

Was wurde bisher gemacht?

- Website scrapen
- Protokolle fetchen
- Protokolle parsen
 - Redner ermitteln
 - Redebeiträge extrahieren
 - Ordnungskommentare filtern
 - Kommentare extrahieren
 - Fehler in Redner Tabelle bereinigen

Wie ist der Zustand aktuell?

Protokolle sind heruntergeladen und geparsed in einer großen Tabelle. Es gibt noch ein paar Issues, aber das meiste was wir haben funktioniert auch soweit.

Was muss noch gemacht werden?

- Parsen verbessern (muss nicht alles gemacht werden, aber wäre vielleicht schön)
 - Kommentare in Kategorien einteilen, z.B. Kategorie Beifall mit Liste der applaudierenden Parteien
 - Metadaten vom Anfang extrahieren
 - Fehler weiter bereinigen (in talks, reden)
- Error Handling beim Fetchen
- Daten auf bestimmte Fragestellungen untersuchen
 - Wer redet am häufigsten / längsten?
 - Was ist typisches Vokabular für einzelne Fraktionen?
 - Welche Fraktion gibt / bekommt den meisten Beifall / die meisten Zwischenrufe?
 - Ab wann wird Pandemie-Vokabular (Virus, Corona, Maske, ...) genutzt?
 - Hat die Pandemie andere Themen wie Klimawandel verdrängt?
- Ergebnisse visualisieren

Wie sieht der Plan für die nächsten Wochen aus?

-24.07. Klausurenphase

25.07.-31.07. Datenanalyse

01.08.-07.08. Visualisierung, Bugfixes und Improvements

08.08.-11.08. Präsentation machen