MUNI
FI

MASARYKOVA UNIVERZITA
FAKULTA INFORMATIKY
BOTANICKÁ 68A, 602 00 BRNO
IČ: 00216224
DIČ: CZ00216224

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

Student:	Petr Zelina
Program:	Informatika
Obor:	Umělá inteligence a zpracování přirozeného jazyka
Garant oboru:	prof. PhDr. Karel Pala, CSc. (UMI)
Vedoucí práce:	doc. RNDr. Aleš Horák, Ph.D.
Katedra:	Katedra strojového učení a zpracování dat
Název práce:	Pretraining and Evaluation of Czech ALBERT Language Model
Název práce anglicky:	Pretraining and Evaluation of Czech ALBERT Language Model
Zadání:	A new powerful language model called ALBERT was presented by Google in 2019 when it reached state-of-the-art results in various NLP tasks for English (question answering, language understanding). ALBERT is based on the idea of transfer learning, where it requires a huge text dataset for pretraining general language representations and in successive separate training it can be fine tuned for a specific task. The goal of this thesis is to pretrain a ALBERT model(s) for the Czech language, evaluate the model efficiency in several extrinsic tasks and compare the results to other current state-of-the-art Czech NLP language models. The submitted thesis will consist of a thesis text and a practical part. The textual part will include a detailed description of ALBERT and other related language modelling approaches and information about available pretrained models suitable for comparison, a description of the pretraining process, and the details of the evaluation and comparison. The practical part will contain all code and data for replicating the results with the respective licenses for publication where possible.
Literatura:	Lan, Zhenzhong, et al. ”Albert: A lite BERT for self-supervised learning of language representations.” arXiv preprint arXiv:1909.11942 (2019). Devlin, Jacob, et al. ”Bert: Pre-training of deep bidirectional transformers for language understanding.” arXiv preprint arXiv:1810.04805 (2018). Liu, Yinhan, et al. ”Roberta: A robustly optimized BERT pretraining approach.” arXiv preprint arXiv:1907.11692 (2019). Peters, Matthew E., et al. ”Deep contextualized word representations.” arXiv preprint arXiv:1802.05365 (2018). Radford, Alec, et al. ”Better language models and their implications.” OpenAI blog (2019). Zellers, Rowan, et al. ”Defending Against Neural Fake News.” arXiv preprint arXiv:1905.12616 (2019).

Zadání bylo schváleno prostřednictvím IS MU.