

Exploring NLP Technologies for Cultural Knowledge Graph Engineering

Objective of this work:

With the advent of the World Wide Web, the world is being overloaded with huge data, that carries potential information, which, once extracted, can be of a big use for both scientists and general public. The important role among this data play historical records and articles, that contain information on important people, locations and events. Since the process of manual information tracking in such resources is very laborious and time-consuming, the technologies for automatic information extraction have become popular in the academia.

In this work, existing technologies for digitizing textual documents, challenges of entities and relation identification, as well as the problems of cultural Knowledge Graph construction will be investigated.

The dataset of the thesis contains articles in German from all areas of Nuremberg's history published in the annual "Mitteilungen des Vereins für Geschichte der Stadt Nürnberg" (MVGN) [1]. The data is not available in digital textual form, thus the first step of the work will be to use the existing OCR technologies for digitization. In order to identify important entities, relations, and classes (e.g. persons, organisations, locations) the digital natural language text will then be structured using state-of-the-art NLP technologies, e.g. Named Entity Recognition, Relation Extraction, Type Inference, etc. As the final contribution the obtained data will be linked to external resources, e.g. Wikidata [2] and GND [3], and integrated into the Knowledge Graph.



[1] https://www.nuernberg.de/internet/stadtarchiv/vgn_publicationen_mvgn.html

[2] <https://www.wikidata.org/>

[3] https://www.dnb.de/EN/Professionell/Standardisierung/GND/gnd_node.html

The project work will be supervised by **Prof. Dr. Harald Sack, Tabea Tietz and Oleksandra Vsesviatska, Information Service Engineering at Institute AIFB, KIT, in collaboration with FIZ Karlsruhe.**

Keywords:

Knowledge Graphs, Cultural Heritage, NLP

Pre-requisites:

Knowledge of Programming with Python

Contact persons:

Tabea Tietz

tabea.tietz@fiz-karlsruhe.de

Oleksandra Vsesviatska

oleksandra.vsesviatska@fiz-karlsruhe.de