Automatic Sampling and Analysis of YouTube Data

Recap - Outlook - Practice

Julian Kohne
Johannes Breuer
M. Rohangis Mohseni

2021-02-25

Course Recap (1)

Session	Example content
Introduction	Why is YouTube data interesting for research?
The YouTube API	API access, API requests, quota limits
Collecting data with the tuber package for R	Collecting channel/video stats & viewer comments
Processing and cleaning user comments	Character encoding, string operations, emoji dictionaries

Course Recap (2)

Session	Example content
Basic text analysis of user comments	Counting and visualizing the frequencies of words and emojis in comments
Sentiment analysis of user comments	Assigning sentiment scores to words and emojis
Excursus: Retrieving video subtitles	Retrieving and parsing YouTube video subtitles

Where to go From Here?

Some topics that we did not cover or only briefly touched upon that you might want to explore next/further:

- Analyses for more than one video: use for-loops, functions from the `apply` family or `map` functions from the `purrr` package
- Advanced text mining and NLP (going beyond bag-of-words approaches): check out the introductions/tutorials mentioned in the session on basic text analysis or this presentation by Cosima Meyer
- Alternatives to dictionary-based approaches for sentiment analysis: See the publications by Boukes et al., 2019 and van Atteveldt et al., 2021
- Supervised machine learning for text analysis: The online book Supervised Machine Learning for Text Analysis in R by Emil Hvitfeldt and Julia Silge is an excellent resource here
- Topic models (unsupervised ML): To get started you can, e.g., have a look at the introductions/tutorials by Rachael Tatman, Julia Silge, or the *Pew Research Center*

Acknowledgements ♡

All slides were created with the `R` package `xaringan` which builds on `remark.js`, `knitr`, and `R Markdown`. The exercises were created with the `unilur` package.

The original inspiration for our emoji parsing and analyses came from a blog post by Jessica Peterka-Bonetta. The `workshop.css` file we used for the layout of the slides includes elements from `CSS` files for `xaringan` presentations by Frederik Aust and David Zimmer.

We thank Janina Götsche and the *GESIS* Training team for taking good care of the organization of this workshop, and all of you for participating!

Any final questions or comments?

Practice time

You now have some time to start or continue working on your own *YouTube* data analysis project. We'll be around, so feel free to ask questions while you work on or get started with your projects.