

Announcement

09.11.2020

BACHELOR OR MASTER THESIS:

Automatic Summarization and Summary Evaluation for German

DESCRIPTION

In recent years, the need for multi-document summarization has increased enormously trying to process the ever-growing volume of information on the world wide web. Because of the fact that manual creation of text summaries by professionals is labor intensive, difficult and expensive to repeat, alternative methods like crowd-based summarization or automated summarization has been tackled in the last decade. Especially in case of automatic summarization and summarization evaluation, most currently available automated tools suffer from low summary quality. To overcome this challenge, the aim is in this thesis to provide an overview of all the existing automatic summarization tools and summary quality evaluation tools for The German, as well as to pre-train BERT language model for The German summarization based on crowdsourcing data.

The goal of the thesis is to develop an automatic summarization tool for The German language. To realize this, following steps are required:

- Literature analysis of current automatic summarization and summary evaluation tools.
- Extension of the BERT: Design and implementation of an automatic summarization and summary evaluation tool for the German language based on the BERT Language Model
- Testing this new tool with a real crowdsourcing task.

The thesis will be written in English.

REQUIREMENTS

- Good statistical knowledge
- Fluent in Python, good knowledge of machine learning libraries
- NLP knowledge is a plus
- Experience with RESTful API and JSON
- Interest in crowdsourcing
- Course of study: computer science, computer engineering, industrial engineering, information systems management, media computer science
- English is required, German is a plus.

CONTACT

Neslihan Iskender neslihan.iskender@tu-berlin.de, Dr.Ing Tim Polzehl tim.polzehl1@tu-berlin.de