Г. В. Грожик
(ГГУ имени Ф. Скорины, Гомель)
Науч. рук. **Е. И. Сукач**, канд. техн. наук, доцент

ПРЕОБРАЗОВАНИЕ РЕЧИ В ТЕКСТ С ПОМОЩЬЮ GOOGLE SPEECH TO TEXT API

Мессенджеры и социальные сети являются активно развивающейся частью цифровой сферы в современном мире. Они применяются для работы, общения с друзьями и коллегами, получения профессиональных навыков, ведения научной, социальной, экономической деятельности, продвижения различного рода товаров и услуг [1]. Простой и удобный интерфейс чата и полезные функции, которые обеспечивают эффективность коммуникации и обмена информацией, служат основой востребованности у потенциальных пользователей.

Google Speech-to-Text API – облачный сервис, предоставляющий возможность пользователям преобразовывать аудио в текстовые расшифровки. Он основывается на современных моделях и алгоритмах машинного обучения. Сервис поддерживает свыше 120 языков и их вариантов, успешно распознавая различные диалекты, акценты и контексты. Он также может работать в шумной обстановке, с несколькими динамиками и сложной терминологией.

Некоторые функции и преимущества Google Speech-to-Text API:

– настройка: пользователи могут настроить модель распознавания речи в соответствии со своими конкретными потребностями и областями, такими как медицина, юриспруденция, финансы и другие сферы. Они также могут предоставлять подсказки, фразы или слова, имеющие отношение к их аудиоконтенту, чтобы повысить точность и релевантность транскрипции;

– адаптация: пользователи могут включить функцию адаптации речи, которая позволяет API учиться на отзывах и данных пользователя и соответствующим образом корректировать транскрипцию. Это уменьшит количество ошибок и улучшит качество вывода с течением времени;

– потоковая передача: пользователи могут передавать аудиоданные в API в режиме реального времени и получать транскрипцию по мере обработки звука. Это может включить субтитры в реальном времени, субтитры или расшифровку событий, встреч или звонков;

– метаданные: пользователи могут получить доступ к дополнительной информации о транскрипции, такой как показатели достоверности, временные метки слов, дневниковое описание говорящего, пунктуация и использование заглавных букв. Данная функция может помочь пользователям анализировать, редактировать или же форматировать текстовый вывод по мере необходимости;

– интеграция: пользователи могут легко интегрировать API с другими сервисами Google, такими как Google Cloud Storage, Google Translate, Google Dialogflow или Google Assistant. Это может позволить пользователям создавать бесшовные и интерактивные приложения, такие как голосовой поиск, голосовые команды, голосовой чат или голосовая аналитика.

Google Speech-to-Text API можно применять в разных задачах и случаях, таких как:

– расшифровка аудио- и видеозаписей. Пользователи могут загружать свои аудио- и видеофайлы в Google Cloud Storage и использовать API для расшифровки их в текст. Это может быть полезно для создания субтитров, подписей, стенограмм или аннотаций подкастов, лекций, интервью или документальных фильмов;

– перевод аудио- или видеопотоков в реальном времени. Пользователи могут передавать аудио- и видеоданные в реальном времени в API, а также получать транскрипцию в режиме реального времени. Это может быть полезно для предоставления

титров в реальном времени, субтитров или же стенограмм вебинаров, конференций, презентаций или трансляций;

– преобразование речи в текст с переводом. Пользователи могут использовать API в сочетании с Google Translate для преобразования речи с одного языка в текст на другом языке. Данная функция может быть полезна для облегчения межкультурного общения, образования или туризма;

– анализ речи в текст. Пользователи могут использовать API в сочетании с Google Dialogflow, Google Natural Language или Google Cloud Speech-to-Text для анализа текстового вывода системы распознавания речи. Эта функция может быть полезна для извлечения идей, намерений, настроений или сущностей из речевых данных и предоставления соответствующих ответов или рекомендаций;

– преобразование речи в текст. Пользователи могут использовать API в сочетании с преобразованием текста в речь Google, Google Cloud Text-to-Speech или Google Assistant, чтобы улучшить вывод текста при распознавании речи. Это может быть полезно для предоставления обратной связи, разъяснения, подтверждения или же помощи говорящему, а также для создания синтетической речи из текста.

