



Международный журнал информационных технологий и энергоэффективности

Сайт журнала: <http://www.openaccessscience.ru/index.php/ijcse/>



УДК 004.522

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ВИРТУАЛЬНЫХ АССИСТЕНТОВ ДЛЯ РАЗРАБОТКИ ПРОГРАММ

¹Федулов Я.А., ¹Тычинская А.М., ¹Якушев В.А., ²Федулова А.С.

¹Филиал Национального исследовательского университета «МЭИ» в г. Смоленске, Смоленск, Россия (214013, г. Смоленск, Энергетический пр-д, 1), e-mail: fedulov_yar@mail.ru

²Национальный исследовательский университет «МЭИ», Москва, Россия (111250, г. Москва, ул. Красноказарменная, 14), e-mail: voitsitskay@mail.ru

В работе представлены распространенные виртуальные ассистенты, позволяющие упростить ввод запроса для поиска информации или запуска приложений. Отражена статистика использования виртуальных помощников с функциями необходимыми для разработки новых программ. Приведены положительные и отрицательные стороны использования виртуальных ассистентов, а также выделены основные требования к данным помощникам. Тестирование каждого из способов ввода запросов, позволяет сделать вывод о необходимости создания модели более обобщенного и интеллектуального виртуального ассистента, выполняющего распознавание звукового сигнала с использованием статистической модели для оптимальной работы приложения, что позволит существенно облегчить ввод запросов. Сделаны выводы о возможностях использования разработанных моделей для создания реальных виртуальных ассистентов, которые позволят облегчить использование поисковых систем и приложений.

Ключевые слова: виртуальный ассистент, интерфейс прикладного программирования, поиск информации, голосовое управление.

COMPARATIVE ANALYSIS OF VIRTUAL ASSISTANTS FOR PROGRAM DEVELOPMENT

¹Fedulov Ya.A., ¹Tychinskaya A.M., ¹Yakushev V.A., ²Fedulova A.S.

¹Branch of the National Research University "MEI" in Smolensk, Smolensk, Russia (214013, Smolensk, Energetichesky pr-d, 1), e-mail: fedulov_yar@mail.ru

²National Research University "MEI", Moscow, Russia (111250, Moscow, Krasnokazarmennaya st., 14), e-mail: voitsitskay@mail.ru

The paper presents common virtual assistants that make it easier to enter a query to search for information or launch applications. The statistics of the use of virtual assistants with the functions necessary for the development of new programs is reflected. The positive and negative aspects of using virtual assistants are given, and the basic requirements for these assistants are highlighted. Testing each of the methods for entering queries allows us to conclude that it is necessary to create a model of a more generalized and intelligent virtual assistant that performs audio signal recognition using a statistical model for optimal operation of the application, which will significantly facilitate the input of queries. Conclusions are made about the possibilities of using the developed models to create real virtual assistants, which will facilitate the use of search engines and applications.

Keywords: virtual assistant, application programming interface, information retrieval, voice control.

В современном мире большое количество информации хранится в сети Интернет, для получения доступа к которой существует ряд способов, таких как ввод запросов вручную с помощью устройств ввода, или использование виртуальных голосовых ассистентов [1]. Применение виртуальных помощников встречается в различных областях человеческой жизни, как игровой индустрии, так и в повседневной деятельности для установки или запуска приложений.

Актуальность рассматриваемой в данной работе темы обуславливается необходимостью учета технических и программных особенностей персональных компьютеров или смартфонов для установки и использования виртуальных ассистентов при решении различных прикладных и научных задач.

Целью исследования является изучение распространенных и малоизвестных виртуальных ассистентов, а также экспериментальное подтверждение преимуществ виртуальных помощников с открытым исходным кодом, которые подходят для решения широкого класса частных задач и предоставляют возможности внедрения улучшений.

