

nccr 
on the move

National Center of Competence in Research -
The Migration-Mobility Nexus

nccr-onthemove.ch

unine[•]
Université de Neuchâtel

Methods Workshop in Quantitative Text Analysis

3-7 June 2024

University of Neuchâtel

Doctoral students and interested NCCR Fellows are asked to
register [here](#) until 31 May 2024



Swiss National
Science Foundation

The National Centres of Competence in Research (NCCRs)
are a funding scheme of the Swiss National Science Foundation

Methods Workshop in Quantitative Text Analysis

This workshop is part of the fourth block of the Doctoral Program of the nccr – on the move (methods, contents, and skills within the own discipline). We welcome all members of the NCCR community and doctoral students enrolled in a CUSO program to join us.

General Description

The workshop aims to bring together NCCR PhD students, fellows and project leaders that use, or aim to use, quantitative text analysis in their projects. The workshop will be segmented with different topics on different days, this way participants can participate on the specific day(s) where the program best overlaps with their research interests and competence. The first two days are optional introductions to the programming languages used in the workshop and to text pre-processing, to make sure we have a baseline level for all participants. The two introduction days (3-4 June) will be open to doctoral students enrolled in a CUSO program.

On 5-6 June several approaches and methods to text analysis will be introduced in a conceptual lecture format. On 6 June we will enjoy a public lecture by expert Dr. Jisu Kim followed by a dinner for workshop participants. Finally, on 7 June, participants will be divided into thematic groups where they will discuss and present their own plans to analyze text to further their NCCR related research.

Requirements

This workshop is open to all interested in learning about the possibilities in text analysis. Because the workshop covers mainly quantitative approaches, attendance on June 3-4 is recommended for newcomers to coding who want to benefit fully from the workshop. The workshop will be taught in Python, but it will be relevant to R users and non-coders as well.

For participation on the final day (7 June), we ask that you submit an abstract detailing a research agenda that includes using text analysis methods on NCCR related research. This can include qualitative approaches to text analysis (frame analysis, discourse analysis) and use of non-coding programs such as NVivo etc.

Practical Information

The course takes place in presence (no online participation) from 3 to 7 June 2024 at the University of Neuchâtel, Rue Abram-Louis Breguet 2, room 2.310. Doctoral students and interested NCCR Fellows are asked to register [here](#) until 30 May 2024. Applications by members of the nccr – on the move will be prioritized, and registrations will be treated on a first come first served basis. Lunch is included for all the participants, and travel expenses to and from Neuchâtel will be reimbursed for members of the nccr – on the move. If you have further questions, please contact Robin Stünzi via email at robin.stunzi@nccr-onthemove.ch.

Timetable of the course

Monday, 3 June: Introduction to Python*

10:15	Setting up
10:30	Basic syntax
12:00	Lunch break
13:30	I/O and project management
15:00	Coffee break
15.30	Basic data manipulation and graph generation

Tuesday, 4 June: Introduction to Text Mining and Text Processing*

10:15	Setting up
10:30	Text pre-processing-Introduction to relevant toolkits
12:00	Lunch break
13:30	Text pre-processing-Feature engineering
15:00	Coffee break
15.30	Hands-on session on Text pre-processing

*Also open to CUSO members

Wednesday, 5 June: Advanced Text Analysis: (Un)Supervised Approaches

10:15	Setting up
10:30	Unsupervised approaches in TA (Clustering, Word embeddings, topic modelling, sentiment analysis etc.)
12:00	Lunch break
13:30	Supervised approaches in TA (Naive Bayes, Support Vector Machines, topic classification, sentiment analysis, etc.)
15:00	Coffee break
15.30	Hands-on session on advanced text analysis

Timetable of the course

Thursday, 6 June: Transformer Models/ChatGPT

10:15	Setting up
10:30	Introduction to Large Language Model (transformer/GPT)
12:00	Lunch break
13:30	Hands on exercises with Large Language Models
15:00	Coffee break
15.30	Public lecture “Unveiling Immigrant Identities in the Digital Sphere“ by Dr. Jisu Kim
18:00	Dinner

Friday, 7 June: Discussion on Specific Projects in Groups

10:15	Introduction to group work
10:30	Parallel group work sessions: text analysis in NCCR related research
12:00	Lunch break
13:30	Presentations of text analysis research agendas

About the Speakers

External Expert

Jisu Kim, Research Scientist, Max Planck Institute for Demographic Research (MPIDR)

Jisu Kim, having earned her PhD in Data Science from Scuola Normale Superiore in Italy, currently works as a research scientist in the Department of Digital and Computational Demography at the Max Planck Institute for Demographic Research. She has been working on exploring and establishing novel methods to improve relevant statistics of international migration using social media data. Her research focuses on the intersection of migration sciences, development economics, social networks, and data-driven algorithms.

Internal Team

Maarja Vollmer (III_IP40): experience with text analysis (LDA combined with topic relevance). Some knowledge about dictionary-based methods also.

Didier Ruedin (III_IP40): experience with coding in R and using text-as-data models

Vestin Hategekimana (III_IP42): Statistic / Text Mining / Web Scraping / Network analysis

Oliver Pedersen (III:47): qualitative approaches; psychology

Karin Vaagland (III_IP39): archival; qualitative; frame analysis, discourse analysis, NVivo

Public Lecture

Unveiling Immigrant Identities in the Digital Sphere

Jisu Kim, Research Scientist, Max Planck Institute for Demographic Research (MPIDR)

Abstract

In today's digital age, understanding online behaviors and identities is of paramount importance. The growing presence of individuals online not only signifies the expanding reach of the digital realm but also presents a distinct opportunity to delve into their behaviors and identities. This extends to studying immigrants online, which yields real-time insights into migration patterns and demographic shifts, thereby informing policy-making and resource allocation decisions. Moreover, this avenue of study offers invaluable insights into integration processes, access to resources, and cultural dynamics, enriching our understanding of societal trends and needs. With this in mind, the presentation introduces a fresh method for identifying immigrants online, employing Twitter data analysis to detect distinctive patterns associated with immigrant status.

The proposed approach encompasses a comprehensive data collection strategy, incorporating both structured and unstructured data sources. While geolocation data are pivotal in international migration studies, this also involves extracting linguistic cues, cultural references, and migration-related keywords from textual data. Additionally, it entails examining network structures and interactions within online communities. Furthermore, ethical considerations and potential challenges related to this approach are addressed, including privacy issues and algorithmic biases. Strategies for mitigating these challenges are explored, such as anonymization methods, transparent data collection practices, and rigorous validation procedures.

By delving into the realm of digital sphere, this presentation aims to deepen our understanding of immigrant experiences online, illuminating their behaviors, interactions, and identities. This has broad implications across disciplines like sociology, anthropology, and immigration studies, offering insights into the integration processes and social dynamics within immigrant communities in the digital era.