

Entwicklung einer grafischen Benutzeroberfläche zur Analyse von Kommunikation und Sprache

Hintergrund

Kommunikation ist ein elementarer Bestandteil des alltäglichen Lebens. Sprache lässt uns miteinander kommunizieren und ist sehr vielfältig. Um Texte zu analysieren, müssen diese von Menschen gelesen und verstanden werden. Gleichzeitig stellt es sich als schwierig und aufwendig heraus, verschiedene Texte miteinander zu vergleichen. Eine Möglichkeit diese Texte zu vergleichen bietet die Part-of-Speech (PoS) Analyse, sowie Bag-of-Words (BoW) für einen direkteren Vergleich von Wörtern. Diese beiden Analysen evaluieren wie die Wortverteilung ist und welche Wörter besonders häufig vorkommen. Diese beiden Analysen sind bisher nur als Konsolenanwendung verfügbar.

Aufgabe

In dieser Bachelorarbeit soll nun eine grafische Benutzeroberfläche (GUI) erstellt werden, welche es ermöglichen soll Datensätze in verschiedenen Dateiformaten, wie z.B. .csv, einzulesen und anschließend soll die Anwendung die Möglichkeit bieten, einzelne Texte/Sätze PoS oder BoW – Tags zuzuordnen. Dazu soll es möglich sein, die zugeordneten Tags deskriptiv zu beschreiben und auszuwerten (grafisch, tabellarisch etc.). Die Ergebnisse sollen im Anschluss exportierbar sein in ausgewählte Formate. Das Tool soll anschließend für verschiedene Datensätze evaluiert werden und herausgefunden werden, ob sich Unterschiede bei Texten erkennen lassen oder sogar Gemeinsamkeiten bei der Verteilung der Wörter.

1. Einarbeitung in die Literatur zur PoS und BoW
2. Anforderungserhebung zur grafischen Oberfläche
3. Programmieren der GUI und den Analysen zur PoS und BoW
4. Evaluierung verschiedener Datensätze zum Testen des Tools

Organisatorisches

Betreuer: Alexander Specht
Prüfer: Prof. Dr. Schneider
Beginn: ab sofort möglich

¹: PoS-tagging: <https://www.nltk.org>

²: BoW-tagging: https://www.tensorflow.org/text/guide/word_embeddings