First Outline

Max Feng

Introduction

The introduction section gives out a brief description of the project. This includes identifying the key terms of the project (semantic change, word embeddings, etc.) and explaining the rationale for this project (lack of research in semantic change of technology words).

Methodology

Building the Annual Corpora

This section introduces the first step of the project: scraping news and building corpora for each year. I will include the most important chunks of code and explain how the code works. I will also provide visualizations (a sample dataframe) to give an idea of what an annual corpus looks like.

Word Embedding Models

Once we have the preprocessed corpora, we are ready to train the word embedding models. In this section, I will explain how a separate model is created for each annual corpus. Also, I will align the models using orthogonal Procrustes technique. This section is expected to be the most technical part of the project as it includes a complicated modeling approach and an alignment technique.

Results

The results section presents the results of the semantic change of each technology word via visualizations. For each word, I will identify whether it has undergone a huge semantic change and try to determine the reason for the semantic change. I will try to relate to real world events that may have an impact on the shift of the meaning.

Dashboard Tutorial

This section gives a tutorial to the Shiny dashboard I will create as the output of this project. I will give an example of a tech word that is of interest. Then I will explain how to use the dashboard to find its visualization of semantic change as well as its analysis of why this happened. It will be mainly step-by-step screen shots of the dashboard. I will also include the link to the dashboard for the readers to experience themselves.

References

The final section gives a list of references for this paper. They will be formatted in a proper way using BibTex.