DOI 10.52209/1609-1825_2022_4_443 УДК 004.896

Технологии распознавания естественного языка для разработки голосового ассистента

[1]***СПИРИНА Елена Александровна,*** *к.п.н., ассоциированный профессор, sea_spirina@mail.ru,*
[1]***ЖУМАБЕКОВ Адильхан Ералыулы,*** *магистрант, sk1llahh2@gmail.com,*
[1]***ГОРБУНОВА Надежда Александровна,*** *к.п.н., ассистент профессора, ant_nadezhda@mail.ru,*
[1]***КАЗИМОВА Динара Ашубасаровна,*** *к.п.н., профессор, dinkaz73@mail.ru,*
[2]***СЕЛЕЗНЕВА Руслана,*** *к.т.н., профессор, seleznevaruslana@gmail.com,*
[1]НАО «Карагандинский университет имени Е.А. Букетова», Казахстан, Караганда, ул. Университетская, 28,
[2]Высшая школа Хуманитас, Польша, Сосновец, 41-200, ул. Яна Килиньского, 43,
**автор-корреспондент.*

Аннотация. *В работе рассматриваются базовые технологии распознавания естественного языка и особенности создания голосового ассистента (помощника) на основе облачного сервиса DialogFlow. В качестве примера разработано приложение – голосовой помощник «IT-абитуриент», который отвечает на вопросы абитуриентов об образовательных программах в сфере IT Карагандинского университета им. академика Е.А. Букетова. В статье проанализированы принципы работы популярных голосовых ассистентов Google Assistant, Amazon Echo (Alexa), Microsoft Cortana, Yandex Alisa. Выявлено, что платформа для распознавания естественного языка Google DialogFlow является удобным инструментом для разработки диалогового пользовательского интерфейса в мобильных приложениях. Рассмотрен алгоритм разработки голосового ассистента на основе облачного сервиса DialogFlow.*

Ключевые слова: *искусственный интеллект, распознавание естественного языка, Web-приложение, голосовой ассистент, голосовой помощник, Google Alisa, облачный сервис DialogFlow, алгоритм разработки голосового помощника.*

Введение

Технологии распознавания естественного языка активно используются для различных областей человеческой деятельности. Голосовые ассистенты (помощники) – это боты, которые работают на основе искусственного интеллекта (ИИ), технологиях машинного обучения, распознавания голоса и обработки естественного языка, чтобы отвечать на вопросы, вести разговоры, осуществлять запуск простых задач [1]. Попытки распознать голос при помощи интеллектуальных технологий, начались ещё в 1930 г. Тогда учёные старались запрограммировать машину на распознавание одночленных звуков, а сейчас наши смартфоны и компьютеры понимают целые предложения. Данный скачок возник в результате развития машинного обучения, которое заставляет нейронные сети самостоятельно анализировать контекст и эффективно определять основной источник звука [2].

Целью исследования являются анализ технологий распознавания естественного языка и разработка голосового ассистента (помощник) «IT-абитуриент», который отвечает на вопросы абитуриента об образовательных программах, помогает разобраться в особенностях IT-специальностей Карагандинского университета им. Е.А. Букетова.

Материалы и методы исследования

Существуют разные подходы к построению систем ИИ: логический, структурный, эволюционный, имитационный. Для решения задач распознавания образов эффективно применяется структурный подход, в рамках которого появились модели, известные под общим наименованием «нейронные сети» (НС). Модели эти отличаются по строению отдельных нейронов, топологии связей между ними и методам обучения. Между наиболее популярными вариантами НС выделяют НС с обратным распространением ошибки, сети Хопфилда и стохастические нейронные сети. Нейронные сети более успешно применяются в задачах определения образов, в том числе нечётких [3, 4].

Для обработки естественного языка (англ. Natural Language Processing, NLP) используются различные модели нейронных сетей и технологии обработки естественного языка с поддержкой глубокого обучения (англ. Deep Learning, DL): рекуррентные нейронные сети (англ. Recurrent Neural

Network, RNN), свёрточные нейронные сети (англ. Convolutional Neural Networks, CNN). Свёрточные нейронные сети в основном применяются в компьютерном зрении, однако также используются в обработке естественного языка, к примеру, в задачах генерации (англ. speech generation) и определения речи (англ. speech recognition). Одной из известных архитектурных нейросетей считается Attention, в основе которой использована двухсторонняя рекуррентная нейронная сеть [4]. Развитием архитектуры Attention является разработка компании Google, которая именуется Transformer. Создатели ликвидировали из архитектуры все рекуррентные (RNN) и свёрточные (CNN) слои, и ввели конструкцию self-attention [5]. Компанией Microsoft Google используется подход Word2Vec, который заключается в преображении слов в вектор, отражающий семантические характеристики текста [6].

