CENTRE OF BIOLOGICAL ENGINEERING
UNIVERSIDADE DO MINHO

I B B
INSTITUTE FOR BIOTECHNOLOGY AND BIOENGINEERING

Cross-cutting computational strategies to genome-scale modelling

Anália Lourenço, Sónia Carneiro, Isabel Rocha, Eugénio C. Ferreira

IBB-Institute for Biotechnology and Bioengineering, Centre of Biological Engineering, Universidade do Minho, Campus de Gualtar, 4710-057 Braga, Portugal

Hereby, the aim is to present some of our research efforts towards the reconstruction of genome-scale models. Namely, we focus on the development of cross-cutting computational strategies for the integration and validation of heterogeneous data in support to traditional manual curation and, describe application scenarios on the model organism *E. coli*.

We address the systematic comparison of database contents and the harvest and extraction of contents from scientific literature. Aiming to help researchers assess the gains and losses to be accounted for in biological repositories and thus, choose the most content-bearing repositories for each particular integration problem/domain, we have implemented a Web-alike report tool [1]. This tool analyses the contents of well-known repositories under user-specified integration scenarios considering the coverage of main biological entities (genes, proteins and compounds) and the evaluation of standard nomenclatures, common names and repository cross-links as elements of integration. Also, acknowledging that most biological data still lays on scientific literature and requires extensive and time-consuming manual curation, we have been developing literature screening and processing tools [2]. The goal is to systematise the search of relevant literature based on user-specified keywords and the extraction of relevant information by applying statistical approaches that exploit simple pattern matching, machine learning and ontological enrichment.

Considering the wide scope of current applications that can benefit from the analysis of large amounts of data, all our tools are publicly available through our group's Web pages (http://biopseg.deb.uminho.pt).

References

[1] Lourenço A, Carneiro S, Rocha M, Ferreira EC, Rocha I, "Challenges in Integrating *E. coli* Molecular Biology Data". Accepted in Briefing in Bioinformatics (2010).
[2] Lourenço A, Carreira R, Glez-Peña D, Méndez JR, Carneiro S, Rocha LM, Díaz F, Ferreira EC, Rocha I, Fdez-Riverola F, Rocha M, "BioDR: Semantic Indexing Networks for Biomedical Document Retrieval". Expert Systems with Applications (2010), 37(4), 3444-3453.