МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ

РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение

высшего образования

«ТЮМЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

ШКОЛА КОМПЬЮТЕРНЫХ НАУК

Кафедра программного обеспечения

ОТЧЕТ  
О РАЗРАБОТКЕ ПРОЕКТА ПО ДИСЦИПЛИНЕ  
«Интеллектуальные системы поддержки принятия решений»

ТЕМА ПРОЕКТА

«Разработка программного решения для транскрибации и реферирования аудиозаписей совещаний для предприятия “Студия 42”»

| Выполнил  обучающийся 4 курса,  МОиАИС-20.03 группы | (подпись) | Чупеев Андрей Дмитриевич |
| --- | --- | --- |
| Выполнил  обучающийся 4 курса,  МОиАИС-20.03 группы | (подпись) | Силин Никита Валерьевич |
| Научный руководитель проекта  Доцент | (подпись) | Сергеев Иван Николаевич |

«Тюмень 2024»

**ОГЛАВЛЕНИЕ**

[**ВВЕДЕНИЕ 3**](#_euovzu6u1uai)

[**ГЛАВА 1. ЦЕЛИ И ТРЕБОВАНИЯ ПРОЕКТА. 5**](#_t8ecryajtseb)

[1.1. Цели проекта 5](#_48055nljwrho)

[1.2. Ограничения проекта 7](#_18qnanytiwbi)

[**ГЛАВА 2. ТЕОРЕТИЧЕСКОЕ ОПИСАНИЕ ИДЕИ РЕШЕНИЯ 8**](#_th4mxjux9opz)

[2.1. Основа проекта 8](#_5gst832ldc4n)

[2.2. Собственные практические решения 9](#_6k0i6t1uxjjl)

[2.3. Возможные улучшения 11](#_rzgd4l4jcm28)

[**ГЛАВА 3. РЕАЛИЗАЦИЯ СЕРВИСА. 13**](#_vbfud4b4m6el)

[3.1. Веб-сервис 13](#_9gof6mq9iz1k)

[3.2. Рабочий процесс (worker) 14](#_ltkcsjt5yue1)

[**ГЛАВА 4. ФУНКЦИОНАЛЬНЫЕ ВОЗМОЖНОСТИ И ПОЛЬЗОВАТЕЛЬСКИЙ ИНТЕРФЕЙС ПРОЕКТА. 17**](#_grap2ggfu1a9)

[4.1. Функциональные возможности 17](#_eyy4on5fvkh4)

[4.2. Пользовательский интерфейс 18](#_6ocuew54aze4)

[**ГЛАВА 5. ОРГАНИЗАЦИЯ ТЕСТИРОВАНИЯ ПРОЕКТА. 21**](#_cqgt315k5hur)

[**ЗАКЛЮЧЕНИЕ 25**](#_tb9jw2ytbkxy)

[**СПИСОК ЛИТЕРАТУРЫ 28**](#_1g1tw8q309gh)

# 

# ВВЕДЕНИЕ

Современные деловые процессы, включающие регулярные совещания и конференции, требуют значительных временных затрат на последующий анализ и документирование обсуждений, что может становится узким местом в общей эффективности работы организаций. В частности, для предприятия "Студия 42" актуальной задачей является создание инструмента, который позволит автоматизировать процесс преобразования аудиозаписей совещаний в текстовые протоколы и краткие отчеты, сокращая время, необходимое для подготовки документации.

Основной сложностью при разработке таких решений является неудовлетворительная эффективность существующих технологий транскрибации в условиях русского языка, особенно при наличии фоновых шумов, тихой речи или нечеткой дикции. Дополнительно ограничение большинства подобных систем заключается в отсутствии функционала для реферирования текста, что усложняет выделение ключевых моментов из текста.

Задачи проекта включают проведение анализа требований и изучение существующих решений для выявления их сильных и слабых сторон. Важным этапом является разработка и интеграция системы, а также ее тестирование с реальными аудиозаписями различной сложности. Завершающим этапом будет внедрение готового решения в рабочий процесс компании.

Таким образом, цель проекта заключается в создании программного решения, которое будет способно транскрибировать аудиозаписи, преобразуя их в текстовые протоколы. Это решение позволит ускорить процесс подготовки документации, избавляя сотрудников от необходимости вручную записывать каждое слово, и сделает возможным автоматическое извлечение ключевых моментов из разговора. Важной частью решения станет функция реферирования, которая будет сокращать объем текста, выделяя только наиболее значимые моменты и обеспечивая краткие отчеты, что позволит эффективно использовать полученную информацию для принятия решений [Нестерова, c. 128-130].

Кроме того, проект будет нацелен на удобство использования для корпоративных пользователей, что гарантирует интуитивно понятный интерфейс и простоту в интеграции в существующие рабочие процессы компании "Студия 42". Удовлетворение этих потребностей позволит повысить производительность, улучшить качество и оперативность работы, снизив временные и трудозатратные издержки на документирование и анализ информации.

# **ГЛАВА 1. ЦЕЛИ И ТРЕБОВАНИЯ ПРОЕКТА.**

## 1.1. Цели проекта

Проект направлен на достижение целого ряда важных целей, каждая из которых вносит свой посильный вклад в общее развитие и совершенствование системы. К числу ключевых задач можно отнести улучшение качества транскрибации, которое подразумевает не только повышение точности распознавания речи, но и способность эффективно обрабатывать сложные по содержанию, шумовые или нечетко записанные аудиоматериалы. Это особенно актуально в тех случаях, когда качество исходного звука оставляет желать лучшего или когда голосовые данные поступают из разнородных источников, имеющих различные технические характеристики. При этом критически важно стремиться к минимизации количества ошибок в итоговом текстовом варианте, чтобы данные могли быть использованы с наибольшей пользой и достоверностью.

