



**FACULTY
OF MATHEMATICS
AND PHYSICS**
Charles University

MASTER THESIS

Bc. Petra Vysušilová

Czech NLP Processing with Contextualized Embeddings

Institute of Formal and Applied Linguistics

Supervisor of the master thesis: RNDr. Milan Straka, Ph.D.

Study programme: Computer science

Study branch: Artificial intelligence

Prague 2020

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In date

signature of the author

Dedication.

Title: Czech NLP Processing with Contextualized Embeddings

Author: Bc. Petra Vysušílová

Institute: Institute of Formal and Applied Linguistics

Supervisor: RNDr. Milan Straka, Ph.D., Institute of Formal and Applied Linguistics

Abstract: Recently, several methods for unsupervised pre-training of contextualized word embeddings have been proposed, most importantly the BERT model (Devlin et al., 2018). Such contextualized representations have been extremely useful as additional features in many NLP tasks like morphosyntactic analysis, entity recognition or text classification.

Most of the evaluation have been carried out on English. However, several of the released models have been pre-trained on many languages including Czech, like multilingual BERT or XLM-RoBERTa (Conneau et al, 2019). Therefore, the goal of this thesis is to perform experiments quantifying improvements of employing pre-trained contextualized representation in Czech natural language processing.

Keywords: key words

Contents

Introduction	2
1 Title of the first chapter	3
1.1 Title of the first subchapter of the first chapter	3
1.2 Title of the second subchapter of the first chapter	3
2 Title of the second chapter	4
2.1 Title of the first subchapter of the second chapter	4
2.2 Title of the second subchapter of the second chapter	4
Conclusion	5
Bibliography	6
List of Figures	7
List of Tables	8
List of Abbreviations	9
A Attachments	10
A.1 First Attachment	10

Introduction

1. Title of the first chapter

An example citation: Anděl (2007) Zonky (2018) Devlin et al. (2018)

1.1 Title of the first subchapter of the first chapter

1.2 Title of the second subchapter of the first chapter

2. Title of the second chapter

2.1 Title of the first subchapter of the second chapter

2.2 Title of the second subchapter of the second chapter

Conclusion

Bibliography

- [Anděl 2007] ANDĚL, J.: *Základy matematické statistiky*. Praha : Matfyzpress, 2007 (Druhé opravené vydání). – ISBN 80-7378-001-1
- [Devlin et al. 2018] DEVLIN, Jacob ; CHANG, Ming-Wei ; LEE, Kenton ; TOUTANOVA, Kristina: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. (2018). – URL <https://arxiv.org/abs/1810.04805>
- [Zonky 2018] ZONKY: *PARAMETRY SEKUNDÁRNÍHO TRHU (Secondary market parameters)*. URL https://zonky.cz/downloads/Zonky_Parametry_sekundarniho_trhu.pdf?v=9615aae593630e00d21588ea61a921c8924e52b93c5beb5cf7a371d86e090fdb, 2018

List of Figures

List of Tables

List of Abbreviations

A. Attachments

A.1 First Attachment