gesis Leibniz Institute for the Social Sciences

hhu Heinrich Heine Universität Düsseldorf

Master Thesis

Short-Text Classification: The Aboutness of Survey Questions

GESIS maintains one of the largest archives of social science surveys, containing over 3000 surveys with over 50.000 questions. To allow easy searching and filtering within the question database, we want to enrich all questions with meaningful attributes. Researchers could, for example, be interested in filtering the topic of the question, what type of information it is asking for, or the answer type for the following question:

Question: Some vaccinations are compulsory. Do you consider this ..?
Answer Options: Perfectly acceptable
Fairly acceptable
Not very acceptable
Not at all acceptable
DK (Don't know)
Not ascertained

Topic: Public Health
Asking for: an evaluation
Answer type: ordinal

First attempts to automatise the attribution prediction process have shown promising results; however, the precision of prediction models is not yet at a level acceptable for a productive system. We have a manually annotated subset of 6000 data points from the question database serving as the basis for our exploration. In this project, we want to focus on the abstract concept of "aboutness" of survey questions [2] on two levels (3 and 17 categories, respectively).

This project includes manually analyzing the dataset, preprocessing the data with state-of-the-art NLP techniques, choosing, implementing, and training prediction models, and evaluating the model performances statistically and intellectually.

Contact:
Andrea Papenmeier
GESIS - Leibniz Institute for Social Sciences
Unter Sachsenhausen 6-8, 50667 Cologne
E-Mail: andrea.papenmeier@gesis.org
Phone: +49 (221) 47694-229

Prof. Dr. Stefan Dietze
Heinrich Heine University Düsseldorf - Data & Knowledge Engineering
Universitätsstr. 1 Building: 25.12 Floor/Room: 01.39, 40225 Düsseldorf
E-Mail: stefan.dietze@hhu.de
Phone: +49 (211) 81-13785

Introductory Readings:

[1] Friedrich, T., & Siegers, P. (2016*). The Ofness and Aboutness of Survey Data: Improved Indexing of Social Science Questionnaires.* In Analysis of Large and Complex Data (pp. 629-638). Springer, Cham.