

**D** DIGIT / РЕЧЕВОЙ ИНТЕЛЛЕКТ / 

**ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ**

**«Digit Речевой интеллект»**

**Описание функциональных характеристик**

Листов 14

## **АННОТАЦИЯ**

Настоящий документ содержит описание функциональных характеристик программного обеспечения «Digit Речевой интеллект», предназначенного для автоматического преобразования аудио и видеозаписей в текст. Превращает записи в удобные протоколы: выделяет ключевые идеи, фиксирует поставленные задачи и принятые решения. Система работает на собственных серверах организации, обеспечивая высокую конфиденциальность данных без обращения во внешние сервисы.

## СОДЕРЖАНИЕ

-----	1
1 ЦЕЛИ И АВТОМАТИЗИРУЕМЫЕ ФУНКЦИИ -----	5
1.1 Цели создания программного обеспечения -----	5
1.2 Автоматизируемые функции программного обеспечения -----	5
1.3 Потенциальные пользователи программного обеспечения -----	6
2 КОМПОНЕНТЫ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ -----	7
2.1 Сервис распознавания аудио и видеоматериалов -----	7
2.2 Сервис обработки и распознавания речи -----	7
2.3 Сервис обработки и распознавания речи -----	8
3 ХАРАКТЕРИСТИКА ФУНКЦИОНАЛЬНОЙ СТРУКТУРЫ -----	9
4 АППАРАТНЫЕ И ПРОГРАММНЫЕ ТРЕБОВАНИЯ К ПО -----	10
4.1 Требования к серверной части модуля «Digit Речевой интеллект» (базовый функционал) -----	10
4.2 Требования к клиентской части -----	11
5 ЧИСЛЕННОСТЬ И КВАЛИФИКАЦИЯ ПОЛЬЗОВАТЕЛЕЙ СИСТЕМЫ -----	12
6 РЕЖИМ ФУНКЦИОНИРОВАНИЯ СИСТЕМЫ -----	13

## СПИСОК ТЕРМИНОВ И СОКРАЩЕНИЙ

Термин, сокращение	Значение
ПО	Программное обеспечение
ASR (Automatic Speech Recognition)	Технология автоматического распознавания речи, предназначенная для преобразования устной аудио- или видеозаписи в человеко-читаемый текст
VAD (Voice Activity Detection)	Технология цифровой обработки сигналов, предназначенная для обнаружения речи в непрерывном аудиопотоке и отделения её от фонового шума или тишины
LLM (Large Language Model)	Большая языковая модель, основанная на нейронной сети с множеством параметров (миллиарды весовых коэффициентов и более), которая проходит предварительное обучение на обширных массивах неразмеченного текста методами самообучения)
Стенограмма (транскрипция)	Полный дословный письменный текст устной речи, полученный в результате преобразования аудио- или видеозаписи
Транскрибация	Автоматизированный процесс распознавания и записи устной речи из аудио- или видеоматериалов в текстовый формат
Резюмирование	Процедура создания структурированного краткого изложения основных итогов встречи или разговора на основе полной стенограммы
Суммаризация	Автоматическое формирование сжатой аннотации обсуждения (несколько предложений), передающей суть разговора в концентрированной форме
Протокол	Формализованный документ по заранее заданному шаблону, составляемый на основе транскрипта; включает полный или сокращённый текст разговора и, при необходимости, резюме или ключевые выводы

# **1 ЦЕЛИ И АВТОМАТИЗИРУЕМЫЕ ФУНКЦИИ**

## **1.1 Цели создания программного обеспечения**

Целью создания программного обеспечения «Digit Речевой интеллект» является обеспечение автоматизации процессов преобразования аудио- и видеозаписей в текстовый вид, а также повышение эффективности работы пользователей с речевой информацией в рамках деятельности организации.

Программное обеспечение обеспечивает решение следующих задач:

- автоматизация процессов распознавания устной речи и формирования текстовых представлений аудио- и видеоматериалов;
- повышение эффективности обработки информации, полученной в ходе совещаний, переговоров и иных коммуникаций, за счёт автоматического формирования стенограмм, резюме и протоколов;
- сокращение временных затрат пользователей на ручную обработку аудио- и видеозаписей;
- повышение качества фиксации результатов обсуждений за счёт автоматического выделения ключевых идей, поставленных задач и принятых решений;
- обеспечение централизованной и структурированной работы с текстовыми результатами обработки речевой информации;
- обеспечение конфиденциальности обрабатываемых данных за счёт функционирования программного обеспечения в изолированном контуре организации без передачи информации во внешние сервисы.

