

В. В. Язовских¹✉, Р. В. Гришин¹

Веб-сервис транскрипции и анализа телефонных разговоров

¹Сибирский государственный университет геосистем и технологий,
г. Новосибирск, Российская Федерация
e-mail: YAzovskih-VV2022@sgugit.ru

Аннотация. В статье представлены результаты разработки веб-сервиса для автоматической транскрипции и интеллектуального анализа телефонных разговоров. Актуальность работы связана с высокой трудоемкостью ручной обработки аудиозаписей и потребностью в инструментах, позволяющих получать не только текстовую расшифровку разговора, но и интерпретируемый аналитический результат. В работе использованы методы анализа предметной области, проектирования реляционной базы данных, построения REST API и интеграции с внешними сервисами распознавания речи и языкового анализа. Разработанный сервис обеспечивает загрузку аудиофайлов, хранение данных, транскрипцию через Yandex SpeechKit, анализ транскрипта через Yandex GPT, работу с настраиваемыми критериями оценки и отображение результата пользователю. Полученное решение может быть использовано для автоматизации анализа телефонных переговоров в организациях, а также в учебных и исследовательских целях.

Ключевые слова: веб-сервис, транскрипция, телефонные разговоры, распознавание речи, интеллектуальный анализ, база данных

V. V. Yazovskikh¹✉, R. V. Grishin¹

Web service for transcription and analysis of telephone conversations

¹Siberian State University of Geosystems and Technologies, Novosibirsk, Russian Federation
e-mail: YAzovskih-VV2022@sgugit.ru

Abstract. The article presents the results of developing a web service for automatic transcription and intelligent analysis of telephone conversations. The relevance of the study is determined by the high labor intensity of manual audio processing and the need for tools that provide not only a text transcript but also an interpretable analytical result. The study applies methods of domain analysis, relational database design, REST API development, and integration with external speech recognition and language analysis services. The developed service supports audio file upload, data storage, transcription using Yandex SpeechKit, transcript analysis using Yandex GPT, work with customizable evaluation criteria, and presentation of the result to the user. The proposed solution can be used to automate the analysis of telephone conversations in organizations, as well as for educational and research purposes.

Keywords: web service, transcription, telephone conversations, speech recognition, intelligent analysis, database

Введение

Цифровизация прикладных процессов обработки данных приводит к активному использованию интеллектуальных методов в задачах, где ранее доминировал ручной анализ. Одним из наиболее сложных и одновременно практически значимых видов данных являются телефонные разговоры, широко применяемые

в продажах, клиентском сервисе и сопровождении пользователей. Такие разговоры содержат не только факт контакта, но и информацию о цели коммуникации, потребностях клиента, ходе диалога, достигнутых договоренностях и качестве работы сотрудника. В информационных системах, ориентированных на поддержку принятия решений, подобные сведения имеют высокую прикладную ценность [1, 2].

На практике обработка телефонных переговоров во многих случаях остается ручной. Прослушивание аудиозаписей, составление кратких выводов и оценка качества общения требуют значительных временных затрат и плохо масштабируются при росте объема звонков. Дополнительной проблемой является субъективность оценки: два специалиста могут по-разному интерпретировать одни и те же реплики, по-разному выделять смысловые акценты и формулировать рекомендации. В результате у организации отсутствует единый воспроизводимый механизм анализа аудиокоммуникаций.

Современные средства автоматического распознавания речи позволяют переводить аудиосигнал в текстовую форму, а языковые модели дают возможность интерпретировать содержание диалога, выделять роли участников и формировать структурированные аналитические выводы. Однако в реальных прикладных задачах востребовано не изолированное применение отдельных технологий, а их объединение в едином веб-сервисе, обеспечивающем полный цикл обработки: от загрузки аудиозаписи до получения итогового аналитического результата.

Целью исследования является разработка веб-сервиса, обеспечивающего автоматическую транскрибацию и интеллектуальный анализ телефонных разговоров. Для достижения поставленной цели были решены задачи анализа предметной области, выбора средств реализации, проектирования структуры данных и построения клиент-серверного решения, интегрированного с внешними сервисами распознавания речи и языкового анализа.

Методы и материалы

В основе разработанного решения лежит клиент-серверная архитектура, позволяющая разделить пользовательский интерфейс, серверную бизнес-логику и внешние сервисы обработки аудиоданных. Серверная часть реализована на языке Python, который широко применяется для интеграционных задач и построения веб-приложений благодаря развитой экосистеме библиотек и средств разработки [3, 4, 5, 18]. Для организации API выбран фреймворк FastAPI, обеспечивающий построение REST-маршрутов, типизацию запросов и удобную реализацию механизмов аутентификации [6, 7, 20].

