

Gregor Wiedemann

Abstract

Integrating Text Mining into Qualitative Data Analysis for Social Sciences

Two developments in computational text analysis widen opportunities of performing qualitative data analysis in social sciences: 1. amounts of digital text worth investigating are growing rapidly, and 2. improvements in algorithmic extraction of semantic structures, also called text mining, allows for further bridging the gap between qualitative and quantitative text analysis. The key factor here is the inclusion of context into computational linguistic models which extends conventional computer-assisted content analysis towards the extraction of meaning. But, to benefit from the heterogeneous set of text mining applications in the light of social science requirements, there is a demand for

- a) conceptual integration of consciously selected methods,
- b) systematic optimization of algorithms and workflows, and
- c) methodological reflections with respect to empirical social research.

This dissertation introduces an integrated workflow of text mining applications to support qualitative data analysis of large scale document collections. To answer abstract social science research questions, it combines several technologies to approach three distinctive sub-tasks:

1. retrieval of relevant documents from large corpora,
2. inductive exploration of retrieved sub-collections, and
3. (semi-)automatic classification of text with content analytic categories.

Technologies used in these workflows are adapted and optimized with respect to requirements from the social science perspective. Their usefulness is shown in an example study to answer a political science question on the subject of “democratic demarcation” in Germany. For this, a corpus of around 600,000 newspaper documents is analyzed. Finally, insights based on requirements, implementation and application of the exemplary analysis workflow are generalized to a methodological framework to support qualitative data analysts by employing various types of text mining. The proposed V-TM framework covers research design recommendations along with evaluation requirements on hierarchical abstraction levels considering technical, methodical and epistemological aspects.