В качестве программно-аппаратной платформы для проведения экспериментов с использованием реальных виртуальных ассистентов выступал ноутбук, на котором установлены параллельно две операционные системы Windows 8 и Ubuntu 2020. Различные операционные системы и установленные браузеры позволили провести необходимые эксперименты и сделать выводы о популярных и малораспространенных виртуальных ассистентах, а также осветить их особенности. В данной статье проводится анализ существующих виртуальных ассистентов, рассматриваются их основные функции, в том числе возможности определения частей для модификации и программного использования.

Виртуальные ассистенты – это программы, работающие на основе искусственных нейронных сетей, при этом запросы могут быть представлены в письменном виде или распознавания голоса и дальнейшей обработки полученного сообщения [2]. Использование искусственных нейронных сетей позволяет данным программам вести диалог, выполнять простые задачи. Чаще всего виртуальные ассистенты представлены голосовыми помощниками, разработанные с использованием искусственного интеллекта, технологий машинного обучения, а также распознавания голоса.

Идея распознавания голосовых сообщений с использованием информационных технологий зародилась еще в середине XIX века, ученые разрабатывали машину для распознавания одночленных звуков. В XXI веке вычислительные машины позволяют находить целые предложения, это стало возможным после проведения обучения и автоматизации, что показывает высокий уровень развития технологий, которые позволяют искусственным-нейронным сетям проводить анализ контекста и определять источник звукового сообщения.

Использование виртуальных помощников может быть различным, как представлено на рисунке 1.

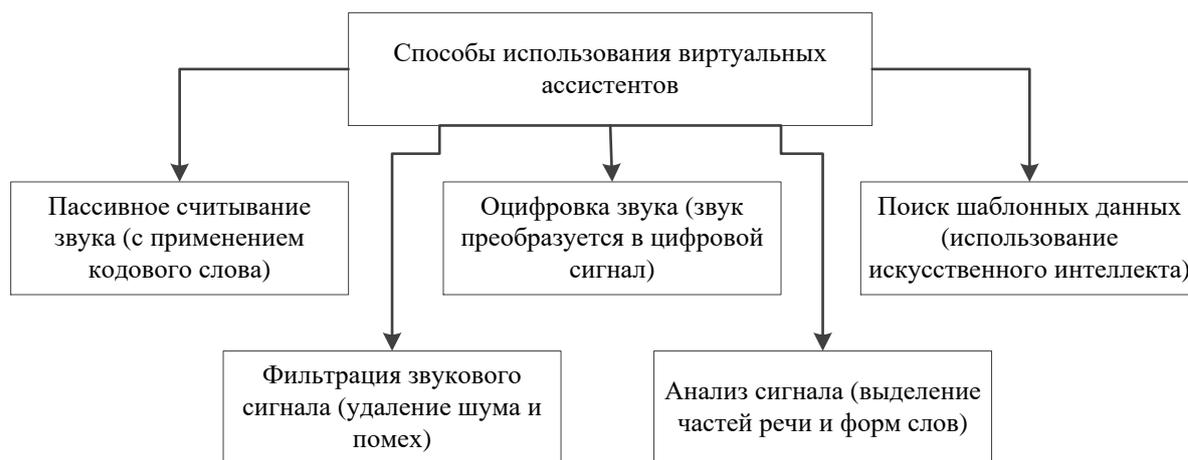


Рисунок 1 – Способы использования виртуальных ассистентов с пояснениями

Для реализации виртуального ассистента может быть выбран любой из существующих языков программирования, и на сегодняшний момент одним из самых распространенных является Python, поскольку он ориентирован на повышение производительности программиста и читаемости кода, а также его стандартная библиотека имеет множество полезных функций.

Для получения необходимых данных о возможности применения готовых решений и дальнейшей разработки программного обеспечения на их основе были проведены следующие эксперименты и сделаны соответствующие выводы. Была поставлена цель установки голосового ассистента на компьютер, для тестирования его свойств.

