



**COMPTExT  
CONFERENCE**

# 6th Annual COMPTExT Conference

## WORKSHOPS

**Pre-Conference  
Workshops**

THURSDAY,  
MAY 2ND

**Friday**

FRIDAY, MAY  
3RD

**Saturday**

SATURDAY,  
MAY 4TH

# Welcome

We are delighted to welcome you to the 2024 COMPTExT meeting at the Vrije Universiteit (VU), Amsterdam. The 2024 meeting is being held at the VU's NU building from May 2-4, 2024.



Enjoy your conference in Amsterdam! You can either use public transportation, which is generally good in Amsterdam and requires no previous action. You can just use Google or Apple pay or contactless payment credit or debit cards to enter the metro stations, trams or buses. We recommend renting a bike, such as MacBike.

## **The COMPTExT 2024 Organising Committee consists of:**

- Mariken A.C.G. van der Velden (Vrije Universiteit Amsterdam)
- Roan Buma (Vrije Universiteit Amsterdam)
- Alona O. Dolinsky (Vrije Universiteit Amsterdam)
- Johannes B. Gruber (Vrije Universiteit Amsterdam)
- Kasper Welbers (Vrije Universiteit Amsterdam)
- Miklós Sebők (Centre for Social Sciences, Budapest)

# PROGRAM

## PRE-CONFERENCE WORKSHOPS, THURSDAY, MAY 2

**9:30 - 12:30** Morning Session

**12:30 - 13:30** Lunch Break

**13:30 - 16:30** Afternoon Session

### MORNING SESSION, THURS 2 MAY 09:30-12:30 - NU BUILDING

#### **Session 1: Intro to text analysis with R - Denise Roth - NU 02A53**

In this introductory workshop, participants will learn the basics of utilizing the programming language R for analyzing textual data. R is a powerful open-source tool for text analysis due to its extensive libraries and packages tailored for statistical computing and natural language processing. Attendees will be introduced to fundamental concepts such as data importation, text preprocessing, sentiment analysis, and machine learning. Through hands-on exercises and demonstrations, participants will gain practical skills to manipulate and analyze text data effectively, empowering them to extract valuable insights from textual sources using R. While having some general basics in R is helpful, it is not necessarily required, as the workshop is designed to be accessible to beginners. Participants are welcome to bring their own data and are encouraged to follow along using their laptops; it is recommended to install both R and R Studio on their computers beforehand to facilitate active participation.

#### **Session 2: Intro to data visualization with R - Alona Dolinsky - NU-02A55**

This workshop will provide an accessible introduction to visualisation techniques in R focusing on the ggplot2 package and relying on the broader tidyverse structure. The materials will cover a wide range of approaches including distributions, frequencies, proportions, associations between variables (+ interaction variables), time series and more complex visualisations as time permits. Participants should have basic knowledge of R and are encouraged to bring their own data to the workshop to use while following along the materials.

#### **Session 3: Intro to web scraping - Johannes Gruber - NU-02A59**

A cornerstone of computational social science has always been to work with data that was not specifically designed for data analysis, but left behind as traces of human actions and societal processes. This workshop provides a practical overview of techniques to gather web data by extracting it from the web and reshaping it into a usable form -- a process usually referred to as web scraping. Web scraping has become significantly more important, as large swaths of the formally open web are now obstructed by technological means. This workshop provides an overview of simple to advanced techniques that can be used to collect essentially all content from the web that is accessible by a human.

## **AFTERNOON SESSION, THURS 2 MAY 13:30-16:30 - NU BUILDING**

### **Session 5: Validation in Automated Text Analysis (especially in Topic Models) - Jana Bernhard - NU 02A53**

Validation is fundamental to scientific inquiry, especially when dealing with extensive unseen data. As researchers, ensuring the trustworthiness of the models used to present results and formulate recommendations is imperative. In this workshop, we will look at different methods that can be used to validate models, their advantages and potential drawbacks, and discuss what it means for a model to be valid.

### **Session 6: Practical Applications of Running Your Code in the Cloud - Kasper Welbers - NU-02A55**

Today, it is easier than ever to run your code in the cloud, and learning how to do so opens many doors as a computational social scientist. For instance, you can perform heavy and long running computations on a server, automate web scrapers to run daily, create your own API, or host a web application for conducting online experiments. In this workshop we will look at various options, including some platforms with a generous free tier, that you can get started with right away.

### **Session 7: Leveraging Parameter-Efficient, Small-Scale Models with Adapters for Social Science Research using Python - Christopher Klamm - NU-02A59**

Language Models have become an essential tool in social science research for analyzing vast amounts of data. However, their training requires significant computing resources, which can present monetary and environmental challenges. In this workshop, we will discuss the potential of using small-scale, parameter-efficient models for social science research. We will demonstrate how to use the Adapter framework for parameter-efficient models using Python and show a practical hands-on example with topic classification.

### **Session 8: MEXCA - Introduction to a tool for Multimodal Emotional eXpression Capture - Gijs Schumacher and Malte Luken - TBD**

MEXCA is a full pipeline for extracting facial expressions, vocal characteristics and text sentiment from debates or conversations between multiple people. The pipeline distinguishes who talks and who is shown in the frame. This way it is well-suited to analyze raw video materials of debates and conversations between politicians or citizens. In the workshop we present the scientific goals of MEXCA and we will do a hands-on introduction of the software.