РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ ПЕДАГОГИЧЕСКИЙ УНИВЕРСИТЕТ им. А. И. ГЕРЦЕНА

ФАКУЛЬТЕТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

КАФЕДРА ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ

**ДИПЛОМНАЯ РАБОТА**

Разработка голосового ассистента для людей с ограниченными возможностями по здоровью

**Студента 4 курса**

**дневного отделения,**

**Галкина Антона Владимировича**

**Руководитель: к.п.н., доцент**

**Авксеньтьева Елена Юрьевна**

**к.ф-м.н., доцент**

**Жуков Николай Николаевич**

Санкт-Петербург

2020 г.

**Оглавление**

[**Введение** 3](#_Toc41150786)

[**Глава 1. Анализ научных публикаций существующих голосовых ассистентов** 6](#_Toc41150787)

[**1.1 Голосовые ассистенты** 6](#_Toc41150788)

[**1.2 Возможности** 8](#_Toc41150789)

[**1.3 Принцип работы** 9](#_Toc41150790)

[**1.4 Недостатки** 13](#_Toc41150791)

[**1.5 Будущее виртуальных помощников** 14](#_Toc41150792)

[**1.6 Выбор языка программирования** 14](#_Toc41150793)

[**1.7 Описание требований и решаемых подготовкой задач** 16](#_Toc41150794)

[**1.7.1 Функционал и технологии голосового ассистента** 16](#_Toc41150795)

[**1.7.2 Выбор библиотек для разработки голосового ассистента** 18](#_Toc41150796)

[**Выводы по главе 1** 21](#_Toc41150797)

[**Глава 2. Проектирование и разработка голосового ассистента** 22](#_Toc41150798)

[**2.1 Функция справки** 22](#_Toc41150799)

[**2.2 Функция прогноз погоды** 22](#_Toc41150800)

[**2.3 Функция произношения текущего времени** 23](#_Toc41150801)

[**2.4 Функция открывания браузера** 24](#_Toc41150802)

[**2.5 Функция приветствие и завершение** 25](#_Toc41150803)

[**2.6 Функция запуска приложения** 26](#_Toc41150804)

[**2.7 Функция перевода** 27](#_Toc41150805)

[**2.8 Функция чата** 27](#_Toc41150806)

[**2.9 Функция поиска в Википедии** 28](#_Toc41150807)

[**2.10 Функция отправки email указанному пользователю** 28](#_Toc41150808)

[**2.11 Функция просмотра новостей** 29](#_Toc41150809)

[**2.12 Функция запуска песни** 30](#_Toc41150810)

[**Выводы по главе 2** 31](#_Toc41150811)

[**Заключение** 32](#_Toc41150812)

[**Список литературы** 33](#_Toc41150813)

# **Введение**

При изучении истории виртуальных помощников, последние несколько лет показали радикальные изменения в развитии и успехе в этой области. В настоящее время большинство крупных технологических компаний имеют собственного виртуального помощника или, по крайней мере, начали его разрабатывать. Научно-фантастическая фантазия о том, чтобы иметь умный компьютер, с которым можно разговаривать, находится уже не в далеком будущем, а в настоящем. Современные голосовые ассистенты пытаются анализировать и учиться на поведении пользователя, чтобы со временем стать более эффективными и удобными для пользователя.

Есть несколько популярных умных помощников, выстроившихся в очередь, чтобы помочь нам, но как можно сказать, какой из них наиболее подходит для них? Поиск идеального помощника ИИ начинается с принятия решения о том, для чего пользователь хочет использовать этот помощник. После получения четкого представления о необходимых навыках начинается исследование. Таким образом, что можно сделать, если они не могут найти продукт, который они ищут? Ответ заключается в том, чтобы построить его.

Личные помощники с голосовым управлением на дому обладают большим потенциалом для оказания помощи людям с ограниченными возможностями по здоровью, а также пожилым людям. Этот тезис исследует аспекты проектирования человеко-компьютерного интерфейса, в частности голосового помощника, чтобы помочь людям управлять своим личным здоровьем и упрощать жизнь. В моей текущем исследовательском проекте фокусируемся на разработке потребительского голосового ассистента для людей с ограниченными возможностями по здоровью и их назначенных доверенных лиц.

Для достижения нашей цели создания простого в использовании и понимании интерфейса данных для людей с ограниченными возможностями по здоровью, исследуются голосовые, домашние персональные помощники в качестве решения. Для внедрения и тестирования были выбран язык программирования Python, потому что для Python существует большое количество библиотек, которые позволяют решить некоторые задачи, возникающие при разработке ассистента.

Актуальность - в настоящее время процесс развития голосовых ассистентов достиг этапа, когда машины понимают, что мы говорим и могут поговорить с нами. Голосовой ассистент, которого хотел бы купить каждый, обязан быть неотличим от голосового помощника-человека. Полезность существующих голосовых помощников определяется в первую очередь качеством распознавания речи и интегрированностью в информационные сети, сервисы и системы «умного дома» и в меньшей степени ориентированностью на специфику конкретного пользователя (или определенной группы пользователей), для которой голосовой помощник создается. Подтверждением этого довода также можно считать недостаточное распространение систем умного дома, основанных на работе голосового помощника для людей с ограниченными возможностями по здоровью. Разумеется, они могут пользоваться голосовым поиском, но на этом функционал использования ассистента заканчивается. Для полноценного использования ассистента таким группам людей необходимы специальные функции управления и взаимодействия, например, функции отслеживания состояния здоровья и контроля за домом. [1]

Таким образом, предметом исследования является – структура и принципы функционирования голосового ассистента, а объект исследования – использование информационных технологий людьми с ограниченными возможностями по здоровью

Для достижения поставленной цели необходимо решить следующие задачи:

1. Проанализировать существующие голосовые ассистенты и их функционал
2. Выявить недостатки существующих голосовых ассистентов
3. Выявить требования для реализации голосового ассистента для людей с ограниченными возможностями
4. Выявить наиболее удобный язык для разработки голосового ассистента
5. Спроектировать модель голосового ассистента.
6. Реализовать разработанную модель с использованием оптимальных технических средств.

Результатом бакалаврской работы является готовый к использованию голосовой ассистент для людей с ограниченными возможностями по здоровью.

Структура бакалаврской работы. Бакалаврская работа состоит из введения, двух глав, заключения, библиографии и приложений. Она содержит «36» страниц и «12» рисунков.

# **Глава 1. Анализ научных публикаций существующих голосовых ассистентов**

## **1.1 Голосовые ассистенты**

Виртуальные голосовые ассистенты — это программные приложения для взаимодействия с людьми посредством голоса и выполнения их команд. Крупные технологические компании имеют своих собственных интеллектуальных виртуальных помощников для различных целей и платформ. Поначалу основной целью такого помощника было ответить на вопросы пользователя, но теперь наступил момент, когда пришло время еще больше расширить его возможности.

В России статистика использования голосовых ассистентов не так хорошо изучена по сравнению с другими странами, где исследование рынка проводятся регулярно [6]. Только Яндекс проводит исследования и публикует результаты. Так, например по предложенным данным в среднем голосовым ассистентом пользуются примерно миллион пользователей, а за месяц число запросов превышает один миллиард.

Одно из исследований компания Just AI выявило рост продаж голосовых ассистентов в 8 раз, основываясь на этом можно сделать вывод о их активном использовании пользователями, а также о быстром развитие рынка и возможностей голосовых ассистентов. Поэтому разработка программного обеспечения очень перспективна [6].

По статистике компании Яндекс приблизительно 28 % пользователей Алисы используют ее в поездках [6]. Важно отметить, что ассистент Алиса — это главный инструмент по коммуникациям с сервисами компании Яндекс. При этом даже как навигатор она не только прокладывает путь и направляет по нему, но и параллельно способна отвечать на вопросы или выполнять другие действия. [3, 6]

С момента появления ассистента Алисы в России она заняла лидирующее место среди конкурентов. По данным компании более 70% пользуются ею каждый день, но важно отметить, что часть опрошенные пользуются не одним голосовым ассистентом.

Что касается конкурентов Алисы, то голосовым ассистентом компании Google пользуются примерно 7 миллионов пользователей, а Siri от Apple 8 миллионов.

Приведу еще один пример, в ноябре 2019 года компания Just AI организовало исследование среди 2000 человек из крупных городов России, которые активно пользовались мобильными приложениями. Анализ проводил голосовой робот вовремя телефонного звонка. Результаты показали, что 71% аудитории хотя бы один раз пользовались голосовыми ассистентами и 78% продолжают использовать и общаться с ними регулярно [3, 6, 8].

Специалисты полагают, что в 2020 году конкуренция платформ и брендов приведет к росту узнаваемости голосовых ассистентов, а также что произойдет повышение информированности Россиян о том, для чего необходимы ассистенты и как ими пользоваться [8].

## **1.2 Возможности**

Голосовые технологии существуют уже много лет - от телефонных систем с автоматическим распознаванием голоса, которые не понимали акценты, до диктофонов с преобразованием голоса в текст, которые производили неточное копирование, но сбои в этих системах помешали широкому распространению.

Тем не менее, достижения в области искусственного интеллекта означают, что возможности голосового помощника наконец становятся реальностью, на сегодня ассистенты – это:

* Поиск в Интернете;
* Поиск новостей;
* Прослушивание музыки;
* Использование навигатора;
* Запуск приложений;
* Отправка сообщений;
* Виртуальный собеседник

Видно, что функционал голосовых ассистентов строится на решении повседневных задач, например, получение данных о погоде, режиме работы того или иного места и.т.д [7]. Важно отметить, что голосовые ассистенты зачастую имеют свои персональные особенности, а набор их функций увеличивается с выходом новых обновлений.

## **1.3 Принцип работы**

Основная цель любого голосового ассистента сводиться к распознаванию речи пользователя. Распознавания речи — это способность восстановить по сигналу то, что человек говорил.

Этот процесс также является самым затруднительным этапом при разработке голосового ассистента, поскольку одни и те же фразы, произнесённые людьми в разных условиях, будут давать различные результаты. Для решения этой проблемы голосовой ассистент разделяет слова на звуковые фрагменты, называемые фонемами. [3, 7] Фонема — это минимальная единица звуковой структуры языка. Фонема не имеет самостоятельного лексического или грамматического значения, но она служит для различения и идентификации значимых единиц языка [3, 7]. На рисунке 1 представлен пример дробления слов на фонемы.

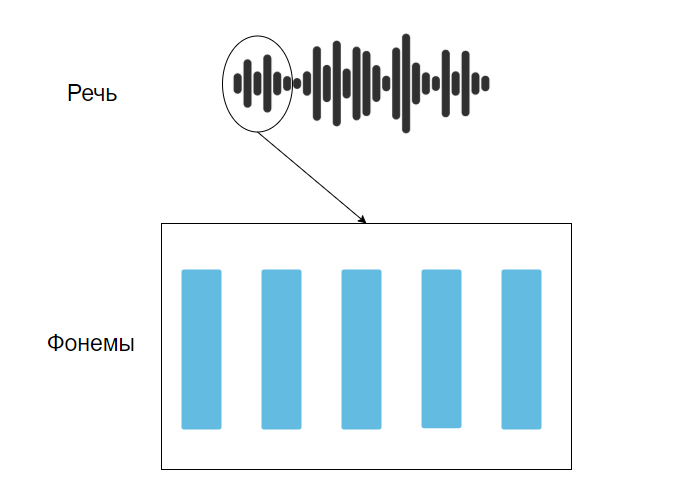


Рис. 1. Пример дробления слов на фонемы

Далее голосовой ассистент работает с числовым представлением фонем, полученных при произношении запроса пользователем. Он отправляет полученные данные на сервера компании разработчика, где запись делится на множественные фреймы из них, извлекаются наиболее важные признаки, то есть происходит удаление личностных характеристик голоса, к примеру тембр или дефекты речи. Фрейм — это фрагмент речи длиной в сотые доли секунды.

Затем фреймы проходят математическускую обработку после чего получается частотная характеристика с помощью, которой ассистент может предположить к какой фонеме принадлежит фрейм. Важно отметить, что ассистент именно предполагает, что было сказано, потому что не всегда получается определить звуки и необходимо достраивать слова.

Для этого используется нейронная сеть, она, основываясь на полученных ранее данных, формирует вероятность появления фонемы, а также пытается предугадать ее появление в фрейме. Вся информация проходит через декодер, учитывающий подключенную к нему базу данных также известную как таблица вероятности, содержащую вероятность появления того или иного символа и используемую языковую модель. Что позволяет достроить слова. На рисунке 2 представлена полная схема распознавания речи

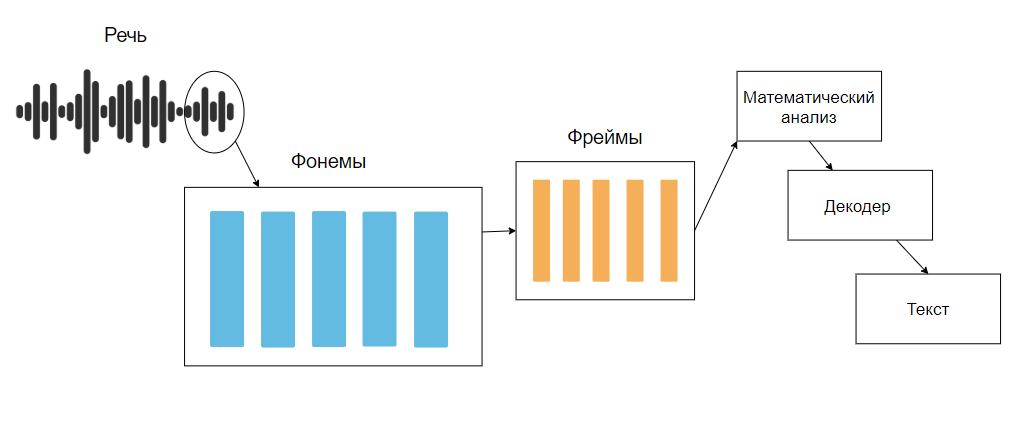


Рис. 2. Полная схема распознавания речи

Завершив распознание голоса, необходимо понять, что пользователь намерен получить от ассистента, например одним из вариантов на данном этапе это отсортировать запрос по намерению (классификатора намерений). Второй, действовать по заранее определённому шаблону.

Определить, что человек имеет в виду своей фразой является основной задачей классификатора намерений. [7, 9] То есть пользователь запросил курс валют, то ассистент должен понять, что речь идёт о поиске в интернете курса валют. Принцип работы классификатора намерений строится на деление запроса на слова и пунктуационные знаки после чего применяются эмбеддинги - сопоставление элемента речи числовому вектору, позволяющие понять смысл слов в запросе. [4, 7, 9] На рисунке 3 показан пример работы классификатора намерений.

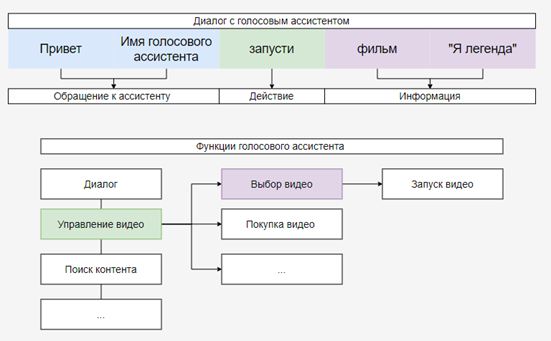


Рис. 3. Пример работы классификатора намерений

Однако голосовой ассистент может дать ответ на вопрос только при наличии достаточного количества информации в запросе, что является редкостью поскольку, пользователи не всегда проговаривают всю нужную для ответа информацию [1, 2, 7]. Приведу пример на голосовом ассистенте Алиса, если запрос не прошел по одному из шаблонов, то он попадает в раздел «собеседник», где искусственный интеллект либо выдаст ответ из наиболее популярных ответов или сообщит о непонимание запроса. Это позволяет сделать голосового помощника более интересным.

Последним важным аспектом голосового ассистента является – личность. Она развивается посотенно по мери использования за счет нейросети. Которые обучаются на различных материалах из интернета, что может привести к редким, но грубым ответам. Это нужно что бы ассистент знал о существование ненормативной лексики и мог не отвечать бездумно. На рисунке 4 приведина полная упрощенная схема работы голосового ассистента.



Рис. 4. Полная схема работы голосового ассистента

## **1.4 Недостатки**

Однако, как было сказано выше, не всегда голосовой помощник может распознавать звуки так как важны еще и обстоятельства запроса, например:

* фоновые шумы,
* четкость сказанного
* скорость речи

Все перечисленное влияет на точность распознавания и эти факторы зачастую сложно учесть при разработке. Поэтому приходиться достраивать слова основываясь на похожих запросах других пользователей.

Еще одна проблема — это хранение информации о пользователи и контроль того, кто имеет доступ к голосовому ассистенту и его информации. Все это приводит к проблемам конфиденциальности: кто имеет доступ к персональным данным пользователя и к помощнику в целом?

