

## WWZ Research Seminar

Tuesday, May 21, 2019, 12:30pm

**„Is text the new data? First experience with conference abstracts“**

**Peter Winker – 25.4.2019**

### Abstract

The use of textual information gained interest and momentum over the last years also in economics. Sometimes, text is considered as the new data in fields covering financial markets (sentiment analysis based on analysts statements, perception of communication of central banks), innovation activities (patent abstracts, newsticker data, websites), or to the history of economic science (topics covered in journal articles) to name only a few recent applications.

However, in order to draw meaningful conclusions from this type of data, a substantial set of steps in processing and analyzing the data has to be taken. Some of them are based on previous experience, some on statistical methods, and some might depend on human judgement. Thus, issues present when dealing with conventional quantitative data (e.g. questionnaire design, response rates, validity, seasonal adjustment, trend extraction etc.) will not cease to exist when considering textual information. They might just be disguised differently, while new challenges show up.

The talk will try to provide an overview of relevant steps in using textual data in a time series context. The abstracts of a conference series in computational finance and statistics will serve as an example. In particular, the following steps will be addressed: 1) selection of appropriate sources (corpora) and establishing access, 2) preparation of the text data for further analysis, 3) identification of themes within documents, 4) quantifying the relevance of themes in different documents, 5) aggregating relevance information over time. Finally some remarks on the use of the generated indicators in further analysis will be provided and a substantial set of open issues regarding, e.g. computational complexity and robustness of the methods might serve as input for the discussion.