УДК 004.934.8

П.В. Травничева, А.А. Козлова

Витебский государственный университет им. П.М. Машерова, г. Витебск, Беларусь

РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ, ПРЕДНАЗНАЧЕННОГО ДЛЯ ОПРЕДЕЛЕНИЯ ПРИНАДЛЕЖНОСТИ ФРАГМЕНТОВ ГОЛОСОВОЙ ИНФОРМАЦИИ

В современном мире голосовые технологии становятся неотъемлемой частью множества сфер деятельности, включая безопасность, искусственный интеллект, анализ данных и судебную экспертизу. С увеличением объемов голосовых данных возрастает потребность в эффективных методах их анализа и идентификации. Развитие технологий искусственного интеллекта и машинного обучения позволяет значительно повысить точность обработки и классификации аудиозаписей, что особенно актуально в области биометрической идентификации и судебной лингвистики.

Современные системы распознавания речи преимущественно ориентированы на конвертацию аудиофайлов в текст, автоматический перевод и анализ речи [1]. Однако они не всегда способны идентифицировать источник голосовой информации, что может быть критически важным в юридических и криминалистических исследованиях, а также в системах защиты персональных данных. В связи с этим возникает необходимость в разработке программного обеспечения, способного определять принадлежность голосовых фрагментов конкретным говорящим.

Цель работы – разработка программного обеспечения для распознавания речи, обеспечивающего точный анализ голосов, автоматическую генерацию временных меток, повышение точности распознавания, соответствие современным требованиям безопасности и универсальность применения в различных сферах.

Материал и методы. Материал исследования – аудиозаписи, содержащие диалоги участников. В работе используются следующие методы: анализ и синтез, практическая реализация.

Результаты и их обсуждение. Перед разработкой было изучено множество существующих технологий распознавания речи, включая алгоритмы машинного обучения, методы обработки сигналов и библиотеки.

В результате изучения были выбраны следующие технологии:

– SpeechBrain;

– Logistic Regression;

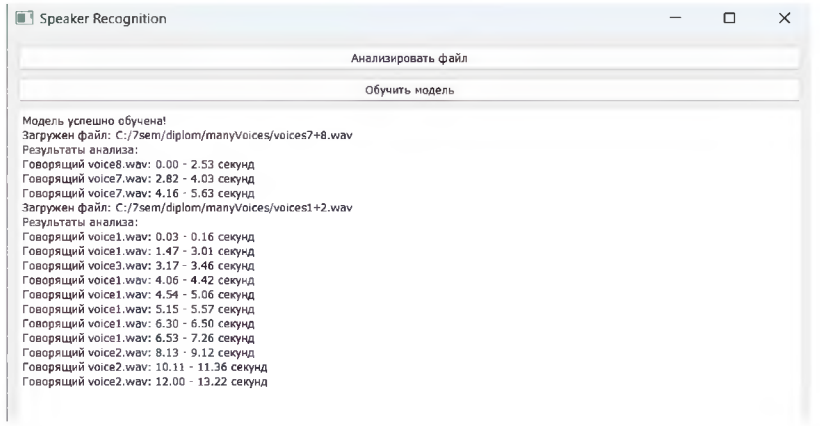
– PyQt5;

SpeechBrain: эта библиотека предоставляет предобученные модели, такие как ECAPA-TDNN, которые эффективны для извлечения эмбеддингов голосов. Она была выбрана за свою высокую точность и возможность работы с многоканальными аудиоданными [2].

Logistic Regression: для классификации говорящих была выбрана логистическая регрессия. Этот алгоритм прост в реализации и обеспечивал хорошую производительность на малых наборов данных [3].

PyQt5: библиотека была выбрана за свою гибкость и широкие возможности для создания адаптивных интерфейсов.

Интерфейс приложения для распознавания речи разработан с акцентом на удобство и интуитивность использования (рис.).



Графический интерфейс приложения

Кнопка «Загрузить аудиофайл» позволяет пользователю выбрать и загрузить WAV-файл из файловой системы. При нажатии открывается диалоговое окно, где пользователь может выбрать нужный файл.

Текстовое поле QTextEdit используется для отображения результатов анализа и сообщений об ошибках.

Кнопка «Обучить модель» позволяет пользователю инициировать процесс обучения модели на заранее подготовленных аудиозаписях.

Заключение. В разрабатываемом программном обеспечении предпринята попытка создать удобный и эффективный инструмент для распознавания речи и идентификации говорящих на основе аудиозаписей.

Литература

1. *Костюков И.В.* Разработка программного обеспечения для распознавания речи и идентификации говорящих. М.: Наука, 2020.
2. *Шафеев А.П., Соловьев Д.В.* Технологии обработки аудиоданных для распознавания речи. СПб.: БХВ-Петербург, 2019.
3. *Лебедев С.Н.* Программирование на Python для обработки сигналов. М.: ДМК Пресс, 2021.