Приведу несколько примеров. В 2016 году голосовой ассистент Google Home уже мог поддерживать несколько пользователей, однако все они были привязаны к одной учетной записи – это означало, что помощник собирал данные о интересах не одного, а нескольких пользователей и мог выдать личную информацию. Позднее Google нашло решение и исправила уязвимость. [26]

При выпуске ассистента Alexa от компании Amazon были случаи срабатывания на звук телевизора и поскольку некоторые из них не имели пароля для совершения покупок была вероятность случайной покупки товара [25]. Поэтому нежелательные голосовые команды являются самым простым способом взломать ассистента, были также случаи активации голосового помощника домашними питомцами [13].

Однако есть еще одна проблема с ассистентами, которые подключены к интернету или другим устройствам. Они могут записывать разговоры и отправлять их в облако без ведома пользователя, а так же если домашняя сеть плохо защищена ее могут взломать хакеры, что так же может вызвать пропажу личных данных [10].

## **1.5 Будущее виртуальных помощников**

Далеко не все люди пользуются голосовыми ассистентами для некоторых требуется время, чтобы привыкнуть к программному обеспечению с человеческим голосом, чтобы знать о них все, в то время как другие уже активно используют своего виртуального помощника. Так или иначе, рынок голосовых растет быстрыми темпами, как в используемых технологиях, так и в пользовательской базе.

## **1.6 Выбор языка программирования**

Какой язык программирования выбрать? Это главный вопрос при разработке голосового ассистента. От его выбора зависит сложность разработки, поэтому рассмотрим наиболее популярные и подходящие варианты.

Java – отличный выбор при разработке сложных проектов, язык обладает большой базой инструментов, которые отлично подходят для обработки естественного языка.

Достоинства:

* Многоплатформенность - то есть может работать на разных операционных системах
* Удобен для совместной работы

Недостатки:

* Сложный код – необходимо писать длинный код, который трудно читать
* Низкая производительность

Prolog – язык, состоящий из логических выражений, популярен при разработке приложений с логическими способностями, благодаря встроенным алгоритмам.

Достоинства:

* Легко читаемый код
* Удобен для совместной работы

Недостатки:

* Сложный для реализации

Python - достаточно простой и понятный язык, а при всей своей простоте Python позволяет реализовать как простой искусственный интеллект, так и сложный.

Достоинства:

* Легко писать и читать код
* Большой набор библиотек для решения различных задач

Недостатки

* Не удобен для реализации мобильных приложений

Подводя итог, наиболее интересным из представленных языков программирования является Python. Так как он наиболее популярен и прост в использование, а использование языков Java и Prolog не целесообразно.

## **1.7 Описание требований и решаемых подготовкой задач**

Требования определяют функциональность программного обеспечения, то есть описывают, что необходимо реализовать и какие возможности должен предоставлять разработанный голосовой ассистент. Перед началом разработки нужно определиться с несколькими вещами:

* какие функции должен уметь выполнять голосовой ассистент
* какие технологии нужно реализовать для работы голосового ассистента
* какие библиотеки необходимо использовать для реализации вышеперечисленного

### **1.7.1 Функционал и технологии голосового ассистента**

Функционал любого голосового ассистента обычно основывается на повседневных запросах пользователей, но для людей с ограниченными возможностями по здоровью нужны еще и дополнительные функции в зависимости от их заболеваний [16 - 24]. При разработке учесть абсолютно все потребности этой группы людей очень сложно, но эту проблему можно решить последовательным добавлением функций голосового ассистента и со временем реализовать их все. Однако ключевой функцией для большинства будет именно возможность озвучивания информации с экрана компьютера [11 - 15].

На момент выхода первой версии ассистента для людей с ограниченными возможностями по здоровью необходимо реализованы следующие функции:

* прогноз погоды
* произношения текущего времени
* открывания браузера и поиска в нем
* начала и завершения работы ассистента
* запуска приложений
* перевода
* чата
* справки о голосовом ассистенте
* поиска в Википедии
* отправки email
* запуска песни
* просмотра новостей

Представленный функционал будут полезны как для людей-инвалидов, так и для их попечителей.

Но по мимо самих функций необходимо еще разработать основные технологии, необходимые для работы голосового ассистента. На рисунке 5 изображены все основные технологии, которые внедрены в голосовые ассистенты и необходимы для его работы.

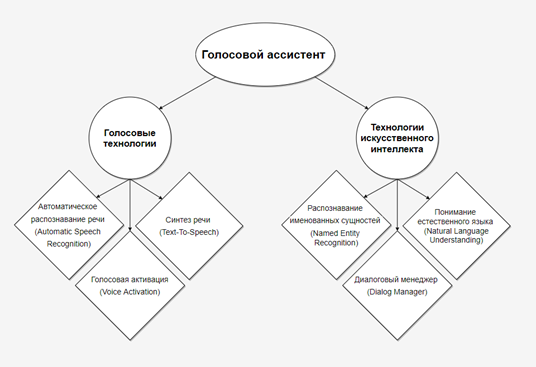


Рис. 5. Технологии, которые необходимы для с работы голосового ассистента

### **1.7.2 Выбор библиотек для разработки голосового ассистента**

Основа голосового ассистента — это движок. Их существует множество вариантов, как платных, так и бесплатных, но для доступности и кроссплатформенности выберем три.

* eSpeak - платный вариант, с поддержкой большинства языков и диалектов
* Microsoft Speech API - бесплатный, встроенный речевой интерфейс
* pyttsx3 - бесплатный, небольшая библиотека

eSpeak — это компактный программный синтезатор речи для английского и других языков, распространяется на Linux и Windows. Библиотека использует метод «формантного синтеза». Это позволяет предоставлять много языков в небольшом размере, речь ясна и может использоваться на высоких скоростях, но она не такая естественная или плавная, как у более крупных синтезаторов, основанных на записях человеческой речи.

Достоинства

* Включает в себя различные тембры, чьи характеристики могут быть изменены.
* Инструменты разработки доступны для производства и настройки данных фонем.

Недостатки

* Доступность – распространяется на платной основе
* Плохое распознавание чисел

Microsoft Speech API является разработкой компании Microsoft для распознавания речи и синтеза речи. А также является свободно распространяемым контентом и не требует дополнительных библиотек для работы при условии использования на операционной системе Windows 10, MacOS и Linux.

Достоинства

* Все основные компоненты легко загружаются и не требуется дополнительные библиотеки.
* Качественная запись и произношение.

Недостатки

* Речевой ввод не всегда может сработать

pyttsx3 — это кроссплатформенная библиотека для перевода текста в речь (Text To Speech - TTS) в приложениях на Python 3. Она использует разные системы синтеза речи в зависимости от текущей операционный системы пользователя.

Достоинства

* Возможность работы без подключения к интернету
* Быстрая работа без задержек между получением текста и его произношением
* Распространение на бесплатной основе

Недостатки

* Очень редко может не понять слово.

Из перечисленных синтезаторов речи для дальнейшей работы будем использовать pyttsx3, но не зависимо какой движок будет выбран, потребуется библиотека pywin32 для доступа к возможностям TTS (Text To Speech).

Далее необходимо выбрать библиотеку, отвечающую за перевод речи в текст. Тут лучше взять SpeechRecognition, решение компании Google. Ее преимущества это

* Малый вес
* Небольшая нагрузка на компьютер, в частности оперативную память
* И быстрая обработка речи

Однако здесь возможно использовать pocketsphinx. Она может распознавать речь не только с микрофона, но и из файла, работает в офлайн режиме, однако для работы ей нужны дополнительные библиотеки и скорость работы ниже, чем у SpeechRecognition.

Также для решения проблемы с распознаванием речи в шумных помещениях необходимо установить библиотеку Fuzzywuzzy для нечеткого сопоставления речи. Нечеткое сопоставление строк — процесс поиска строк, соответствующих заданному шаблону. Библиотека использует расстояние Левенштейна для расчета различий между последовательностями. Что в итоге позволит улучшить точность голосового ввода.

Важно отметить, что разрабатываемого голосового ассистента зовут «Тоша». Выбор имени является важным аспектом разработки так как обращаться к голосовому ассистенту с трудновыговариваемым именем затруднительно, а также существует возможность что ассистент просто не поймет, что обращались именно к нему. Далее рассмотрим реализацию основных функций голосового помощника.

## **Выводы по главе 1**

1. На основе отечественной и зарубежной литературы, был проведен анализ существующих голосовых ассистентов и принципа их работы.
2. В результате анализа были рассмотрены основные возможности и недостатки существующих голосовых ассистентов, а также перспективы их развития
3. Основываясь на результатах анализа, были выявлены основные функции необходимые для людей с ограниченными возможностями по здоровью.
4. В результате анализа были сделаны выводы о нецелесообразности использования языков программирования Java и п Prolog при разработке голосового ассистента и необходимости использования Python в качестве основного языка.
5. Так же в результате анализа были выбраны основные библиотеки необходимые для функционирования голосового ассистента.

# **Глава 2. Проектирование и разработка голосового ассистента**

## **2.1 Функция справки**

Предоставляет пользователю список справок для каждой функции, которые есть в голосовом ассистенте. Вовремя как меню справки открыто, приводятся объяснение того, как работать с различными функциями.

## **2.2 Функция прогноз погоды**

Позволяет пользователю узнать метеопрогноз для любого указанного города мира. В ответе выдается такая информация как

* Максимальная температура
* Минимальная температура
* Состояние погоды (облачно, дождливо и так далее)

Командой активации запроса являются различные вариации вопроса – «какая сейчас погода в Петербурге?» При ее произношении Тоша начнет искать название города при помощи регулярного выражения - re.search () из библиотеки языка pyowm . После начинается поиск данных, связанных с погодой через:

* get\_status () – выражение для поиска погодных условий
* get\_tempera () – поиск максимальной и минимальной температуры

На рисунке 6 представлена логика работы функция прогноз погоды.

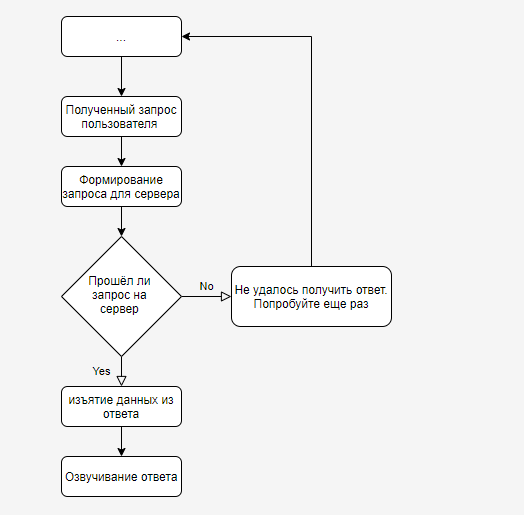


Рис. 6. Логика работы функция прогноз погоды

## **2.3 Функция произношения текущего времени**

Позволяет пользователю узнать текущее время. Ее можно запустить при помощи команды «Тоша, ты можешь сказать который час?» или любой другой вариации этой фразы. Тоша скажет вам текущее время вашего часового пояса.

На рисунке 7 представлена логика работы функции текущего времени.

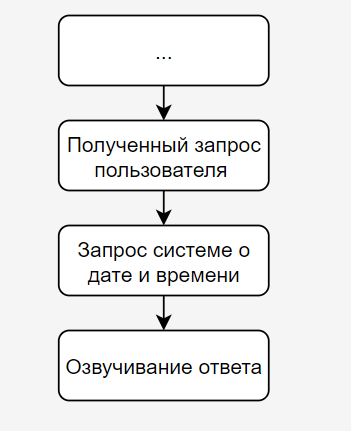


Рис. 7. Логика работы функции произношения текущего времени

## **2.4 Функция открывания браузера**

Позволяет пользователю открывать веб-сайт, при произношении адреса сайта или его названии, например через команду - «Тоша, открой Вконтакте» или другие похожие фразы, но ключевым словом является «открой», если оно присутствует в речи, то Тоша начнет поиск названия сайта или его адрес через метод re.search (). После чего ассистент будет использовать модули браузера и добавит протокол и домен сайта, если это необходимо и откроет нужный сайт.

