ПАО «ВымпелКом»
ул. 8 Марта, д. 10, стр. 14, г. Москва, 127083
+7 (495) 783 07 00
ОКПО 17337364, ОГРН 1027700166636
ИНН 7713076301, КПП 997750001

beeline.ru

РУКОВОДСТВО ПОЛЬЗОВАТЕЛЯ

Программное обеспечение Система транскрибации разговоров и подсчета характеристик

Листов 10

г. Москва, 2024г.

ПАО «ВымпелКом»
ул. 8 Марта, д. 10, стр. 14, г. Москва, 127083
+7 (495) 783 07 00
ОКПО 17337364, ОГРН 1027700166636
ИНН 7713076301, КПП 997750001

Оглавление

ПАО «ВымпелКом»
ул. 8 Марта, д. 10, стр. 14, г. Москва, 127083
+7 (495) 783 07 00
ОКПО 17337364, ОГРН 1027700166636
ИНН 7713076301, КПП 997750001

Общая информация о документе

Настоящий документ содержит информацию, необходимую для эксплуатации экземпляра информационной системы «Система транскрибации разговоров и подсчета характеристик» со стороны пользователя.

В данном документе представлено краткое описание функционала и методово работы с сервисом.

ПАО «ВымпелКом»
ул. 8 Марта, д. 10, стр. 14, г. Москва, 127083
+7 (495) 783 07 00
ОКПО 17337364, ОГРН 1027700166636
ИНН 7713076301, КПП 997750001

1. Общая информация о ПО

1.1. Описание системы

"Система транскрибации разговоров и подсчета характеристик" позволяет автоматически распознавать или переводить речь в текст с учетом разных параметров, таких как частота дискретизации, битность, домен (общий или профессиональный), постобработка (нормализация, пунктуация), программное обеспечение может анализировать аудио на предмет тишины, перебиваний, пола, эмоций и других метрик. Аудио может быть распознано в потоковом, синхронном или асинхронном режиме. Программное обеспечение имеет возможность подключения к различным ИТ-системам.

Основные функциональные характеристики ПО включают:

- Транскрибацию разговоров с использованием нейросетевых моделей;
- Изменение бизнес функциональных настроек системы;
- Подсчет характеристик разговоров (например длительность, перебивания, эмоциональный окрас);
- Хранение сырых метрик от нейронной сети и аналитических данных;
- Взаимодействие через REST API и gRPC для интеграции с другими системами;
- Обеспечение безопасности данных и конфиденциальности с помощью современных методов шифрования и защиты.

1.2. Требования к квалификации для работы с ПО

Системе не имеет ролевой модели и является сервисом для сервисов. Для настройки и начала использования ПО требуется высококвалифицированные специалисты, обладающие следующими навыками:

- владеть навыками работы с технической документацией;
- обладать знаниями о стандартных протоколах передачи данных;
- обладать знаниями о средствах мониторинга ПО;
- обладать знаниями о средствах резервного копирования и восстановления данных в Системе;
- уметь определять источник сбоя функционирования Системы и производить его описание;
- владеть знаниями и навыками администрирования сетевых инструментов;
- осуществлять управление УЗ пользователей, производить назначение прав доступа и блокировку пользователей.

1.5 Перечень документации, с которой необходимо ознакомиться пользователям

Для работы с ИС «Система транскрибации разговоров и подсчета характеристик» пользователю необходимо ознакомиться с:

- руководством администратора;

- руководством пользователя;
- инструкцией по установке экземпляра ПО;
- описанием функциональных характеристик;
- описанием процессов, обеспечивающих поддержание жизненного цикла ПО.

2. Назначение и условия применения

Программное обеспечение "Система транскрибации разговоров и подсчета характеристик" представляет собой гибридное решение, доступное как в формате облачного сервиса, так и в виде решения, устанавливаемого внутри инфраструктуры заказчика. Доступ к сервису осуществляется через gRPC или REST API, что обеспечивает интеграцию с внешними системами и различными приложениями. ПО предназначено для построения систем анализа голосовых данных, которые могут использоваться для контроля персонала в учреждениях, обслуживающих население.

