

Translational Text Mining

Ulf Leser

Humboldt-Universität zu Berlin

Berlin, Germany

leser@informatik.hu-berlin.de

Abstract

Text Mining has become an important tool for many areas of biomedical research. Nevertheless, its impact actually is still surprisingly small, given the enormous body of knowledge published every day and the difficulties of biomedical researchers to keep an overview of relevant developments even only in their very specific fields of research. Why has the usage of some form of semantic search engine or large-scale information extraction pipeline not yet become the standard procedure for being up-to-date wrt related work? In this talk, I highlight some of the key issues text mining faces when it tries to become "translational", i.e., invade daily biomedical research. Examples are (a) the misleading focus on "correct extraction" where it should be "correct biological fact", (b) the widespread negligence of full texts and patents, and (c) the inaccessibility of typical machine learning models for end users. Notwithstanding some technical barriers, I argue that the community must invest more efforts to move closer to its users to achieve proper recognition in the field.

Biography

Ulf Leser studied computer science at the Technische Universität München and did his PhD at the Technische Universität Berlin. After positions in research institutes and in the private sector, 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 Bioinformatics analysis, topics he typically approaches in interdisciplinary projects with biologists and medical doctors. He is speaker of the graduate school SOAMED (Service-oriented architectures for medical applications) and a board mem-

ber of the Berlin School for Integrative Oncology (BSIO).