## **1.2 Автоматизируемые функции программного обеспечения**

ПО «Digit Речевой интеллект» позволяет автоматизировать:

- загрузка пользователем аудио- и видеоматериалов в систему с поддержкой распространённых форматов файлов (MP3, WAV, OGG, FLAC, MP4, M4A, AVI, MOV, MKV, WMV, FLV, MPEG, MPG, WEBM, 3GP, AVCHD, HEVC, VOB, TS, M4V, OGV, MPEG4) с ограничением размера загружаемого файла до 1 ГБ;
- автоматическое распознавание устной речи с преобразованием аудио- и видеозаписей в текстовый формат;
- автоматическая постобработка распознанного текста, включая расстановку пунктуации и исправление ошибок;
- формирование полной стенограммы (транскрипции) аудио- или видеозаписи;

- автоматическое резюмирование содержания записи;
- автоматическое выделение ключевых элементов обсуждения (идеи, задачи, решения);
- формирование протоколов на основе результатов обработки аудио- и видеоматериалов;
- хранение и обработка данных в локальном контуре организации (on-premise) без передачи информации во внешние системы.

### **1.3 Потенциальные пользователи программного обеспечения**

Потенциальными пользователями программного обеспечения «Digit Речевой интеллект» являются:

- сотрудники организаций, осуществляющие работу с аудио- и видеоматериалами (записи совещаний, переговоров, интервью и иных коммуникаций);
- специалисты, ответственные за подготовку стенограмм, протоколов и аналитических материалов на основе речевой информации;
- руководители и сотрудники подразделений, использующие результаты обработки речевых данных для контроля исполнения задач и анализа принятых решений;
- сотрудники, участвующие в совещаниях и заинтересованные в фиксации и последующем использовании их результатов;
- иные пользователи, в функциональные обязанности которых входит обработка, анализ и документирование устной информации.

## 2 КОМПОНЕНТЫ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ

Программное обеспечение «Digit Речевой интеллект» реализовано на основе клиент-серверной архитектуры и представляет собой совокупность взаимосвязанных сервисов, обеспечивающих обработку аудио- и видеоматериалов и последующий анализ речевой информации.

Система развёртывается в инфраструктуре Заказчика (on-premise) и функционирует без использования внешних облачных сервисов.

В состав программного обеспечения входят следующие сервисы:

### 2.1 Сервис распознавания аудио и видеоматериалов

Сервис предназначен для приёма, хранения и первичной обработки загружаемых пользователем аудио- и видеофайлов.

В рамках сервиса используются следующие компоненты:

- объектное хранилище (S3-совместимое) на базе MinIO для хранения исходных файлов и результатов обработки;
- in-memory хранилище Redis для организации очередей задач;

Сервис обеспечивает:

- приём и сохранение загружаемых файлов;
- управление задачами обработки;
- хранение результатов обработки.

### 2.2 Сервис обработки и распознавания речи

Сервис представляет собой программную обёртку над моделями обработки речевой информации и обеспечивает выполнение ключевых этапов анализа аудиосигнала.

В состав сервиса входят следующие модели:

- VAD-модель (Voice Activity Detection) на базе Silero, обеспечивающая выделение участков аудиосигнала, содержащих речь;
- ASR-модель (Automatic Speech Recognition) GigaAM-v3, обеспечивающая преобразование речевого сигнала в текстовый формат.

Сервис обеспечивает:

- сегментацию записи на речевые фрагменты;
- преобразование аудио в текст (транскрибацию);

- передачу результатов в последующие этапы обработки.

### **2.3 Сервис обработки и распознавания речи**

Сервис предназначен для аналитической обработки текстовых данных, полученных в результате распознавания речи.

В основе сервиса используется LLM-модель Qwen.

Сервис обеспечивает:

- постобработку текста (улучшение читаемости, структурирование);
- автоматическое резюмирование содержания;
- выделение ключевых идей, задач и принятых решений;
- формирование структурированных протоколов.

### **3 ХАРАКТЕРИСТИКА ФУНКЦИОНАЛЬНОЙ СТРУКТУРЫ**

Программное обеспечение «Digit Речевой интеллект» представляет собой программный продукт, реализованный на основе клиент-серверной архитектуры и предназначенный для обработки аудио- и видеоматериалов с последующим преобразованием речевой информации в текстовый вид.

Функциональная структура программного обеспечения включает следующие основные уровни:

- пользовательский уровень, обеспечивающий взаимодействие пользователей с системой посредством веб-интерфейса;
- прикладной уровень, включающий сервисы загрузки, хранения и управления обработкой аудио- и видеоматериалов;
- уровень обработки данных, включающий сервисы распознавания речи, сегментации аудиосигнала и интеллектуальной обработки текста;
- уровень хранения данных, обеспечивающий хранение исходных файлов, промежуточных данных и результатов обработки.

