Introduction to the German Text Summarization Challenge

With the rise of deep learning, automatic text summarization has made promising progress. However, it is still an unsolved problem and an open research topic. At SwissText 2019 we aimed to explore ideas and solutions regarding summarization of German texts. For that we invited participants to the 1st German Text Summarization Challenge. The results were presented at the conference.

With this challenge we created a basis for interesting discussions and hope to have taken the NLP community for German text understanding a step further.

For the challenge, we provided the participants with 100,000 texts together with reference summaries extracted from the German Wikipedia. The aim was to generate abstractive summarizations. In order to avoid steering the development towards certain metrics, we did not release our evaluation metrics until the end of the challenge.

For our evaluation we adapted the English ROUGE-package to the German language. Specifically, we use the same processing with German stemming and stop words and additionally split up compound words. The resulting scores offer an estimate of the summary quality. However, it cannot replace human judgment and must not be viewed as precise measurement. ROUGE does not adequately capture frequent abstractive summarization errors such as word repetitions or false facts. For this reason, we did not rank the submissions. In the following table are the scores of all participants.

Team	ROUGE-1	ROUGE-2
Shantipriya Parida, and Petr Motlicek (s2)	40.16	22.17
Dmitrii Aksenov, Georg Rehm, Julian Moreno Schneider	40.35	21.86
Nikola Nikolov	34.66	19.33
Valentin Venzin, Jan Deriu, Didier Orel, Mark Cieliebak	39.78	23.41
Pascal Fecht	40.89	23.46

The following sections contain the descriptions submitted by the participants. They aim not to represent scientific papers but are systems descriptions of the developed methods.