На рисунке 8 представлена логика работы функции открытия браузера

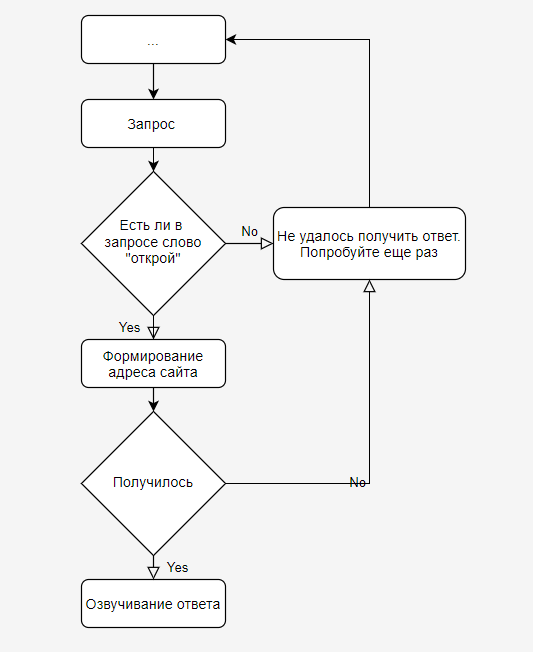


Рис. 8. Логика работы функции открытия браузера

## **2.5 Функция приветствие и завершение**

Позволяет пользователю начать или остановить работу голосового ассистента. Для этого необходимо соответственно произнести: «Здравствуй, Тоша» и «Заверши работу, Тоша» Так же в зависимости от времени суток помощник будет реагировать по-разному. Например, если вы начинаете работу 12 утра и до 6 вечера и говорите «привет», то ассистент ответит «Добрый день».

На рисунке 9 представлена логика работы функции приветствия

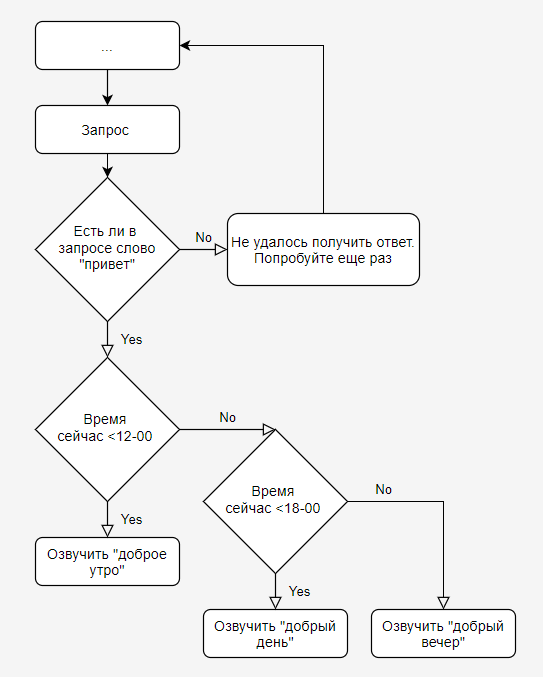


Рис. 9. Логика работы функции приветствия

## **2.6 Функция запуска приложения**

Функция запуска приложений активируется командой «Тоша, запусти имя приложения» ключевым словом тут является – запусти, но можно использовать и открой. После получения команды ассистент начнет поиск имени приложения если оно установлено на вашем персональном компьютере. Для этого все также используется метод re.search(), а также добавление к названию приложения суффикса .exe или .app в зависимости от операционной системы пользователя. После чего приложение запустится при помощи функции Popen (), являющейся одним из процессов Python.

На рисунке 10 представлена логика работы функции запуска приложения

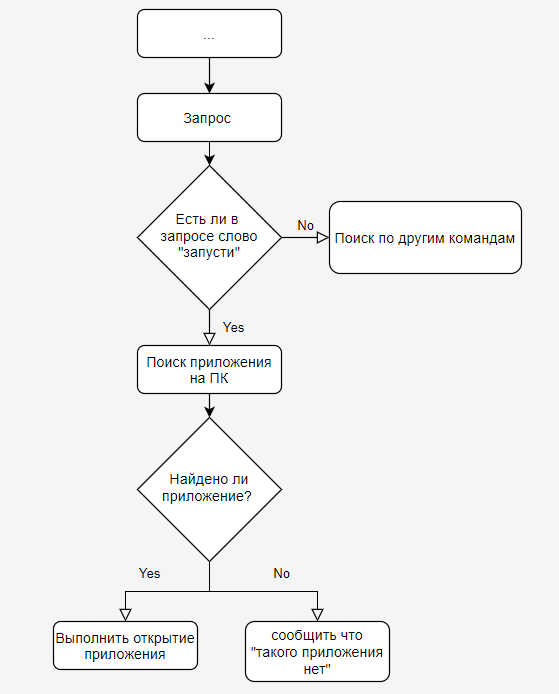


Рис. 10. Логика работы функции запуска приложения

## **2.7 Функция перевода**

Предоставляет пользователю как оригинальный текст, так и переведенный текст в зависимости от объективного языка, заданного пользователем. Пользователь дает исходный текст и объектный язык, который он хочет; затем переводчик вернет результат переведенного текста на основе исходного текста и требуемого языка. Между тем, в базе данных было сохранено 25 объектных языков, которыми пользователь может пользоваться.