Описание экспериментов

Для работы были установлены последние версии браузеров, а также скачаны и установлены необходимые для экспериментов виртуальные помощники: Алиса (Yandex), Siri (Apple), Google-ассистент (для телефонов на Android), CMU Sphinx и Julius (Linux).

Голосовой помощник Алиса (Yandex)

Алиса – виртуальный русский голосовой ассистент, разработан программистами компании «Яндекс» в 2017 году. Алиса считается одной из самых популярных приложений среди виртуальных ассистентов, так как имеет большое количество функций: распознавание естественной речи, имитация живого диалога, может решать прикладные задачи.

Работа приложения представлена следующими пунктами:

- 1) речь пользователя транслируется на сервер распознавания;
- 2) сервер преобразует полученный звуковой сигнал в текстовый;
- 3) полученный текст отправляется в сервис в определенную сущность – классификатор интенгов;
- 4) каждому интенту соответствует специальная модель (семантический теггер).

Основная задача модели – выделение необходимой информации для себя из сказанной пользователем фразы. Обработанная информация в модуль dialog manager, в котором хранится контекст диалога, а также что предшествовало данному запросу. В данный модуль поступают результаты разбора сказанной реплики, и система должна принять решение о дальнейших действиях с ним.

Выполнение запроса с использованием виртуального ассистента «Алиса» можно представить схематично, как показано на рисунке 2.



Рисунок 2 – Обработка звукового запроса виртуальным ассистентом

Данный виртуальный помощник функционирует на смартфонах и компьютерах, а также на информационно-развлекательной платформе для автомобилей (Яндекс-Авто), в «умной» колонке со встроенным голосовым помощником «Алиса».

Установка и работа с данным виртуальным ассистентом проста и понятна для пользователя смартфона и компьютера, что можно считать основным достоинством данного приложения, но привязанность к определенному браузеру делает его негибким и ограничивает возможности для дальнейшего использования.

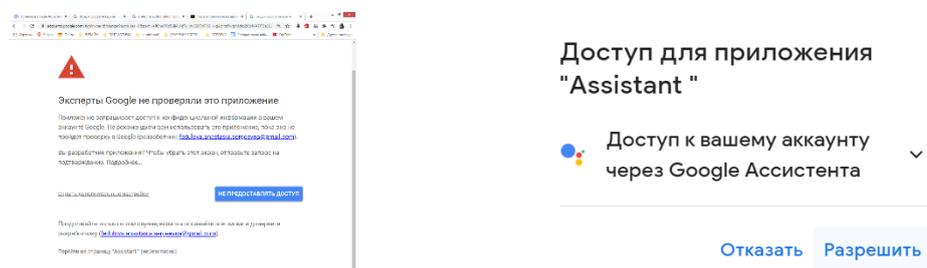
Интерфейс прикладного программирования (Application Programming Interface, API) в приложении Алиса позволяет разработать собственный навык и опубликовать его в специализированном сервисе «Яндекс.Диалоги» и он будет доступен для дальнейшей работы с ассистентом. Разработка навыка начинается с определения задачи, которая должна быть решена с помощью него, при этом задействуются определенные инструменты для интеграции навыка в Алису. Выделяются следующие типы навыков: навыки общего типа, навыки общего типа, навыки «Синтезатор».

Google-ассистент (на платформе Android)

Google Assistant – виртуальный голосовой ассистент, разработан программистами компании «Google» в 2016 году. Предпочтительнее работает на смартфонах с операционной системой Android, но для компьютеров разработан ассистент, устанавливаемый с помощью Google Cloud Platform - Tools & Modern Applications, но для этого необходимо иметь Google-аккаунт.

