

From Word Counts to LLMs: An Introduction to Computational Text Analysis

Prof. Dr. Matthias R. Hastall

Course Description

Computational text analysis (CTA) offers powerful opportunities to uncover the content, meaning, and linguistic structures within large document collections that would remain beyond the reach of manual coding. This hands-on course introduces practical skills for analyzing textual data using a range of computational methods. Key techniques include text preprocessing, dictionary-based approaches, topic modeling, and classification using large language models (LLMs).

Each day combines concise lectures with guided coding exercises and applied tasks. Particular emphasis is placed on comparing methods with respect to their strengths, limitations, and appropriate use cases. Students will work with real-world data such as political speeches, news articles, or social media content. By the end of the course, students will be able to design, implement, and critically evaluate computational text analysis workflows.

No prior programming experience is required, although basic familiarity with quantitative research methods is beneficial. All analyses will be conducted using free and accessible software tools. Participants are encouraged to use their own computers; a Windows-based system is recommended, as most tools are readily available for this platform.

Length

The course comprises five sessions of six hours each (30 hours in total).

Difficulty

This is an introductory course. No prior knowledge of computational text analysis is required. A basic understanding of quantitative research methods is helpful.

Expectations

Active participation in all sessions is expected. This includes attending at least four out of five days, completing assignments, and engaging in exercises. To support interaction, participants are expected to keep their cameras on for approximately 80% of the session time where possible.

Please note that sessions will not be recorded, in order to foster an open and collaborative learning environment. Comprehensive course materials and documentation will be provided after each session.

Contact Information

For questions regarding the course or specific requirements, please contact:
matthias.hastall@tu-dortmund.de

Preliminary Schedule

Please note that this schedule is preliminary. We will make adjustments depending on the course participants' research interests and course progress.

Morning sessions focus on theoretical foundations, while afternoon sessions emphasize hands-on practice and discussion. Assigned readings should be completed in advance of each session.

The course takes place daily from August 31 to September 4, 2026, via Zoom. Sessions begin at 09:00 and end at approximately 15:30.

Please note: All course times refer to the Istanbul time zone (TRT, UTC+3).

Date, Time		Topics
Aug. 31, 2026	Morning Session	Introduction to Computational Text Analyses history concepts common methods methodological foundations strengths and limitations ethical considerations
	Afternoon Session	<i>Practice Session 1</i> preprocessing standardization exploratory applications (KWIC analyses, n-grams, word frequency comparisons)
Sep. 01, 2026	Morning Session	Dictionary-based Approaches introduction types and possibilities strengths and pitfalls
	Afternoon Session	<i>Practice Session 2</i> applying existing dictionaries building custom dictionaries sentiment analysis
Sep. 02, 2026	Morning Session	Topic Modeling Approaches methodological foundations possibilities limitations
	Afternoon Session	<i>Practice Session 3</i> running topic models interpreting and validating topics
Sep. 03, 2026	Morning Session	Large Language Model (LLM) Approaches methodological foundations prompt design limitations
	Afternoon Session	<i>Practice Session 4</i> one-shot and few-shot classifications annotations validation
Sep. 04, 2026	Morning Session	Further Options & Future of Computational Text Analysis recent developments method combinations advanced topics summary & outlook
	Afternoon Session	<i>Practice Session 5</i> previously covered or new techniques, as requested

Literature/Reading Assignments

Literature assignments and recommendations are posted in our ABU eLearning system.