

Text Mining in Biograph

Walter Daelemans

CLiPS, University of Antwerp, Belgium

Walter.Daelemans@ua.ac.be

The Biograph project is a biomedical knowledge discovery project combining graph data mining with structured biomedical information and with text mining on medline abstracts. It is a cooperation between the molecular genetics, data mining, and computational linguistics research groups of the University of Antwerp. In this talk, I will outline the general architecture of the system, which is currently applied to the problem of ranking genes according to relevance for diseases (gene prioritization). I will describe the text mining component (rule induction combined with rule certainty and confidence measurement using among others modality and negation analysis), and show results both on text mining benchmarks and gene prioritization benchmarks. The approach is more generally applicable than for gene prioritization only; I will show how the approach can be interpreted as a general question answering engine.