ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

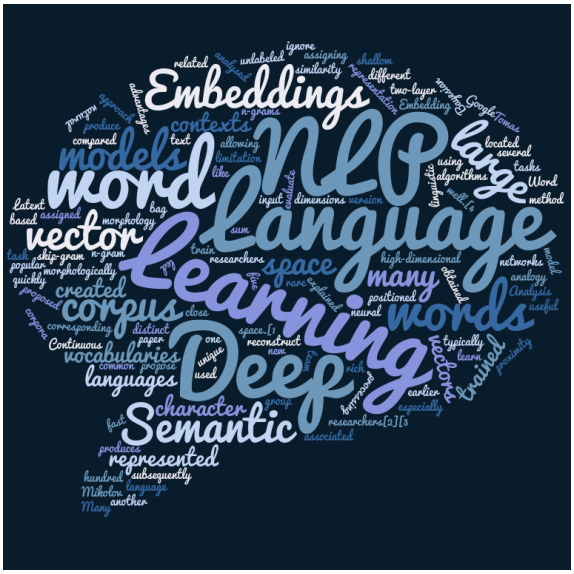
Distributed Computing

Prof. R. Wattenhofer

Improving Word Embeddings for NLP

Word embeddings are an important technique in natural language processing, and have been shown to significantly outperform previous methods. Word embeddings such as word2vec also exhibit interesting semantic properties, such that words with similar meaning lie close together in embedding space. The directions in embedding spaces can also correspond to semantic features, such that one can perform vector algebra on embedding vectors with very interesting results. For example: $King - Man + Woman = Queen$.

There are many different NLP tasks that profit from good embeddings, such as translation, question answering, summarization, and so on. Thus, the continuous development of new word (or sentence, paragraph, etc.) embeddings is an important research topic.



In this thesis we build on previous methods for computing word embeddings. Our aim is to achieve better results in, e.g., word disambiguation.

Requirements: Creativity and programming skills are an advantage. The student(s) should be able to work independently!

Interested? Please contact us for more details!

Contacts

- Gino Brunner: brunnegi@ethz.ch, ETZ G63
- Oliver Richter: richtero@ethz.ch, ETZ G63