Methods to Analyze Text as Data

Doctoral Course, 7.5 ECTS, Jönköping International Business School, Spring 2025

Maximum number of participants: 15

Course schedule

All sessions are broadcast on Zoom (https://ju-se.zoom.us/j/66228720267) but attendance on site is preferable for didactic reasons.

Session 1: Tuesday, April 29, 10:00 – 16:00, B5002

- Lecture: Introduction and methods overview
- Tutorial: basic commands in R
- Prior to class: read Gentzkow et al. (2019) and Garz (2020); install R and R Studio on your laptop from https://posit.co/download/rstudio-desktop/

Session 2: Tuesday, May 13, 10:00 – 16:00, B5002

- Tutorial: basic commands in R (continued)
- Seminar: basic text statistics
- Prior to class: read Goes et al. (2014), read Quanteda quick start guide: https://quanteda.io/articles/pkgdown/quickstart.html

Session 3: Wednesday, May 14, 10:00 – 16:00, B5002

- Seminar: text lookup via word lists, text classification based on similarity
- Prior to class: read Ammann et al. (2014) and Beattie (2020)

Session 4: Tuesday, May 20, 10:00 – 16:00, B5002

- Seminar: sentiment analysis, topic modeling
- Prior to class: read Armelius et al. (2020) and Taeuscher et al. (2021)

Session 5: Wednesday, May 21, 10:00 – 16:00, A5002

- Seminar: supervised learning
- Coding lab: work on-site, discuss rough idea
- Prior to class: read Lee et al. (2018)

Session 6: Thursday, June 12, 10:00 – 16:00, B5015

- Live demonstration of own R scripts

The paper presentations will be organized during Session 1. A paper might be presented by multiple students, depending on the number of course participants:

Paper	Focus	Field
Goes et al. 2014: "Popularity Effect" in User-Generated Content: Evidence from Online Product Reviews. *Information Systems Research*, 25.	Basic text statistics	Information Systems
Ammann et al. 2014: Do Newspaper Articles Predict Aggregate Stock Returns? *Journal of Behavioral Finance*, 15.	Text lookup via word lists	Finance
Beattie 2020: Advertising and Media Capture: The Case of Climate Change. *Journal of Public Economics*, 188.	Text classification based on similarity	Economics
Armelius et al. 2020: Spread the Word: International Spillovers from Central Bank Communication. *Journal of International Money and Finance*, 103.	Sentiment analysis	Economics
Taeuscher et al. 2021: Gaining Legitimacy by Being Different: Optimal Distinctiveness in Crowdfunding Platforms. *Academy of Management Journal*, 64.	Topic modeling	Management
Lee et al. 2018: Advertising Content and Consumer Engagement on Social Media: Evidence from Facebook. *Management Science*, 64.	Supervised learning	Marketing

Further literature

Garz 2020: Quantitative methods. In von Rimscha (ed.): *Management and Economics of Communication*, pp. 109 – 127.

Gentzkow et al. 2019: Text as Data. *Journal of Economic Literature*, 57.

Jurafsky and Martin 2021: *Speech and Language Processing – An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Third edition.

Examination

The course is graded on a pass/fail scale. A pass requires active participation, presentation of a paper, and presentation of a self-written R script.

Paper presentation (20%)
The paper presentation should last ca. 45 minutes and cover the following aspects:

- Summary of the paper's core elements:
    - Motivation and research question
    - Theory
    - Data
    - Methods
    - Results and their academic and practical implications
- Description of how the text-as-data method in question is applied in the paper (consult the appendix of the paper if necessary)
- Description of complementary methods used in the paper

Written assignment in the form of an R script (80%)
The R script needs to implement one of the text-as-data methods discussed in class and provide a summary / visualization of the results. The proper execution of the script will be verified through a live demonstration (max. 15 minutes) during Session 6.

Teacher

Marcel Garz
Associate Professor of Economics
Room: B5015
marcel.garz@ju.se

marcelgarz.com
mediabiasworkshop.org
datamethodsinitiative.org

Marcel's work focuses on the economics of news markets, especially from an empirical perspective. He is particularly interested in issues with political implications, such as media slant and opinion diversity, as well as the role of social platforms for news supply and news consumption. His research often involves text-as-data techniques and methods supporting causal inference from observational data.