3. Контракт gRPC

```
syntax = "proto3";

package ru.beeline.stt.v2;

import "google/protobuf/empty.proto";
import "prosody.proto";

service Utils {
  rpc getSSUToUpload(google.protobuf.Empty) returns (GetSSUToUploadResponse);                    // Получаем самоподписанный URL для загрузки
}

service Service_v2 {
  rpc AsyncTranscribationSend(RecognizeRequestAsync) returns (RecognizeResponseAsync);           // Асинхронное распознование, отправить файл на распознование
  rpc AsyncTranscribationStatus(BeelD) returns (RecognizeResponseAsync);                         // Асинхронное распознование, получить статус
  rpc AsyncTranscribationCancel(BeelD) returns (RecognizeResponseAsync);                         // Асинхронное распознование, отменить команду send
  rpc AsyncTranscribationDeleteResult(BeelD) returns (RecognizeResponseAsync);                   // Асинхронное распознование, удалить файл с результатом распознавания

  rpc OfflineTranscribation(RecognizeRequest) returns (RecognizeResponse);                       // Офлайн распознование
  rpc StreamOfflineTranscribation(stream RecognizeRequestStream) returns (RecognizeResponse);    // Потоковое распознование офлайн
  rpc StreamOnlineTranscribation(stream RecognizeRequestStream) returns (stream ScorerTuple);    // Потоковое распознование онлайн
}

enum AudioEncoding {
  ENCODING_UNSPECIFIED = 0;
  LINEAR16 = 1;
  reserved "FLAC"; reserved 2;
  MULAW = 3;
  reserved "AMR"; reserved 4;
  reserved "AMR_WB"; reserved 5;
```

ПАО «ВымпелКом»
ул. 8 Марта, д. 10, стр. 14, г. Москва, 127083
+7 (495) 783 07 00
ОКПО 17337364, ОГРН 1027700166636
ИНН 7713076301, КПП 997750001


```
  reserved "OGG_OPUS"; reserved 6;
  reserved "SPEEX_WITH_HEADER_BYTE"; reserved 7;
  ALAW = 8;
  LINEAR32F = 9;
  reserved "OGG_VORBIS"; reserved 10;
  RAW_OPUS = 11;
  MPEG_AUDIO = 12;
}

message GetSSUToUploadResponse {
  string url = 1;                          // URL для загрузки аудиофайла
  BeeID id = 2;                            // Id загружаемого файла в S3 Storage
  sint64 expire = 3;                       // Время действия ссылки для загрузки (в днях)
}

message BeeID {
  string value = 1;
}

message RecognizeResponseAsyncFailed {
  BeeID id = 1;
  string error = 2;
}

message RecognizeResponseAsyncNotFoundAudio {
  BeeID id = 1;
}

message RecognizeResponseAsyncCreated {
  BeeID id = 1;
}

message RecognizeResponseAsyncInQueue {
  BeeID id = 1;
}

message RecognizeResponseAsyncInProgress {
  BeeID id = 1;
  uint64 elapsedTime = 2;
}

message RecognizeResponseAsyncCompleted {
  BeeID id = 1;
  RecognizeResponse transcribe = 2;
}

message RecognizeResponseAsyncCanceled {
  BeeID id = 1;
}

message RecognizeResponseAsyncNotCanceled {
  BeeID id = 1;
  string status = 2;
}

message RecognizeResponseAsyncDeleted {
  BeeID id = 1;
}
```