Наиболее известными голосовыми ассистентами являются: Google Assistant, Siri от Apple, Amazon Echo (Alexa) от Amazon, Cortana от Microsoft, Алиса от Yandex. В работе [7] рассмотрены принципы работы некоторых из них. Так, например, Alexa построена на основе процедуры преобразования речи в отдельные слова, звуки и предложения. Потенциал обработки естественного языка в Cortana получены от Tellme Networks и соединены с базой данных семантического поиска под наименованием Satori.

Голосовой ассистент Алиса работает на основе технологии SpeechKit, которая использует клиентскую библиотеку для интерпретации текста и голоса. Для анализа речевого потока применяется языковая модель определенной предметной области, которая позволяет подбирать лучшие варианты для перевода голоса в текст. Далее запрос передается Алисе, и она пытается определить категории предметной области. Данный модуль является классификатором «намерений» [8].

В дальнейшем для формирования «намерений» в Алисе стал применяться облачный сервис распознавания естественного языка от Google DialogFlow, который поддерживает различные языки, в том числе русский. DialogFlow используется для разработки диалогового пользовательского интерфейса в мобильных приложениях, веб-приложениях, устройствах, ботах и т.п. [9, 10].

Результаты и их обсуждение

В работе [1] определен обобщенный алгоритм работы голосовых ассистентов, который будет использоваться нами для разработки приложения «IT-абитуриент» на основе облачного сервиса DialogFlow.

Этап 1. Построение концептуальной модели приложения. Целью приложения является предоставление абитуриентам конкретной информации об образовательных программах в сфере IT, которые предлагает Карагандинский университет имени Е.А. Букетова: 6B06101-Информатика, 6B06103-Информационные системы, 6B06104-Математическое и компьютерное моделирование.

Для определения процессов, построим диаграммы деятельности UML. Выделим пользователей (актёров), которые будут взаимодействовать с программным продуктом:

- администратор программы – специалист, который будет управлять всем контентом и функциями в системе;

- пользователь – абитуриенты университета, но в дальнейшем приложением может пользоваться и студент.

Определим основные прецеденты: просмотр новостей, контактов, информация об образовательных программах.

Исполнитель: администратор – производится в панели администратора веб-приложения. Описание шагов:

- авторизоваться в системе;

- перейти к добавлению или редактированию информации.

Исполнитель: пользователь – производится в веб-приложении. Описание шагов:

- запустить голосовой помощник;

- начать диалог с ботом.

Процесс ответа на запрос пользователя представим в виде диаграммы деятельности UML (рисунок 1).

Этап 2. Выбор инструментов. Для создания Web-приложения с голосовым ассистентом использованы инструменты:

- HTML/CSS – язык разметки и кастомизация UI элементов интерфейса;

- JavaScript и библиотеки;

- Node js – среда выполнения JavaScript, основанная на движке V8;

- Редактор Visual Studio Code – среда для разработки веб- и облачных приложений;

- UNIT E2 система тестирования.

- DialogFlow – облачный сервис распознавания естественного языка от Google;

- Speech recognition – модуль Node JavaScript для распознавания речи.

Этап 3. Разработка серверной части приложения голосового помощника. Для разработки голосового помощника «IT-абитуриент» выполним регистрацию в сервисе DialogFlow, для которой понадобится аккаунт в Google. Далее создаем в DialogFlow агента с помощью команды «Create agent», который будет выполнять основные шаги по общению с пользователем, задаем ему название, выбираем русский язык.

Чтобы агент начал обрабатывать запросы пользователя, создаем для каждого агента его намерения (Intents, цели). Для интеграции приложения требуется получить его токен (ключ) в google cloud platform и Project ID. После регистрации продукта создаем для веб-приложения сервисный аккаунт, который необходим для получения json файла – JSON-ключ для авторизации при обращении к сервису через API.