Алгоритм установки [3] данного приложения на компьютер достаточно трудоемкий. Далее представлен процесс установки Google Assistant на персональный компьютер (рисунок 3). Как видно из рисунка, необходимо давать полный доступ к компьютеру, чтобы приложение

начало функционировать на компьютере. Далее представлен результат работы с установленным приложением, рисунок 4.



```
Please visit this URL to authorize this application:
https://accounts.google.com/o/oauth2/auth?response_type=code&client_id=371305482358-9cdeak2aodbmvehkao54bujk563ig6vl.apps.googleusercontent.com&redirect_uri=urn%3Aietf%3Aawg%3Aoauth%3A2.0%3Aaob&scope=https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fassistant-sdk-prototype+https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fassistant-sdk-prototype&state=WhrizNlQDwJgell2pV6bIsPOLdMmT4&prompt=consent&access_type=offline
Enter the authorization code: 4/1AY0e-g7D-7g7W1Xu8CyViCbe6R-DrKWZALiC6gXUCLJqt8zZAURNvDydKME

googlesamples-assistant-devicetool --project-id primal-context-299110 register-model --manufacturer "Assistant SDK developer" --product-name "Assistant SDK light" --type LIGHT --model "371305482358-9cdeak2aodbmvehkao54bujk563ig6vl.apps.googleusercontent.com_OAuth"
py -m googlesamples.assistant.grpc.pushtotalk --device-model-id "371305482358-9cdeak2aodbmvehkao54bujk563ig6vl.apps.googleusercontent.com_OAuth" --project-id primal-context-299110
```

Рисунок 3 – Установка Google-Assistant на компьютере с Windows

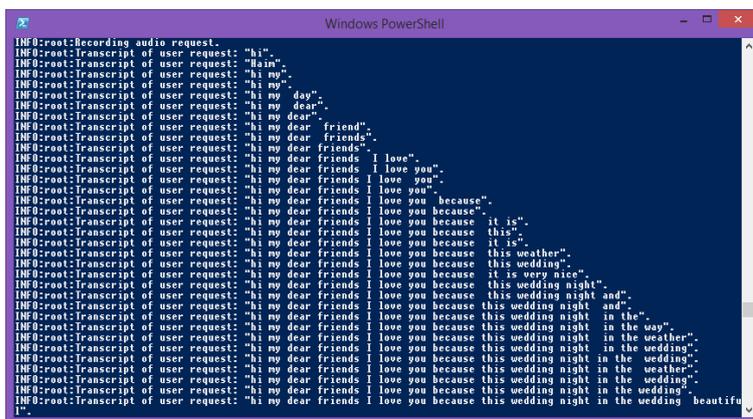


Рисунок 4 – Работа приложения на компьютере

Виртуальный помощник Google Assistant используется в смартфонах, включен в приложение для мгновенного обмена сообщениями, в умный голосовой Wi-Fi динамик для управления умным домом, в программное обеспечение умных часов от Google. Основная проблема в установке на персональный компьютер, виртуальный помощник на компьютере воспринимает только английские слова.

Код данного виртуального ассистента частично открыт для модификаций. Для выполнения и добавления запросов и интеграции с Ассистентом, разработчики предоставляют Google доступ к определенным данным, таким как:

- 1) информация необходимая, чтобы установить связь между аккаунтами (пользователи могут взаимодействовать со сторонними функциями через виртуального помощника);
- 2) библиотеки данных, а также коллекции контента (позволяют приложению выполнять запросы с использованием сервиса разработчика);

3) данные для взаимодействия с пользователями (информация передается в Google разработчиками, использующими Actions on Google).

4) пользовательские запросы и данные, которые связаны через запрос (информация отправляется в Google устройствами внешних производителей со встроенным виртуальным ассистентом (колонки, телевизоры) для того, чтобы помощник смог осуществлять дальнейшую работу).

Для разработки собственных приложений с использованием предоставляемых голосовым ассистентом возможностей, были рассмотрены возможности работы с использованием Google API Assistant. Служба Google Assistant предоставляет низкоуровневый API, который позволяет вам напрямую управлять байтами звука в запросе и ответе Assistant. Привязки для этого API могут быть созданы для таких языков, как Node.js, Go, C ++, Java, для всех платформ, поддерживающих gRPC. Python предоставляется справочный код для захвата звука, воспроизведения звука и управления состоянием разговора.

