



UNIVERSITÄT
LEIPZIG

Auftaktveranstaltung

Text Mining Übung

16.11.2020

Christian Kahmann



ORGANISATORISCHES

- Website: <http://asv.informatik.uni-leipzig.de/courses/288>
- BBB: <https://conf.informatik.uni-leipzig.de/b/kah-wrk-6q4>
- Ansprechpartner:
 - Christian Kahmann: kahmann@informatik.uni-leipzig.de

ZIEL

- Anwendung von Text Mining Verfahren
- Verwendet wird hierbei die Programmiersprache R und IDE RStudio
- Jede Woche wird ein Thema gemeinsam implementiert
- Die Folien sowie benötigte Daten und Materialien werden auf der Abteilungswebsite hochgeladen
- Es wird ca. 4 Übungsblätter geben, die Sie selbstständig bearbeiten sollten und deren Lösung in der darauffolgenden Woche zu Beginn besprochen wird

CREDITS

- Die Übungssessions orientieren sich an den Vorarbeiten von Gregor Wiedemann und Andreas Niekler
 - *Wiedemann, Gregor; Niekler, Andreas (2017): Hands-on: A five day text mining course for humanists and social scientists in R. Proceedings of the 1st Workshop on Teaching NLP for Digital Humanities (Teach4DH@GSCL 2017), Berlin.*
- Die Aufgaben basieren auf Vorarbeiten von Maciej Janicki

AGENDA

UNTER VORBEHALT

- 16.11.20 Einführung in R
- 23.11.20 Import von Texten, Textstatistik, Crawling
- 30.11.20 Zeitreihen, Sentiments
- 7.12.20 Keyword Extraktion
- 14.12.20 Kookkurrenzanalyse
- 4.1.20 Klassifikation
- 11.1.21 Topic Modelle
- 18.1.21 POS & NER Tagging
- 25.1.21 spaCy / Anwendung von Verfahren auf Filmdatensatz
- 1.2.21 Dokumentähnlichkeiten für Filmempfehlungen

FRAGEN



UNIVERSITÄT
LEIPZIG

Christian Kahmann

kahmann@informatik.uni-leipzig.de