Fundamentals of Quantitative Text Analysis

Online course learning objectives

This course teaches learners how to analyze large amounts of textual data by applying R programming skills to an efficient, powerful and easy-to-use method - quantitative text analysis. This course is perfect for social scientists who want to understand the theory and assumptions that underpin quantitative text analysis, whilst developing their R programming skills via practical examples of analysis with real texts.

This course will help learners to:

- Understand the theoretical basis for Quantitative Text Analysis.
- Survey methods for systematically extracting quantitative information from text for social scientific purposes.
- Identify texts and units of texts for analysis.
- Convert texts into matrices for quantitative analysis.
- Analyze these matrices to generate inferences using quantitative or statistical methods.

Language: English
Time to complete: 15 hours
Instructor: Professor Jonathan Slapin

Online course full syllabus

MODULE ONE: INTRODUCTION TO TEXT ANALYSIS AND CONCEPTUAL FOUNDATIONS

This module introduces students to the types of questions that text analysis can answer, the tools that the course will use to answer them, and offers examples of analyses using text-as-data.

- Introduction explaining course purpose: goals and objectives.
- Conceptual foundations of text analysis.
- Quantitative text analysis as a field and the development of the field.
- Logistics and software - required setup and work files.
- A basic example of performing a text analysis.

MODULE TWO: THE BASICS OF WORKING WITH TEXTUAL DATA

This module discusses how to obtain textual data and how to get it into a format that is suitable to analyze. It finishes with an example of using complexity and readability measures.

- Where to obtain textual data.
- Formatting and working with text files.
- Practical considerations of indexing and metadata.
- Units of analysis: strategies for selecting units of analysis.
- Overview and examination of complexity and readability measures.

Fundamentals of Quantitative Text Analysis

MODULE THREE: EXAMINING INDIVIDUAL WORD OCCURRENCES

This module teaches students how to summarize texts in a corpus by looking at the occurrence of individual words using tools such as keywords in context and dictionaries.

- Keywords in context: coverage and examples of KWIC.
- Consideration of concordance and dictionaries.
- Detecting and identifying collocations.
- Stemming: An in-depth discussion of text types, tokens, and equivalencies.
- Stop words and feature weighting: An in-depth discussion of text types, tokens, and equivalencies.

MODULE FOUR: COMPARING ACROSS TEXTS

This final module introduces students to models that allow for comparison across texts within a corpus by examining their word usage, including building dictionaries and creating scales.

- Euclidean distance and its use in comparing texts.
- Cosine similarity and its use in comparing texts.
- General principles and rationale for dictionaries.
- External dictionaries: How to add a third-party dictionary.
- How to create your own dictionary.
- Overview of wordscores.
- Implementing in R a basic model.