

IPSA Online Summer School on Research Methods, Antalya 2024

Introduction to Text-as-Data

(Draft Syllabus)

Instructor: Dr. Melike Ayşe Kocacık Şenol

Email: makocacik@sabanciuniv.edu

Course Topics:

Public speeches, social media posts, or legislative documents are important information sources which are currently used increasingly in political science and international relations studies. This course will introduce you with the skills and computational tools needed to extract meaningful information from texts. The course will introduce the steps of text-as-data approaches: acquiring texts, transform it to data and analyze it for different questions. Through hands-on exercises and practical examples, you will learn how to use the R programming language.

Course Objectives:

- Understand the fundamentals of text analysis and its relevance to political science research.
- Gain proficiency in preprocessing, analyzing, and visualizing textual data using R.
- Explore various applications of text analysis in political science, including sentiment analysis, and topic modeling.
- Apply text-as-data techniques to real-world political texts and research questions.
- Develop critical thinking skills for interpreting and communicating results derived from text analysis.

Schedule:

Morning sessions are for lectures, the afternoon sessions are for discussions and introduction to the software. Assigned readings should be done before the respective session.

Date, Time	Topics
Day 1	Morning Session <u>Introduction to Text-as-Data:</u> What are we doing here?
	Afternoon Session <u>Practice Session 1:</u> Introduction to R and R Studio
Day 2	Morning Session <u>The bag of words approach:</u> Preprocessing Data Frame Matrix
	Afternoon Session <u>Practice Session 2</u>
Day 3	Morning Session <u>Modelling Bag of Words:</u> Word Embeddings
	Afternoon Session <u>Practice Session 3</u>
Day 4	Morning Session <u>Supervised and Semi-Supervised Classification Techniques:</u> Dictionary Analysis Sentiment Analysis
	Afternoon Session <u>Practice Session 4</u>
Day 5	Morning Session <u>Unsupervised Classification Techniques:</u> Topic Models LDAs Word Counts
	Afternoon Session <u>Practice Session 5</u>

General Reading Recommendations: