|  |  |
| --- | --- |
| Курсовая работа | |
| Титульный Лис | |
|  | Студент: Хан Денис Андреевич |

Содержание

[Введение 4](#_Toc184802856)

[Актуальность работы 4](#_Toc184802857)

[Проблема исследования 4](#_Toc184802858)

[Цель исследования 5](#_Toc184802859)

[Объект исследования 5](#_Toc184802860)

[Методы исследования 5](#_Toc184802861)

[Анализ литературы и открытых источников: 5](#_Toc184802862)

[Сравнительный анализ: 6](#_Toc184802863)

[Метод систематизации и классификации: 6](#_Toc184802864)

[Проектирование базы данных: 6](#_Toc184802865)

[Тестирование: 6](#_Toc184802866)

[Новизна работы 6](#_Toc184802867)

[Обзор существующих технологий распознавания речи 6](#_Toc184802868)

[История развития технологий распознавания речи 6](#_Toc184802869)

[Современные инструменты и платформы 6](#_Toc184802870)

[Сравнение функциональных возможностей инструментов 6](#_Toc184802871)

[Сбор данных для базы данных 7](#_Toc184802872)

[Критерии выбора инструментов 7](#_Toc184802873)

[Источники информации 7](#_Toc184802874)

[Описание параметров для базы данных 7](#_Toc184802875)

[Проектирование базы данных 7](#_Toc184802876)

[Разработка структуры базы данных 7](#_Toc184802877)

[Моделирование данных: схемы и связи 7](#_Toc184802878)

[Технические требования и используемые технологии 7](#_Toc184802879)

[Реализация базы данных 7](#_Toc184802880)

[Инструменты для создания базы данных 7](#_Toc184802881)

[Создание таблиц и реализация связей 7](#_Toc184802882)

[Реализация функций поиска и фильтрации данных 7](#_Toc184802883)

[Наполнение и тестирование базы данных 7](#_Toc184802884)

[Наполнение базы данных реальными данными 7](#_Toc184802885)

[Проверка корректности работы базы данных 7](#_Toc184802886)

[Тестирование на примерах использования 7](#_Toc184802887)

[Анализ и обсуждение результатов 8](#_Toc184802888)

[Анализ собранных данных 8](#_Toc184802889)

[Выводы по использованию базы данных 8](#_Toc184802890)

[Возможные улучшения и доработки 8](#_Toc184802891)

[Заключение 8](#_Toc184802892)

[Выводы по проделанной работе 8](#_Toc184802893)

[Дальнейшие перспективы и развитие темы 8](#_Toc184802894)

[Список литературы 8](#_Toc184802895)

# ****Введение****

## ****Актуальность работы****

В современном мире технологии распознавания речи играют ключевую роль в улучшении взаимодействия человека с машинами и системами. С развитием искусственного интеллекта и машинного обучения, решения для распознавания речи применяются в различных областях: от голосовых помощников и автоматизированных систем управления до медицинских и образовательных приложений. Такие платформы, как Google Assistant, Siri, и Alexa, становятся неотъемлемой частью повседневной жизни, а также инструментами для повышения производительности в бизнесе.

Тем не менее, разнообразие инструментов распознавания речи и различие их характеристик (точность, поддерживаемые языки, возможность интеграции в системы, стоимость) создают сложность для их выбора и применения в разных проектах. В связи с этим возникает потребность в создании базы данных, которая позволит систематизировать информацию о различных решениях для распознавания речи. Такая база данных будет полезна разработчикам, исследователям и компаниям, которые хотят внедрить распознавание речи в свои продукты, но сталкиваются с трудностью выбора наиболее подходящего инструмента.

## ****Проблема исследования****

Технологии распознавания речи стремительно развиваются, и на рынке появляется множество инструментов и платформ, предназначенных для преобразования речи в текст и выполнения голосовых команд. Однако существует проблема недостаточной систематизации информации об этих инструментах. Из-за широкого разнообразия решений, различий в их характеристиках (например, поддерживаемые языки, точность, производительность, способы интеграции) пользователи, разработчики и исследователи сталкиваются с трудностями при выборе подходящего инструмента для своих нужд.

Отсутствие единого ресурса, где можно было бы быстро и удобно получить сравнение характеристик различных систем, приводит к увеличению временных и трудовых затрат при анализе и выборе технологий распознавания речи. Это особенно критично для компаний, стремящихся внедрить данную технологию в свои продукты, и для исследователей, работающих в области улучшения голосовых интерфейсов.

Таким образом, проблема исследования заключается в необходимости создания удобной и структурированной базы данных, которая объединит информацию о существующих инструментах распознавания речи и предоставит пользователям возможность легко сравнивать их по ключевым параметрам. Решение этой проблемы позволит ускорить процесс выбора подходящих технологий и облегчит их внедрение в различные сферы деятельности.