Хранение структурированных данных организовано в реляционной базе PostgreSQL. Для взаимодействия с БД использована ORM SQLAlchemy, позволяющая описывать сущности предметной области в виде программных моделей и выполнять операции создания, чтения, изменения и удаления данных без ручного формирования SQL-запросов в основной части логики приложения [8, 9]. В базе данных предусмотрено хранение учетных записей пользователей, сведений

о звонках, сегментов транскрипта, критериев анализа, общих результатов анализа и детальных результатов оценки по каждому критерию.

Аудиофайлы не сохраняются непосредственно в базе данных, а размещаются во внешнем S3-совместимом объектном хранилище. Для этого используется сервис Yandex Object Storage и библиотека boto3, что позволяет централизованно хранить медиафайлы и передавать ссылки на них во внешние сервисы распознавания речи [14, 19]. Аутентификация реализована на основе токенов доступа JWT, что обеспечивает защищенный доступ к функциям системы и разграничение прав пользователя и администратора [7, 15].

Для транскрибации телефонных разговоров применяется сервис Yandex SpeechKit, позволяющий преобразовывать аудиозапись в текст и получать сегменты реплик, привязанные к условным спикерам [10, 11]. После завершения транскрибации текст разговора и активные критерии оценки передаются в сервис Yandex GPT, который возвращает структурированный аналитический результат: определение ролей участников, краткий вывод по разговору, договоренности, риски, следующий шаг, общую оценку и рекомендации [12, 13]. Пользовательский интерфейс реализован с применением стандартных веб-технологий HTML, CSS и JavaScript, а тестирование HTTP-запросов выполнялось в среде Postman [16, 17].

Результаты

В результате работы был разработан веб-сервис, обеспечивающий полный цикл обработки телефонного разговора: загрузку аудиофайла, сохранение ссылки на объект в хранилище, запуск транскрибации, формирование транскрипта, интеллектуальный анализ содержания разговора и отображение результатов пользователю. Архитектура сервиса приведена на рис. 1.

Логика обработки данных реализована последовательно. После загрузки файла создается запись о звонке и фиксируется текущий статус обработки. На следующем этапе система инициирует транскрибацию и сохраняет в базе данных текст реплик и временные границы сегментов. Затем выполняется интеллектуальный анализ транскрипта, в ходе которого учитываются активные критерии оценки, заданные администратором. Итог работы сервиса представляет собой не только расшифровку разговора, но и формализованный аналитический результат, пригодный для дальнейшего использования.

Ключевые функциональные возможности разработанного веб-сервиса представлены в табл. 1.

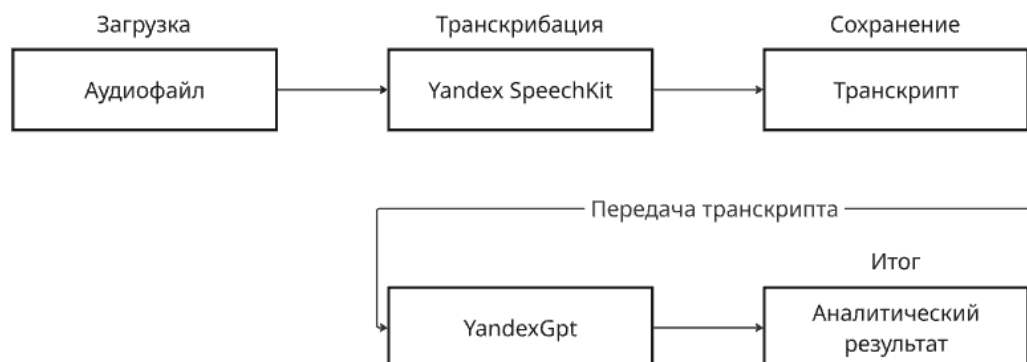


Рис. 1. Архитектура веб-сервиса транскрибации и анализа телефонных разговоров

В табл. 1 представлены основные функциональные модули системы и результаты их работы.

Таблица 1. Функциональные модули разработанного веб-сервиса

Модуль	Назначение	Входные данные	Результат
Загрузка звонка	Прием аудиофайла и создание записи о звонке	Файл, имя клиента, имя менеджера	Ссылка на объект и статус обработки
Транскрибация	Преобразование речи в текст	Ссылка на аудиофайл	Транскрипт и сегменты реплик
Интеллектуальный анализ	Интерпретация содержания разговора	Транскрипт и активные критерии	Роли, итог, риски, оценка, рекомендации
Управление критериями	Настройка параметров анализа	Название, описание, вес, активность	Изменяемый набор критериев
Пользовательский интерфейс	Отображение результатов и управление данными	Ответы API	Страницы звонков, анализа, профиля и критериев

Реализованный сервис включает модуль авторизации, модуль работы со звонками, модуль транскрибации, модуль интеллектуального анализа, модуль критериев оценки и модуль управления пользователями. После авторизации пользователь получает доступ к списку звонков, может загрузить новый аудиофайл, инициировать его обработку и просмотреть результат. Администратор, помимо этого, управляет учетными записями и составом критериев анализа.

Принципиальным результатом разработки является переход от хранения необработанных аудиозаписей к формированию интерпретируемой структуры данных.

