ReBIL : Relating Biological Information through Literature

50-word abstract:

ReBIL aims to improve the efficiency of information extraction systems applied to biological literature. The project is based on the correlation between structural and functional classifications of gene products. We evaluate extracted information by checking if they preserve the correlation. More information about Rebil is available at http://xldb.fc.ul.pt/rebil/.

One Page Abstract:

ReBIL aims to improve the efficiency of information extraction systems applied to biological literature, using the correlation between structural and functional classifications of gene products. The dogma of molecular biology that sequences should be correlated with their biological activity supports our approach. We developed a new method that evaluates extracted information by checking if gene products from a common family match a common set of biological properties. To evaluate the method, we developed an information extraction system that automatically annotates carbohydrate-active enzymes issued from CAZy (whose public interface available at http://afmb.cnrs-mrs.fr/CAZY) with functional properties extracted from literature. CAZy attributes each carbohydrate-active enzyme to one or more families of catalytic and carbohydrate-binding modules according to its modular structure. A biological ontology (GO) structures the functional properties as a graph. To compute the relatedness between functional properties, we implemented a semantic similarity measure in GO. So far, we measured a correlation between the modular structures and functional properties of annotations automatically extracted from literature. This result shows that our method is a viable approach for automatic validation of extracted biological information. More information about Rebil project is available at http://xldb.fc.ul.pt/rebil/.