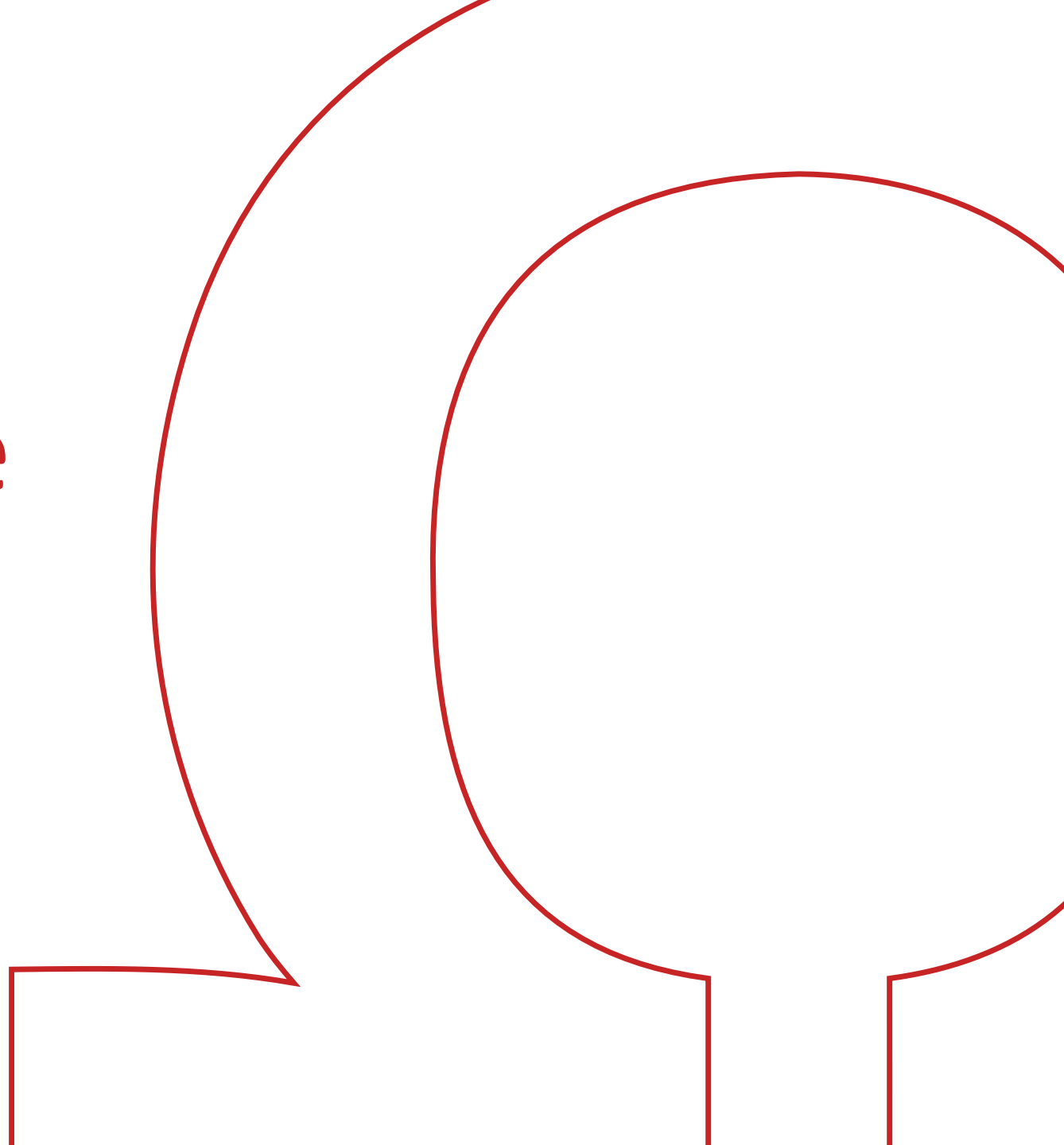


Political Language Processing

Julian Dreykorn
Simon Zeitler
Jan Tischner



Datensatz

Daten in JSON-Form von API des deutschen Bundestags

- Versuch: Selbst scrapen und verarbeiten der Plenarprotokolle des 20. Bundestages
- Fallback: Bereits segmentierte Daten des 19. Bundestages von bundestags-mine.de

Statistik-Task

Redezeiten-Analyse

- Unterschiede zwischen den Parteien erkennen
- Verteilung der Redezeit zwischen den Geschlechtern auswerten
- Lösungsansatz: Auswertung von Wort- und Zeichen-Anzahlen in den Reden

ML-Task 1: Klassifikation

Automatische Erkennung der Parteizugehörigkeit einer Rede

- Lösungsansatz: Aufbau eines Klassifikationsmodells
- Evaluierung: Accuracy und Precision der Klassifikationsergebnisse

ML-Task 2: Topic-Modeling

Untersuchung der Hauptthemen der Reden

- Lösungsansatz: Identifikation von Themenclustern in den Reden
- Evaluierung: Bewertung der Kohärenz und Unterscheidbarkeit der gefundenen Themen

Political Language Processing