Master Project / Master Thesis:

Investigation of machine learning techniques for classification of scientific texts

Description:

The amount of scientific literature is steadily increasing. While this also increases the complexity of finding articles of a particular research topic, it allows to gather knowledge from this literature with techniques of text mining and natural language understanding. One question of scientific interest is, whether it is possible to choose literature based on their topic without specifying all necessary keywords.

Goal of this project is to investigate, whether it is possible to estimate the topic of a scientific article from the text itself. Depending on the particular type of project (Master or Bachelor thesis, Master project, . . . ), the following subtasks shall be accomplished:

a) Identification and rating of relevant previous work
b) Preparation and clean-up of an appropriate dataset
c) Comparison and selection of appropriate preprocessing methods (word2vec, glove, . . .
d) Implementation of different machine learning algorithms based on existing libraries
e) Evaluation of the classification
f) Documentation of the project

The implementation should be done either in Python or R by use of executable notebooks (Jupyter or R Markdown).

Chair: Prof. Dr.-Ing. Sascha Spors
Supervisor: Dr.-Ing. Frank Krüger
Email: frank.krueger2@uni-rostock.de