Curatable Named-entity Recognition using Semantic Relations

Yi-Yu Hsu 1 and Hung-Yu Kao 2

1 the Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan

2 the Department of Computer Science and Information Engineering, National Cheng Kung University, Taiwan

Abstract

Named-entity Recognition (NER) plays an important role in the development of biomedical databases. However, the existing NER tools produce multifarious named-entities which may result in both curatable and non-curatable markers. To facilitate biocuration with a straightforward approach, classifying curatable named-entities is helpful with regard to accel-erating the biocuration workflow. Co-occurrence Interaction Nexus with Named-entity Recognition (CoINNER) is a web-based tool that allows users to identify genes, chemicals, diseases, and action term mentions in the Comparative Toxicogenomic Database (CTD). To further discover interactions, CoINNER uses multiple advanced algorithms to recognize the mentions in the BioCreative IV CTD Track. CoINNER is developed based on a prototype system that annotated gene, chemical, and disease mentions in PubMed abstracts at BioCreative 2012 Track I (literature triage). We extended our previ-ous system in developing CoINNER. The pre-tagging results of CoINNER were developed based on the state-of-the-art named entity recognition tools in BioCreative III. Next, a method based on conditional random fields (CRFs) is proposed to predict chemical and disease mentions in the articles. Finally, action term mentions were collected by latent Dirichlet alloca-tion (LDA). At the BioCreative IV CTD Track, the best F-measures reached for gene/protein, chemical/drug and disease NER were 54% while CoINNER achieved a 61.5% F-measure.

System URL: http://ikmbio.csie.ncku.edu.tw/coinner/introduction.htm