

Knowledge discovery in neuroinformatics

Bartłomiej WILKOWSKI^{a,b}

^a*DTU Informatics, Technical University of Denmark*

^b*Center for Integrated Molecular Brain Imaging*

Abstract. Traditionally, the process of turning data into biomedical knowledge has involved manual, meta-analysis of results reported in journals. Since the amount of the scientific data produced nowadays in neuroscience is increasing very dramatically, a resultant expansion of the medical databases has created a significant potential for design of new data modeling and information retrieval tools and services enabling faster data processing, analysis and dissemination among a highly interdisciplinary community of researchers.

Keywords. Neuroinformatics, data sharing, medical ontology

This PhD study aims in the discussion and design of necessary tools, methods and software, which can help in turning data into biomedical knowledge (meta-analysis of results reported in journals and information retrieval) [3], efficient medical data modeling and integration of various databases and repositories facilitating the data exchange between researchers from the whole neuroscience field.

The Center for Integrated Molecular Brain Imaging (CIMBI) with which this PhD study is associated, has established a large database of behavioral, genetic and imaging data. The key challenge of the research to be carried out during the PhD study is to develop the methods for integration of the CIMBI and related distributed databases including literature, biobanks and DTU's functional imaging database "Brede" in order to create an intelligent service for efficient information retrieval. Such a service is likely to become important not only for extracting information but also for an assistance in various aspects of research such as discovery of new facts, identification of previously undiscovered associations followed by proposal of new functions [1].

Medical ontologies (e.g. UMLS) and formal and statistical methods are considered to be a key for both database integration as well as the development of a process which is referred to as "knowledge discovery". Such a process can be understood as a pipeline of methods and techniques which include: text-mining of the scientific publications and further information retrieval (keyword extraction) [2], automatic interpretation of findings, discovery of new relationships and even design of new experiments.

References

- [1] M. Krallinger and A. Valencia, Text-mining and information-retrieval services for molecular biology, *Genome Biol*, 6(7), 2005
- [2] B. Wilkowski et al., Bridging the gap between coordinate- and keyword- based search of neuroscientific databases by UMLS-assisted semantic keyword extraction, *Human Brain Mapping 2009*, 2009.
- [3] B. Wilkowski et al., Coordinate-based meta-analytic search for the SPM neuroimaging pipeline – The BredeQuery plugin for SPM5, *Proceedings - HEALTHINF 2009*, 2009.