Система сохраняет не только исходный файл и сплошной текст транскрипта, но и сегменты разговора, результаты анализа по каждому критерию, а также общий вывод по звонку. Это позволяет использовать сервис как инструмент накопления аналитических данных для последующей оценки качества коммуникации.

Обсуждение

Полученные результаты показывают, что объединение сервисов распознавания речи и языкового анализа в рамках единого веб-приложения позволяет перейти от фрагментарной обработки телефонных разговоров к последовательному и воспроизводимому сценарию анализа. В отличие от подхода, при котором пользователь получает только текстовый транскрипт, предложенное решение формирует структурированный результат, пригодный для контроля качества коммуникации и дальнейшей интерпретации.

Важным преимуществом разработанного сервиса является поддержка настраиваемых критериев анализа. Критерии не зафиксированы жестко в программном коде, а хранятся в базе данных и управляются через интерфейс администратора. Это позволяет адаптировать систему под разные прикладные сценарии без изменения серверной логики и делает решение более гибким по сравнению с системами, использующими неизменяемые правила оценки.

Вместе с тем качество аналитического вывода зависит от точности транскрибации исходного аудио и от структуры формируемого запроса к языковой модели. Перспективой дальнейшего развития работы является расширение набора критериев анализа, интеграция сервиса с CRM-платформами, использование дополнительных средств визуализации результатов и проведение сравнительного тестирования на более широком наборе аудиозаписей.

Заключение

В статье представлены результаты разработки веб-сервиса для автоматической транскрибации и интеллектуального анализа телефонных разговоров. Выполнен анализ предметной области, обоснован выбор клиент-серверного подхода, реализованы серверная и пользовательская части приложения, а также интеграция с Yandex SpeechKit, Yandex GPT и S3-совместимым объектным хранилищем.

Разработанный сервис обеспечивает получение транскрипта разговора и формирование аналитического результата, включающего определение ролей участников, общий итог, оценку по критериям и рекомендации. Практическая значимость работы заключается в возможности применения разработанного решения для автоматизации анализа телефонных переговоров в организациях, а также в учебных и исследовательских целях.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Подчукаев, В. А. Теория информационных процессов и систем : учеб. пособие для вузов / В. А. Подчукаев. – Москва : Гардарики, 2007. – 207 с.
2. Российская Федерация. Законы. Об информации, информационных технологиях и о защите информации : федер. закон N 149-ФЗ : принят Государственной Думой 8 июля 2006 г.

- : одобрен Советом Федерации 14 июля 2006 г. : послед. ред. // КонсультантПлюс : сайт. – URL: http://www.consultant.ru/document/cons_doc_LAW_61798/ (дата обращения: 26.05.2026).
3. Лутц, М. Изучаем Python. – Санкт-Петербург : Символ-Плюс, 2021.
 4. Рамальо, Л. Fluent Python. Ясное, лаконичное и эффективное программирование. – Санкт-Петербург : Питер, 2022.
 5. Лубанович, Б. Простой Python. Современный стиль программирования. – Санкт-Петербург : Питер, 2021.
 6. FastAPI Documentation. – URL: <https://fastapi.tiangolo.com/>.
 7. FastAPI | Security – OAuth2 with Password (and hashing), Bearer with JWT tokens. – URL: <https://fastapi.tiangolo.com/tutorial/security/oauth2-jwt/>.
 8. SQLAlchemy Documentation. – URL: <https://docs.sqlalchemy.org/>.
 9. PostgreSQL: Documentation. – URL: <https://www.postgresql.org/docs/>.
 10. Yandex SpeechKit | Документация. – URL: <https://yandex.cloud/ru/docs/speechkit/>.
 11. Yandex SpeechKit | Распознавание речи. – URL: <https://yandex.cloud/ru/docs/speechkit/quickstart/stt-quickstart-v1>.
 12. Yandex Foundation Models | Документация. – URL: <https://yandex.cloud/ru/docs/foundation-models/>.
 13. Yandex GPT API, совместимый с OpenAI. – URL: <https://yandex.cloud/ru/docs/foundation-models/concepts/openai-compatibility>.
 14. Yandex Object Storage | Документация. – URL: <https://yandex.cloud/ru/docs/storage/>.
 15. RFC 7519: JSON Web Token (JWT). – URL: <https://datatracker.ietf.org/doc/html/rfc7519>.
 16. MDN Web Docs. – URL: <https://developer.mozilla.org/ru/>.
 17. Postman Learning Center. – URL: <https://learning.postman.com/docs/getting-started/introduction/>.
 18. Python Documentation. – URL: <https://docs.python.org/3/>.
 19. boto3 Documentation. – URL: <https://boto3.amazonaws.com/v1/documentation/api/latest/index.html>.
 20. Pydantic Documentation. – URL: <https://docs.pydantic.dev/>.

© В. В. Язовских, Р. В. Гришин, 2026