

УДК 004.934

ИСПОЛЬЗОВАНИЕ МЕТОДОВ АНАЛИЗА ГОЛОСА ДЛЯ РЕАЛИЗАЦИИ ФУНКЦИОНАЛЬНОСТИ ПО ТРАНСКРИБИРОВАНИЮ

Антоненко А.Е. (ИТМО)

Научный руководитель – кандидат технических наук Болдырева Е.А.
(ИТМО)

Введение. В связи с развитием технологий и ростом использования аудио-контента, задача по транскрибированию звуковых записей и последующему анализу этих данных приобретает всё более актуальный характер. Традиционные методы оказываются трудоемкими, поэтому требуются методы автоматического или полуавтоматического анализа голоса и распознавания речи для эффективного транскрибирования. Использование средств интеллектуального анализа аудио данных может обрабатывать большие объемы информации, применяясь в интерактивных сервисах, таких как подкасты.

Обзор существующих методов транскрибирования, сравнительный анализ сервисов для создания транскрипций, а также проектирование системы интеграции сервиса транскрибирования, направлены на выявление оптимальных решений и обеспечение эффективной интеграции. Этот комплексный подход не только позволяет оценить текущие методы, но и создать основу для новых, более точных и автоматизированных решений в сфере транскрибирования. Исследование вносит вклад в техническую область и имеет практическое значение, цель которого – интеграция методов голосового анализа с веб-приложениями для расширения возможностей по взаимодействию с медиа-контентом.

Основная часть. Существуют различные методологии, а также “speech-to-text” [1] сервисов для создания транскрипций аудио-контента. Основной целью данной задачи является распознавание голосового содержимого аудиофайла и преобразование в текстовый формат. Кроме того, дополнительные задачи включают в себя создание временных меток слов или предложений, выделение ораторов и определение основной тематики текста. Решение этих задач может осуществляться с применением ручных, автоматизированных методов или их комбинации [2].

В настоящей статье проведен анализ четырёх сервисов для транскрибирования аудио – Vosk, SymlAI, AssemblyAI и RevAI. Анализ осуществлен с точки зрения функционала в обработке различных языковых особенностей и возможностей интеграции. Исследование базируется на анализе публично доступных данных, технической документации и личном опыте использования [3]. Основные параметры анализа включают точность транскрибирования, способность обработки разнообразных языковых контекстов и гибкость интеграции. Исследование позволяет выявить преимущества и ограничения каждого инструмента, что поможет определить наилучший выбор в конкретных сценариях использования.

Работа над проектированием системы была разделена на несколько этапов: проектирование архитектуры, разработка базы данных, создание интерфейса пользователя, разработка API серверного приложения и окончательная реализация веб-приложения. Были выделены основные бизнес-сущности, спроектирована инфраструктура [4] и база данных приложения, а также был разработан дизайн интерфейса, который предоставляет пользователям легко загружать аудио, создавать, просматривать транскрипции, и выполнять поиск по контенту. Для реализации необходимого функционала по транскрибированию было разработано смешанное решение на базе AssemblyAI.

Выводы. Проведённый анализ сервисов транскрибирования подтвердил наличие многочисленных инструментов для этой задачи, многие из которых являются платными. Необходимо тщательно выбирать инструменты в зависимости от конкретных требований задачи, учитывая их функциональные возможности, стоимость и качество предоставляемых

услуг. Предложенное проектирование приложения по транскрибированию аудио предоставляет комплексное и удобное решение для пользователей. Архитектура приложения, интерфейс пользователя и функциональность спроектированы с учетом современных тенденций в разработке приложений и предоставляют пользователю инструменты для управления и взаимодействия с контентом. Ожидается, что данное приложение будет способствовать повышению доступности и вовлеченности аудитории в мир подкастов.

Список использованных источников:

1. Trivedi A., Pant N., Shah P., Sonik S., “Speech to text and text to speech recognition systems-Areview” // Journal of Computer Engineering (IOSR-JCE) . – 2018. С. 36-43.
2. Лапицкая С.А. Разработка алгоритмов распознавания интонаций речи //В сборнике: 77-я научная конференция студентов и аспирантов Белорусского государственного университета. материалы конференции в 3 ч.. Минск. – 2020. С. 83-87.
3. Вашкевич Г.С., Азаров И.С. Детектирование речевой активности на фоне шумов при помощи свёрточной нейронной сети //В сборнике: Цифровая обработка сигналов и ее применение. dsra - 2020. Доклады на 22-ой Международной конференции. Сер. "Цифровая обработка сигналов и её применение" Москва. – 2020. С. 546-550.
4. Martin Fowler Microservices [Электронный ресурс] URL: <https://martinfowler.com/articles/microservices.html> (дата обращения: 26.01.2023).