

Workshop on Computational Text-Analysis

3rd Term, Academic Year 2022-2023

Dates: 11, 19, 20, 21 April + 11 May

Instructor: Jessica di Cocco (Max Weber Fellow)

Teaching Assistants: Rebecca Kittel (SPS Researcher) & Alsen Kokalari (SPS Researcher)

Sponsoring Professor: Ellen Immergut

Administrative Assistant: Charlotte Bufano

Syllabus

Workshop Description

An increasing number of studies are based on text-as-data, representing a vast reservoir of information for varied research. This workshop responds to the growing social science interest in using this type of data by highlighting their advantages and limitations. It aims to provide a general overview of computational textual analysis and the opportunities arising from machine learning. The course will provide students with coding skills they will implement in practice through hands-on sessions and labs in R.

In every lesson, we will start with a short introduction to the analysis tool explored that day, also mentioning the most updated papers using those methods. Subsequently, we will work together with examples, looking at the primary analysis tools, such as topic modelling, sentiment analysis, emotion detection, supervised machine learning, and manual coding. No programming knowledge is required to attend this workshop. However, a basic understanding of R and (or) Python may be helpful. During the labs, participants can start and complete the assignments or test their knowledge on different corpora, replicating the coding introduced during the lessons.

Requirements for Credits

Attendance at the 5 lectures and completing the 3 practical assignments; the lab is not compulsory but we hope it will be helpful for starting and completing the hands-on assignments.

Programme

April 11th 2023: Online

Lecture, 10:00 – 12:00:

- Introductory Session, general overview of text-as-data approaches, how it works and where it is applied, current state of the art.

April 19th 2023:

Lecture, 10:00 – 12:00: Seminar Room 3

- Possible topics: Sentiment Analysis and/or dictionary analysis

Lab Session, 13:00 – 16:00: Sala Del Capitolo

- Applying the code that was discussed during the class, e.g. applying to a new/different corpus
- Starting with the assignments

April 20th 2023: Seminar Room 2

Lecture, 10:00 – 12:00:

- Topic Modelling (e.g. LDA topic models)

Lab Session, 13:00 – 16:00:

- applying the code that was discussed during the class, e.g. applying to a new/different corpus
- starting with the assignments

April 21st 2023: Seminar Room 2

Lecture, 10:00 – 12:00:

- Web-Scraping or any other topic that is of interest according to the questionnaire (e.g. emotion detection)

Lab Session, 13:00 – 14:00 (possibility of extending to 16:00):

- applying the code that was discussed during the class, e.g. applying to a new/different corpus
- starting with the assignments

May 11th 2023

Lecture, 10:00 – 12:00: Seminar Room 2

- Coding exercises with machine learning applications
- Concluding session covering topics that have not been touched upon or where more information is needed (e.g. brief overview on image-as-data)
- Q&A session