

Services for annotation of biomedical text

Jörg Hakenberg
Dept. Computer Science and Engineering,
Arizona State University, Tempe, AZ/85281-8809, USA

4 April 2008

Motivation: Text mining in the biomedical domain in recent years has focused on the development of tools for recognizing named entities and extracting relations. Such research resulted from the need for such tools as basic components for more advanced solutions. Named entity recognition, entity mention normalization, and relationship extraction now have reached a stage where they perform comparably to human annotators (considering inter-annotator agreement, measured in many studies to be around 90%). Many tools have been made available, through web-interfaces or as downloadable software using non-standardized formats for in- and output. To advance progress in text mining, solutions are needed to both provide and combine the results of 'basic' information retrieval and extraction tools.

Results: Our groups at Technical University Dresden, Humboldt-Universität zu Berlin, and Arizona State University developed systems for named entity recognition, normalization, and relationship extraction. As evaluated during and after the BioCreative 2 challenge, recognition of proteins achieves 86% f-measure, normalization of gene mentions 85%, and extraction of protein-protein interactions including mapping to UniProt 25%.

Conclusions: We consider the BioCreative meta-service an ideal framework to make available information extraction tools to a variety of users: researchers from the biomedical domain, database curators, and researchers in text mining who can use the services as input for subsequent analyses. At the time of writing this abstract, twelve groups provide their tools as services to the BCMS server. We currently participates with tools for recognizing names of genes/proteins and species, normalization of gene mentions to EntrezGene, protein mentions to UniProt, mentions of species to NCBI Taxonomy, as well as classifying abstracts for protein-protein interactions.

Availability: For more information, please refer to <http://alibaba.informatik.hu-berlin.de/bcms/>. BCMS is available at <http://bcms.bioinfo.cnio.es/>.