

## Development of a Text Mining Framework for Systems Biology

Andre Lamurias<sup>1</sup>, Luka A. Clarke<sup>2</sup>, Francisco M. Couto<sup>1</sup>

<sup>1</sup>LaSIGE, Faculdade de Ciências, Universidade de Lisboa, Portugal

<sup>2</sup>BioISI: Biosystems & Integrative Sciences Institute, Faculdade de Ciências,  
Universidade de Lisboa, Portugal

Literature is one of the major sources of current biomedical information, in form of papers, patents or other types of written reports. To understand a biological system, it is necessary to integrate the results of various studies, which is highly time-consuming due to the increasing amount of published studies. One possible approach to solving this problem is through text mining. Text mining methods aim at automatically extracting relevant information from the literature. These methods can identify entities that are mentioned in the document, as well as the relations described between entities. However, it is necessary to organize the information extracted so that it can be efficiently used to answer user queries.

The objective of this study is to develop a text mining framework for regulatory networks of genes and micro RNAs. The regulation of micro RNAs has been linked to several diseases, such as cancer and obesity. The developed framework was able to identify interactions between micro RNA and genes described in papers related to cystic fibrosis. These results are a starting point to the development of a human disease knowledge graph.

The scope of the framework will be expanded to include information about specific diseases. The final objective is to develop a knowledge graph about human disease based on large-scale datasets. This knowledge graph will be used to generate disease networks, answer biomedical questions and generate hypotheses, improving current knowledge of human diseases.

**Preference for presentation:** Poster or talk

**Location:** University of Minho

**Author for Correspondence:** [alamurias@lasige.di.fc.ul.pt](mailto:alamurias@lasige.di.fc.ul.pt)