Программный модуль «Speech to text»

**Описание программы**

Инв. № подл.

Подп. и дата

Взам. инв. №

Инв. № дубл.

Подп. и дата

**Листов\_17**

2023

**АННОТАЦИЯ**

Программный модуль «Speech To Text» –это автоматизированная система, разработанная для помощи сотрудникам оперативных служб, осуществляющих контроль нарушений в телефонных переговорах лиц, содержащихся на территории учреждений ФСИН России, которые производятся при помощи специализированной закрытой телефонной сети.

На данный момент нарушениями считается удержание вызова, переход в разговоре на иностранную речь и использование стоп-слов. Типы нарушений добавляются в систему по мере обучения нейронной сети.

Разговор анализируется в режиме реального времени. Результат обнаружения нарушений сохранятся в журнале звонков и доступен для дальнейшего анализа.

Модуль «Speech To Text» функционирует совместно с программой «МИНОС».

ОГЛАВЛЕНИЕ

[1. ОБЩИЕ СВЕДЕНИЯ 4](#_Toc135384907)

[1.1 Обозначения и наименования программы 4](#_Toc135384908)

[1.2 Программное обеспечение, необходимое для функционирования программы 4](#_Toc135384909)

[1.3 Языки программирования, на которых написана программа 4](#_Toc135384910)

[2. ФУНКЦИОНАЛЬНЫЕ НАЗНАЧЕНИЯ 5](#_Toc135384911)

[2.1 Назначение программы 5](#_Toc135384912)

[2.2 Сведения о функциональных ограничениях на применение 5](#_Toc135384913)

[3. ОПИСАНИЕ ЛОГИСТИЧЕСКОЙ СТУКТУРЫ 6](#_Toc135384914)

[3.1 Сервисная структура 6](#_Toc135384915)

[3.2 Алгоритм функционирования 6](#_Toc135384916)

[3.3. Структура программы с описанием функций составных частей и связи между ними 7](#_Toc135384917)

[3.4 Связи программы с другими программами 10](#_Toc135384918)

[4. ИСПОЛЬЗУЕМЫЕ ТЕХНИЧЕСКИЕ СРЕДСТВА 10](#_Toc135384919)

[4.1 Пользовательский компьютер следующей конфигурации: 10](#_Toc135384920)

[4.2 Аппаратно-программные ресурсы выделенного сервера: 10](#_Toc135384921)

[5. ВЫЗОВ И ЗАГРУЗКА 10](#_Toc135384922)

[6. ВХОДНЫЕ ДАННЫЕ 11](#_Toc135384923)

[7. ВЫХОДНЫЕ ДАННЫЕ 11](#_Toc135384924)

[8. ПЕРЕЧЕНЬ СОКРАЩЕНИЙ 11](#_Toc135384925)

# 1. ОБЩИЕ СВЕДЕНИЯ

# 1.1 Обозначения и наименования программы

Наименование программы – программный модуль «Speech To Text»

# 1.2 Программное обеспечение, необходимое для функционирования программы

Персональный компьютер с установленной операционной системой:

* Windows 7 с пакетом обновления 3 и выше (32/64-разрядная, Windows 7, Windows 8, Windows 10);
* Mac OS X Версия 10.6-10.8 и выше (Snow Leopard / Lion / Mountain Lion);
* Linux – Ubuntu 12.04 и выше.

# 1.3 Языки программирования, на которых написана программа

Язык API предоставляемого API: gRPC

* Операционная система: Linux Debian
* Язык программирования: Go
* Используемая СУБД: PostgreSQL

Микросервис для принятия решения обрыва связи:

* Операционная система: Linux Debian
* Язык программирования: Go

База стоп-слов/словосочетаний:

* Операционная система: Linux Debian
* Язык программирования: Go
* Используемая СУБД: PostgreSQL

Интерфейс управления в «МИНОС»:

* Операционная система: Linux Debian
* Язык программирования: Perl(5),Php(7.1)
* Используемая СУБД: PostgreSQL

Интерфейс для модерирования стоп-слов для внутренних сотрудников

* Операционная система: Linux Debian
* Язык программирования: Perl(5),Php(7.1)
* Используемая СУБД: PostgreSQL

# 2. ФУНКЦИОНАЛЬНЫЕ НАЗНАЧЕНИЯ

# 2.1 Назначение программы

Программный модуль «Speech to text» позволяет:

* распознавать переход речи осужденных с русского на иностранный язык;
* распознавать удержание вызова в аудио или видео разговорах осужденного;
* автоматически обрывать звонок при включенном автообрыве для ИУ в случае выявления нарушений;
* выявлять слова и речевые конструкции, несущие в себе противоправные действия в аудио или видео разговорах осужденного;
* фиксировать все диалоги осужденных в текстовом виде с помощью транскрибации аудио записей.

