

The Biomedical Ontology Applications (BOA) framework

Bruno Tavares, Hugo P. Bastos, Daniel Faria, João D. Ferreira,
Tiago Grego, Catia Pesquita, Francisco M. Couto

Faculty of Sciences, University of Lisboa, Portugal
{btavares,joao.ferreira,cpesquita}@xldb.di.fc.ul.pt

Abstract. The Biomedical Ontology Applications (BOA) framework consists of a set of web applications that aim at an effective exploration of biological information and knowledge discovery using Biomedical Ontologies. This paper presents three web applications: ProteInOn for semantic similarity and protein set characterization based on Gene Ontology (GO); CMPSim for semantic similarity of chemical compounds and metabolic pathways using the Chemical Entities of Biological Interest ontology (ChEBI); and GRYFUN for the visualization, filtering and analysis of GO functional annotation profiles of a given protein family. The web tools and their documentation, including videos, are freely available at <http://xldb.di.fc.ul.pt/wiki/BOA>

Keywords: Biomedical Ontologies, Semantic Similarity, Web Applications

1. Introduction

The Biomedical Ontology Applications (BOA) framework integrates biomedical ontologies and databases with a number of algorithms and visualization tools, focusing on ontology-based semantic similarity. The BOA web tools, ProteInOn, CMPSim and GRYFUN, are interconnected, and currently support functional analyses that encompass chemical compounds, gene products and metabolic pathways.

2. BOA Web Tools

ProteInOn (Protein Interactions & Ontology) integrates data from the Gene Ontology (GO) [1], GOA, IntAct [2] and UniProt [3] databases. It provides eight GO-based semantic similarity measures [4,5] that can be applied to proteins or GO terms across the three GO aspects. It also supports the identification of the most meaningful functional annotations of a given set of proteins, based on the probability of annotation of each GO term, and the retrieval of interacting proteins. The integration

of these features supports complex analysis such as candidate gene identification [6].

CMPSim (Chemical and Metabolic Pathway Similarity) uses the ChEBI ontology (Chemical Entities of Biological Interest) [7] to calculate semantic similarity between chemical compounds and between pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG) [8] by comparing the sets of ChEBI terms associated with them. The integration of ChEBI-based semantic similarity has been shown to improve existing chemical compound classification systems [9]. CMPSim is also able to find the most meaningful chemical compound classes in a set of pathways, highlighting their common properties.

GRYFUN (GRaph anaLYzer of FUNctional annotation) supports the visualization of functional profiles of protein sets. These profiles are structured as subgraphs of GO, showing the terms annotated to the protein set and allowing a global view of its functional broadness and specificity. They also provide detailed statistics and relevant metrics by integrating data from GO and UniProt, and support the selection of protein subsets sharing a specific profile. GRYFUN is currently restricted to a dataset of carbohydrate-active enzymes.

Acknowledgments

This work was partially supported by the Fundação para a Ciência e Tecnologia through the Multiannual Funding Programme and the grants ref. SFRH/BD/29797/2006, SFRH/BD/36015/2007, SFRH/BD/42481/2007, SFRH/BD/48035/2008 and SFRH/BD/69345/2010. It was also supported by the European Commission through the EPIWORK project under the FP7 (Grant #231807).

References

1. The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. *Nature Genetics*. 25, 2010, pp. 25–29.
2. Aranda, B., Achuthan, P., Alam-Faruque, Y., et al.: The IntAct molecular interaction database. *Nucleic Acids Research*. 2010, Vol. 38, pp. D525-D531, (2010).
3. The UniProt Consortium. The Universal Protein Resource (UniProt). *Nucleic Acids Research*. 2010, vol. 38, pp. D142-D148, (2010).
4. Pesquita, C., Faria, D., Bastos, H., Ferreira, A., Falcao, A., Couto, F.: Metrics for GO based protein semantic similarity: a systematic evaluation. *BMC bioinformatics*, 9(S5), S4, (2008).
5. Pesquita, C., Faria, D., Falcão, A. O., Lord, P., Couto, F. M.: Semantic similarity in biomedical ontologies. *PLoS Computational Biology*, 5(7), (2009).
6. Bastos, HP, Tavares B, Pesquita C, Faria D, Couto FM: Application of Gene Ontology to Gene Identification In Silico Tools for Gene Discovery. Springer, (2011).
7. Degtyarenko, K., Matos, P., Ennis, M., et al. ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic acids research*. (2007).
8. Kanehisa, M., Goto, S., Hattori, M., et al.: From genomics to chemical genomics: new developments in KEGG. *Nucleic acids research*. 2006, Vol. 34, p. D354, (2006).
9. Ferreira, J. D., Couto, F. M.: Semantic Similarity for Automatic Classification of Chemical Compounds. *PLoS Computational Biology*, 6(9), (2010).