Взаимодействие компонентов системы осуществляется в рамках единого технологического процесса, включающего загрузку данных, их обработку и предоставление пользователю результатов в виде стенограмм, резюме и протоколов.

Доступ пользователей к функциональным возможностям программного обеспечения осуществляется через веб-интерфейс.

## 4 АППАРАТНЫЕ И ПРОГРАММНЫЕ ТРЕБОВАНИЯ К ПО

### 4.1 Требования к серверной части модуля «Digit Речевой интеллект» (базовый функционал)

Для развертывания программного обеспечения «Digit Речевой интеллект» используется серверная инфраструктура Заказчика.

Программное обеспечение устанавливается на виртуальный или физический сервер и функционирует под управлением серверной операционной системы.

Рекомендуемые минимальные конфигурации серверного оборудования, а также перечень необходимого системного и прикладного программного обеспечения для серверной части «Digit Речевой интеллект» приведены в таблице 1.

Таблица 1 - Требования к серверной части «Digit Речевой интеллект»

№ п/п	Наименование параметра	Значение параметра
Сервер приложений сервиса распознавания аудио и видеоматериалов, и сервиса обработки и распознавания речи		
1.	CPU - Количество ядер процессора, шт. (с поддержкой инструкций AVX, AVX2 или AVX512)	16
2.	RAM - Объем оперативной памяти, ГБ.	16
3.	HDD – Объем дисков, ГБ	40 (ОС) + 512 (ПО)
Операционная система:	«Astra Linux SE»	
ПО, которое будет установлено на сервер:	Redis, S3 хранилище MinIO, Docker Compose, Docker Engine	
Сервер приложений сервиса обработки и распознавания речи		
1.	CPU - Количество ядер процессора, шт. (с поддержкой инструкций AVX, AVX2 или AVX512)	16
2.	RAM - Объем оперативной памяти, ГБ.	16
3.	HDD – Объем дисков, ГБ.	40 (ОС) + 512 (ПО)
4.	GPU – Объем видеопамяти, ГБ. (рекомендуемый минимум)	40
Операционная система:	«Astra Linux SE»	
ПО, которое будет установлено на сервер:	Docker Compose, Docker Engine, vLLM	

## 4.2 Требования к клиентской части

Для работы с программным обеспечением «Digit Речевой интеллект» используется рабочее место пользователя, оснащённое современным веб-браузером и доступом к сети, обеспечивающей взаимодействие с серверной частью системы.

Клиентская часть не требует установки специализированного программного обеспечения и функционирует через веб-интерфейс.

Минимальные требования к клиентскому рабочему месту включают:

- наличие персонального компьютера с доступом в сеть;
- наличие одного из современных веб-браузеров (например, Google Chrome, Mozilla Firefox, Microsoft Edge или их аналогов);
- обеспечение стабильного сетевого соединения с серверной частью программного обеспечения;
- наличие устройств ввода-вывода, необходимых для работы с системой (клавиатура, мышь и др.).

Дополнительных требований, связанных с установкой специализированных компонентов или расширений, не предъявляется.

## **5 ЧИСЛЕННОСТЬ И КВАЛИФИКАЦИЯ ПОЛЬЗОВАТЕЛЕЙ СИСТЕМЫ**

Количество пользователей программного обеспечения «Digit Речевой интеллект» определяется текущими потребностями Заказчика и не ограничивается архитектурными или функциональными возможностями ПО.

Для работы с программным обеспечением «Digit Речевой интеллект» пользователи должны обладать базовыми навыками работы с персональным компьютером, современными веб-браузерами и пользовательскими интерфейсами информационных систем.

Специальная техническая подготовка для использования функционала ПО «Digit Речевой интеллект» не требуется.

## **6 РЕЖИМ ФУНКЦИОНИРОВАНИЯ СИСТЕМЫ**

В основном режиме функционирования программное обеспечение «Digit Речевой интеллект» обеспечивает:

- функционирование в круглосуточном режиме (24 часа в сутки, 7 дней в неделю),
- выполнение заявленных функциональных возможностей в полном объёме, за исключением периодов проведения регламентных, профилактических и иных технических работ, а также времени устранения нештатных ситуаций.

В случае возникновения нештатных ситуаций предусмотрена возможность восстановления работоспособности программного обеспечения «Digit Речевой интеллект» путём возврата к последней работоспособной версии с сохранением целостности и непротиворечивости данных на уровне системы управления базами данных.