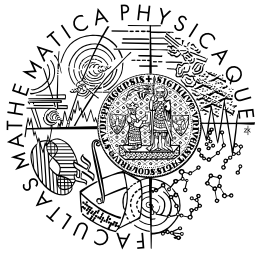

FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

MASTER THESIS

Bc. Jan Bílek

Genres classification by means of machine learning

Department of Theoretical Computer Science and Mathematical Logic

Supervisor of the bachelor thesis:	Mgr. Roman Neruda Csc.
Study programme:	Computer Science
Study branch:	Artificial Intelligence

Prague 2018

I declare that I carried out this bachelor 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: Genres classification by means of machine learning

Author: Bc. Jan Bílek

Department: Department of Theoretical Computer Science and Mathematical Logic

Supervisor: Mgr. Roman Neruda Csc., Institute of Computer Science, The Czech Academy of Sciences

Abstract: Abstract.

Keywords: key words

Contents

1. Introduction

Approx. two pages wrappinng up following 3 sections.

Motivation

Goals

Outline

2. Background and Related Work

Word2vec[1], doc2vec[2], deep learning[3].

2.1 Text Analysis

2.2 Project Gutenberg

What is Project Gutenberg. Metadata - author and genre classification.

3. Design

Describing experiments and approaches of the thesis.

3.1 Experiment 1

3.2 Experiment 2

4. Evaluation

5. Implementation

Describes details of the implementations. We used gensim[4].

- Which algorithm is used?
- How is it deployed? Flask, Heroku $\cdots$

6. Summary

6.1 Summary and Conclusions

6.2 Future Work

Bibliography

[1] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781, 2013.

[2] Quoc V. Le and Tomas Mikolov. Distributed representations of sentences and documents. *CoRR*, abs/1405.4053, 2014.

[3] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[4] Radim Řehůřek and Petr Sojka. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta, May 2010. ELRA. `http://is.muni.cz/publication/884893/en`.