## **2.8 Функция чата**

Позволяет пользователю общаться с голосовым ассистентом, чтобы весело провести время. Режим чата изначально закрыт и потребуется активировать его с помощью соответствующей команды. После входа в режим чата помощник будет выдавать текстовый ответ всякий раз, когда пользователь говорит с ним; однако ответ был заранее определен и сохранен в базе данных. Для каждого запроса программа определяет категорию запроса и случайным образом выбирает ответ из пула ответов в зависимости от категории запроса.

## **2.9 Функция поиска в Википедии**

Позволяет голосовому ассистенту получить и озвучить любую информацию с сайта Википедии. Активация происходит по команде «расскажи о …». К примеру, Тоша, расскажи о языке программирования python. Для реализации была использована библиотека wikipedia — это библиотека Python, которая позволяет получать и анализировать данные из Википедии, после найденная информация озвучивается. Но тут приходится устанавливать ограничение на 600 символов иначе будет озвучена вся страница, что не очень удобно.

## **2.10 Функция отправки email указанному пользователю**

Предоставляет возможность отправки писем голосом, но есть ограничение – это работает только для почты gmail. Для реализации функции необходима библиотека Smtplib именно она отвечает за отправления писем, а ее модули помогут определить получателя и соответствия его адреса протоколу SMTP (SMTP – это один из протоколов необходимый для передачи писем в сети). После инициализации SMTP gmail с помощью метода smtplib.SMTP () идет распознавание сервера через функцию ehlo () и кодируется сеанс starttls (). Далее голосовой в целях безопасности спросит логин и пароль от почты и отправит сообщение.

На рисунке 11 представлена логика работы функции отправки email

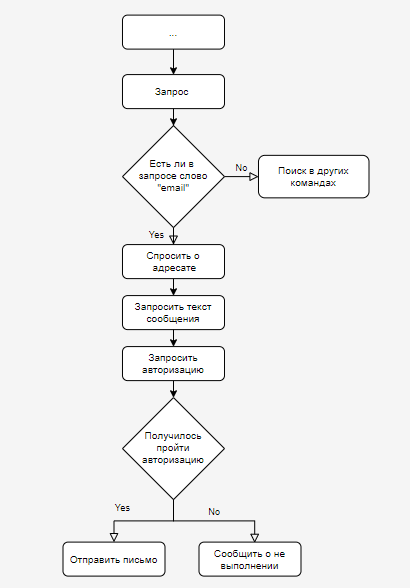


Рис. 11. Логика работы функции отправки email

## **2.11 Функция просмотра новостей**

Озвучивает пользователю последние новости для активации необходимо сказать: «Тоша, расскажи последние новости». Ассистент воспользуется сервисами Google и с их помощью зачитает последние 10 событий в мире.

На рисунке 12 представлена логика работы функции просмотра новостей

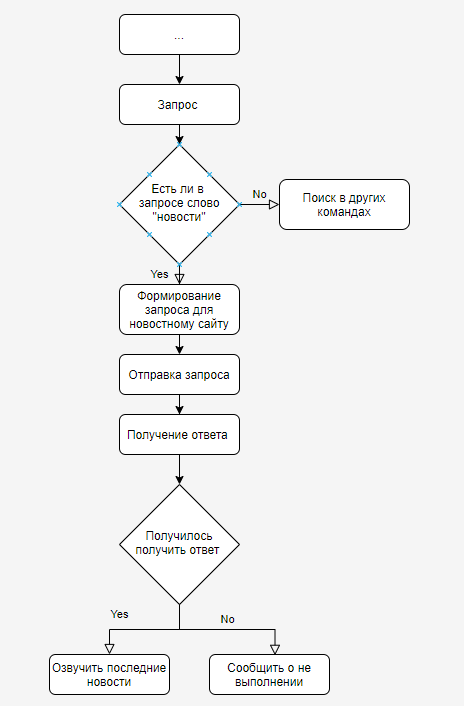


Рис. 12. Логика работы функции просмотра новостей

## **2.12 Функция запуска песни**

Позволяет голосовому ассистенту воспроизводить указанную песню через медиаплеер VLC. Активация происходи по команде «Тоша, сыграй мне песню» после чего ассистент попросит уточнить какую песню следует воспроизвести. Самостоятельно найдет ее через интернет или из указанной папки, если песни нет в указанном месте, то он загрузит на компьютер. В целях сохранения свободного места на диске компьютера песни скаченные из интернета удаляются после прослушивания.

Для создания функции были использованы и модуль

* библиотека youtube\_dl – для поиска музыки на сервисе Youtube
* библиотека VLC – для воспроизведения установленной песни

## **Выводы по главе 2**

На основании требований был спроектирован и разработан голосовой ассистент для людей с ограниченными возможностями по здоровью.

# **Заключение**

В результате разработки голосового ассистента для людей с ограниченными возможностями по здоровью были решены следующие задачи:

1. Проанализировать существующие голосовые ассистенты
2. Провести анализ функционала существующих голосовых ассистентов
3. Выявить требования необходимые для реализации голосового ассистента для людей с ограниченными возможностями

На основе вышесказанного получилось разработать и создать удобного и простого в использовании голосового ассистента - Тошу, который является кроссплатформенным приложением, то есть может быть установлен на такие операционные системы как Windows 10, MacOS и Linux. При этом он не требовательный к ресурсам персонального компьютера, а также имеет возможность работы, как при подключении к сети интернет, так и без подключения.

Также был реализован основной функционал, включающий 15 функций, достаточных для использования в повседневной жизни людьми инвалидами.

В случае дальнейшего развития приложения возможно добавление нового функционала, без изменения принципов работы голосового ассистента, а также размещение приложения на базе мобильной станции, а не персонального компьютера.

