

Identifying Interactions Between Chemical Entities in Text*

Andre Lamurias and Francisco M. Couto

LaSIGE, Department of Informatics, Faculty of Sciences, University of Lisbon

We developed Identifying Chemical Entities (ICE 1.0) which recognizes the chemical named entities mentioned in a given text and performs resolution of each entity to the Chemical Entities of Biological Interest [2](ChEBI) ontology [3].

Since interactions between chemical entities are also frequently reported in biomedical texts, we aim at developing a new method (ICE 2.0) capable of identifying interactions between entities in the same text, as proposed by the task 9.2 of SemEval 2013 [6]. For this task, an interaction is defined as when one chemical compound or drug influences the level or activity of another. We intend to adapt some of the approaches used for this competition, developing a new method that used semantic similarity for identifying and validating interactions.

The chemical entities recognized and mapped by ICE 1.0 would be the starting point for this new method. Then, the semantic similarity between each pair can be calculated to verify if both entities participate in a chemical process. We have implemented three semantic similarity measures in our system that can be calculated between every two terms of the ChEBI ontology. These semantic similarity measures we have implemented are Resnik's similarity [5], simUI[1] and simGIC [4].

We propose a new method that uses semantic similarity, along with lexical features extracted from the original text, to train classifiers to predict if a pair of ChEBI terms represents an interaction. Our assumption is that two ChEBI terms that participate in an interaction have also a semantic relation defined in the ontology. We intend to make this method accessible from a web tool and web service.

Acknowledgements

This work was financially supported by FCT through the financial support of the SPNet project (PTDC/EBB-EBI/113824/2009) and through funding of LaSIGE Strategic Project, ref. PEst-OE/EEI/UI0408/2014.

References

- [1] R. Gentleman. Visualizing and distances using GO. URL <http://www.bioconductor.org/docs/vignettes.html>, 2005.
- [2] J. Hastings, P. de Matos, A. Dekker, M. Ennis, B. Harsha, N. Kale, V. Muthukrishnan, G. Owen, S. Turner, M. Williams, et al. The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. *Nucleic acids research*, 41(D1):D456–D463, 2013.
- [3] A. Lamurias, T. Grego, and F. M. Couto. Chemical compound and drug name recognition using CRFs and semantic similarity based on ChEBI. In *BioCreative Challenge Evaluation Workshop vol. 2*, volume 489, page 75, 2013.
- [4] C. Pesquita, D. Faria, H. Bastos, A. Falcão, and F. Couto. Evaluating GO-based semantic similarity measures. In *Proc. 10th Annual Bio-Ontologies Meeting*, pages 37–40, 2007.
- [5] P. Resnik. Using information content to evaluate semantic similarity in a taxonomy. *arXiv preprint cmp-lg/9511007*, 1995.
- [6] I. Segura-Bedmar, P. Martínez, and M. Herrero-Zazo. 2013 SemEval-2013 Task 9: Extraction of drug-drug interactions from biomedical texts. In *In Proceedings of the 7th International Workshop on Semantic Evaluation (SemEval)*. Citeseer, 2013.

*Corresponding Author: Andre Lamurias. Address: LASIGE, Faculdade de Ciências da Universidade de Lisboa, Departamento de informática, Edifício C6 Piso 3, Campo Grande 1749 - 016 Lisboa. E-mail: alamurias@lasige.fc.ul.pt.