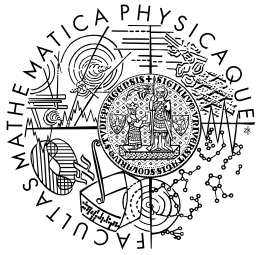

FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

MASTER THESIS

Zuzana Šimečková

Entity Relationship Extraction

Institute of Formal and Applied Linguistics

Supervisor of the master thesis: RNDr. Milan Straka, Ph.D.
Study programme: Computer Science
Study branch: IUI

Prague 2020

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources. It has not been used to obtain another or the same degree.

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 ....

Author's signature

Dedication.

Title: Entity Relationship Extraction

Author: Zuzana Šimečková

Institute: Institute of Formal and Applied Linguistics

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

Abstract: Abstract.

Keywords: key words

Contents

Introduction

There has been made noticeable progress in natural language processing since the first deep neural networks attempts. With multiple new approaches and inventions such as multitask learning, word embeddings, RNN, attention and the transformer architecture. Last year Devlin et al. [2018] created BERT and managed to achieve state-of-the-art performance in eleven natural language processing tasks, including GLUE (7.7% point absolute improvement), MultiNLI accuracy (4.6% absolute improvement) and SQuAD problems.

not a sentence

In this thesis, we will try to use those novel approaches to predict relation between two entities based on a Czech sentence. First part of this thesis will be focused on data. We will introduce some existing English datasets for Entity Relation Extraction. Than we will describe how we prepared data for Czech version of this task using distant supervision on Czech Wikipedia and Wikidata. Second part

o čem bude druhá část

1. Datasets

An example citation: Anděl [2007]

1.1 tacred dataset

1.2 SEMEVAL 2010 task 8 dataset

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

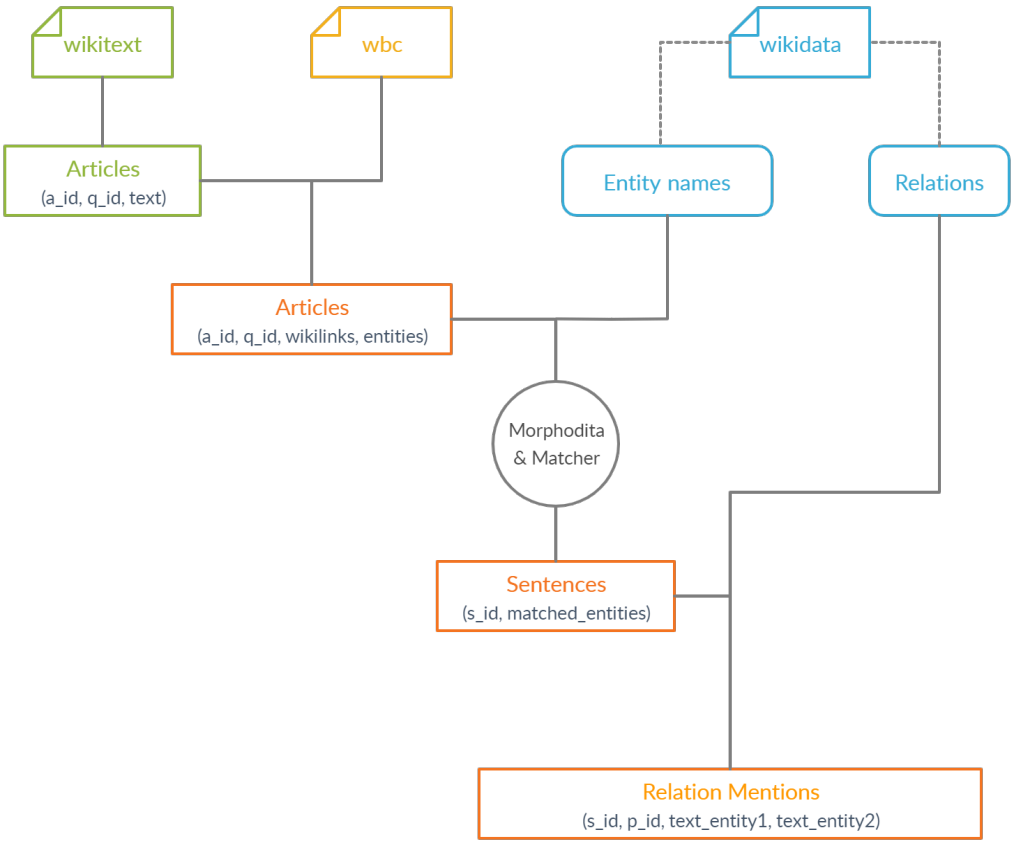

Figure 2.1: Zjednodušený diagram výroby korpusu

Conclusion

Bibliography

J. Anděl. *Základy matematické statistiky*. Druhé opravené vydání. Matfyzpress, Praha, 2007. ISBN 80-7378-001-1.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

List of Figures

List of Tables

List of Abbreviations

A. Attachments

A.1 First Attachment