Разработанный голосовой помощник размещен в репозитории платформы Github. Для доступа к приложению нужно перейти по адресу <https://github.com/TradesMark/Jarvi-assistant>

# **Список литературы**

1. Хлопенкова А.Ю., Белов Ю.С. Методы обработки естественного языка в виртуальных голосовых помощниках // E-Scio. 2019 №11. С.167-173.
2. Поляков Е. В., Мажанов М. С., Качалова М. В., Поляков С. В. Разработка интеллектуального голосового ассистента и исследование обучающей способности алгоритмов распознавания естественного языка // Системный администратор. 2017. № 12. С. 80-85.
3. Леонид К. Как устроена Алиса. Лекция Яндекса // [Электронный ресурс] Режим доступа: https://habr.com/ru/company/yandex/blog/349372 (дата обращения: 20.04.2020).
4. Павел М. Обзор мобильных Text-To-Speech движков [Электронный ресурс]. Режим доступа: https://habr.com/ru/post/102199/ (дата обращения: 15.05.2020).
5. RUSBASE. 10 виртуальных ассистентов: обзор [Электронный ресурс]. Режим доступа: https://rb.ru/list/from-siri-to-ozlo (дата обращения: 14.05.2020).
6. Robogeek. Рынок умных колонок и голосовых ассистентов: Россия и мир. Прогнозы и тренды [Электронный ресурс]. Режим доступа: http://www.robogeek.ru/analitika/rynok-umnyh-kolonok-i-golosovyh-assistentov-rossiya-i-mir (дата обращения: 15.05.2020).
7. Академия Яндекса. Как устроена работа голосовых помощников [Электронный ресурс]. Режим доступа: https://academy.yandex.ru/posts/kak-ustroena-rabota-golosovykh-pomoschnikov (дата обращения: 14.05.2020).
8. Lumpics.ru. Встречайте: Алиса — голосовой помощник от Яндекса. [Электронный ресурс]. Режим доступа: http://lumpics.ru/meet-alice-voice-assistant-from-yandex (дата обращения: 20.04.2020).
9. Techopedia. Intelligent Virtual Assistant. [Электронный ресурс]. Режим доступа: https://www.techopedia.com/definition/31383/intelligent-virtual-assistant (дата обращения: 14.05.2020).
10. T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A survey of socially interactive robots," Robotics and Autonomous Systems, vol. 42, no. 3, pp. 143-166, 2003/03/31/ 2003.
11. C. Nass, Y. Moon, and P. Carney, "Are People Polite to Computers? Responses to Computer-Based Interviewing Systems1," vol. 29, no. 5, pp. 1093-1109, 1999.
12. M. McTear, Z. Callejas, and D. Griol, The Conversational Interface: Talking to Smart Devices. Springer Publishing Company, Incorporated, 2016, p. 422.
13. Voicebot.ai, "Voice assistant consumer adoption report 2018," Voicebot.ai2018, Available: https://voicebot.ai/voice-assistant-consumer-adoption-report-2018/, Accessed on: 2018.
14. T. W. Bickmore, L. Caruso, K. Clough-Gorr, and T. Heeren, "‘It’s just like you talk to a friend’ relational agents for older adults," Interacting with Computers, vol. 17, no. 6, pp. 711-735, 2005.
15. W. L. Cheong, Y. Jung, and Y.-L. Theng, "Avatar: a virtual face for the elderly," presented at the Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry, Hong Kong, China, 2011.
16. M. E. Kamali, L. Angelini, M. Caon, G. Andreoni, O. A. Khaled, and E. Mugellini, "Towards the NESTORE e-Coach: a Tangible and Embodied Conversational Agent for Older Adults," presented at the Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, Singapore, Singapore, 2018.
17. G. Riva, "ALFRED: A Personalized, Fully Interactive, and Mobile Assistant for Independent Living," vol. 21, no. 3, pp. 212-213, 2018.
18. S. Anderson, N. Liberman, E. Bernstein, S. Foster, E. Cate, B. Levin, and R. Hudson, "Recognition of elderly speech and voice-driven document retrieval," in 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No. 99CH36258), 1999, vol. 1, pp. 145-148: IEEE.
19. M. Vacher, S. Caffiau, Fran, #231, o. Portet, B. Meillon, C. Roux, E. Elias, B. Lecouteux, and P. Chahuara, "Evaluation of a Context-Aware Voice Interface for Ambient Assisted Living: Qualitative User Study vs. Quantitative System Evaluation %J ACM Trans. Access. Comput," vol. 7, no. 2, pp. 1-36, 2015.
20. Carroll, C. Chiodo, A. X. Lin, M. Nidever, and J. Prathipati, "Robin: Enabling Independence For Individuals With Cognitive Disabilities Using Voice Assistive Technology," presented at the Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, Denver, Colorado, USA, 2017.
21. Hassoon, J. Schrack, D. Naiman, D. Lansey, Y. Baig, V. Stearns, D. Celentano, S. Martin, and L. Appel, "Increasing Physical Activity Amongst Overweight and Obese Cancer Survivors Using an Alexa-Based Intelligent Agent for Patient Coaching: Protocol for the Physical Activity by Technology Help (PATH) Trial," (in eng), JMIR Res Protoc, vol. 7, no. 2, p. e27, Feb 12 2018.
22. E. Corbett and A. Weber, "What can I say?: addressing user experience challenges of a mobile voice user interface for accessibility," presented at the Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services, Florence, Italy, 2016.
23. D. Beirl, N. Yuill, and Y. Rogers, "Using Voice Assistant Skills in Family Life," 2019.
24. S. Turkle, Reclaiming conversation: The power of talk in a digital age. Penguin, 2016.
25. I. Amazon.com. (2019). Build Skills with the Alexa Skills Kit. Available: https://developer.amazon.com/docs/ask-overviews/build-skills-with-the-alexaskills-kit.html
26. A. o. Google. (2019, 01/31/2019). Responses | Actions on Google | Google Developers. Available: https://developers.google.com/actions/assistant/responses