Голосовой помощник Siri (Apple)

Siri – виртуальный голосовой ассистент, разработан программистами Международного центра искусственного интеллекта (SRI) в 2007 году, названного самым большим проектом, в котором задействован искусственный интеллект [**Ошибка! Источник ссылки не найден.**], который в 2010 был выкуплен компанией Apple.

Особенности данного виртуального ассистента [5]:

1) использует обработку естественной речи, для ответа на вопросы, а также для построения рекомендации;

2) приспосабливается индивидуально под каждого пользователя, изучает предпочтения в течение долгого периода времени.

Распознавание речи возможно с использованием голосовых технологий компании Nuance Communications [6].

Действия и ответы формируются с использованием разработок партнёров:

- Citysearch, BooRah, Yelp Inc, Yahoo Local, Yandex, ReserveTravel – решение деловых вопросов;
- Eventful, StubHub и LiveKick – события и концертная информация;
- MovieTickets.com, Rotten Tomatoes, The New York Times информация о фильмах;
- Yahoo Weather для информации о погоде;
- Bing, Yahoo, Google для веб-поиска.

Минусы использования данного ассистента.

1. Использование только на продуктах компании Apple.

2. Запросы могут быть не обработаны.

3. Настройка может быть сделана на прием запросов, также при заблокированном телефоне (иные лица могут использовать данный телефон для отправки СМС-сообщений, также сервиса СМС-банкинга для получения баланса карты владельца телефона или кражи денег).

Для понимания работы приложения Siri необходимо рассмотреть API – SiriKit, которое работает с использованием набора доменов, которые предоставляют связанные области функциональности, такие как обмен сообщениями. В каждом домене есть набор намерений, которые представляют конкретные задачи, решаемые пользователем с помощью Siri. Обработка

языка в приложении сводится к тому, что SiriKit решает, какое намерение и приложение запрашивает пользователь, и код проверяет, что пользователь просит, а затем выполняет это.

Системы распознавания речи CMU Sphinx и Julius (Linux)

CMU Sphinx – группа систем распознавания речи, разработанных программистами в Университете Карнеги-Меллона. Данный набор содержит в себе различные распознаватели речи, такие как Sphinx 2–4, а также тренажер акустических моделей – SphinxTrain. В 2000 году группа Sphinx открыла исходный код некоторых компонентов распознавателя речи, речевые декодеры поставляются с примерами приложений [7]. Для обучения акустической модели необходимо дополнительно программное обеспечение, компилятор языковой модели, а также общедоступный словарь произношения – cmudict.

CMU Sphinx имеет следующие особенности: современные алгоритмы распознавания речи для эффективного распознавания речи, инструменты CMU Sphinx разработаны специально для платформ с низким уровнем ресурсов, гибкий дизайн, большое количество инструментов для функций, связанных с распознаванием речи.

Julius – высокопроизводительное двухпроходное программное обеспечение (ПО) для исследователей и разработчиков, связанных с речью, используется, чтобы проводить декодирование непрерывной речи, для этого имеет обширный словарь.

Выполняет декодирование в реальном времени на большинстве современных персональных компьютерах, с использованием триграммы слов и контекстно-зависимой скрытой марковской модели. В рассматриваемую систему включены основные методы поиска.

Основной платформой является Linux, а также иные рабочие станции Unix, отдельные версии ассистента поддерживают работу на операционной системе Windows . Julius – бесплатное ПО с открытым исходным кодом.

В Julius версии 3.4 встроен грамматический анализатор распознавания Julian . Julian – модификация Julius, где языковая модель – разработанный вручную тип конечного автомата, называемый грамматикой детерминированного конечного автомата, который может использоваться для создания системы голосовых команд со своим словарем или для различных задач голосовой диалоговой системы.