# 2.2 Сведения о функциональных ограничениях на применение

Для взаимодействия сотрудников учреждений с программным модулем, необходим логин и пароль для авторизации в программе «МИНОС» и персональный компьютер, с установленным браузером и имеющий подключение к сети интернет.

Модули и компоненты программы не будут функционировать без предустановленных ОС на выделенных серверах и развёрнутых баз данных.

# 3. ОПИСАНИЕ ЛОГИСТИЧЕСКОЙ СТУКТУРЫ

# 3.1 Сервисная структура

Программный модуль «Speech to text» состоит из нескольких функциональных частей:

* API Asterisk Gateway Interface (AGI);
* Микросервис для принятия решения обрыва связи;
* Микросервис для выявления наличия стоп-слов в транскрибации звонка;
* База стоп-слов/словосочетаний;
* Интерфейс управления в «МИНОС»;

# 3.2 Алгоритм функционирования

Обработка речи происходит в процессе разговора, в режиме реального времени. В процессе функционирования модуль производит:

* выявление нарушений в процессе звонка;
* автообрыв звонка при выявлении нарушений;
* транскрибацию речи и сравнение её с имеющейся базой стоп-слов и словосочетаний.

В случае обнаружения нарушения, модуль выдаёт оповещение в «МИНОС».

Сотрудник имеет возможность прервать звонок, с последующим указанием причины нарушения.

В журнале звонков нарушения выделяются специальными метками.

Мониторинг нарушений доступен в трёх рабочих разделах, программы «МИНОС»:

* «Учет Спецконтингента» - при просмотре журнала звонков в «карточке» конкретного лица.
* «Контроль звонков» - при просмотре журнала звонков.
* «Онлайн контроль переговоров» - в процессе онлайн контроля звонка.

# 3.3. Структура программы с описанием функций составных частей и связи между ними

**API Asterisk Gateway Interface (AGI)**



Рисунок 1.

Структура программы «Приложение сбора и потокового распознавания данных» (Рисунок 1):

Таблица – Структура программы

|  |  |
| --- | --- |
| **Название** | **Наименование** |
| Буфер данных | Системная библиотека |
| Встроенный сервер | Программный сервис |
| HTTP-frontend | Программный модуль |
| GRPC-frontend | Программный модуль |
| Inference  | Программный сервис |

Буфер данных - построен на базе Redis v5 или v6. Для очередей блоков данных и уведомлений используются очереди. Для хранения контекстов простое key-value хранилище.

Встроенный сервер – процесс, порождающий при старте свои (форки) копии, в идеале общим числом по числу ядер ЦПУ машины. Процесс выполняет следующие шаги:

* выбирает с TCP-стека блоки данных размером до 64Кбит;
* выполняет декодирование из формата, указанного в content-type заголовке, в pcm\_s16le;
* записывает декодированные pcm-фреймы в очередь блоков потока в буфер;
* записывает уведомление о появлении данных в очередь инферера по правилу inferrer\_id = request.uuid % num\_inferers;
* записывает метаданные аудиофайла, в частности – число каналов;
* асинхронно слушает очередь Inference и записывает результаты потоков в тело ответа.

HTTP – **frontend** – модуль поддержки HTTP-протокола для потокового распознавания аудио.

GRPC – **frontend** – модуль поддержки gRPC протокол для потокового распознавания аудио.

Inference – процесс, порождающий при старте свои копии (форки), в идеале, общим числом по количеству ГПУ машины. Процесс выполняет следующие шаги:

* слушает, блокируя очередь уведомлений;
* при появлении обновлений в каком-либо потоке читает из буфера его блоки данных и контекст;
* склеивает блоки данных и конвертирует бинарную строку во float32-тензоры фреймов поканально;
* вычисляет вывод ML-модели плавающим по данным тензорам окном размером в 1 с (размер окна зависит от конкретной модели);
* складывает хвосты тензоров в буфер в контекст потока для последующей склейки с новыми блокам данных.

