

Web-Scale Domain-Specific Information Extraction

Ulf Leser

Humboldt University, Berlin, Germany
leser@informatik.hu-berlin.de

Information Extraction (IE) from unstructured texts is a technology with growing importance in many applications. Three important challenges to IE are the achievement of high quality results, scalability of methods to very large corpora, and integration of IE results with other data for downstream analysis. In this talk, we will highlight recent advances and open questions in these areas by drawing from extensive experiences in developing and applying IE for biomedical research.

Biography. Ulf Leser studied computer science at the Technische Universität München and obtained his PhD in Data Integration and Query Planning from Technische Universität Berlin. After positions in research institutes and in the private sector, in 2002 he became a professor for Knowledge Management in Bioinformatics at Humboldt-Universität zu Berlin. His research focuses on scientific data management, statistical Bioinformatics, biomedical text mining and infrastructures for large-scale biomedical analysis and is typically carried out in interdisciplinary projects with domain scientists, especially from Medicine and Biology. He is speaker of the DFG-funded graduate school “SOAMED - Service-oriented architectures for medical applications”, chairman of the coordinated BMBF project “PREDICT - Comprehensive Data Integration for Personalized Ontology”, PI of the DFG research unit Stratosphere, and a board member of the DFG-excellence graduate school “BSIO - Berlin School for Integrative Oncology”.