**Центр научно-технической информации и библиотек – филиал ОАО «РЖД»**

**Дифференцированное Обеспечение Руководства**

**97/2025**

# Интеллектуальная система для транскрибации речи

Транскрибация, или преобразование речи в текст, стала частью рабочего процесса для многих специалистов, в том числе работников железнодорожного транспорта. Так как машинистам, поездными диспетчерам, дежурным по станции, начальником пассажирского поезда, работниками инфраструктуры и другим специалистам приходится многократно между собой вести переговоры, стало необходимым разработать технологический инструмент, облегчающий данный процесс.

Интеллектуальный транскрибатор – это инновационное устройство для автоматического преобразования устной речи из аудиозаписей или разговоров в текстовый формат с использованием современных алгоритмов распознавания речи и искусственного интеллекта. Он может сохранять записи важных переговоров, улучшать коммуникацию и взаимодействие работников на линии, что будет способствовать более эффективному реагированию на оперативные изменения в поездной обстановке.

Действие транскрибаторов базируется на машинном обучении и искусственном интеллекте и практически не требует вмешательства человека. При этом качество распознавания речи варьируется в зависимости от языка, акцента, произношения и состояния аудиозаписи.

Работа интеллектуального транскрибатора представляет собой последовательный и логичный процесс, начинающийся с базовой подготовки (рис. 1). На подготовительном этапе программа загружает предварительно обученную модель и формирует словарь аудиофайлов, содержащий ожидаемые тексты и ключевые слова для последующей проверки. После этого система запрашивает у пользователя идентификационный номер аудиофайла. Следующим шагом становится проверка допустимости введенного номера,

которая служит важным барьером защиты от ошибок ввода и гарантирует продолжение работы только с существующим файлом.

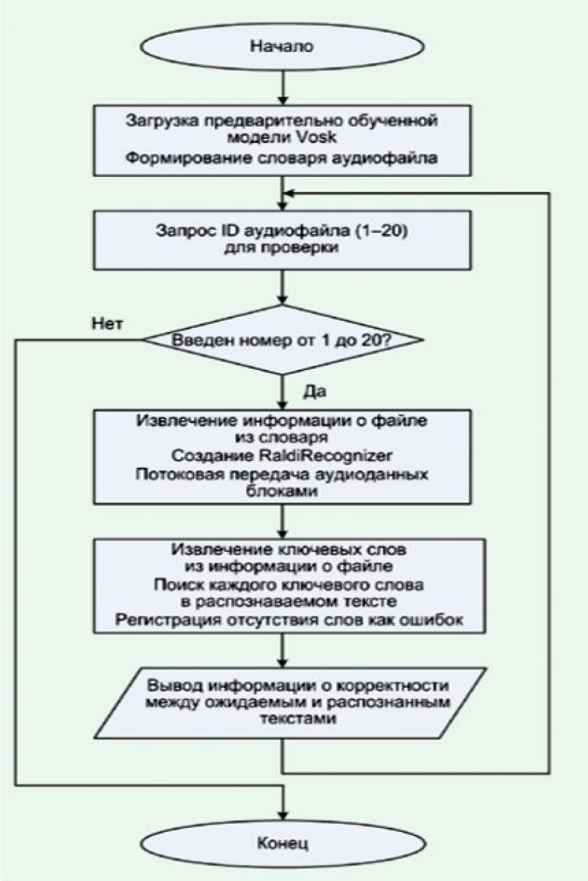


*Рис. 1. Работа транскрибатора*

После успешной валидации запускается ядро процесса. На этом этапе программа извлекает информацию о выбранном файле из словаря и создает экземпляр KaldiRecognizer[1]. Аудиоданные считываются и передаются блоками, что позволяет эффективно обрабатывать файлы любого размера.

Полученный текст сравнивается с эталоном. Система автоматически извлекает из словаря ключевые слова для текущего файла и проверяет их наличие в распознаваемом тексте. Каждое отсутствующее ключевое слово регистрируется как ошибка. Завершающей фазой служит формирование детального отчета, где пользователю представляется сводка, включающая файл с ожидаемым и распознанным текстом, а также список найденных ошибок. После этого программа возвращается к запросу нового номера. Возврат создает непрерывность процесса, благодаря чему пользователь проверяет файлы один за другим без перезапуска приложения. Чтобы завершить программу, пользователь вводит соответствующий идентификатор.

[1] KaldiRecognizer – основной движок распознавания библиотеки транскрибатора Vosk, который принимает поток сырых аудиоданных и, опираясь на загруженную модель, преобразует их в текстовую транскрипцию.

Ниже представлены актуальные технологические разработки и перспективы применения транскрибаторов на сети железных дорог.

