

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
МЕХАНИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ
Кафедра дифференциальных уравнений и системного анализа

ПОЛОЙКО

Борис Игоревич

ОБРАБОТКА ЕСТЕСТВЕННЫХ ЯЗЫКОВ НА ОСНОВЕ
МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Аннотация магистерской диссертации
Специальность 1-31 80 03 «Математика и компьютерные науки»

Научный руководитель:
кандидат физ.-мат. наук, доцент
А. Э. Малевич

Минск, 2024

В магистерской диссертации 45 страниц, 7 иллюстраций, 13 источников, 3 приложения.

Ключевые слова: ОБРАБОТКА ЕСТЕСТВЕННЫХ ЯЗЫКОВ, ГЕНЕРАЦИЯ КРАТКОГО СОДЕРЖАНИЯ, МОДЕЛЬ ЭНКОДЕРА-ДЕКОДЕРА, ТРАНСФОРМЕРЫ

Объектом исследования диссертации являются алгоритмы генерации краткого содержания текста.

Целью работы является изучение и реализация алгоритмов генерации краткого содержания текста на русском языке и сравнение их эффективности.

Для достижения поставленной цели были использованы: литература, язык программирования Python, библиотека тензорных вычислений TensorFlow, библиотека построения моделей трансформеров для обработки естественных языков HuggingFace.

В диссертации получены следующие результаты:

1. Реализованы абстрактные и экстрактные модели для генерации краткого содержания текста на русском языке.
2. Построены метрики эффективностей реализованных моделей.

Магистерская диссертация выполнена автором самостоятельно.

The master thesis is presented in the form of an explanatory note of 45 pages, 7 figures, 13 references, 3 applications.

Keywords: NATURAL LANGUAGE PROCESSING,
AUTOMATIC TEXT SUMMARIZATION, ENCODER-DECODER MODEL,
TRANSFORMERS

The research object of the thesis are algorithms that generate summary of a given text.

The purpose of this work is to study and implement algorithms that generate summary of a given text in russian and compare their effectiveness.

To achieve the goal Python programming language, the framework for tensor computing TensorFlow and HuggingFace library for building transformer models were used.

The main results of the thesis are as follows:

1. Abstractive and extractive summarization algorithms that process text in russian are implemented.
2. Metrics are calculated for the implemented algorithms.

The master thesis was done solely by the author.