

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

Аннотация к дипломной работе

ВЕКТОРНОЕ ПРЕДСТАВЛЕНИЕ ТЕКСТОВЫХ ДАННЫХ
И ИХ ИСПОЛЬЗОВАНИЕ В ПОИСКОВОЙ СИСТЕМЕ

Лабун Екатерина Александровна

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

2023

В дипломной работе 32 страницы, 12 иллюстраций, 7 источников.

ПОИСКОВАЯ СИСТЕМА, ВЕКТОРНОЕ ПРЕДСТАВЛЕНИЕ ТЕКСТОВЫХ ДАННЫХ, ГЛУБОКИЕ НЕЙРОННЫЕ СЕТИ, РЕКУРРЕНТНЫЕ НЕЙРОННЫЕ СЕТИ, ДОЛГАЯ КРАТКОСРОЧНАЯ ПАМЯТЬ, ТРАНСФОРМЕРЫ, КОСИНУСНОЕ СХОДСТВО

Объектом исследования дипломной работы является векторное представление текстовых данных

Целью дипломной работы является изучение способов применения векторного представления в поисковой системе.

Для достижения поставленной цели были использованы: язык программирования Python, инструмент для создания моделей и приложений Hugging Face, фреймворк SentenceTransformers.

В дипломной работе получены следующие результаты:

1. Изучены способы получения векторного представления текстовых данных для последующего использования в поисковой системы.
2. Найдены способы решения поставленных задач для реализации конечного сервиса.

Дипломная работа является завершенной, поставленные задачи решены в полной мере, присутствует возможность дальнейшего развития исследований.

Дипломная работа выполнена автором самостоятельно.

Thesis project is presented in the form of an explanatory note of 32 pages, 12 figures, 7 references.

SEARCH ENGINE, VECTOR REPRESENTATION, DEEP NEURAL NETWORKS, RECURRENT NEURAL NETWORKS, LONG SHORT-TERM MEMORY, TRANSFORMERS, COSINE SIMILARITY

The research object of this thesis project is to study methods of vector representation of text data.

The purpose of this work is to find an approach to use vector representation in search engine.

To achieve the goal we used Python programming language, an instrument of creating models and applications HuggingFace, framework SentenceTransformers.

The main results of the thesis project are as follows:

1. The methods of vector representation and their use in search engine are studied.
2. The methods of creating services for search engine are found.

The thesis project is complete, all tasks have been successfully done, there is a possibility for further research and development.

The thesis project was done solely by the author.