# Introduction to web scraping and text analysis in R

| | |
|---|---|
| **Tutor** | Dr Nicolas Attalides |
| **Organization** | Digital Skills, University of Lucerne |
| **Language** | English |
| **ECTS-Points** | |
| **Contact** | nadia.buehler@unilu.ch |
| **Dates and time** | Online<br>Friday 2nd June (Day 1)<br>Friday 9th June Friday (Day 2)<br>9:30 – 16:30 |
| **Content** | The R programming language offers a huge variety of statistical analysis solutions with over 16000 packages available to install and continues to expand in areas like visualization, text analysis and machine learning.<br><br>This course focuses on collecting data from the web applying a technique called web scraping. You will learn how to use R packages that specialise in processing HTML to gather data. During the course, we will explore the area of Natural Language Processing and perform data manipulations and semantic analysis on the text data that we collected.<br><br>This course is split into two parts:<br>Part 1:<br>• Introduction to the basic concepts of HTML structures of a webpage<br>• Learn how to collect text data using web scraping R packages<br>Part 2:<br>• Introduction to Natural Language Processing (NLP) and some key concepts<br>• Learn how to do text mining and text manipulations<br>• Learn how to use tokenization and stop words and how to perform sentiment analysis |
| **Prerequisites/ Materials** | Course participants are expected to have some basic knowledge of the R programming language. Prior experience in basic data analysis (such as data manipulation and visualisation) would help the learning experience but is not required.<br><br>Participants should have their own laptop with R, RStudio and the relevant packages installed. Instructions for the technical setup will be circulated by the instructor before the course. Learning material such as slides, code and solutions to exercises will be circulated by the instructor after the course. |
| **Teaching method** | This course includes a range of activities such as web-scraping demos, live-coding sessions, interactive quizzes, and practical exercises to work individually or in a group. Active participation and contribution are recommended. |