

Text Mining for Relationship Extraction

Summary

In qualitative research natural language texts are often annotated to extract common themes or concepts. The relationships between these concepts are usually only informally defined, if defined at all. This thesis shall first define types of relationships that frequently occur between the concepts (codes) in a code system within the context of qualitative data analysis (QDA) and then explore options within the field of natural language processing (NLP) and information retrieval (IR) to extract such relationships.

Work Results

- Literature review
 - The basics of QDA should be reviewed.
 - State of the art on information extraction (IE)
- Research approach and execution
 - Analyse multiple code systems in regards to what types of relationships may exist
 - Create relationships manually, and validate these through user feedback
 - Let multiple people extract relationships for the same code system and identify common themes
 - Document a manual extraction method
 - Evaluate whether the typical IE methods can be applied
 - Build a taxonomy using lexical databases, grammatical or statistical rules or other NLP related tools, focusing on the relationships between codes
 - Evaluate the performance against manually generated relationships
- Research results
 - A well defined concept for relationships within code systems
 - A tool capable of extracting these relationships

Supervisor

Andreas Kaufmann, M. Sc. andreas.kaufmann@fau.de

Prof. Dr. Dirk Riehle, dirk.riehle@fau.de

Open Source Research Group
Computer Science Department
Friedrich-Alexander University

More information: <http://osr.cs.fau.de/theses/resources/>