**Микросервис для принятия решения обрыва связи**

При обнаружении нарушения в звонке микросервис автоматически прерывает соединение для осужденного и родственника в соответствии с выявленным типом нарушения. Это достигается путем взаимодействия с сервисом Asterisk через протокол [AMI](https://www.pbxware.ru/wiki/asterisk_managment_interface_ami). В параметрах SIP передается команда "hangup" для соответствующего канала, что приводит к обрыву звонка. Результаты автоматического обрыва звонка отображаются сотруднику ФСИН в разделе "Журнал звонков" «МИНОС».

**База стоп-слов/словосочетаний, разрешенных слов/словосочетаний**

Микросервис производит поиск стоп-слов и словосочетаний в полученной транскрибации звонка, используя базу данных стоп-слов и словосочетаний. Если при проверке обнаруживается соответствие с записями в базе данных, транскрибация звонка сохраняется для последующего анализа сотрудником ФСИН в разделе «Журнал звонков» «МИНОС». Наполнение базы данных осуществляется на основе письменных запросов, предоставляемых сотрудниками ФСИН в технический отдел.

# 3.4 Связи программы с другими программами

«МИНОС» - Личный кабинет сотрудника ФСИН России в АПК «Зонателеком».

# 4. ИСПОЛЬЗУЕМЫЕ ТЕХНИЧЕСКИЕ СРЕДСТВА

# 4.1 Пользовательский компьютер следующей конфигурации:

Персональный компьютер с установленной операционной системой:

* Windows 7 с пакетом обновления 3 и выше (32/64-разрядная, Windows 7, Windows 8, Windows 10);
* Mac OS X Версия 10.6-10.8 и выше (Snow Leopard / Lion / Mountain Lion);
* Linux – Ubuntu 12.04 и выше.

Комплектация персонального компьютера, не хуже:

* Монитор с минимальным разрешением экрана 1024 х 768 точек;
* Сетевой интерфейс Ethernet 10/100 Base-T(порт RJ-45);
* 512 МБ ОЗУ;
* Процессор Intel Core 2 Duo с частотой 2,0 ГГц, процессоры Intel (Universal Binary);
* Клавиатура;
* Манипулятор «мышь».

# 4.2 Аппаратно-программные ресурсы выделенного сервера:

Операционная система: Linux Debian.

Используемая СУБД: PostgreSQL

# 5. ВЫЗОВ И ЗАГРУЗКА

Сервисы и компоненты программы запускаются автоматически, после загрузки операционной системы на сервере.

Для мониторинга нарушений необходимо авторизоваться в «МИНОС». Для этого нужно пройти по ссылке <https://fsin.zonatelecom.ru>

# 6. ВХОДНЫЕ ДАННЫЕ

Входными данными для программного модуля «Speech to text»:

* Аудиозвонок;
* Видеозвонок;
* Список стоп-слов/словосочетаний.

# 7. ВЫХОДНЫЕ ДАННЫЕ

Выходными данными для программного модуля «Speech to text»:

* Журнал звонков с данными:
	+ Выявленные нарушения в звонках;
	+ Данные об автообрыве звонков с выявленными нарушениями;
	+ Данный о выявленных стоп-словах или словосочетаниях в звонках;
* Транскрибация звонков с найденными стоп-словами или словосочетаниями
* Отчеты с данными о правонарушениях.

# 8. ПЕРЕЧЕНЬ СОКРАЩЕНИЙ

ФСИН России– федеральная служба исполнения наказаний.

ОС – операционная система

СУБД – система управления база данных

ОЗУ – оперативное запоминающее устройство

МИНОС- Личный кабинет сотрудника ФСИН России.

|  |
| --- |
| Лист регистрации изменений |
| Номера листов (страниц) | Всего листов (страниц)в докум. | № документа | Входящий № сопрово-дительного документа и дата | Подп. | Дата |
| Изм. | изменен-ных | заменен-ных | новых | аннули-рован-ных |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |