Computational Text Analysis and the Politics of Education

Winter Term 2023/24
University of Marburg

Instructor: Milan Thies, Researcher, European University Institute
Email: milan.thies@eui.eu

The syllabus and course content can and will be adapted to the level of prior knowledge of the students.

Course Description and Rationale	In this seminar, we will explore the exciting field of computational text analysis, also known as text-as-data. As our world becomes increasingly digital, vast amounts of legal, political, scientific, and journalistic text data have become available for analysis. This course is designed to introduce advanced undergraduate and graduate students to foundational concepts and practical applications of studying computational text analysis. Politics is about conflict and cooperation between societal groups. Harold Lasswell once defined it as "*who gets what, when, how*?". Actors in the political arena, be it politicians, lobbyists, or mass movements, engage in (distributional) conflicts through language — be it spoken out loud in speeches, dialogues, and protest chants, or written down in position papers, protocols, and regulations. Throughout the seminar, students will learn how to use computational methods to explore the traces of conflict and cooperation preceding political change. We will focus on a policy field that not only affects every university student personally but also one that has been at the forefront of global change: Education. Over the past century, education has expanded from being a privilege of the few to an almost universal and global experience. In Europe, this trend is now increasingly leading to educational upgrading and a massive expansion of higher education. But how has this quiet revolution come about? And what are the central conflicts of educational governance today? The seminar is organized in three stages of the learning process. First, students will learn about the *foundations* of text-as-data in political science, as well as

	central theories of education politics. Further, they will study how to manually analyze texts without computational assistance. Next, we will focus on the *methods* of computational text analysis. Students will learn about the theoretical underpinnings of computational text analysis as well as the use of text-as-data methods through hands-on exercises using the R statistical programming language. We will also discuss recent examples of empirical research using computational text analysis. Last but not least, students will *apply* the learned methods in an in-class project. Small groups of students will be provided with textual data on a policy process and subsequently implement one or several of the techniques learned throughout the seminar
Objectives	The course is designed as an introduction to computational text analysis. Students will: **Theoretical Foundations**: I. Learn about the theoretical underpinnings of text-as-data approaches and their relevance in social and political science research. II. Explore the role of texts as manifestations of political conflict and cooperation. **Practical Application**: III. Develop practical skills in implementing text-as-data methods using the R statistical programming language through hands-on exercises and in-class projects. **Empirical Insights**: IV. Learn about the global process of educational expansion and understand how current political conflicts shape education policies.

Sessions	
Foundations of Computational Text Analysis	
Day 1 **Session 1**	**Texts as Traces of Political Conflict** • (Re-)visiting the policy process

	• Textual manifestation of political conflict • Textual manifestations of political change • Overview of the course
Day 1 **Session 2**	**The Politics of Education** • The development of education systems over the long run • The partisan politics of education • Interest groups and education policy • Current debates in education politics: Privatization versus state provision, stratification versus inclusion, technology and education
Day 1 **Session 3**	**Qualitative Text Analysis** • Inductive and deductive approaches • Strengths and weaknesses of qualitative text analysis
Methods of Computational Text Analysis	
Day 2 **Session 4**	**A Newcomer's Guide to Computational Text Analysis** • Fundamental assumptions of computational text analysis • Varieties of text-as-data techniques • Strengths and weaknesses of computational text analysis
Day 2 **Session 5**	**Working with R: Introduction to R, Quanteda & Descriptive Analyses** • Basics of the statistical programming language R and Rstudio (depending on level of student exposure to R) • Representation of text in Quanteda • Importing textual data • Preprocessing and cleaning textual data • Descriptive analyses (word frequencies, keywords in context, keyness, wordclouds)

Day 3 **Session 6**	**Computational Text Analysis Without Machine Learning** • Word counts in political science research • Dictionaries as a text-as-data technique
Day 3 **Session 7**	**Supervised- and Unsupervised Machine Learning for Text Analysis** • Overview: unsupervised and supervised machine learning • Unsupervised topic models • Semi-supervised topic models
Application	
Day 3 **Session 8**	**In-Class Project I: Introduction to Cases and Data** • Students will be familiarized with recent reforms in European education systems. They will subsequently form small groups to analyse textual sources on one of the reforms.
Day 4 **Session 9**	**In-Class Project II: Supervised Group Work** • Students apply one or several techniques learned throughout the seminar. • During the project all students will be assisted in overcoming difficulties by the instructor while engaging deeper with both method and subject.
Day 4 **Session 10**	**Presentation, Reflection and Outlook** • Presentation: Most interesting finding of each in-class project. • Reflection: Learning from the seminar & how computational text analysis techniques could be used in upcoming coursework/theses. • Outlook: Recent advances in text-as-data research.