

Bachelor thesis



**Czech
Technical
University
in Prague**

NLP Trolls

Luka Peraica

**Supervisor: Ing. Radek Mařík, CSc.
April 2024**

Acknowledgements

We thank the CTU in Prague for being a very good *alma mater*.

Declaration

I declare that this work is all my own work and I have cited all sources I have used in the bibliography.

Prague, April 16, 2024

Prohlašuji, že jsem předloženou práci vypracoval samostatně, a že jsem uvedl veškerou použitou literaturu.

V Praze, 16. dubna 2024

Abstract

Keywords: manual, degree project,
 \LaTeX

Supervisor: Ing. Radek Mařík, CSc.

Abstrakt

V záplavě mnoha zdrojů a množství mediálních zpráv není jednoduché se zorientovat i pro profesionální mediální analytiku. Výrazem demokracie je i možnost se ke zprávám vyjadřovat a tříbit si názory v diskusních příspěvcích dílčích zpráv. Diskuse však vytváří prostor i pro osoby, jejichž cílem je z rozmanitých důvodů diskuse narušovat a překrucovat. Cílem práce je vytvořit komponenty systému, který umožní sledovat linie vývoje tématu a identifikovat příspěvky narušitelů, tzv. trollů.

Klíčová slova: manuál, závěrečná práce, \LaTeX

Contents

1 Introduction	1
1.1 Problem Statement	1
1.2 Structure of the Thesis	1
2 Theoretical Background	3
2.1 Sentiment Analysis	3
2.2 Topic Detection Techniques	3
2.3 Text Style Analysis	3
2.4 Troll Detection	3
3	5
4	7
5	9
6	11
A Bibliography	13

Figures

Tables



Chapter 1

Introduction

1.1 Problem Statement

In today's flood of diverse media sources and information, even professional media analysts find it challenging to navigate and filter reliable content. A key aspect of democracy is the ability to express opinions and refine perspectives through discussions on news articles. However, these online discussions also create opportunities for individuals whose goal is to disrupt and manipulate conversations for various reasons. The rise of online trolling has become a significant issue, as trolls deliberately provoke, mislead, and incite conflict, thereby spreading misinformation and fostering hostility in digital spaces.


The internet, as a central platform for communication, information sharing, and community building, is increasingly affected by this phenomenon. Studies, such as that by Fornacciari et al.[FMP⁺18], demonstrate that different types of trolls display unique behavioral patterns, emphasizing the need for diverse and adaptive detection methods. Natural Language Processing (NLP) has emerged as a crucial tool in addressing this challenge, offering methods to automatically identify and mitigate the impact of trolls. This thesis aims to develop components of a system capable of tracking the evolution of discussion topics and identifying disruptive contributions from trolls. It provides an overview of various NLP techniques for troll detection, including stylometry, topic modeling, deep learning, and transformer models.

1.2 Structure of the Thesis



Chapter 2

Theoretical Background

-  2.1 Sentiment Analysis
-  2.2 Topic Detection Techniques
-  2.3 Text Style Analysis
-  2.4 Troll Detection



Chapter 3



Chapter 4



Chapter 5



Chapter 6



Appendix A

Bibliography

- [FMP⁺18] Paolo Fornacciari, Monica Mordonini, Agostino Poggi, Laura Sani, and Michele Tomaiuolo, *A holistic system for troll detection on Twitter*, Computers in Human Behavior **89** (2018), 258–268.