Помимо повышения точности распознавания, проект предусматривает оптимизацию использования ресурсов. В рамках этих мероприятий особое внимание уделяется снижению потребления оперативной памяти, что позволит работать более эффективно внутри существующего ограничения в 16 ГБ. Это не просто технический показатель — эффективное использование памяти обеспечит большую устойчивость системы при работе с интенсивными и ресурсоёмкими задачами. Одновременно с этим важно повысить производительность обработки на GPU, так как современные алгоритмы, занимающиеся транскрибацией, нередко опираются на параллельные вычислительные мощности для ускорения процессов распознавания речи. Оптимизация GPU-производительности будет способствовать более гибкому и быстрому обращению с большим массивом аудиоданных, что, в свою очередь, повысит общую пропускную способность системы.

Далее, существенную роль играет обновление технологической базы проекта. Под этим подразумевается не только устранение устаревших зависимостей, которые могут негативно влиять на стабильность и совместимость программного обеспечения, но и сокращение уровня фрагментации используемых библиотек, утилит и иных компонентов. Все эти технические улучшения, производимые в рамках обновления технологического стека, призваны повысить устойчивость, актуальность и предсказуемость работы системы. Современные, регулярно обновляемые технологии обеспечат более ровное и предсказуемое функционирование, а также снизят риск возникновения непредвиденных ошибок, связанных с неактуальными или конфликтующими зависимостями.

Помимо внутренних технологических аспектов, особое значение уделяется и пользовательскому опыту. Цель состоит в том, чтобы сделать весь процесс взаимодействия с системой максимально удобным и прозрачным. Предполагается обеспечить своевременное уведомление пользователей о каждом этапе обработки файлов, что позволит им понимать, что именно происходит в данный момент и сколько времени может занять выполнение определенных операций. Для ещё большего удобства планируется предоставление результатов транскрибации через разнообразные каналы — в частности, удобный веб-интерфейс и электронную почту, что значительно расширит спектр вариантов доступа к итоговым данным. Важным штрихом станет корректное отображение текста с учетом распределения по различным участникам, что имеет значение в ситуациях, где в аудио задействовано несколько говорящих. Такая детализация делает расшифровку более понятной и структурированной для всех заинтересованных лиц.

## 1.2. Ограничения проекта

В ходе реализации проекта необходимо принять во внимание ряд ограничений, напрямую влияющих на возможности и сроки выполнения поставленных задач. Среди ресурсных ограничений ключевым является объем оперативной памяти в 16 ГБ, который не может быть превышен. Это создает определенные рамки при использовании тяжелых и ресурсоёмких алгоритмов, заставляя разработчиков подходить к выбору методов и инструментов более осознанно. Кроме того, возможности GPU, доступные в рамках проекта, также имеют свои границы, что вынуждает искать компромисс между скоростью и точностью, а в отдельных случаях — внедрять дополнительные механизмы оптимизации и распределения нагрузки. При работе со сложными, интенсивно зашумленными или большими по длительности аудиофайлами риск избыточной нагрузки на память и вычислительные ресурсы возрастает, поэтому особое внимание должно быть уделено грамотному управлению ресурсами и оптимизации производительности.

Наконец, нельзя забывать о временных ограничениях. Проект должен быть завершен к установленному сроку — до 13 декабря 2024 года. Наличие чёткого дедлайна стимулирует концентрировать усилия на планировании этапов работ, их приоритизации и постоянном мониторинге прогресса. Соблюдение временных рамок критически важно, поскольку оно напрямую влияет на общую эффективность всего проекта и уровень доверия со стороны заинтересованных сторон, будь то клиенты, пользователи или руководство. Таким образом, каждый из перечисленных аспектов — начиная от улучшения качества транскрибации и заканчивая соблюдением ресурсных и временных ограничений — выступает в качестве кирпичика в фундаменте комплексного, масштабного подхода к развитию и совершенствованию системы транскрибации.

# **ГЛАВА 2. ТЕОРЕТИЧЕСКОЕ ОПИСАНИЕ ИДЕИ РЕШЕНИЯ**

## 2.1. Основа проекта

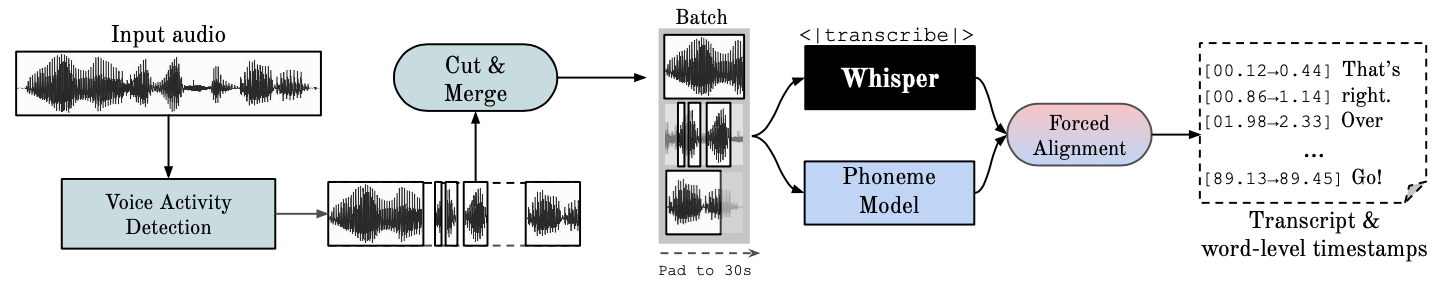
За основу транскрибирующего ядра легла система WhisperX, которая, как показано на рисунке 1, представляет собой многоэтапный подход к автоматическому распознаванию речи, точной сегментации текстового материала и получению детализированных временных меток на уровне отдельных слов. Данный инструментальный набор призван не только ускорить процесс транскрибации, но и существенно повысить точность результатов, а также обеспечить гибкую адаптацию к специфике входных данных и потребностям конечных пользователей. Использование WhisperX в качестве фундамента проекта было обусловлено его многосторонней функциональностью, актуальностью, а также способностью гармонично сочетать несколько подсистем, начиная от детекции голосовой активности (Voice Activity Detection, VAD) и заканчивая продвинутыми техниками форсированного выравнивания (Forced Alignment). Начальный этап обработки аудиофайлов в рамках решения включает процедуру детекции голосовой активности, призванную отделить сегменты с фактической речью от участков тишины или фонового шума. Это не только гарантирует, что система сфокусируется на значимых фрагментах звука, но и позволяет избежать ложных срабатываний или так называемых "галлюцинаций" моделей, при которых алгоритм может порождать несоответствующий исходной записи текст[Yoshioka, T., Abramovski, I., Aksoylar, C., с.1-6].

