if they are (a) selected in OBOEdit, (b) have a certain lexical overlap with the new candidate concept, (c) are contained in the specified definition, or (d) evidence for an relationship could be found in any of the OBO listed ontologies.

With the novel Ontology Generation plug-in for OBO-Edit 2, we contribute to the community by increasing the tool support for the development and maintenance of biomedical ontologies.

**Availability:** The plug-in is available as part of the beta53 release of OBO-Edit.