Call for Master Thesis

Concept Detection in Biomedical Documents

June 2014

Abstract
Detecting biomedical concepts in text documents or queries is a useful tool for improving biomedical information retrieval or navigating biomedical document collections. Known successful techniques for biomedical concept detection rely on natural language processing and/or machine learning and incur a substantial processing overhead at the document indexing stage compared to keyword-only indexing.
The goal of this master thesis is to evaluate a novel efficient concept detection algorithm based on position-dependent keyword matching on a recent public biomedical data set. Concepts are taken from a biomedical thesaurus called Medical Subject Headings (MeSH). Results should be compared to the accuracy achieved by the well-known MetaMap concept detector.

Required work

- Implement the evaluation framework for concept detection on the given data set.
- Integrate novel concept detection algorithm and assess its performance.
- Deploy MetaMap and apply it to the given data set.
- Perform a literature search on concept detection techniques.
- Describe the algorithm, evaluation experiments, and results in the master thesis.
- A condensed version of the master thesis may be submitted to an appropriate conference for publication.

Available resources

- Java implementations of novel concept detection algorithms for MeSH concepts.
- Data set of ~75,000 biomedical documents.
- MetaMap implementation.

Required knowledge and skills

- Java programming skills
- Introductory knowledge or interest in information retrieval

Desired duration: 6 months, starting as soon as possible

Advisors

- Prof. Laszlo Böszörmenyi
- Mario Taschwer

Contact: Mario Taschwer <mario.taschwer@aau.at>