

Zadání diplomové práce

Student:

Bc. Jan Kusák

Studijní program:

N2647 Informační a komunikační technologie

Studijní obor:

2612T025 Informatika a výpočetní technika

Téma:

Detekce strojově přeložených textů pomocí strojového učení
Machine Learning Detection of Machine Translated Texts

Jazyk vypracování:

čeština

Zásady pro vypracování:

Cílem práce je popsat a navrhnout klasifikátor pro textová data, který bude schopen detekovat strojově přeložené texty pomocí Google Translate a DeepL. Tento klasifikátor realizovat pomocí vhodných technologií a ověřit jeho funkčnosti vzhledem ke zvolenému datasetu.

Práce bude obsahovat:

1. Popis problematiky klasifikace textových dat.
2. Návrh vhodného klasifikátoru nebo více klasifikátorů.
3. Implementace klasifikátoru pomocí vhodných technologií a frameworků.
4. Ověření fungování navrženého postupu a porovnání výsledků s jinými metodami.

Seznam doporučené odborné literatury:

- [1] VASWANI, Ashish, Noam SHAZEER, Niki PARMAR, Jakob USZKOREIT, Llion JONES, Aidan N. GOMEZ, Łukasz KAISER a Illia POLOSUKHIN. Attention is all you need. In: Advances in neural information processing systems. 2017, s. 5998-6008.
- [2] ROTHMAN, Denis. Transformers for natural language processing: build innovative deep neural network architectures for NLP with Python, PyTorch, TensorFlow, BERT, RoBERTa, and more. Birmingham: Packt Publishing, 2021. Expert insight. ISBN 978-1800565791.
- [3] ATIENZA, Rowel. Advanced Deep Learning with TensorFlow 2 and Keras: Apply DL, GANs, VAEs, deep RL, unsupervised learning, object detection and segmentation, and more. 2nd edition. Packt Publishing, 2020. ISBN 978-1800568273.

Formální náležitosti a rozsah diplomové práce stanoví pokyny pro vypracování zveřejněné na webových stránkách fakulty.

Vedoucí diplomové práce: **prof. Ing. Jan Platoš, Ph.D.**

Datum zadání: 01.09.2022

Datum odevzdání: 04.07.2023

Garant studijního oboru: prof. RNDr. Václav Snášel, CSc.

V IS EDISON zadáno: 02.06.2023 08:32:23