

Course: Text Analysis with Digital and Computational Tools

Abstract

The course emphasizes foundational skills in computational text analysis and equips students to independently utilize digital tools for this purpose. Traditional humanities research is often rooted in ways of thinking, with text serving as a pivotal element in human culture, spanning from ancient writings to modern literature and journalism. As textual collections transition from physical bookshelves to digital formats, there is a shift from close reading of rare texts to "remote reading" of expansive digital repositories. This transformation each corpus of textual knowledge to be treated as data that can be analyzed and reinterpreted from new perspectives. The course highlights unique opportunity presented at this moment to revolutionize humanities, with GenAI tools emerging as the next generation of analytical instruments. Students will learn to conduct basic computational research through hands-on practice with corpus analysis tools and AI tools available to everyone. At the same time, the results of the digital analysis will be critically examined, together with acknowledging the challenges that accompany computational processing of textual data and the importance of maintaining the principles of scientific research..

Teaching details

Designed for the fall semester 2025 for research students in the faculties of humanities, social sciences, and law.

Course number: 1886.4001

Lecturer: Dr. Vered Silber-Varod (TAD Center) in collaboration with guest lecturers.

Teaching Assistant: Ms. Stav Klein

Day and hour: Tuesday, 14:15-15:45

Grade composition – submission of exercises (20%) and final course assignment (80%).

Syllabus

Lesson 1	What is a word? What is Textual Corpus?
Lesson 2	Introduction to Natural Language Processing: From Eliza to Gemini; From the Turing Test to Artificial Intelligence.
Lesson 3	Text transformation: lemmatization, morphological analysis, part-of-speech labeling, entity identification. How linguistic knowledge helps analyze linguistic trends and patterns of varied texts?
Lesson 4	Corpus analysis Part A
Lesson 5	Corpus analysis Part B
Lesson 6	Word embeddings
Lesson 7	Identifying writing styles (stylometry)
Lesson 8	Topic modeling: whether and how can topics found in documents be identified without reading the documents individually?

Lesson 9	Semantic networks: connections between and within texts.
Lesson 10	LLM doesn't scare me – How can large language model-based tools be harnessed for research?
Lesson 11	Unique characteristics in analyzing text originating from speech - principles and work processes.
Lesson 12	Examples of studies that used the methods studied so far.
Lesson 13	Summary and presentation of project proposals.

Bibliography

1. William J. Turkel and Alan MacEachern, [The Programming Historian](https://programminghistorian.org/en/about) 1st edition (Network in Canadian History & Environment: 2007-2008). <https://programminghistorian.org/en/about>
2. מילים שקולות – צעדים ראשונים במחקר הספרות החישובי / איתי מרינברג-מיליקובסקי. (2022). הוצאת למדא האוניברסיטה הפתוחה
3. מחקר חישובי במדעי הרוח – אסופת מאמרים / אופיר מינץ-מנור, איתי מרינברג-מיליקובסקי. (2022). הוצאת למדא האוניברסיטה הפתוחה