Основными вариантами современных транскрибаторов с учетом качества распознавания и затрачиваемого времени являются следующие.

*Google Speech Recognition* – сервис с большой точностью преобразования устной речи в текстовый формат и пониманием множества языков. Он может быть задействован в различных приложениях и устройствах, отличается значительной производительностью и надежностью, имеет высокий процент распознавания окончаний слов, открытый код и простоту его модификации. К недостаткам сервиса относится возможность неправильного распознавания слов и необходимость доступа к сети интернет.

*Vosk Speech Recognition* – бесплатная и открытая библиотека распознавания речи, предназначенная для встраивания в приложения и системы. Она обеспечивает высокую точность и скорость, поддерживает несколько языков и может работать оффлайн. Из преимуществ можно выделить наличие открытого кода и отсутствие необходимости подключения к интернету.

*Dictation* – приложение, которое позволяет диктовать текст в микрофон своего устройства, а затем автоматически преобразовывать устную речь в письменный текст. Оно встроено в операционную систему, может поддерживать различные языки и акценты, приложения для обработки текста и стороннее программное обеспечение. Текстовый контент создается без необходимости печати, что увеличивает продуктивность при работе с текстом. К недостаткам относится отсутствие открытого кода и длительность распознавания текста.

*HTML Speech Recognition* – технология, помогающая вебсайтам и веб-приложениям распознавать речь с использованием микрофона. Она основана на веб-стандарте и обеспечивает создание голосовых интерфейсов для взаимодействия с веб-содержанием, интегрируется в веб-страницы с помощью языка программирования JavaScript. Эта технология содействует увеличению доступности веб-содержания для людей с ограничениями в моторике и взаимодействию с интернетом через голосовой ввод, что делает ее важной частью современной веб-разработки. К преимуществам относится наличие открытого кода и высокий процент распознавания слов и их окончаний.

Для дальнейшей разработки системы и применения ее на сети железных дорог был выбран транскрибатор Vosk Speech Recognition, так как он обладает наибольшим количеством библиотек для языка программирования Python и совместим с функциями для создания графического интерфейса, а также хорошо зарекомендовал себя в распознавании окончаний слов при наличии шумов в аудиофайле. Поскольку программное обеспечение предполагается

устанавливать на объектах критической инфраструктуры, то одним из решающих факторов является способность модели работать в оффлайн режиме.

Для программного обеспечения, реализующего интеллектуальный транскрибатор Vosk Speech Recognition, были взяты библиотеки для обработки звука, машинного обучения и обработки текста, входящие в стандартную библиотеку Python, такие как:

*vosk* – дает возможность распознавания акустических характеристик речи (аудиофайлов) с помощью моделей Kaldi;

*wave* – предоставляет функциональность формата WAV для хранения аудиофайлов звуковых данных без сжатия;

*json* – формирует возможность работы с данными в формате JSON, используется для обработки результатов распознавания;

*difflib* – обеспечивает функциональность для сравнения распознаваемого и ожидаемого текстов и выявления ошибок.

Чтобы протестировать программное обеспечение, предназначенное для работы с транскрибатором Vosk Speech Recognition, изначально нужно запустить программу. Появляется консольное окно, куда нужно ввести идентификационный номер аудиофайла, который сравнивается с ожидаемым текстом по ключевым словам. Если при сравнении ключевые слова в распознаваемом и ожидаемом текстах совпадают, то компилятор не выявляет ни единой ошибки. Это говорит о том, что текст распознан точно. Если же ключевые слова не совпадают, компилятор выводит на консоль ошибку «несоответствие слов».

К примеру, дежурный по станции говорит машинисту текст: «маршрут следования на станцию на такой-то путь». Машинист должен повторить этот текст. Далее программа сравнивает ключевые фразы из речи дежурного и машиниста и в случае несовпадения (или отклонения) информирует машиниста, что тот неверно понял команду. В этом случае дежурный по станции вновь повторяет свою фразу.

Внедрение интеллектуального транскрибатора на железнодорожном транспорте улучшает коммуникации между диспетчерами или дежурными по станции и машинистами. Использование модели Vosk даст возможность работы в оффлайн режиме и обеспечит устойчивость к шумам, что особенно важно в условиях эксплуатации.

Рассмотренная система способствует снижению числа ошибок, повышению уровня безопасности и созданию предпосылок для дальнейшей автоматизации процессов управления движением поездов.

*Источники: по материалам сайта alphacephei.com, (англ. яз.);*
*по материалам Btw.media, (англ. яз);*
*Автоматика, связь, информатика. – 2025. – № 11. – с. 33-35*