В докладе рассказывается о реализации функции голосового ввода в многопользовательском чате с использованием языка программирования Java и сервиса распознавания голоса Google API. В методе onCreate определяется кнопка для запуска голосового ввода информации. Алгоритм работает следующим образом. Берется id данной кнопки, после чего создается метод setOnClickListener, в котором создается новое намерение Intent для вызова метода ACTION_RECOGNIZE_SPEECH. Этот метод нужен для запуска действия, которое запрашивает у пользователя речь и отправляет ее через распознаватель речи. После чего с помощью EXTRA_LANGUAGE_MODEL˚ и ˚LANGUAGE_MODEL_FREE_FORM идет распознавание языка, на котором говорит пользователь. При этом отображается подсказка, с помощью которой можно понять, что уже можно говорить. Это делается с помощью EXTRA_PROMPT. Далее вызывается переопределенный метод onActivityResult, в котором благодаря EXTRA_RESULTS формируется ArrayList<String> результатов распознавания, то есть текста, который сказал пользователь. Текст появляется в поле editText, после чего его можно отправить участникам чата.

Дальнейшее применение голосового ввода в мессенджерах открывает широкие возможности для улучшения взаимодействия между пользователями. Например, функция автоматического исправления ошибок распознавания может основываться на алгоритмах обработки естественного языка, что позволит системе корректировать неточности, допущенные в процессе преобразования аудио в текст. Это особенно важно в профессиональной среде, где требуется высокая точность передачи информации.

Кроме того, возможности Google Speech-to-Text API позволяют интегрировать дополнительные функции, такие как фильтрация ненормативной лексики, адаптация модели под специфическую лексику и работа с разными источниками звука. С помощью пользовательских словарей и контекстных моделей можно настроить систему так, чтобы она учитывала особенности конкретной отрасли или типа контента, что повышает точность распознавания и качество транскрипций. Например, в медицинских приложениях можно добавить специализированные термины, а в юридических — правовую лексику, что существенно улучшает результаты работы API.

Инновационные разработки в области искусственного интеллекта также предполагают внедрение систем, способных анализировать эмоциональную окраску речи. Анализ тональности позволяет определять настроение говорящего, что может быть использовано для автоматической оценки удовлетворенности клиентов, анализа качества обслуживания или даже для создания адаптивных систем поддержки пользователей. Такие системы могут не только фиксировать текст, но и автоматически выделять ключевые моменты, интонационные изменения и важные акценты, что в дальнейшем может

использоваться для обучения моделей или проведения детального анализа коммуникационных процессов.

Еще одной перспективной областью является интеграция голосового ввода с системами автоматического перевода. Используя Google Translate в паре с Speech-to-Text API, можно создать универсальные решения для межкультурного общения, позволяющие в реальном времени переводить речь с одного языка на другой. Это открывает новые горизонты для международных компаний, образовательных учреждений и туристических сервисов, обеспечивая мгновенное понимание и обмен информацией без языковых барьеров.

Кроме того, голосовой ввод может быть использован в качестве элемента системы безопасности. Внедрение функций голосовой аутентификации позволяет не только вводить текст, но и подтверждать личность пользователя на основе уникальных характеристик его голоса. Это может служить дополнительной мерой защиты, особенно в условиях, когда требуется высокий уровень безопасности данных и контроля доступа к конфиденциальной информации.

Не менее важной является интеграция голосового ввода с другими сервисами Google, такими как Google Cloud Storage, Google Dialogflow и Google Assistant. Это позволяет создавать комплексные решения, в которых распознавание речи становится лишь одним из компонентов общей системы. Такие интегрированные решения могут включать автоматическую обработку звонков, интеллектуальный анализ данных и предоставление пользователю интерактивных возможностей, что существенно повышает удобство и эффективность работы как отдельных приложений, так и всей экосистемы цифровых сервисов.

Исходя из всего вышесказанного, можно сделать вывод, что интеграция голосового ввода на основе Google Speech-to-Text API в многопользовательские чаты и другие цифровые платформы открывает широкий спектр возможностей для развития коммуникационных технологий. Это позволяет не только улучшить качество обслуживания, сделать интерфейсы более доступными и удобными, но и открыть новые пути для инновационных решений в области искусственного интеллекта, автоматизации процессов и обеспечения безопасности. Такой подход делает современные мессенджеры более функциональными, адаптированными к требованиям времени и способными удовлетворить растущие потребности пользователей в условиях цифровой трансформации.

Литература

1. Грожик, Г. В. Разработка многопользовательского чата с использованием инструментов FIREBASE / Г. В. Грожик // XXVII Республиканская научная конференция студентов и аспирантов «Новые математические методы и компьютерные технологии в проектировании, производстве и научных исследованиях», 2024 г. − Гомель : ГГУ им. Ф. Скорины. – С. 178.