## ****Цель исследования****

Целью данного исследования является создание базы данных, которая систематизирует информацию об инструментах распознавания речи, чтобы упростить их анализ и выбор для различных сфер применения. Эта база данных должна предоставлять пользователям возможность сравнивать существующие решения по ключевым параметрам, таким как точность распознавания, поддерживаемые языки, производительность, способы интеграции и другие важные характеристики.

Создание такой базы данных будет способствовать оптимизации процесса выбора инструментов распознавания речи и их более широкому и эффективному внедрению в продукты и сервисы, что, в свою очередь, повысит качество взаимодействия пользователей с технологиями и сократит временные затраты на анализ доступных решений.

## Объект исследования

Объектом исследования являются существующие инструменты и технологии распознавания речи, представленные на рынке. Это программные и аппаратные решения, которые обеспечивают преобразование устной речи в текст или выполнение голосовых команд. В рамках исследования рассматриваются такие параметры, как точность распознавания, поддерживаемые языки, способы интеграции, производительность, стоимость и области применения этих инструментов.

**Предмет исследования**

Предметом исследования является процесс систематизации и структурирования информации о характеристиках и функциональных возможностях инструментов распознавания речи в рамках создания базы данных. В частности, исследование сосредоточено на методах сбора, хранения, анализа и организации данных о таких параметрах, как точность распознавания, поддерживаемые языки, интеграционные возможности, производительность и другие ключевые характеристики данных инструментов.

## ****Методы исследования****

В ходе исследования применяются следующие методы:

### **Анализ литературы и открытых источников**:

Изучение научных статей, технической документации, руководств и обзоров для получения информации о существующих инструментах распознавания речи, их характеристиках и применении.

### **Сравнительный анализ**:

Сравнение инструментов по ключевым параметрам (точность, производительность, поддерживаемые языки, способы интеграции), что позволит выявить сильные и слабые стороны различных решений.

### **Метод систематизации и классификации**:

Систематизация собранных данных для создания удобной и логически структурированной базы данных, позволяющей пользователям легко находить и сравнивать инструменты.

### **Проектирование базы данных**:

Метод проектирования реляционных баз данных для создания схемы и структуры хранения информации, а также для эффективной организации данных об инструментах распознавания речи.

### **Тестирование**:

Тестирование базы данных с реальными данными для проверки её работоспособности, корректности хранения и извлечения информации, а также удобства использования.

# ****Новизна работы****

Новизна данной работы заключается в создании уникальной базы данных, которая систематизирует и упорядочивает информацию о существующих инструментах распознавания речи, объединяя их ключевые характеристики и параметры в одном ресурсе. В отличие от существующих обзоров и статей, которые обычно охватывают отдельные аспекты технологий, данная база данных предлагает комплексный подход, позволяющий пользователям не только анализировать и сравнивать инструменты по таким параметрам, как точность, производительность, поддерживаемые языки и интеграционные возможности, но и легко находить решения, наиболее подходящие для конкретных задач.

База данных представляет собой удобный инструмент для разработчиков, исследователей и компаний, что ускоряет процесс выбора технологии для внедрения и повышает эффективность использования систем распознавания речи в различных областях.

# Обзор существующих технологий распознавания речи

## История развития технологий распознавания речи

**Начало исследований и разработки первых устройств**:

* В 40-х годах XX века начались исследования по анализу речевых сигналов с использованием спектральных анализаторов. В СССР было создано устройство, которое распознавало гласные русского языка на основе анализа энергии в различных частотных полосах.
* **Первое устройство для распознавания речи**:
* В 1952 году было создано первое устройство для распознавания речевых команд. Оно распознавало цифры, произнесённые голосом, с помощью анализа акустических характеристик. Эта технология была примитивной, но дала старт дальнейшему развитию.
* **Достижения в 60-е годы**:
* В 1962 году IBM представила устройство, распознающее 16 произносимых слов, включая цифры, под названием "Shoebox". Это устройство стало одной из первых систем, демонстрирующих возможности автоматического распознавания речи.
* **Развитие в 70-е годы**:
* В 1971 году Агентство DARPA начало финансировать проекты по созданию высокоэффективных систем распознавания речи. Система HARPY, разработанная в университете CMU, стала первой успешной моделью с использованием словаря из 1000 слов и точностью распознавания 90%.
* **Использование скрытых марковских моделей**:
* В конце 70-х – начале 80-х годов была разработана технология скрытых марковских моделей, которые стали основой для построения современных систем акустического анализа речи. Эти модели позволили значительно улучшить точность распознавания.
* **Внедрение в массовое использование**:
* В 1990 году Dragon Systems представила первую коммерческую программу для распознавания речи Dragon Dictate. А в 1993 году IBM выпустила систему Personal Dictation для персональных компьютеров.
* **Современные системы**:
* В начале 2000-х годов Microsoft интегрировала технологии распознавания речи в офисные приложения, а компании IBM и Philips развивали свои решения. Разработка мобильных приложений, таких как Microsoft Voice Command, Speereo Voice Translator и других, расширила возможности использования технологий распознавания речи.

