УДК 004.89

РАЗРАБОТКА СЕРВЕРНОЙ ЧАСТИ ПРИЛОЖЕНИЯ ДЛЯ АННОТИРОВАНИЯ ТЕКСТОВ РАБОЧИХ ПРОГРАММ ДИСЦИПЛИН СЕРВИСА “КОНСТРУКТОР ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ”

Балдина Д.Д. (студент, Университет ИТМО)

Научный руководитель – кандидат физико-математических наук, доцент Хлопотов М.В. (Университет ИТМО)

Введение. В настоящее время из-за огромного количества различных сетевых ресурсов и большого объема данных все сложнее становится поиск нужной информации. По исследованиям Мадлен Сидофф, было выяснено, что люди практически не читают длинные тексты: лишь 60% опрошенных дошли до конца статьи, но время, потраченное на прочитывание ее содержания, показывает, что они скорее бегло просматривали, чем читали [1]. Именно поэтому все более актуальными становятся методы автоматического реферирования, которые позволяют с помощью современных решений в области машинного обучения, выделять ключевую информацию в объемных текстовых документах. Особую важность использования таких современных моделей машинного обучения приобретают системы в области образовательной деятельности [2]. Так, спроектированная и в дальнейшем разработанная система с автоматическим реферированием описаний рабочих программ сервиса “Конструктор образовательной программы” сможет повысить эффективность обработки текстовой информации в образовательной среде университета и поможет студентам более осознанно подходить к выбору своей образовательной траектории.

Основная часть. В качестве информационной системы было решено спроектировать и разрабатывать веб-приложение.

Платформа будет включать в себя возможность реферировать тексты по запросу из сервиса конструктора, сокращая объем и сохраняя смысл. Для реализации этой функции в конструкторе программ обучения будут добавлены графические элементы, позволяющие отправлять запросы на разрабатываемый сервис автореферирования. Кроме того, функционал приложения будет включать в себя сбор актуальных данных по описаниям программ обучения, возвращение результатов на сервис конструктора и возможность замены текущего описания на реферированный.

На данный момент были рассмотрены существующие современные методы автоматического реферирования текста и их классификация, изучен сервис “Конструктора Образовательных Программ” и собран датасет описаний рабочих программ из этого приложения с помощью API, реализованы изученные алгоритмы и модели суммаризации, а также проведен анализ результатов работы моделей и их численных показателей, где в качестве модели автореферирования была выбрана большая языковая модель RuT5.

В ходе разработки планируется использовать следующий технологический стек: DjangoREST framework (DRF) для разработки серверной части, Hugging Face API для подключения модели автореферирования, PostgreSQL с использованием инструмента Django ORM для хранения данных, Docker для развертывания приложения конструктора.

Выводы. Разработанную систему интеллектуального аннотирования описаний рабочих программ можно будет использовать для различных целей - на сайте для абитуриентов, или для выбора дисциплин.

Список использованных источников:

1. How People Read Short Articles [Original Research] [Электронный ресурс] // Madeleine Sidoff – URL: https://speero.com/post/how-people-read-short-articles-original-research (дата обращения: 13.02.2024).

2. Мухамадиева Кибриё Бахадировна. Машинное обучение в совершенствовании образовательной среды. – Узбекистан: Бухарский инженерно-технологический институт, 20. – С. 70-77. (дата обращения: 14.02.2024)