Таким образом, благодаря VAD можно значительно повысить точность последующих этапов и оптимизировать ресурсозатраты системы. Данный шаг особенно важен при работе с долгими аудиозаписями, включающими ненужные паузы, шумы или периоды без речи, что позволяет сконцентрировать вычислительную мощность только на действительно содержательных частях. После того как голосовая активность идентифицирована, соответствующие сегменты аудио собираются в батчи, что упрощает последующую обработку и позволяет эффективно использовать параллельные вычислительные ресурсы. В частности, путем аккуратной нарезки и дальнейшего слияния сегментов обеспечивается более плотная загрузка вычислительных мощностей GPU, ускоряется процесс транскрибации, а общая производительность системы выходит на новый уровень. Возможность подгонки длительности сегментов до 30 секунд нацелена на оптимальный баланс между скоростью и точностью: если сегменты слишком короткие, процесс становится слишком фрагментированным, а если слишком длинные — может возникнуть дополнительная нагрузка и потенциальное снижение точности. Таким образом, продуманный подход к формированию батчей аудио данные снижает риски, связанные с неравномерной нагрузкой и использованием ресурсов.

## 2.2. Собственные практические решения

Ключевым этапом технологической цепочки является непосредственно транскрибация, выполняемая моделью Whisper. Этот инструмент, прошедший обучение на огромных массивах данных, способен обеспечивать высочайшую точность распознавания речи. Важной особенностью Whisper является поддержка пакетной обработки данных, благодаря чему можно добиться увеличения скорости распознавания вплоть до 70-кратного реального времени при использовании модели large-v2. Такие впечатляющие показатели позволяют существенно сократить время от поступления аудио до получения готового текста, делая систему особенно полезной для компаний, которым необходимо быстро и эффективно обрабатывать большие объёмы речевых данных, например, записи совещаний, конференций, вебинаров, интервью или подкастов. Однако базовые возможности Whisper по предоставлению временных меток ограничиваются уровнем высказываний, что может вызывать определенные неточности и затруднять работу с полученным текстом.

Для устранения этой проблемы был внедрен этап форсированного выравнивания. С помощью фонемных моделей, таких как wav2vec2.0, происходит уточнение временных границ до уровня отдельных слов. Это означает, что каждая единица текста будет прочно привязана к конкретному временному интервалу в аудиозаписи, обеспечивая повышенную точность и удобство последующего анализа. Такая детальная привязка текста к аудио особенно важна в аналитических сценариях, когда нужно быстро найти и воспроизвести определенный момент речи, либо провести контентный анализ, учитывая точное время произнесения тех или иных фраз.



*Рис.1. Схема работы ядра.*

Дополнением к этому функционалу является возможность расширить систему за счет диаризации, то есть определения говорящих в записи. Используя pyannote-audio, можно идентифицировать, кто из участников беседы говорит в тот или иной момент. Это открывает новые горизонты для анализа сложных многоголосых записей, например, больших совещаний с множеством участников или дискуссионных панелей. Благодаря такой возможной интеграции система не только расшифровывает текст и привязывает его ко времени, но и структурирует его по спикерам, делая итоговые стенограммы более информативными и удобными для дальнейшей работы.

Для улучшения автономности и снижения зависимости от внешних источников данных мы доработали систему, позволив ей работать полностью локально. Теперь все необходимые модели и зависимости предустанавливаются и размещаются на внутренних ресурсах. Это обеспечивает повышенную безопасность и надежность, поскольку при обработке конфиденциальных или внутренних данных компании можно не беспокоиться о необходимости подключения к сети, скачивании моделей извне или риске недоступности внешних сервисов. Такой подход исключает ситуации, когда отсутствие или нестабильность интернет-соединения замедляли бы или вообще делали бы невозможным процесс транскрибации.

Важной частью модернизации стал пересмотр набора библиотек и зависимостей, используемых в системе. Обновление технологий не просто выполняет формальную функцию, но реально улучшает производительность, снижает вероятность конфликтов версий и обеспечивает совместимость с современными аппаратными и программными платформами. Это также открывает новые возможности масштабирования и внедрения инновационных решений, оптимизирующих как вычислительную нагрузку, так и качество распознавания. Переход на более подходящие к специфике аудиозаписей модели, включая усовершенствованные варианты для русского языка и для условий повышенного шума, повысил точность распознавания и надежность результатов в реалистичных рабочих сценариях компании.

