Literature mining applied to systems biology information retrieval

REFERENCE
S5/05

Lourenço, A. M. G.[1,2], Simões, A. M. B.[1], Almeida, J. J. D.[1], Rocha, M. F. P.[1], Rocha, I. C. P.[2] and Ferreira, E. M. C.[2]

1. Department of Informatics, University of Minho, Braga
2. IBB - Institute for Biotechnology and Bioengineering, Center of Biological Engineering, University of Minho, Braga

analia@di.uminho.pt

Systems Biology is an important research field that generates new data continuously and demands for non-trivial data processing and analysis[2]. Data availability is tremendous, but System Biology data conciliation is quite challenging. It requires the crossing of multiple data sources and the interpretation of different data formats. Although Molecular Biology databases provide relevant information, most of the necessary data lays in scientific publications[3]. Our framework, the SysBio Explorer, addresses the automatic retrieval of potentially relevant documents and produces detailed document annotation and summaries that highlight document primary contents. As it can be applied to any user-specified problem, any task that involves information retrieval and processing can benefit from its use, but it is particularly well-suited for curation tasks and the inference of metabolic and regulatory networks. SysBio retrieves potentially relevant documents from bibliographic databases and accessible journals and uses accessible encyclopaedic knowledge in order to cope with non-standardised terminology and identify biological entities (e.g. genes, proteins, compounds, ...) and relations[1]. User-specified knowledge is also taken into account, addressing term recognition and term disambiguation issues as well as providing problem-specific information.

1. Liu, H. F. *et al.* (2006) *Journal of the American Medical* **13**:497-507.
2. Natarajan, J. *et al.* (2005) *Critical Reviews in Biotechnology* **25**:31-52.
3. Schuemie, M. J. *et al.* (2004) *BMC Bioinformatics* **20**:2597-2604.