### **Современные инструменты и платформы**

В последние годы технологии распознавания речи значительно эволюционировали благодаря стремительному развитию искусственного интеллекта и нейронных сетей. Современные платформы позволяют не только выполнять преобразование устной речи в текст с высокой точностью, но и реализовывать широкий спектр дополнительных функций, таких как идентификация говорящего, анализ интонации и эмоциональной окраски, а также синхронный перевод. В данной секции рассматриваются основные современные решения и инструменты, применяемые в области распознавания речи, а также их функциональные возможности и перспективы развития.

**Основные коммерческие платформы:**

* **Google Cloud Speech-to-Text.** Эта платформа предоставляет высокоточные модели распознавания речи, поддерживающие множество языков. Особенностью сервиса является возможность работы как с потоковыми, так и с записанными аудиоданными, а также наличие инструментов для адаптации модели под специфические задачи заказчика.
* **IBM Watson Speech to Text.** IBM предлагает комплексное решение, ориентированное на корпоративный сегмент. Платформа обладает широкими возможностями по настройке, позволяет реализовывать транскрипцию в реальном времени и интегрироваться с другими сервисами IBM, что делает её привлекательной для крупных бизнес-процессов.
* **Microsoft Azure Speech Services.** Решение от Microsoft отличается высокой масштабируемостью и интеграционными возможностями в экосистему Azure. Помимо стандартной транскрипции, сервис поддерживает функционал синтеза речи, что открывает дополнительные возможности для создания интерактивных приложений.
* **Amazon Transcribe.** Этот сервис от Amazon Web Services ориентирован на автоматизацию обработки аудиоданных, включая транскрипцию звонков и видеоархивов. Особенностью платформы является возможность автоматической пунктуации и разделения речи на спикеров, что упрощает дальнейший анализ данных.

**Открытые и специализированные решения:**

Помимо коммерческих предложений, на рынке активно развиваются и открытые проекты, такие как:

* **Kaldi.** Открытая платформа, используемая в академических исследованиях и разработках. Kaldi предоставляет инструментарий для создания и обучения собственных моделей распознавания речи, что позволяет адаптировать систему под специфические языковые или предметные области.
* **Mozilla DeepSpeech.** Основанная на технологиях глубокого обучения, данная система ориентирована на создание доступных и модифицируемых решений. DeepSpeech активно используется разработчиками для создания кастомизированных моделей, особенно в проектах с ограниченным бюджетом.

Сравнительный анализ, плюсы, минусы, и примеры

| Инструмент | Плюсы | Минусы | Пример использования |
| --- | --- | --- | --- |
| **Google Cloud Speech-to-Text** | • Высокая точность (≈95–97 % на общедоступных наборах)¹ • Поддерживает более 125 языков и диалектов² • Стриминговое API, автоматическая пунктуация, адаптация модели³ | • Платное решение с почасовой тарификацией (от $0.006/мин)⁴ • Требует подключения к облаку Google | Распознавание звонков контакт-центра для автоматического индексирования разговоров и аналитики тональности. |
| **IBM Watson Speech to Text** | • Гибкая настройка под доменные словари⁵ • Корпоративная интеграция с другими сервисами IBM⁶ • Высокая отказоустойчивость | • Более высокая стоимость по сравнению с конкурентами⁷ • Сложнее начать для малых проектов | Транскрипция медицинских консультаций с добавлением пользовательских терминов и аббревиатур. |
| **Microsoft Azure Speech Services** | • Унифицированная платформа «Speech» (распознавание, синтез, перевод)⁸ • Низкая задержка (<300 мс на среднем звуке)⁹ • Глубокая интеграция с экосистемой Azure | • Требует знания Azure SDK и управления ресурсами • Стоимость зависит от уровня SLA и региона | Голосовое управление IoT-устройствами в «умном доме» с использованием локального контейнера (Speech Containers). |
| **Amazon Transcribe** | • Автоматическая диаризация (отделение спикеров)¹⁰ • Поддержка потоковой и пакетной обработки • Авто-пунктуация | • Ограниченные возможности настройки языковых моделей • Тарифы AWS могут быть непрозрачны | Анализ и автоматическое субтитрирование видеолекций для образовательной платформы. |
| **Kaldi (open-source)** | • Полный контроль над архитектурой моделей¹¹ • Поддержка HMM-DNN гибридов • Широкая кастомизация (создание собственных акустических/языковых моделей) | • Сложная кривая обучения, требует глубоких знаний DSP и ML • Нет встроенного облачного API | Исследовательский прототип речевой системы для редкого языка, обученный на собственном корпусе аудио. |
| **Mozilla DeepSpeech (open-source)** | • Основан на end-to-end нейронных сетях (CTC)¹² • Прост в развёртывании и дообучении для новых языков • Активное сообщество | • Ниже точность «из коробки» по сравнению с коммерческими сервисами (≈85–90 %)¹³ • Ограниченная поддержка спикеров | Голосовой интерфейс для оффлайн-приложения на мобильных устройствах с возможностью дообучения на пользовательских данных. |

