

# Posudek oponenta diplomové práce

Autor/autorka práce: **Bc. Michal Sejak**

Název práce: **Summarization of News Articles**

## Obsah práce

The author uses a newly constructed Czech summarization dataset to evaluate the performance of various state-of-the-art neural-network-based systems on the task of extractive summarization of news articles. The models are based on BERT and Longformer architectures, pretrained for Czech language. The authors conclude that for unknown datasets, BERT-based setup is likely more suitable, while when the structural properties of the data (e.g. length and number of sentences) is approximately known in advance, Longformer-based models can achieve better results.

## Kvalita řešení a dosažených výsledků

In the first half of the thesis, which introduces the used concepts and related work, the author shows a clear understanding of the underlying techniques and algorithms and a good overview of recent improvements of the key methods in this research field. As for the experiments, the evaluation metrics are well selected. I welcome the use of the BERT score on top of the classical (but problematic) ROUGE family evaluation measures. For the classification and clustering, the model choices and parameter choices are very well reasoned and alternatives discussed, which shows the author's proficiency in the topic. From the sheer volume of experiments conducted, the work is impressive in terms of how quickly the author was able to grasp and implement the number of techniques needed.

## Formální úroveň

The thesis is very well written, showing proficiency in scientific English writing as well as a clear understanding of the research questions, possibilities to solve them, and choices made along the path of answering the research questions. The work is very well structured and easy to follow.

## Práce s literaturou

An appropriate level of state-of-the-art literature is covered and correctly explained in detail.

## Splnění zadání

The author has fulfilled all the given requirements, in a scope and quality which exceeds the expectations.

## Dotazy k práci

What data was the CZERT pretrained on? Is it possible that some of the news articles in the test set overlap with the documents seen in pretraining? If so, do you think this would be an issue?

How many documents form one summarization event on average? What is the impact of the number of documents on the quality of the summarization?

In Section 3.1 you are listing the size of the training and evaluation datasets. Have you used any separate development set for the training or have you picked the best-performing model directly on the test set? If so, what can be the drawbacks of this?

Navrhuji hodnocení známkou **výborně** a práci doporučuji k obhajobě.

Marburg, 10.6.2022

Prof. Dr. Lucie Flek