API имеет особенности: программное обеспечение с открытым исходным кодом, высокоскоростное и точное распознавание в реальном времени на основе стратегии двух проходов, низкие требования к памяти.

Для дальнейшей работы необходимо получить информацию об использовании голосовых ассистентов в России [8]: «Алиса» – осведомлены 96% респондентов и используется 58,79%. Google Assistant – 20,13% участников исследования используют, а осведомлены 67%. О Siri знают 61% опрошенных, из них 16,29% активно используют. CMU Sphinx и Julius (Linux) знают 15% участников, используют 4,79%, статистика представлена на рисунке 5.

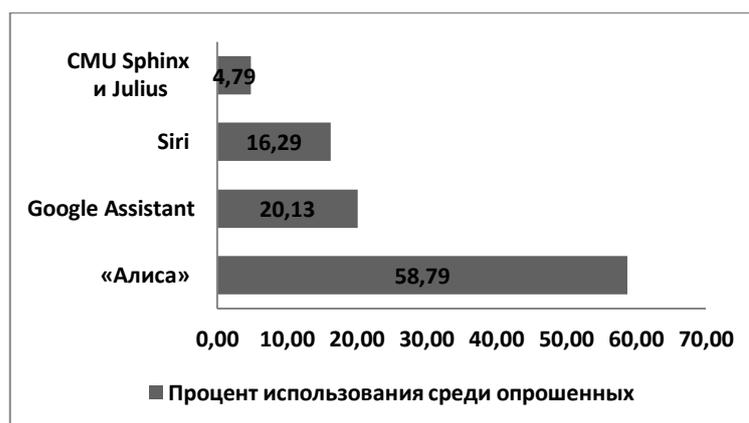


Рисунок 5 – Статистика использования виртуальных ассистентов

По результатам проведенного анализа с учетом потребностей как пользователей, так и разработчиков программного обеспечения были выделены основные характеристики виртуальных помощников, представленные в таблице 1.

Таблица 1 – Характеристики виртуальных помощников

№ п/п	Характеристики	Алиса	Google Assistant	Siri	CMU Sphinx и Julius
1	Операционная система	Microsoft Windows, Android и iOS	Android, Chrome OS, iOS, iPadOS, KaiOS, Linux, Wear OS, Android Auto	iOS 5-13, macOS Sierra, macOS High Sierra, macOS Mojave	Linux, Mac OS, Windows, Android
2	Приветствие или активация	Привет, Алиса	Ок, Google	Привет, Siri	Задается пользователем (ок, компьютер)
3	Понимание произвольной речи и команд	Да	Да	Да	Настраиваемое
4	Применение	Распознавание речи, запуск приложений, реальное общение	Поиск фраз, работа с умными устройствами	Управление телефонными коммуникациями	Задается пользователем
5	Открытость кода	Smart Home API	Python client library or generated bindings for languages like Go, Java, C#, Node.js, and Ruby	API SiriKit	Консольный, API
6	Распознавание	Wake-Up-Word ASR	Wake-Up-Word ASR	Wake-Up-Word ASR	Алгоритм Витерби, bushderby
7	Реализация настраиваемых функций	Нет	Нет	Нет	Да

Исходя из проведенного сравнительного анализа основных голосовых помощников и их API, применительно к задачам распознавания звукового сообщения и разработки программных средств на их основе были сделаны следующие выводы. В случае возникновения необходимости использования виртуального голосового помощника, который будет легко установлен на требуемой платформе, подойдут готовые популярные решения Алиса, Google

Assistant, Siri, которые предоставляют достаточный набор предустановленных функциональных возможностей.