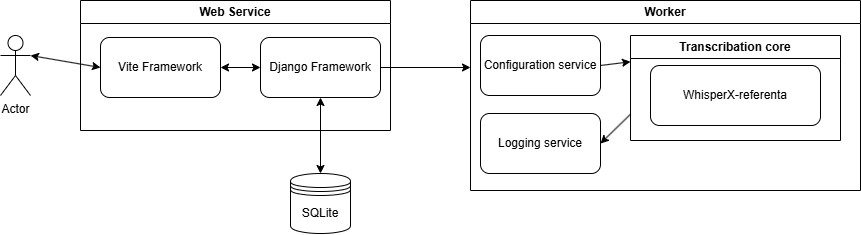
## 2.3. Возможные улучшения

Дальнейшее развитие проекта предполагает масштабирование возможностей системы и улучшение её производительности при работе с большими объёмами данных. Мы планируем внедрить параллельную обработку, что позволит ускорить транскрибацию протяженных записей и значительно сократит общее время, необходимое на получение готовых протоколов, особенно в случаях, когда необходимо обрабатывать сразу несколько больших файлов (например, серию совещаний за неделю или архив вебинаров). Параллельная обработка позволит более эффективно использовать доступные ресурсы, сокращая задержки и делая процесс более плавным и предсказуемым. Кроме того, мы рассматриваем возможность интеграции кастомных моделей диаризации, обученных на внутренних данных компании. Это позволит нам улучшить точность определения говорящих, учитывая их индивидуальные особенности, голосовой тембр, темп речи, словарный запас и типичные сценарии взаимодействия. Такой уровень адаптации будет крайне полезен для крупного бизнеса, где в диалогах часто присутствуют одни и те же сотрудники, и их голоса известны системе. Подобный подход позволит более точно и надежно разделять реплики различных участников, что облегчит дальнейший анализ стенограммы и повысит её практическую ценность. Еще одним направлением развития станет расширение языковых возможностей. Планируется интеграция механизмов автоматического определения языка, а также моделей, способных работать с мультиязычными аудиозаписями, где в рамках одной сессии могут переключаться разные языки. Такая функциональность незаменима для международных компаний, глобальных совещаний и переговоров, в ходе которых участники свободно переходят с одного языка на другой. Добавление полноценной поддержки многоязычности повысит универсальность и гибкость системы. Чтобы сделать процесс транскрибации и последующего анализа данных максимально доступным для всех пользователей внутри компании, мы разрабатываем удобный пользовательский интерфейс. Он будет обладать возможностью визуализировать временные метки, выделять ключевые реплики, проводить поиск по тексту и выполнять иные полезные действия. Это сделает анализ полученных протоколов интуитивно понятным и гораздо более эффективным, поскольку пользователи смогут быстро ориентироваться в стенограммах, находить нужные фрагменты, отмечать значимые детали и при необходимости экспортировать данные для отчетов, презентаций или других форм последующей обработки. На долгосрочной перспективе мы планируем адаптировать модели распознавания речи под специфику внутреннего корпоративного аудиоархива, что даст возможность учесть используемый в компании профессиональный сленг, специфическую терминологию, наименования продуктов, отделов, проектов, а также упоминания брендов и имен собственных, которые могут быть характерны исключительно для внутреннего корпоративного контекста. Такая тонкая адаптация повышает точность распознавания и снижает риск недопонимания или неверной интерпретации важной информации.

Таким образом, WhisperX изначально предоставляет мощный технологический каркас, который мы последовательно расширяем, улучшая автономность, надежность, производительность и точность распознавания речи. В нашем проекте мы стремимся сделать систему не только более быстрой и масштабируемой, но и более дружественной к специфике бизнеса, более устойчивой к потенциальным сбоям и более простой в использовании за счет удобного интерфейса. Все эти шаги в совокупности формируют фундамент для долгосрочного развития внутренней аналитической и документирующей инфраструктуры, позволяя быстро и эффективно извлекать знания из аудиоматериалов, формализовать их в удобной текстовой форме и открывать новые возможности для анализа и принятия решений.

# **ГЛАВА 3. РЕАЛИЗАЦИЯ СЕРВИСА.**

В состав решения входят два ключевых компонента: веб-сервис и отдельный рабочий процесс (worker), каждый из которых несёт собственный набор задач, направленных на достижение максимальной эффективности в работе приложения. (См. рис. 2). Этот подход разделения обязанностей демонстрирует стремление архитектуры к модульности и распределённости, что в свою очередь упрощает внесение изменений и обновлений как в логику фронтенда и бекенда, так и в модули, отвечающие за фоновую обработку данных..



*Рис.2. Схема сервиса.*

## 3.1. Веб-сервис

Веб-сервис состоит из фронтенд- и бекенд-частей, работающих в тесной связке для формирования удобного, интуитивно понятного пользовательского интерфейса и надёжной обработки поступающих с него запросов. Фронтенд-часть, построенная на основе современного фреймворка Vite, обеспечивает высокую производительность и простоту разработки интерфейсов. Благодаря Vite веб-разработчики получают возможность быстро обновлять и тестировать пользовательский интерфейс, гарантируя, что конечные пользователи смогут взаимодействовать с приложением без задержек и ощутимого дискомфорта. Все запросы, инициируемые на фронтенде (например, отправка данных форм, запрос на получение определённой информации или запуск операций по обработке данных), аккуратно передаются на бекенд, который реализован с помощью популярного фреймворка Django. Django берёт на себя ключевые задачи: маршрутизацию входящих запросов, валидацию полученных данных, взаимодействие с базой данных SQLite, а также возвращение ответов на фронтенд. Благодаря Django разработчики могут концентрироваться на бизнес-логике приложения, используя обширный набор встроенных инструментов и функциональных возможностей, упрощающих работу с базой данных, обработку сессий, авторизацию и аутентификацию пользователей, а также другие рутинные задачи серверной части.

