KIT
Karlsruher Institut für Technologie
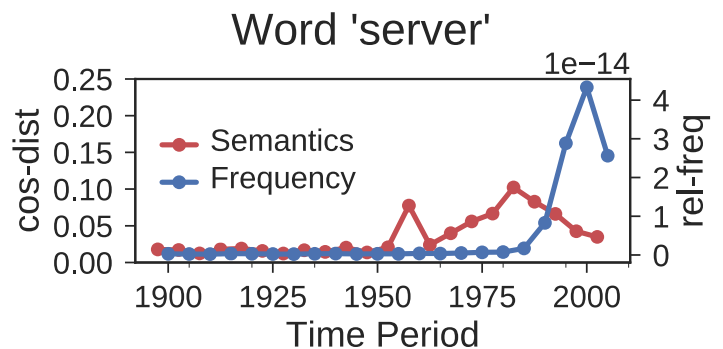
Master's Thesis

Trend Monitoring on Twitter Streams with Semantic Change Analysis

Master's Thesis

Language is constantly evolving. As part of historical linguistics, semantic change analysis examines how the meaning of a word evolves over time. Such semantic awareness is important for web analysis, to adapt to recent hypes or trends. Recent research has developed approaches to analyze semantic changes computationally. Commonly, word embedding models are used to learn

distributed representations of words in a vector space. The movement in the vector space then represents semantic change. By additionally including word frequencies the detection has been further improved. However, the previously developed approaches work from a static analysis point. Including additional text requires learning new embeddings and re-computing the used model. These operations are computationally expensive. Transfering semantic change analysis to an online setting is therefore of particular interest. The ability to monitor social media channels, such as Twitter, in real time would allow immediate reactions and adaptions. At the time of a new Tweet, the embedding update should be computed and used for semantic change detection.

The focus of this thesis is to develop an approach for semantic change analysis on text streams. In particular, the following research questions are of interest:

- What are appropriate window sizes for the detection or can it be performed online?
- Certain events heavily influence the amount of generated text in social media channels. What are the requirements for parallel real time computation relative to input volume?
- Considering limited resources, how close is the detection quality to the optimal solution?
- What are the limitations of such a system compared to the analysis in a static setting?

This results in the following tasks:

- Exploratory analysis on recent advances for training embeddings online
- Design and implementation of an approach to detect semantic changes on text streams
- Experimental evaluation of the approach on synthetic and real world text data including a comparison to the previously used models

Throughout this work, you will acquire a deep knowledge on semantic change detection and stream analysis. You train highly demanded skills in development and evaluation of data-mining algorithms. Knowledge from a lecture such as "Big Data Analytics" is not a prerequisite. However, elementary statistical knowledge, programming skills and the ability to accomplish conceptual work are desired.

Ansprechpartner

| Adrian Englhardt | adrian.englhardt@kit.edu | +49 721 608-48131 | Raum: 341 |
| Am Fasanengarten 5 | 76131 Karlsruhe | Gebäude: 50.34 | |