



## Domain-specific Text Summarization

### Motivation

Text summarization aims at distilling the essential information from a text to produce a shorter version, such as generating headlines for news and subject lines for emails. Recent summarization methods use deep learning models that are trained and evaluated on English and standard corpora. However, to what extent these models can generalize to other languages like German and technical domains with rare words have not been explored. Specifically, summarizing technical texts is challenging as we require expertise and deep understanding in the domain of interest.

This project aims to explore and develop transfer learning methods to generate a sentence summary from a given textual snippet (a news article or an email written in German on a technical domain).

### Task Description

- Literature review: Exploring and investigating unsupervised and transfer learning methods for text summarization
- Implementation: Developing a text summarization method to transfer from pretrained neural summarization models to new languages and domains
- Evaluation: Evaluating the proposed method on benchmark datasets
- System demonstration: Building an interactive demo of the proposed method

### References

- Philippe Laban, Andrew Hsi, John Canny, and Marti A Hearst. 2020. The Summary Loop: Learning to Write Abstractive Summaries Without Examples. 5135–5150.
- Nabil Hossain, Marjan Ghazvininejad, and Luke Zettlemoyer. 2020. Simple and Effective Retrieve-Edit-Rerank Text Generation. 2532–2538.
- Rui Zhang and Joel Tetreault. 2019. This Email Could Save Your Life: Introducing the Task of Email Subject Line Generation.
- LongSumm 2021: The 2nd Shared Task on Generating Long Summaries for Scientific Documents.

### Contact

Analysis



Programming



Literature



Prof. Dr. Iryna Gurevych

Thy Thy Tran, PhD  
Dr. Mohsen Mesgar

[thesis@ukp.informatik.tu-darmstadt.de](mailto:thesis@ukp.informatik.tu-darmstadt.de)