WORKSHOP ON COMPUTATIONAL TEXT ANALYSIS

This intensive one-day workshop in computer-assisted text analysis is organized with the collaboration of the UNESCO Chair on Open Science from the University of Montreal as well as the Danish Center for Studies on Research and Research Policy from the University of Aarhus With the exponential growth of textual documents in digital format, computational models, methods, and tools for text analysis are quickly becoming an essential part of research and investigative procedures. This workshop aims to initiate data analysts working at the Rigsrevisionen to the theoretical, methodological, and implementational aspects of computational text analysis in R. More precisely, its objective is to allow the participants to acquire basic declarative and procedural knowledge of the different operations involved in computational text analysis: constitution of the corpus, preprocessing, modeling, analysis, evaluation and visualization.

Following a general presentation of computational text analysis and the distributional linguistics on which it relies, the workshop will focus in the morning on the first part of the processing pipeline, which begins with corpus constitution and ends with the building of text matrices. In the afternoon, different analysis methods based on vectorial semantics will be presented, with a special focus information retrieval, co-occurrence analysis and topic modelling. For each step of the processing pipeline, theory and examples will be presented to the participants, followed by step-by-step, easy-to-follow exercises tailored to the needs of the participants and allowing them to hone their skills on text data of high relevance to their professional activities. All teaching material will be made available to the participants in a Github repository, which they will be able to access at any given time during and after the workshop.