## 3.2. Рабочий процесс (worker)

Отдельный рабочий процесс (worker) существует для того, чтобы разгрузить веб-сервис от тяжёлых, долгих и ресурсоёмких операций. Вместо того, чтобы выполнять все задачи синхронно в бекенде и тем самым потенциально снижать общую отзывчивость системы, ресурсоёмкая логика вынесена в отдельный компонент, к которому веб-сервис направляет запросы на выполнение. Таким образом, worker может быть масштабирован и оптимизирован независимо от веб-сервиса. Он выполняет объёмные операции в фоновом режиме, получая задания и возвращая результаты по мере готовности. Этот подход обеспечивает более высокую отказоустойчивость системы: даже если worker будет перегружен или временно недоступен, веб-сервис по-прежнему сможет обслуживать часть запросов пользователей, обрабатывать операции, не требующие тяжёлой фоновой работы, и поддерживать минимальный функционал интерфейса. Внутри worker’а имеются конфигурационный модуль и модуль логирования, которые позволяют детально настраивать параметры выполнения операций и отслеживать состояние и события, происходящие в процессе работы фоновых задач. Это может включать в себя гибкую настройку таймаутов, количество одновременно обрабатываемых задач, контроль версий используемых библиотек, а также анализ логов для диагностики и устранения проблем. Наличие такого уровня контроля облегчает поиск и решение проблем, повышает удобство сопровождения и администрирования системы, а также упрощает будущий рефакторинг и интеграцию новых функций. Центральным элементом рабочей подсистемы является модуль «Transcription core», в котором сосредоточена основная логика по преобразованию аудио- или текстовых данных. Он использует библиотеку WhisperX-referenta, специально разработанную для выполнения транскрибации и реферирования текстов[Методические основы…, с.14]. Это означает, что любой сложный процесс, связанный с семантическим анализом, обработкой естественного языка, преобразованием аудио в текст, может быть реализован и улучшен в рамках данного модуля. Благодаря этому все ресурсоёмкие задачи по обработке данных, их модификации и структурированию делегируются worker’у, освобождая веб-сервис от тяжёлой нагрузки и позволяя ему быстро и эффективно взаимодействовать с пользователями. Взаимодействие между веб-сервисом и worker’ом организовано через чётко определённый протокол обмена сообщениями. Такая четкость и формальность в коммуникации необходимы для снижения рисков несогласованности между компонентами. Веб-сервис направляет worker’у задания на выполнение транскрибации, анализ или любую другую сложную обработку данных, при этом worker своевременно возвращает результаты после завершения необходимых вычислений. Данный протокольно-ориентированный подход способствует повышению надёжности всей системы: даже в условиях высокой нагрузки или временных затруднений с выполнением задач, оба компонента остаются независимыми и способны функционировать автономно, обрабатывая доступные для них запросы. Важно подчеркнуть, что данная архитектура предоставляет возможность независимого масштабирования компонентов. Если нагрузка на веб-сервис значительно возрастёт, можно увеличить количество инстансов фронтенда или бекенда, распределив запросы по группе серверов. Аналогично, если интенсивность обработки фоновых операций возрастает, можно запустить несколько экземпляров worker’а, распределяя вычислительную нагрузку между ними. Это даёт дополнительную гибкость при развёртывании и обслуживании системы, позволяя постепенно адаптироваться к меняющимся условиям и потребностям пользователей. В итоге, описываемая архитектура представляет собой целостную и гибкую систему, в которой веб-сервис и рабочий процесс эффективно дополняют друг друга. Веб-сервис сосредоточен на взаимодействии с пользователем, маршрутизации запросов и хранении данных в базе данных SQLite, тогда как worker берёт на себя тяжёлую работу по трансформации аудио- и текстовых данных, а также дополнительные фоновые задачи. Такое распределение обязанностей повышает производительность, устойчивость и удобство сопровождения приложения, делая его готовым к расширению и модернизации в будущем.

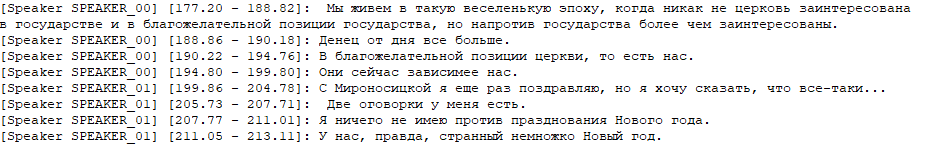
# **ГЛАВА 4. ФУНКЦИОНАЛЬНЫЕ ВОЗМОЖНОСТИ И ПОЛЬЗОВАТЕЛЬСКИЙ ИНТЕРФЕЙС ПРОЕКТА.**

## 4.1. Функциональные возможности

Система обладает функционалом: позволяет пользователям загружать аудиофайлы через удобный веб-интерфейс с поддержкой популярных форматов, таких как MP3 и WAV. Загруженные файлы обрабатываются с минимальными задержками.

Транскрибация речи выполняется автоматически с разделением текста на реплики участников, что упрощает анализ диалогов. Решение эффективно работает с записями различного качества, включая аудио с тихой речью, фоновыми шумами и нечеткой дикцией, гарантируя высокую точность распознавания.

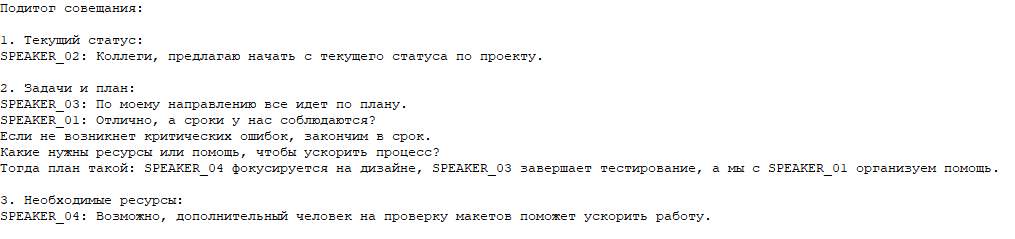
Система поддерживает диаризацию участников, автоматически выделяя голоса разных спикеров и привязывая их реплики к условным идентификаторам, таким как "Спикер 1" или "Спикер 2". Пользователь может редактировать эти идентификаторы, делая результат более персонализированным(рис.3).



*Рис.3. Диаризация участников*

По завершении обработки пользователь получает уведомление по электронной почте с ссылкой на результат, что обеспечивает удобство и оперативный доступ к данным.

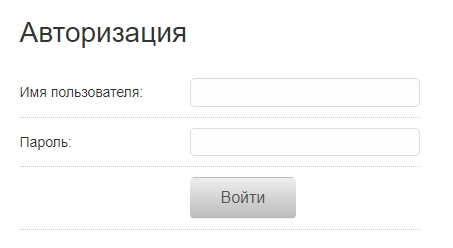
Результаты представлены в структурированном виде через веб-интерфейс с выделением ключевых слов и фраз(рис.4). Пользователь также может скачать архив с обработанным текстом и дополнительными материалами, такими как логи процесса обработки.



