Stephan Mandt
mandt@uci.edu
http://www.stephanmandt.com/

Title: Analyzing and Visualizing Word Semantics over Centuries

Abstract: Language evolves over time and words change their meaning due to cultural shifts, technological inventions, or political events. We consider the problem of detecting shifts in the meaning and usage of words over a given time span based on text data. Capturing these semantic shifts requires a dynamic language model.

Dynamic word embeddings (Bamler&Mandt, ICML 2017) is a probabilistic language model that allows users to measure how individual words change their meaning over time, such of centuries. This analysis is based on analyzing millions of digitized books. Every word in the vocabulary is represented as a vector in a so-called semantic space, where similar words are close to each other. As language evolves over time, these word vectors change their position, reflecting shifts in the semantics and relationships of words. This model, thus, gives rise to interesting data visualizations.

The first goal of this project will be to implement a user-friendly version of the dynamic word embedding training algorithm, following the literature. The second goal is to create an interactive platform that enables different visualizations of the data, such as finding the nearest neighbors of a given target word across all times, or making movies of how the semantic space evolves as a whole. The project will ultimately enable historians and political scientists to use the new methodology for their own research without having to be experts in statistical modeling.