
Bachelor Thesis

Automating the Assignment of SBO-Terms

Elisabeth Fritze

Reviewer: Jun.-Prof. Dr. Andreas Dräger

Supervisor: Alina Renz

1 Background

The **Systems Biology Ontology (SBO)** is a fixed set of terms used to describe the field of systems biology, especially the construction of computational models. SBO-terms are often used to annotate a mathematical model, providing supplementary information about model entities, such as metabolites or reactions. The ontology has a tree-like structure with seven main branches, describing different aspects of the field: *physical entity representation*, *occurring entity representation*, *mathematical expression*, *metadata representation*, *modelling framework*, *participant role*, and *systems description parameter*. An excerpt from the tree is seen in the figure below. All nodes are linked to their parent by an *is_a* relationship, meaning all child nodes are also instances of their parent term. [2]

At this point, the addition of general, top-level SBO-terms to a model can be done automatically. However, the addition of precise descriptions of, e.g., biochemical reactions, such as redox reactions or ionisations, needs to be done manually: All terms have to be determined and added individually to each of their occurrences. This means that a laborious lookup in relevant databases is necessary. Hence, automation of this task would be desirable.

2 Aim

This thesis aims to automate the assignment of SBO-terms within an SBML-model. The focus will be set to the ‘*entity*’ (SBO:0000236, physical entity representation) and ‘*interaction*’ (SBO:0000321, occurring entity representation) branches of the ontology. The task will be solved using an iterative approach, by starting in the upper levels of the tree and then gradually adding more complexity to the automation. Integration into an existing framework will be considered.

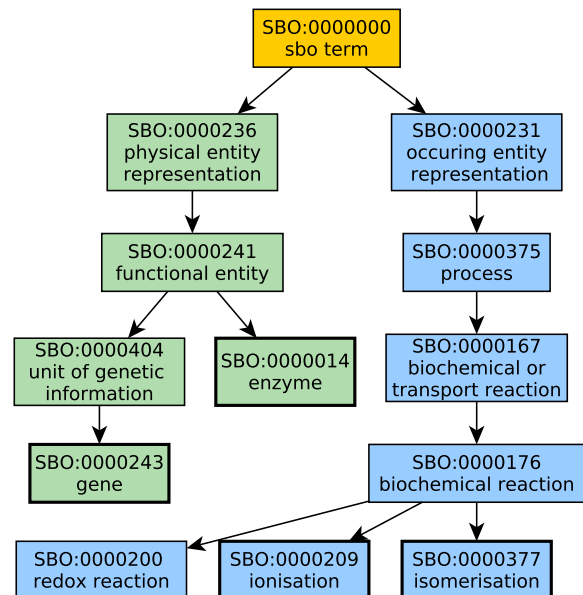


Figure 1: Excerpt from the tree of SBO-terms. Terms marked with a bold border are leaf nodes of the tree.

3 Requirements

Experience in Python and appropriate libraries, such as libSBML [1], for constraint-based reconstruction and analysis of *in-silico* models. Fundamental understanding of (bio)chemistry. Interest in the field of systems biology, especially its ontology.

References

- [1] BORNSTEIN, Keating S. M. Jouraku A. B. J. J. B. J. ; HUCKA, M.: LibSBML: An API Library for SBML. 24 (2008), S. 880–1. <http://dx.doi.org/10.1093/bioinformatics/btn051>. – DOI 10.1093/bioinformatics/btn051
- [2] COURTOT M, Knüpfer C et a. Juty N N. Juty N: Controlled vocabularies and semantics in systems biology. (2011), Oct, S. 543. <http://dx.doi.org/10.1038/msb.2011.77>. – DOI 10.1038/msb.2011.77