

Abschlussbericht

Leon Burgard, Josua Kugler, Christian Merten

11. August 2021

Projektbeschreibung

Unser Projekt „Plenarprotokolle“ stellt mittels dem Paket `hateimparlament` Funktionen zur Analyse der Plenarprotokolle der 19. Wahlperiode des deutschen Bundestages zur Verfügung. Diese Funktionen können in vier Bereiche unterteilt werden:

1. Herunterladen der Protokolle
2. Konvertierung der XML-Dateien in Tibbles
3. Reparieren von Fehlern
4. Analyse

Das Herunterladen der Protokolle gelingt über die Funktion `fetch_all()`, welche auf die Website des deutschen Bundestages zugreift und die XML-Dateien einzeln herunterlädt. Hierzu haben wir das Paket `rvest` verwendet, welches wir bereits in der Vorlesung kennengelernt haben. Durch `read_all()` werden diese heruntergeladenen XML-Dateien in eine benannte Liste mit fünf Tibbles (`speaker`, `speeches`, `talks`, `comments` und `applause`) geschrieben. Allerdings benötigt man diese Tibbles immer wieder und es ist ziemlich zeitaufwändig die XML-Dateien immer wieder neu in Tibbles einzulesen, deshalb haben wir zusätzlich eine Funktion `write_to_csv()` geschrieben, die die fertigen Tibbles als CSV-Dateien speichert. Diese können dann sehr schnell durch `read_from_csv()` eingelesen werden, wodurch viel Zeit gespart wird. Da diese Protokolle kleine Fehler enthalten, müssen diese noch im nächsten Schritt bereinigt werden, was mit `repair()` funktioniert. Hierbei wird das Paket `tidyverse` viel benutzt, welches insgesamt sehr viel in unserem Projekt beansprucht wird, da wir uns mit der Datenanalyse beschäftigen. In `analyse.R` stellen wir noch einige Hilfsfunktionen bereit, die es dem Nutzer vereinfachen die Daten auszuwerten. Beispielsweise steht schon eine Funktion zur Verfügung, die ein Balkendiagramm erstellt, bei dem jede Partei des Bundestages separat ausgewertet wird. Hierbei wird das Paket `ggplot2` verwendet. Im letzten Schritt unseres Projekts haben wir Fragestellungen festgelegt, die wir mithilfe von unserem Paket beantworten wollten. Die Daten und unsere Ergebnisse visualisierten wir mithilfe von `ggplot2` und `tidyverse` in Vignetten.

Organisation des Teams

Während der ersten Projektphase wurden hauptsächlich die Funktionen zum Herunterladen der Dateien und Konvertieren und Reparieren der Tibbles geschrieben. Dies geschah größtenteils in Einzelarbeit, wobei hierbei die gegenseitige Kontrolle und Nachfragen die Funktionen optimiert haben. Zwischendurch wurde immer mal wieder zu einer HeiConf-Konferenz einberufen, um sich selbst den Zwischenstand klar zu machen und die Herausforderungen für die nächsten Wochen zu besprechen. In der zweiten Hälfte des Projekts kümmerten wir uns dann um die Analyse der Daten und stellten unsere Ergebnisse in Vignetten da und erzeugten Dokumentationen für alle Funktionen, die für den Nutzer wichtig sind.

Meine Beteiligung