*Рис.4. Пример выделения ключевых фраз*

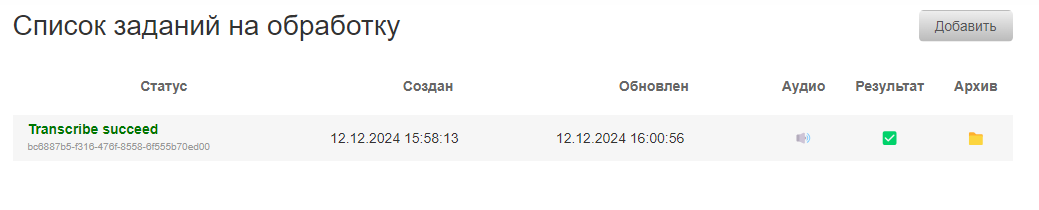
## 4.2. Пользовательский интерфейс

Пользовательский интерфейс системы автоматической транскрибации и реферирования был разработан с акцентом на удобство и интуитивность использования. Основной целью при создании интерфейса было обеспечить простой и быстрый доступ к ключевым функциям, при этом минимизируя необходимость в технических знаниях со стороны пользователей.



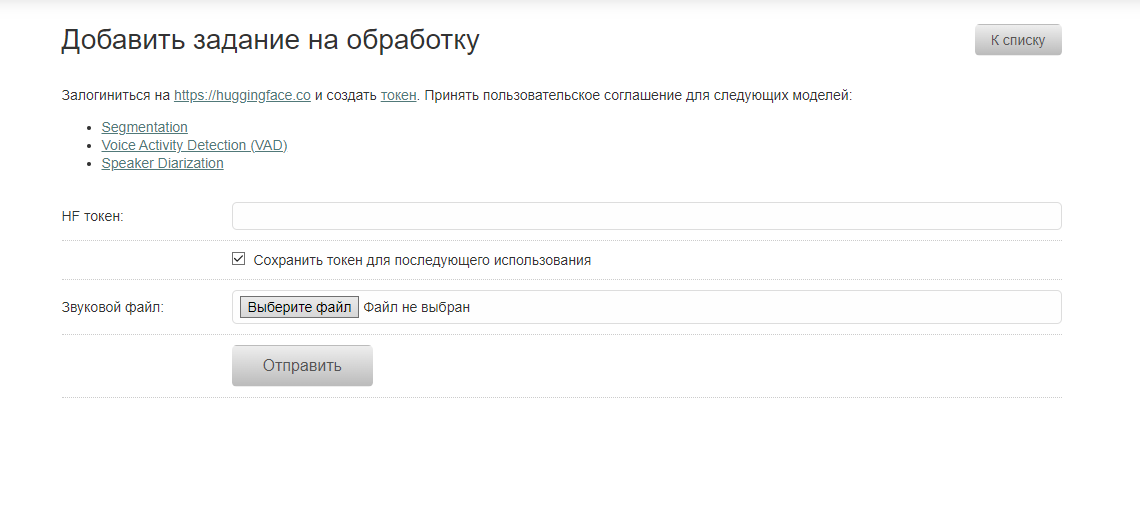
*Рис.5. Авторизация*

Форма для авторизации пользователя (рис.5). В верхней части экрана отображается заголовок "Авторизация". Далее идут два поля ввода: первое для имени пользователя, второе для пароля. Под полями ввода находится кнопка "Войти", которая активирует процесс авторизации. Стиль интерфейса прост и лаконичен, с чётким разделением элементов и визуальной ясностью.



*Рис.6. Отображения информации о задачах*

После прохождения авторизации, пользователь попадает на страницу где интерфейс отображает список задач на обработку (рис.6), где каждая задача имеет несколько атрибутов, таких как статус, время создания и обновления, а также индикаторы, подтверждающие наличие аудиофайла и успешность выполнения задачи. Визуально задачи представлены с указанием времени начала и окончания обработки, а также с возможностью доступа к результатам и архиву. Интерфейс включает элементы управления для добавления новых задач, а также предоставляет пользователю возможность отслеживать прогресс выполнения задач и управлять ими.



*рис.7 добавление заданий на обработку*

На экране отображается заголовок, который информирует пользователя о добавлении задания на обработку (рис.7). Под заголовком размещены инструкции, в которых говорится о том, что перед выполнением задач необходимо зарегистрироваться на сайте и создать токен для доступа к функциям системы. Для доступа имеется поле для ввода HF токена, который необходим для выполнения операций. Также предусмотрена возможность сохранить токен для повторного использования. В нижней части интерфейса расположена кнопка для выбора звукового файла, который требуется обработать. Завершается форма кнопкой “Отправить”, которая подтверждает действие и отправляет файл на обработку.

# **ГЛАВА 5. ОРГАНИЗАЦИЯ ТЕСТИРОВАНИЯ ПРОЕКТА.**

Тестирование системы в рамках компании "Студия 42" представляет собой не просто отдельный технический этап, а целый комплексный процесс, который раскрывает перед сотрудниками и руководством широкий спектр возможностей и перспектив. Сотрудники, непосредственно работающие над проектом, а также те, кто будет использовать систему в ежедневной практике, вовлечены в каждодневное взаимодействие с новым решением, что помогает сформировать максимально целостное представление о его функциональности. Данный подход далеко не ограничивается лишь формальным оцениванием ключевых возможностей: он включает анализ множества критериев, среди которых удобство интерфейса, скорость обработки информации, точность транскрипции, а также способность системы адаптироваться к разнообразным рабочим условиям.

