

ТЕХНИЧЕСКОЕ ЗАДАНИЕ

на создание веб-сервиса автоматической расшифровки и структурирования аудио «Speak2MD»

1. Общие сведения

1.1. Полное наименование системы: Веб-сервис автоматической транскрибации и генерации Markdown-конспектов «Speak2MD».

1.2. Назначение системы: Автоматизация процесса перевода аудиозаписей (рабочих совещаний, лекций, голосовых заметок) в структурированный текстовый формат с выделением заголовков, тезисов и задач.

1.3. Заказчик и Разработчик: Разрабатывается в рамках учебного проектного интенсива.

1.4. Основания для разработки: (из файла “Паспорт проекта «Speak2MD» (версия 1.3)”).

- Потребность в инструменте для быстрой обработки записей встреч без использования дорогостоящих облачных решений с поминутной тарификацией.

1.5. Плановые сроки:

- Реализация минимально жизнеспособного продукта (MVP): 03.10.2025 - 07.11.2025.
- Реализация полноценной системы: 07.11.2025 - 28.11.2025.

2. Цели и назначение создания АС

2.1. Цели:

- Сокращение времени на подготовку протоколов встреч (Meeting Notes) в 8–10 раз по сравнению с ручной расшифровкой.
- Обеспечение точности распознавания речи (метрика WER) не хуже 18% на русском языке.
- Создание единого цифрового архива стенограмм с возможностью поиска и экспорта.

2.2. Критерии успеха:

- Успешная обработка аудиофайлов длительностью до 120 минут без сбоев.
- Корректное форматирование валидного файла разметки (Markdown).
- Корректная работа ролевой модели доступа (Гость, Пользователь, Администратор).

3. Характеристика объекта автоматизации

3.1. Тип системы: Веб-приложение (клиент-серверная архитектура).

3.2. Основные бизнес-процессы:

- Загрузка и запись:** Передача пользователем аудио-потока или файла на сервер.
- Конвейер обработки:** Преобразование формата - Определение голосовой активности (VAD) - Распознавание речи (ASR) - Лингвистическая постобработка (LLM).
- Доставка результата:** Обновление статуса обработки в реальном времени и предоставление готового документа.
- Управление контентом:** Хранение истории расшифровок, систематизация, предоставление общего доступа.

3.3. Участники (Акторы):

- **Гость:** Доступ к публичным страницам (Главная, Вход/Регистрация) и просмотр стенограмм по публичным ссылкам (только чтение).
- **Пользователь:** Загрузка аудиоматериалов, управление личным архивом стенограмм, управление тарифным планом.
- **Администратор:** Мониторинг очереди задач, управление учетными записями пользователей, просмотр системной статистики.

4. Требования к автоматизированной системе

4.1. Требования к структуре и функционированию

Система должна состоять из следующих компонентов:

1. **Клиентская часть (Frontend):** Графический интерфейс пользователя на HTML5/JS.
2. **Серверная часть (API Gateway):** Сервер приложений (на базе FastAPI), обрабатывающий HTTP-запросы и WebSocket соединения.
3. **Служба фоновой обработки (Worker):** Изолированный процесс для выполнения ресурсоемких задач машинного обучения (ASR/LLM).
4. **База данных:** СУБД PostgreSQL для хранения учетных записей, метаданных задач и ссылок.
5. **Файловое хранилище:** Локальная файловая система для хранения исходных аудиозаписей и результирующих артефактов (.wav, .md, .json).

4.2. Функциональные требования (FR)

Группа 1: Аутентификация и Учетные записи

- **FR-AUTH-01:** Регистрация пользователей с указанием логина, пароля и имени. Пароли должны храниться в зашифрованном виде (хеширование bcrypt).
- **FR-AUTH-02:** Аутентификация пользователей с использованием механизма JWT (Access + Refresh токены).
- **FR-AUTH-03:** Предоставление пользователю информации о его профиле, текущем тарифном плане и остатке квот (/api/me).

Группа 2: Работа с аудиоданными

- **FR-CORE-01:** Загрузка аудиофайлов форматов MP3, M4A, WAV через REST API.
- **FR-CORE-02:** Запись аудиосигнала с микрофона непосредственно в браузере (MediaRecorder API) и передача на сервер.
- **FR-CORE-03:** Предварительная валидация файлов: ограничение размера до 50 МБ и длительности до 120 минут (для тарифа "Бесплатный").
- **FR-CORE-04:** Асинхронная обработка. Прием файла должен немедленно возвращать идентификатор задачи (job_id), не блокируя соединение на время распознавания.

