@Note2 – open-source computational tools for biomedical text mining

Hugo Costa[1] and Miguel Rocha[2]
Emails: hcosta@silicolife.com, mrocha@di.uminho.pt
[1]SilicoLife, Braga, Portugal
[2]Centre of Biological Engineering (CEB), University of Minho, Braga, Portugal

The @Note2 project (http://www.anote-project.org) has been providing, over the last years, a number of user-friendly open-source computational tools addressing the main tasks in biomedical text mining (BTM). It is a project developed as a collaboration of the Biosystems group, at the Center of Biological Engineering from the University of Minho (www.ceb.uminho.pt/biosystems) and the company SilicoLife (www.silicolife.com), whose first version has been published in 2009 [1].

Since then, a major reformulation has been developed, including several novel features and the redesign in many others, leading to the current version 2.0, recently released, and fully written in Java, using the MySQL database.

@Note2 provides an end-user application, supported over a set of core libraries with a well-defined API, covering the main tasks in BTM related to Information Retrieval (IR) and Information Extraction (IE). IR main functionalities include loading documents from local folders, performing queries over Pubmed and other resources (e.g. patent or publisher specific databases), managing and updating query results, downloading the relevant documents, PDF to text conversion, relevance assignment and corpora management.

Regarding IE, the main tasks implemented are Named Entity Recognition (NER), Relationship Extraction (RE) and document clustering, being available several methods for all cases, including in-house and third-party algorithms. @Note also provides user-friendly interfaces for the creation of lexical resources (e.g. dictionaries) and a powerful curation environment for NER and RE annotations.

@Note has been developed following the Model-View-Controller (MVC) paradigm, taking advantage on the features of the plug-in based AI Bench framework (http://www.aibench.org), which provides a clear separation of core libraries and GUI tools, facilitating the development of new features in the form of plug-ins and thus inviting the participation of the community.

Current new developments include a novel machine learning (ML) workbench for NER and RE, to allow training and validation of ML models from manually annotated corpora, as well as improved multiuser support.

[1] A. Lourenço, R. Carreira, S. Carneiro, P. Maia, D. Glez-Peña, F. Fdez-Riverola, E. C. Ferreira, I. Rocha, M. Rocha, "@Note: a workbench for biomedical text mining.," J. Biomed. Inform., vol. 42, no. 4, pp. 710–20, Aug. 2009.