Seminar

Einführung in NLP-Anwendungen mit R und Python unter besonderer Berücksichtigung von Twitterdaten

Ablaufplan

	Tag 1 (Fr, 30.04.2021)	Tag 2 (Fr, 07.05.2021)
09:00-10:30	Kick-off IntroNLP Quanteda universe NLP-Analyse-Pipeline Technisches Setup	Word Embeddings Classification Task Hintergrund Machine Learning mlr3 universe
10:30-10:45	*Pause*	*Pause*
10:45-12:00	Scraping Basic Text Cleaning	Analyse der Ergebnisse Visualisierung Ausblick: Machine Learning Pipeline
12:00-13:00	*Pause*	*Pause*
13:00-14:30	Basic Text Cleaning, continued Gewinnung statischer Text-Features	Intro Deep Learning & Transfer Learning BERT: fundamentale Tasks & Anwendungen
14:30-14:45	*Pause*	*Pause*
14:45-16:00	Topic Modeling	Sentimentanalyse mit BERT
		Wrap-up

	Allgemeines
	Basic: NLP mit R
	Advanced: NLP mit BERT

Literatur

Kenneth Benoit (2020): Package ‘quanteda’, URL: https://cran.r-project.org/web/packages/quanteda/quanteda.pdf

Silge, Julia/Robinson, David (2017): Text Mining with R: A Tidy Approach. O’Reilly. (https://www.tidytextmining.com/)

Simon Munzert, Christian Rubba, Peter Meiner, and Dominic Nyhuis (2015): Automated Data Collection with R: A Practical Guide to Web Scraping and Text Mining. Wiley Publishing *(verfügbar über die LMU Online-Bibliothek)*

Margaret E Roberts, Brandon M Stewart, Dustin Tingley, Edoardo M Airoldi, et al. (2013): The structural topic model and applied social science. Advances in neural information processing systems workshop on topic models: computation, application, and evaluation, 1-20.

Miner, G., Elder IV, J., Fast, A., Hill, T., Nisbet, R. and Delen, D. (2012). Practical text mining and statistical analysis for non-structured text data applications, Academic Press.

Manning, C., Raghavan, P. and Schutze, H. (2010). Introduction to information retrieval, ¨ Natural Language Engineering.

Feldman, R. and Sanger, J. (2007). The text mining handbook: advanced approaches in analyzing unstructured data, Cambridge university press.

Hastie, T., Tibshirani, R., Friedman, J. et al. (2002). The elements of statistical learning.