

Computational Text Analysis

Instructors: Mirko Wegemann and Eva Krejcova

Organiser: Simon Hix

Dates: 30-31 May and 3-4 June 2024

Credits: 10

Contact: Jennifer.Dari@eui.eu

Course Outline

In contemporary social science, we are faced with an era of big data. Political actors regularly justify their decisions on various communication channels, institutions publish policy reports, and individuals state their opinions on social media and comment sections of newspaper outlets. But how to make use of these data?

This workshop helps researchers in (1) gathering textual data from publicly accessible webpages, (2) preparing the raw material for analysis, (3) acquiring techniques to analyse the data and (4) understanding recent trends in text- and images-as-data. Thereby, the workshop is structured alongside four input sessions and 2-3 practical sessions.

1. Input + practical session on web-scraping
2. Input + practical session on the “old pipeline” of textual analysis (data preparation, and unsupervised classification/scaling methods)
3. Input [+ practical session] on the “new pipeline” of textual analysis (in the input session: supervised classification using transformer and large language models using Python; in the practical session, classification task in R)
4. Input session on how to use images-as-data in Python [led by Eva Krejcova, PhD]

The workshop provides a mixture of methodological input and practical examples in which the researchers can apply the newly learned skills directly on their data of interest. Its goal is to give researchers a first glimpse of what is possible to do with their textual data (and to a lesser extent, image data). To be most effective, researchers should have profound knowledge of R, knowledge of Python is not necessary since the respective sessions (3 and 4) will be conducted as showcase sessions.

Practical Information

Regarding software, please make sure to have the recent version of R and RStudio installed on your laptops. For the sessions on Python, you can register for free at [Google Colab](https://colab.research.google.com/) that allows you to run Python code without installing it locally.