

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Факультет прикладной математики и информатики

Кафедра математического моделирования и анализа данных

Аннотация к дипломной работе

**«Проектирование сервиса машинного перевода с использованием
нейронных сетей»**

Соломевич Александр Сергеевич

Научный руководитель – заведующий кафедрой математического
моделирования и анализа данных, доктор экономических наук, доцент
Малюгин В. И.

Минск, 2024

РЕФЕРАТ

Дипломная работа, 54 с., 18 рис., 3 табл., 15 источников

Ключевые слова: ОБРАБОТКА ИСКУССТВЕННОГО ЯЗЫКА, МАШИННЫЙ ПЕРЕВОД ТЕКСТА, РЕКУРРЕНТНАЯ НЕЙРОННАЯ СЕТЬ, МЕХАНИЗМ ВНИМАНИЯ, ТРАНСФОРМЕР, PYTORCH

Объект исследования – методики подготовки и предобработки текстовых данных, архитектуры для задач обработки естественного языка.

Цели работы – изучить вопросы, связанные с подготовкой данных в NLP, исследовать возможные архитектуры нейросетей для задачи Seq2Seq и, в частности, машинного перевода, спроектировать сервис для автоматического перевода на языке Python.

Методы исследования – анализ и изучение существующих архитектур и их последующее экспериментальное сравнение.

Результаты работы – результатом работы является выявление наиболее подходящей архитектуры для задачи машинного перевода и её интеграция в веб-сервис, разработанный для эксплуатации конечным пользователем.

Область применения – задачи обработки естественного языка, возможность дообучения модели на более узкоспециализированных текстах и ее дальнейшее применение в выбранном домене.

РЭФЕРАТ

Дыпломная праца, 54 ст., 18 мал., 3 табл., 15 крыніц

Ключавыя словы: АПРАЦОЎКА ШТУЧНАЯЯ МОВЫ, МАШЫННЫ ПЕРАКЛАД ТЭКСТА, РЭКУРЭНТНАЯ НЕЙРОННАЯ СЕТКА, МЕХАНІЗМ УВАГІ, ТРАНСФАРМЕР, PYTORCH

Аб'ект даследавання – методыкі падрыхтоўкі і перадапрацоўкі тэкставых дадзеных, архітэктурны для задач апрацоўкі штучнай мовы.

Мэты працы – вывучыць пытанні, звязаныя з падрыхтоўкай дадзеных у NLP, даследаваць магчымыя архітэктурны нейра-сетак для задачы Seq2Seq і, у прыватнасці, машыннага перакладу, спраектаваць сэрвіс для аўтаматычнага перакладу на мове Python.

Метады даследавання – аналіз і вывучэнне існуючых архітэктур і іх наступнае эксперыментальнае параўнанне..

Вынікі працы – вынікам працы з'яўляецца выяўленне найболей падыходнай архітэктурны для задачы машыннага перакладу і яе інтэграцыя ў вэб-сэрвіс, распрацаваны для эксплуатацыі канчатковым карыстачом.

Вобласць выкараставання – задачы апрацоўкі штучнай мовы, магчымасць данавучання мадэлі на больш вузкаспецыялізаваных тэкстах і яе далейшае ўжыванне ў абраным дамене.

ABSTRACT

Diploma thesis, 54 pp., 18 illustrations, 3 tables, 15 sources

Keywords: NATURAL LANGUAGE PROCESSING, MACHINE TRANSLATION, RECURRENT NEURAL NETWORK, ATTENTION MECHANISM, TRANSFORMER, PYTORCH

Object of the research – methods for preparing and preprocessing text data, architectures for natural language processing tasks.

Purposes of the work – explore issues related to data preparation in NLP, research possible neural network architectures for the Seq2Seq tasks and, in particular, machine translation, and design a service for automatic translation in Python.

Methods of the research – analysis and exploration of architectures and following experimental comparison of them.

The result of the work is revealing of the most suitable architecture for machine translation task and integration of it into web-service.

The scope – natural language processing tasks, the model can be finetuned on narrowly focused texts with subsequent usage in the selected domain.