Процесс тестирования проводится в реальной корпоративной среде, что существенно повышает достоверность получаемых результатов. Система устанавливается на сервере компании, и её функционал проверяется на материале настоящих рабочих совещаний, регулярно проходящих в "Студии 42". Вместо создания искусственных, упрощённых тестовых данных, используется полный набор реальных аудиозаписей, демонстрирующих весь спектр возможных ситуаций: от небольших обсуждений текущих задач до крупных планёрок с множеством участников. В этом многообразии легко проследить, насколько система способна корректно обработать различные типы речевых потоков, адекватно справиться с фоновым шумом, не потерять нить дискуссии и правильно идентифицировать каждого говорящего.

Особое внимание в ходе тестирования уделяется деталям, которые могут быть незаметны при беглом осмотре, но оказываются критически важными для практического использования решения. Например, вопрос диаризации — правильного выделения голосов отдельных участников совещания — становится центральным, если в комнате ведут разговор сразу несколько человек. Система должна не просто транскрибировать речь, но и корректно привязывать конкретные фразы к конкретным участникам. Это может оказаться непростой задачей, особенно если речь идёт о совещании, в котором люди имеют различные речевые особенности, говорят с акцентами, на нескольких языках или при большом объёме фоновых шумов (стук клавиатур, шелест бумаг, звуки улицы за окном, непредвиденные перебивания).

Помимо оценки чисто технических характеристик, тестировщики уделяют внимание и тому, насколько легко и интуитивно понятно работать с системой рядовым сотрудникам, не имеющим глубокого технического бекграунда. Интерфейс должен быть достаточно дружелюбным, чтобы пользователь мог без долгих инструкций и тренировок быстро сориентироваться: найти нужный фрагмент транскрипта, выделить ключевые тезисы, экспортировать подготовленный протокол в удобном формате. Чем проще и быстрее можно выполнить эти повседневные задачи, тем выше ценность решения для компании.

Чтобы оценить экономический эффект от внедрения системы, проводится сравнительный анализ эффективности работы "до" и "после". На одном конце спектра — ручной труд по прослушиванию записей и подготовке протоколов встреч, процесс длительный, часто утомительный и подверженный человеческим ошибкам. На другом — автоматизированное решение, способное в разы сократить время обработки данных и повысить точность итоговой документации. Фактические замеры длительности формирования отчетов, сравнительный анализ качества транскрибации и проверка доступности результатов позволяют получить объективную картину того, насколько новая система реально облегчает повседневную жизнь сотрудников.

Собранная в ходе тестирования обратная связь от сотрудников содержит широкий спектр мнений и предложений. Некоторые отмечают, что система существенно ускоряет сбор итоговых материалов, другие указывают на нюансы, требующие доработки: недостаточно точную диаризацию, определённые проблемы с обработкой речи при наличии сильных акцентов или же необходимость в более гибких настройках интерфейса. Подобный многообразный и разносторонний фидбек предоставляет разработчикам понимание того, что именно нуждается в корректировке, и помогает сформировать приоритеты для дальнейшего развития продукта.

Итогом тестирования становится не только формальное заключение о том, готова ли система к широкому внедрению. В результате мы имеем подробное описание выявленных недочётов, рекомендации по их устранению, а также целый ряд перспективных идей, которые можно реализовать в будущих версиях. Таким образом, тестирование не сводится к разовой проверке: это динамический, живой процесс, в котором "Студия 42" не просто убеждается в работоспособности продукта, но и формирует благодатную почву для постоянного совершенствования решения, делая его более эффективным, удобным и востребованным для всех участников рабочего процесса.

Для более детального понимания функциональности системы было проведено несколько тестов. В ходе тестирования использовались записи совещаний с различным уровнем фона, шумами и акцентами. Выяснилось, что система справляется с лёгким фоном, однако в условиях шумной обстановки (например, в офисах с несколькими разговаривающими людьми) точность транскрибации снижалась на 15-20%.

Тестирование диаризации (определение говорящих) показало, что система демонстрирует хорошие результаты в определении говорящих, однако в случае сильных акцентов или перебиваний точность снижалась. Некоторые говорящие были неправильно идентифицированы, особенно в случае быстрого перехода от одного участника к другому.

Проведенные опросы среди сотрудников показали, что интерфейс системы достаточно интуитивно понятен для пользователей, не обладающих техническим опытом. Однако несколько сотрудников выразили пожелания по улучшению организации поиска нужных фрагментов в протоколах, а также более гибкой настройке экспорта данных.

Замеры времени, необходимого для подготовки протоколов с использованием новой системы, показали, что процесс автоматической транскрибации и реферирования сокращает время на 60-70% по сравнению с ручной обработкой записей. Время на подготовку полного отчета, включающего основную информацию из совещания, уменьшилось с 2-3 часов до 30-40 минут.

Эти тесты дали полное представление о возможностях системы, выявили её сильные стороны и области, которые требуют доработки, что позволит на основе полученных данных сформировать план для дальнейшей оптимизации.

# **ЗАКЛЮЧЕНИЕ**

В ходе разработки программного решения для транскрибации и реферирования аудиозаписей совещаний для предприятия "Студия 42" удалось не только достичь поставленных целей, но и сформировать целостную систему, которая соответствует предъявленным требованиям и даже превышает ожидания по ряду аспектов. Процесс создания данного решения был длительным и многогранным: от первичного анализа технических требований и оценки возможностей внедрения современных алгоритмов обработки звука до кропотливой работы над внедрением и оптимизацией архитектуры программного обеспечения, учитывающей разнообразные сценарии применения.

Созданное программное решение стало результатом использования передовых технологий и подходов к распознаванию речи, что позволило не только преобразовать аудиозаписи в подробные текстовые протоколы, но и обеспечить высокий уровень точности, который сохраняется даже в непростых условиях записи. Эти условия включают в себя наличие фонового шума, нечеткую дикцию участников беседы, а также низкую громкость речи отдельных ораторов. Благодаря адаптивным моделям машинного обучения и комплексным алгоритмам обработки сигнала удалось достигнуть уровня распознавания, при котором итоговые тексты требуют минимальной последующей корректуры со стороны сотрудников.

