Course Description
MEDD8925 Analysing Textual and Documentary Data in Quantitative Research

Course description
This course introduces students to research that uses digitised textual data in the context of educational research. It covers how data may be collected by researchers and digitised, data sources that exist in structured/unstructured pools of data online and how to obtain them, as well as data available in institutional or open repositories in the form of corpora developed by other researchers. This course introduces to students a wide range of methods and research questions for exploring digital textual data quantitively, including but not limited to corpus linguistic approaches, data mining and natural language processing techniques. The course is hands-on and students will be able to use simple natural language processing tools that facilitate quantitative exploration of texts, such as corpus analysis software, simple code and AI.

Course objectives
1. To introduce to the students a broad range of methods and approaches that can be employed in analysing textual data in quantitative research in the discipline of education
2. To develop students' ability to critically read research that employs corpus and natural language processing techniques in the learning sciences
3. To widen students' skills and knowledge of educational research designs
4. To provide students with entry-level research skills training in common techniques for analysing Textual and Documentary Data in Quantitative Research

Course learning outcomes (CLOs)	Aligned programme learning outcomes (PLOs)
1. Demonstrate and communicate up-to-date knowledge about the quantitative use and analysis of textual/documentary data in the broad context of research in education	PLOs 1, 4
2. Apply methodological knowledge to the analysis of textual data	PLO 2
3. Evaluate contemporary research in the learning sciences that uses quantitative textual data techniques and	PLO 1
4. Adhere to ethical and intellectual property obligations when conducting textual research	PLOs 3, 4, 5
5. Create and report on a small-scale class project that analyses textual data using language processing techniques	PLOs 2, 3, 5

Course assessment methods
• Critical Analysis of a Research Article
• Peer Feedback Calibration Activity
• Class Research Paper (analysis of a modest set of textual data)

Course content and topics
• Introduction to computational methods for textual data analysis (Corpus Linguistics/NLP)
• What is a corpus and what is it used for?
• History of the field, big data and AI
• Wordlists and Frequency Distributions
• Where to get documents for analysis
• Intro to desktop corpus software
• Analysis of Grammar and Vocabulary in Text
• Concordance and KWIC analysis
• Collocations
• Multi-word Expressions

- Baseline norming, comparing distributions in texts
- Characterizing what texts are 'about'
- Keywords/Topics in texts
- Tagging (part of speech)
- Lemmatization
- Sentiment and semantic analysis
- REGEX (regular expressions)
- Lexical Density and Lexical Diversity in Texts
- Dispersion and Range
- Corpus based sociolinguistics 3, 4
- Register and genre/Variation in language
- Research design and data collection
- Stylistics: Learning about literary texts using quant tools
- Micro and Macro analysis of authorial style, character and themes.
- What is a computer model? 3, 4
- Building simple models that analyze text.
- Coding with AI support to analyse text
- Gui's vs. ad-hoc code

Required / recommended readings and online materials
Readings will be provided in Moodle.

Other additional course information
Advanced Research Method (ARM) course [Note: Students should bring their laptops to classes and download free tools for textual analysis]