Call for Master Thesis

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

An Analysis of Approaches for Diachronic Embeddings

Objective of this thesis:

Natural language (whether spoken or written) is prone to change due to many factors including cultural, societal, and technological factors. Words take on new meanings and lose their older senses. In the context of analyzing historical texts, understanding these changes is of great importance. The methods proposed for automated lexical change detection have also been utilized for finding similarities between documents over time in long-term archives.

This thesis in particular focuses on a theoretical, as well as practical analysis of the existing algorithms proposed so far for automated lexical change detection. These methods are mostly based on word embeddings, where the vector spaces are generated over different time periods [1]. Finally, the similarity between the words can be measured across these vector spaces. This thesis will mostly be centred on analyzing existing data sets, proposed methods and evaluation metrics used for the comparison of these algorithms. An optional step would be to analyze the possibility of adding temporal information to WordNet [2].

[1] https://nlp.stanford.edu/projects/histwords/
[2] https://wordnet.princeton.edu/

The thesis will be supervised by **Dr. Mehwish Alam and Mahsa Vafaie**, **Information Service Engineering at Institute AIFB, KIT, in collaboration with FIZ Karlsruhe.**

Keywords: Natural Language Processing, Machine Learning
Pre-requisites: Knowledge of Programming with Python.

Please send your CV and questions (if any) to the following e-mail address.

Contact person:
Dr. Mehwish Alam
mehwish.alam@kit.edu

aifb
Institute of
Applied Informatics and
Formal Description Methods
http://www.aifb.kit.edu/

Karlsruhe Institute of Technology