

Grounding ontologies in literature

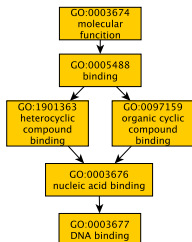
Computational Bioscience Research Center, KAUST

Maxat Kulmanov and Robert Hoehndorf

Introduction and Motivation

- Ontologies are widely used in biology and biomedicine
- AberOWL (<http://aber-owl.net>) - has more than 600 hundred ontologies
- Learning meaning of ontology classes in text will be useful to
 - find novel axioms
 - generate annotations

Identify classes in text



Pax genes have been cloned on the basis of their homology to the Drosophila segmentation gene paired. They share a common domain, the paired domain, that is sufficient to mediate sequence-specific DNA binding. Thus far, nine members have been characterized, which exhibit highly restricted temporal and spatial expression patterns. The analysis of mouse mutants has revealed their crucial role in the formation of a variety of tissues. In particular, they are involved in the regulation of early steps in organ development. They act to define the regional specification of distinct germ layers.

In order to learn meanings of ontology classes from literature we need

- Named Entity Recognition methods
- Semantics from ontologies



Work Plan

- Identify classes in text
 - Use concept recognition method (possibly by BLAH participant)
 - Use the context of annotated entities
- Extend Word2Vec
 - Normalize labels/synonyms into concepts
 - Propagate occurrences over ontology structure
- Add annotations of AberOWL ontology classes to LinkedAnnotation project
- Add embeddings to Bio2Vec project

Thank you!