В случае использования предоставляемого популярными решениями средств разработки API приходится сталкиваться с рядом существенных ограничений, связанных с жесткостью и привязанностью к определенным программно-аппаратным платформам. Возможностей популярных готовых решений голосовых ассистентов не хватает для решения нетипичных задач, что приводит к необходимости создания собственного виртуального ассистента, работающего на персональных компьютерах или мобильных платформах. Для решения такого типа задач голосового управления требуется работать с виртуальными ассистентами с открытым кодом, дописывать и тонко настраивать их работу, и в данном случае открытые проекты CMU Sphinx и Julius предоставляют гораздо более широкие возможности.

На основе проведенного исследования ведется построение собственного усовершенствованного голосового ассистента, позволяющего не только осуществлять запрос по голосовому сообщению, но также осуществлять анализ полученного текста, проводить его распознавание и выдавать ответ. Разработка проводится на основе группы систем распознавания речи, для системы Linux с использованием структуры системы с применением кодового слова, что позволит использовать приложения не только для вывода информации, а также для запуска определенных приложений.

Список литературы

1. Смыслова Л. В. Чат-бот как современное средство интернет-коммуникаций // Молодой ученый. 2018. №9. С. 36-39.
2. Поначугин А.В., Пичужкина Д.Ю., Смекалова Е.С. Голосовой помощник как технология обработки данных. // НАУКА БЕЗ ГРАНИЦ. М.: НГПУ им. К. Минина, 2020 С. 96-100.
3. Установка Google Ассистента в Windows. URL: <https://g-ek.com/ustanovka-google-assistenta-v-windows-10>. Дата обращения: 12 декабря 2020.
4. M. Pinola, Speech recognition through the decades: how we ended up with Siri, [Электронный ресурс] // PCWorld, 2011. URL:http://www.techhive.com/article/243060/speech_recognition_through_the_decades_how_we_ended_up_with_siri.html?page=0.
5. Jain M., Kota R., Kumar P., Patel S.N. Convey: Exploring the Use of a Context View for Chatbots // Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018.
6. Janarthanam S. Hands-On Chatbots and Conversational UI Development: Build chatbots and voice user interfaces with Chatfuel, Dialogflow, Microsoft Bot Framework, Twilio, and Alexa Skills. Packt Publishing, 2017.
7. 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.
8. Наумов М.Ю., Чистяков А.С. Применение систем искусственного интеллекта в различных сферах деятельности // Постулат. 2017. №5.

References

- 1 Smyslova LV Chatbot as a modern means of Internet communications // Young scientist. 2018. No. 9. pp. 36-39.
 - 2 Ponachugin A.V., Pichuzhkina D.Yu., Smekalova E.S. Voice assistant as a data processing technology. // SCIENCE WITHOUT BORDERS. М.: NGPU im. K. Minina, 2020 pp. 96-100.
 - 3 Installing Google Assistant in Windows. URL: <https://g-ek.com/ustanovka-google-assistent-a-v-windows-10>. Date of treatment: December 12, 2020.
 - 4 M. Pinola, Speech recognition in decades: how we came to Siri, [Electronic resource] // PCWorld, 2011. URL: http://www.techhive.com/article/243060/speech_recognition_through_the_decades_how_we_ended_up_with_siri.html? Page = 0.
 - 5 Jain M., Kota R., Kumar P., Patel S.N. Convey: Exploring the Use of Context View for Chatbots // Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018.
 - 6 Janarthnam S. Hands-On Chatbots and Conversational UI Development: Build chatbots and voice user interfaces with Chatfuel, Dialogflow, Microsoft Bot Framework, Twilio, and Alexa Skills. Packt Publishing, 2017.
 - 7 Polyakov E.V., Mazhanov M.S., Kachalova M.V., Polyakov S.V. Development of an intelligent voice assistant and study of the learning ability of natural language recognition algorithms // System Administrator. 2017. No. 12.
 - 8 Naumov M.Yu., Chistyakov A.S. Application of artificial intelligence systems in various fields of activity // Postulate. 2017. No. 5.
-