Группа 3: Мониторинг и Результаты

- **FR-STAT-01:** Отслеживание прогресса обработки в реальном времени через протокол WebSocket. События: «прогресс», «смена статуса», «завершено», «ошибка».
- **FR-RES-01:** Генерация итогового документа в формате Markdown, содержащего заголовок, основной текст и структурированные списки.
- **FR-RES-02:** Возможность скачивания результата в виде файлов .md (текст) и .json (структурные данные).

- **FR-RES-03:** Предпросмотр отформатированного текста непосредственно в окне браузера.

Группа 4: Управление стенограммами и доступом

- **FR-TRN-01:** Просмотр списка личных стенограмм с возможностью просмотра md-текста и фильтрации.
- **FR-TRN-02:** Удаление стенограмм (каскадное удаление записи в БД и связанных файлов на диске).
- **FR-TRN-03:** Добавление и редактирование текстовых тегов для классификации записей.
- **FR-SHARE-01:** Генерация уникальной публичной ссылки для предоставления доступа к результату третьим лицам без регистрации.
- **FR-SHARE-02:** Возможность принудительного отзыва публичной ссылки владельцем.

Группа 5: Администрирование и Тарификация

- **FR-ADM-01:** Просмотр администратором реестра всех пользователей с отображением их статуса (Активен/Заблокирован).
- **FR-ADM-02:** Просмотр реестра всех задач обработки в системе с фильтрацией по статусу выполнения.
- **FR-PLAN-01:** Автоматический учет потребления ресурсов (минут распознавания) за календарный месяц.
- **FR-PLAN-02:** Механизм смены тарифного плана пользователем (Free, Pro, Team) с автоматическим обновлением лимитов.

4.3. Требования к внешним интерфейсам (IF)

- **IF-01 (Протоколы):** Взаимодействие между Клиентской и Серверной частями осуществляется по протоколу HTTP/1.1 (REST API) и WebSocket.
- **IF-02 (LLM-интеграция):** Взаимодействие с сервисом больших языковых моделей (LLM) для задач суммаризации текста.
- **IF-03 (ASR-модели):** Использование локально развернутых моделей распознавания речи (Faster-Whisper / PaddleSpeech).

4.4. Системные требования и ограничения

- **NFR-01 (Производительность):** Время обработки аудиозаписи не должно превышать 50% от её длительности (коэффициент x0.5) на эталонном оборудовании.
- **NFR-02 (Надежность):** Система должна корректно обрабатывать сбои (неверный формат, ошибки модели) и возвращать информативные коды ошибок HTTP (4xx/5xx).
- **CR-01 (Технологический стек):** Серверная часть на языке Python (фреймворк FastAPI), база данных PostgreSQL.
- **CR-02 (Информационная безопасность):** Разграничение доступа на уровне ролевой модели (RBAC): пользователь имеет доступ только к своим данным, администратор — ко всем.

5. Состав и содержание работ (Этапы реализации)

В связи с итеративным подходом к разработке, создание системы разбито на функциональные блоки. Текущий статус реализации:

Этап 1. Базовое ядро (MVP Core) — Реализовано

- Разработка API для загрузки файлов.

- Реализация конвейера конвертации и распознавания речи.
- Настройка WebSocket для отображения прогресса.

Этап 2. Хранение и Структурирование — Реализовано

- Подключение базы данных PostgreSQL.
- Реализация моделей данных для задач и артефактов.
- Интеграция с LLM для интеллектуального форматирования текста.

Этап 3. Пользовательские функции и Безопасность — Реализовано

- Реализация модуля регистрации и авторизации.
- Личный кабинет пользователя со списком расшифровок.
- Функционал «Поделиться результатом».
- Панель администратора для управления системой.

6. Требования к документированию

По результатам выполнения работ должен быть представлен комплект документации:

1. **Спецификация API:** Описание методов программного интерфейса в формате OpenAPI (файл speak2md-mvp.yaml и API_Documentation.md).
2. **Инструкция по развертыванию:** Руководство по запуску системы в контейнерах (в файле README.md).
3. **Руководство пользователя:** Описание сценариев работы с веб-интерфейсом.

Составил: Сергеев А.П. Дата 13.11.2025

Обновлено: Сергеев А.П. Дата 24.11.2025 (Смена статуса реализации этапов на “Реализовано”)