ПАО «ВымпелКом»
ул. 8 Марта, д. 10, стр. 14, г. Москва, 127083
+7 (495) 783 07 00
ОКПО 17337364, ОГРН 1027700166636
ИНН 7713076301, КПП 997750001


```
message RecognizeResponseAsyncNotDeleted {
  BeeID id = 1;
  string status = 2;
}

/**
 * Ответы метода RecognizeAsync
 */
message RecognizeResponseAsync {
  oneof result {
    RecognizeResponseAsyncNotFoundAudio   not_found = 1;
    RecognizeResponseAsyncFailed          failed = 2;
    RecognizeResponseAsyncCreated         registered = 3;
    RecognizeResponseAsyncInQueue         in_queue = 4;
    RecognizeResponseAsyncInProgress      in_progress = 5;
    RecognizeResponseAsyncCompleted       completed = 6;
    RecognizeResponseAsyncCanceled        canceled = 7;
    RecognizeResponseAsyncNotCanceled     not_canceled = 10;
    RecognizeResponseAsyncDeleted         deleted = 8;
    RecognizeResponseAsyncNotDeleted      not_deleted = 9;
  }
}

/**
 * Параметры для анализа нелексических характеристик аудиофайла
 */
message CallMetrics {
  bool overlap_speech = 1;                  // Процент одновременной речи
  bool interrupting = 2;                    // Детектирование перебивания
  bool speech_speed = 3;                    // Скорость речи
  bool silence_duration = 4;                // Длительность молчания в обоих каналах одновременно
  bool pause_count = 5;                     // Паузы в каждом канале
}

/**
 * Параметры распознования
 */
message RecognitionConfig {
  ASRService service = 1;                   // Сервис для стороннего распознавания аудио
  string model = 2;                         // Модель распознавания
  AudioEncoding encoding = 3;               // Формат аудио
  uint32 sample_rate_hertz = 4;             // Частота дискретизации (Гц)
  string language_code = 5;                 // Язык речи для распознавания
  uint32 max_alternatives = 6;              // Максимальное количество альтернатив для финальных и промежуточных
результатов
  bool enable_automatic_punctuation = 7;    // Простановка знаков препинания (запятые, точки и знаки вопроса) и
прописных букв. true — вернется текст со знаками препинания и прописными буквами, false — вернется текст без знаков
препинания и прописных букв
  bool enable_gender_identification = 8;    // Определить вероятность пола говорящего: мужской или женский.
  bool enable_age_identification = 9;       // Определить вероятность возрастной группы говорящего.
  bool enable_emotions_identification = 10; // Определить вероятность эмоционального окраса речи говорящего.
  bool lexical_sentiment_analytics = 11;    // Определять эмоции по лексике
  bool profanity_filter = 12;               // Включить фильтр ненормативной лексики
  bool text_normalization = 13;             // Включить нормализацию текста
  bool summarization = 14;                  // Включить суммаризацию текста
  uint32 num_channels = 15;                 // Количество каналов в аудио. Для онлайн распознавания
StreamOnlineTranscribation - мы можем принять только 1 канал, в остальных случаях - 2 канала
  bool do_not_perform_vad = 16;             // Отключить разбиение текста на фразы. true — распознанный текст вернется
одной фразой. false — текст в ответе будет разбит на фразы
```