## Сравнение функциональных возможностей инструментов

| Функция | Google Cloud STT | IBM Watson STT | Azure Speech | Amazon Transcribe | Kaldi | DeepSpeech |
| --- | --- | --- | --- | --- | --- | --- |
| Поддержка языков (>100) | ✔︎ | ≈10–20 | ≈85 | ≈50 | любая¹¹ | любая¹² |
| Стриминговое распознавание | ✔︎ | ✔︎ | ✔︎ | ✔︎ | через gRPC | нет (CLI) |
| Автоматическая пунктуация | ✔︎ | ✕ | ✔︎ | ✔︎ | нет | нет |
| Диаризация (speaker-ids) | ✕ | ✕ | плагин¹⁴ | ✔︎ | да | да¹² |
| Кастомизация словаря | ✔︎ | ✔︎ | ✔︎ | частично | полная | полная |
| Поддержка адаптации под шум | ✔︎ | ✔︎ | ✔︎ | базовая | manual | manual |
| Реaltime-латентность (мс) | <300 | <500 | <300 | <400 | >500 | >500 |
| Оффлайн-режим | SDK (Edge) | Edge SDK | Containers | AWS GP3 | native | native |
| Стоимость (прим. $/час) | 0.36 | 0.40 | 0.40 | 0.24 | свободно | свободно |

# ****Сбор данных для базы данных****

## Критерии выбора инструментов

**Точность распознавания** (Word Error Rate, WER)

* **Поддерживаемые языки и диалекты**
* **Реaltime-латентность**
* **Возможности кастомизации** (доменные словари, шумоустойчивость)
* **Интеграционные опции** (REST/gRPC, контейнеры, SDK)
* **Цена и модель тарификации** (pay-as-you-go, подписка, бесплатные квоты)
* **Лицензирование** (проприетарное vs. open-source)
* **Уровень поддержки и сообщества**
* **Безопасность и конфиденциальность** (sensitive data handling)
* **Дополнительные функции** (diarization, punctuation, emotional tone)

## Описание параметров для базы данных

| Поле | Тип | Описание |
| --- | --- | --- |
| tool\_id | INT (PK) | Уникальный идентификатор инструмента. |
| name | VARCHAR | Название (например, «Google Cloud Speech-to-Text»). |
| vendor | VARCHAR | Поставщик (Google, IBM, Microsoft, AWS, Open-source). |
| licensing | VARCHAR | Лицензия (GPL, Proprietary, freemium и т. д.). |
| languages\_supported | TEXT | Перечисление поддерживаемых языков/диалектов. |
| wer | FLOAT | Средний Word Error Rate (%). |
| latency\_ms | INT | Средняя задержка распознавания в мс. |
| streaming\_support | BOOLEAN | Поддержка стримингового режима (TRUE/FALSE). |
| punctuation | BOOLEAN | Автоматическая пунктуация (TRUE/FALSE). |
| diarization | BOOLEAN | Разделение по спикерам (TRUE/FALSE). |
| customization | BOOLEAN | Возможность добавления доменных словарей и обучения модели (TRUE/FALSE). |
| deployment\_options | TEXT | Опции развёртывания (облако, контейнеры Docker/Kubernetes, локально). |
| pricing\_model | TEXT | Описание тарифов (pay-as-you-go, подписка, free tier и т. д.). |
| last\_updated | DATE | Дата последнего обновления API или версии SDK. |
| reference\_url | VARCHAR | Ссылка на официальную документацию или релевантную публикацию. |

# ****Проектирование базы данных****

## Разработка структуры базы данных

------

## Моделирование данных: схемы и связи

------

## Технические требования и используемые технологии

------

# **Реализация базы данных**

## Инструменты для создания базы данных

-----

## Создание таблиц и реализация связей

------

## Реализация функций поиска и фильтрации данных

------

# ****Наполнение и тестирование базы данных****

## Наполнение базы данных реальными данными

-------

## Проверка корректности работы базы данных

-------

## Тестирование на примерах использования

--------

# ****Анализ и обсуждение результатов****

## Анализ собранных данных

-----

## Выводы по использованию базы данных

------

## Возможные улучшения и доработки

------

# ****Заключение****

## Выводы по проделанной работе

------

## Дальнейшие перспективы и развитие темы

------

# ****Список литературы****