Функциональность проекта не ограничивается лишь базовой транскрибацией. Важно отметить, что система включает диаризацию участников, что позволяет четко определить, кто именно из присутствующих высказал ту или иную реплику. Это особенно актуально для совещаний, где в дискуссии может участвовать большое количество людей, и впоследствии необходимо отслеживать индивидуальный вклад каждого из них. Кроме того, реализована удобная визуализация результатов: пользователи могут просматривать расшифрованные тексты в интуитивно понятном веб-интерфейсе, где информация представлена структурированно, с акцентами на ключевых моментах. Предусмотрена также возможность экспорта данных в различные форматы, что облегчает дальнейшую работу с материалом, например, интеграцию с внутренними системами документооборота или анализ собранных сведений с использованием сторонних инструментов.

Все эти возможности сделали разрабатываемую систему универсальным и гибким инструментом, способным повысить общую эффективность внутренних процессов компании. В ходе детального и многоэтапного тестирования было выявлено, что использование системы не только уменьшает нагрузку на сотрудников, вынужденных ранее вручную расшифровывать и структурировать аудиозаписи, но и позитивно сказывается на общем качестве подготовки протоколов совещаний. Сокращается время, затрачиваемое на рутинные операции, снижается вероятность человеческой ошибки, и, как следствие, повышается скорость распространения достоверной информации внутри организации.

Обратная связь от сотрудников компании "Студия 42" продемонстрировала, что внедренное решение не только удобно в использовании, но и практически ценно для выполнения ежедневных задач. Благодаря простому и понятному интерфейсу, а также гибким настройкам, стало гораздо легче ориентироваться в массиве данных, быстро находить нужные фрагменты и формировать краткие отчеты. Все это создает условия для развития культуры прозрачного и оперативного информационного обмена в рамках предприятия.

Однако на достигнутом останавливаться не планируется. Разработанное программное обеспечение имеет значительный потенциал для дальнейшего совершенствования и масштабирования. В перспективе возможна интеграция системы с другими корпоративными инструментами: например, с системами планирования ресурсов, управления проектами, CRM-платформами или даже собственными аналитическими модулями, которые позволят на основе полученных данных строить прогнозы, выявлять закономерности и принимать более взвешенные бизнес-решения. Кроме того, планируется дальнейшее улучшение моделей транскрибации и алгоритмов диаризации, чтобы повысить точность работы в самых разнообразных условиях: будь то международные совещания с участием носителей разных языков или специфические отраслевые совещания с обилием профессиональной терминологии.

Таким образом, реализация данного проекта продемонстрировала, что современные технологии обработки аудиоданных и текстовой информации могут играть ключевую роль в решении актуальных прикладных задач бизнеса. Применение подобных инструментов способствует оптимизации рабочих процессов, повышению производительности сотрудников и улучшению качества принимаемых решений. С каждым этапом развития подобных систем предприятия, подобные "Студии 42", получают дополнительное конкурентное преимущество, выражающееся не только в экономии времени и ресурсов, но и в более четком, ясном и своевременном доступе к критически важной информации. Всё это ведет к укреплению позиции компании на рынке, росту ее репутации, а также созданию благоприятной среды для дальнейшего инновационного развития и применения новых, еще более продвинутых технологий анализа и обработки данных.

# 

# **СПИСОК ЛИТЕРАТУРЫ**

1. **Нестерова, Н. М.**, **Герте, Н. А.** Реферирование как способ извлечения и представления основного содержания текста // *Российская и зарубежная филология*. — 2013. — № 4. — С. 45–52. — URL:<https://cyberleninka.ru/article/n/referirovanie-kak-sposob-izvlecheniya-i-predstavleniya-osnovnogo-soderzhaniya-teksta> (дата обращения: 15.12.2024).
2. **Соловова, Е. Н.**, **Басова, И. А.** Необходимость обучения студентов профильных языковых вузов (факультетов) реферированию звучащих академических текстов // *Вестник Московского университета. Серия 19. Лингвистика и межкультурная коммуникация*. — 2015. — № 2. — С. 98–105. — URL:<https://cyberleninka.ru/article/n/neobhodimost-obucheniya-studentov-profilnyh-yazykovyh-vuzov-fakultetov-referirovaniyu-zvuchaschih-akademicheskih-tekstov> (дата обращения: 15.12.2024).
3. **Yoshioka, T.**, **Abramovski, I.**, **Aksoylar, C.**, и др. Advances in Online Audio-Visual Meeting Transcription // *arXiv preprint arXiv:1912.04979*. — 2019. — 12 с. — URL:<https://arxiv.org/abs/1912.04979> (дата обращения: 15.12.2024).
4. **Методические основы обучения письменному реферированию аудиотекстов в профильных языковых вузах** : дис. ... канд. пед. наук : 13.00.02 / **Соловова Е. Н.** — Москва, 2015. — 189 с. — URL:<https://nauka-pedagogika.com/pedagogika-13-00-02/dissertaciya-metodicheskie-osnovy-obucheniya-pismennomu-referirovaniyu-audiotekstov-v-profilnyh-yazykovyh-vuzah> (дата обращения: 15.12.2024).
5. **Теория и практика аудиовизуального перевода: отечественный и зарубежный опыт** // *Вестник Челябинского государственного университета*. — 2014. — № 1 (341). — С. 114–119. — URL:<https://cyberleninka.ru/article/n/teoriya-i-praktika-audiovizualnogo-perevoda-otechestvennyy-i-zarubezhnyy-opyt> (дата обращения: 15.12.2024).
6. **Аннотирование и реферирование научного текста** : учебное пособие / **Соловова Е. Н.**, **Басова И. А.** — Астрахань : Астраханский государственный медицинский университет, 2023. — 56 с. — URL:<https://astgmu.ru/wp-content/uploads/2023/02/Annotirovanie-i-referirovanie-nauchnyh-tekstov-1.pdf> (дата обращения: 15.12.2024).