ПАО «ВымпелКом»
ул. 8 Марта, д. 10, стр. 14, г. Москва, 127083
+7 (495) 783 07 00
ОКПО 17337364, ОГРН 1027700166636
ИНН 7713076301, КПП 997750001


```
  optional CallMetrics call_analytics = 17;          // Параметры для анализа нелексических характеристик аудиофайла
}

/**
 * Параметры запроса метода AsyncTranscribationSend
 */
message RecognizeRequestAsync {
  RecognitionConfig config = 1;
  oneof audio {
    bytes content = 2;                        // Аудио в виде массива байт
    string uuid = 3;                          // ID исходного аудиофайла на S3 storage
    S3FileLocation fileLocation = 4;                // bucket и id объекта в S3 совместимом хранилище, где хранится аудиофайл
    string url = 5;                           // preSigned GET URL аудио файла
  }

  message S3FileLocation {
    string bucket = 1;
    string key = 2;
  }
}

message RecognizeRequest {
  RecognitionConfig config = 1;
  bytes content = 2;
}

message RecognizeRequestStream {
  oneof command {
    RecognitionConfig config = 1;
    bytes content = 2;
  }
}

message WordInfo_v2 {
  float start_time = 1;                       // Время начала слова в исходном аудио
  float end_time = 2;                         // Время конца слова в исходном аудио
  string word = 3;                            // Отдельное слово внутри фразы
}

message PhrasesResponse {
  float start_time = 1;                         // Временная метка начала фразы
  float end_time = 2;                           // Временная метка конца фразы
  string phrase = 3;                            // Фраза
  repeated EmotionProbability emotion_identification_proba = 4; // Список классов эмоций для фразы
  repeated WordInfo_v2 words = 5;                     // Список отдельных слов внутри фразы
  uint32 channel = 6;                           // Канал аудио
}

message ChannelMetricsResponse {
  repeated TimeInterval speech_intervals = 1;           // Список отрезков с нагрузкой
  float speech_percent = 2;                     // Процент нагрузки в канале (в %)
  optional float silence_percent = 3;                   // Процент пауз в канале (в %)
  repeated TimeInterval interrupting_intervals = 4;      // Список отрезков перебиваний, в которые данный канал перебивал
собеседника
  optional float speech_speed_wpm_result = 5;                 // Скорость речи, слов в минуту
  optional float speech_speed_spm_result = 6;                 // Скорость речи, символов в минуту
  optional int32 pauses_result = 7;                     // Количество пауз в канале (единиц)
  optional float silence_duration = 8;                  // Длительность пауз (в секундах)
  float gender_identification_female_proba = 9;         // Вероятность, что пол - женский
```

```
  float gender_identification_male_proba = 10;          // Вероятность, что пол - мужской
  repeated AgeProbability age_identification_proba = 11;  // Вероятность возрастной группы
}

/**
 * Классы эмоций, и их вероятность.
 */
message EmotionProbability {
  prosody.EmotionProba.EmotionType type = 1;            // Наиболее вероятный класс эмоции
  float proba = 2;                                      // Вероятность класса
}

/**
 * Классы возрастов, и их вероятность.
 */
message AgeProbability {
  prosody.AgeProba.AgeType type = 1;                    // Наиболее вероятный диапазон возраста
  float proba = 2;                                      // Вероятность класса
}

/**
 * Время начала и завершения одного отрезка аудио данных в канале.
 */
message TimeInterval {
  float begin = 1;                                      // Начало интервала (в секундах)
  float end = 2;                                        // Конец интервала (в секундах)
}

message CallMetricsResponse {
  repeated ChannelMetricsResponse channels = 1;         // Метрики для определенного канала
  // Общие метрики для обоих каналов
  float duration = 2;                                   // Длительность сигнала (в секундах)
  optional float mutual_silence_duration = 3;           // Длительность молчания в обоих каналах одновременно (в секундах)
  optional float overlap_speech_duration = 4;           // Длиность одновременной речи (в секундах)
  repeated TimeInterval overlap_speech_intervals = 5;   // Список отрезков с одновременной речью
  optional float overlap_speech_percent = 6;            // Процент одновременной речи в звонке
}

/**
 * Параметры ответа методов OfflineTranscribation и StreamOfflineTranscribation
 */
message RecognizeResponse {
  repeated PhrasesResponse phrases = 1;
  optional CallMetricsResponse call_metrics_results = 2;
  optional string summary = 3;                          // результат суммаризации
}

message ScorerTuple {
  string text = 1;
  repeated WordTimes word_times = 2;
}

message WordTimes {
  string word = 1;
  int32 start_ms = 2;
  int32 end_ms = 3;
}
```

ПАО «ВымпелКом»
ул. 8 Марта, д. 10, стр. 14, г. Москва, 127083
+7 (495) 783 07 00
ОКПО 17337364, ОГРН 1027700166636
ИНН 7713076301, КПП 997750001


```
enum ASRService {
  INTERNAL = 0;                    // Beeline inner transcribe service}
```

4. Методы API

Каждый метод требует передачи заголовка authorization.

Необходимые content-type:

- для строковых значений (url, id, S3 bucket, S3 path): text/plain
- для передачи файла: application/octet-stream
- для конфига: application/json