

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Факультет прикладной математики и информатики

Кафедра биомедицинской информатики

Аннотация к дипломной работе

“Построение семантических векторных представлений текстовых документов”

Синявский Тимур Владимирович

Научный руководитель - ст. преподаватель кафедры БМИ Николаев Г. И.

Минск, 2021

РЕФЕРАТ

Дипломная работа: 43 страницы, 20 рисунков, 3 таблицы, 15 источников, 2 приложения.

Ключевые слова: ЭМБЕДДИНГИ, ТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ, КЛАСТЕРИЗАЦИЯ, КЛЮЧЕВЫЕ СЛОВА.

Объект исследования: текстовая информация.

Цель работы: реализация пайплайна тематического моделирования для данных белорусских источников.

Результат: проведено сравнение различных алгоритмов, из них выбран оптимальный, разработана система сбора статей и применения пайплайна тематического моделирования, а также сайт для визуализаций результатов.

Область применения: новостные агрегаторы, разовый анализ любого корпуса, анализ трендов.

ABSTRACT

Diploma thesis: 43 pages, 20 figures, 3 tables, 15 sources, 2 applications.

Keywords: EMBEDDINGS, TOPIC MODELING, CLUSTERING,
KEYWORDS.

Object of research: text data.

Objective: implementation of a topic modeling pipeline for data from
belarusian sources.

Result: a comparison of various algorithms was conducted and the optimal
one was selected, a system for collecting articles and applying a topic modeling
pipeline was developed, as well as a site for results visualization.

Scope: news aggregators, one-time analysis of any corpus, trend analysis.