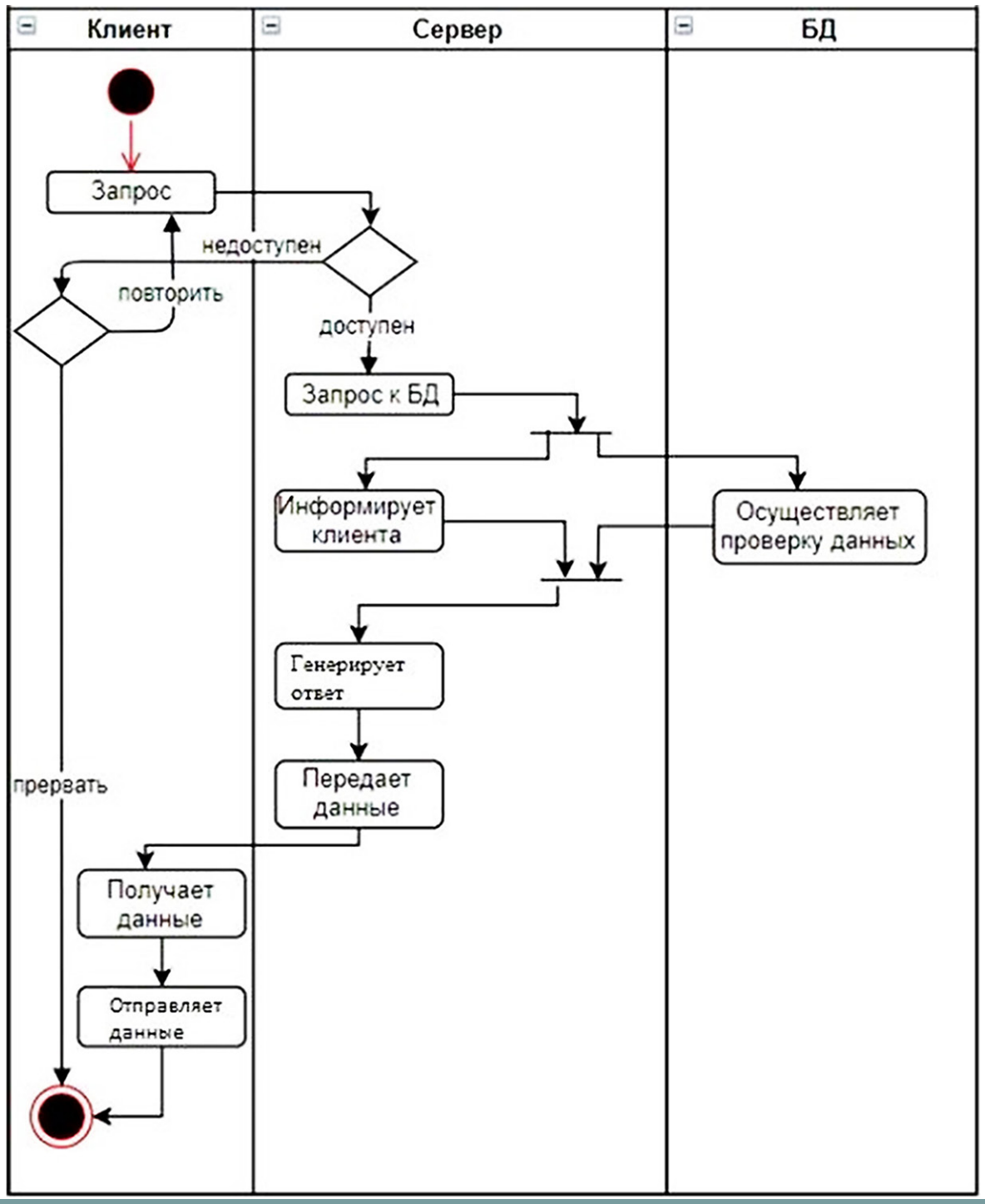

Рисунок 1 – Диаграмма деятельности процесса

Заполнение данных для искусственного интеллекта является важным элементом в создании приложения. В любом из намерений (intent) настраиваются «признаки», по которым будет происходить переход именно в него. Для этого создаются тренировочные фразы (Training phrases), на основе которых DialogFlow определяет то или иное намерение пользователя.

Затем редактируются intent Welcome, т.е. описываем все варианты в Response. В поле Intent name вводится название намерений. В блоке Training phrases вводятся примеры фраз, на которые помощник должен отреагировать нужным образом, а в блоке Responses добавляются один или несколько ответов, которые голосовой ассистент будет произносить в ответ на вопрос пользователя.

Для корректной работы голосового ассистента создается нужное количество намерений (intent) с вопросами и ответами для составления диалога пользователя с приложением (рисунок 2).

Этап 4. При реализации клиентской части Web-приложения разрабатываются все основные элементы интерфейса с применением JS скриптов, которые определяют интерактивность.

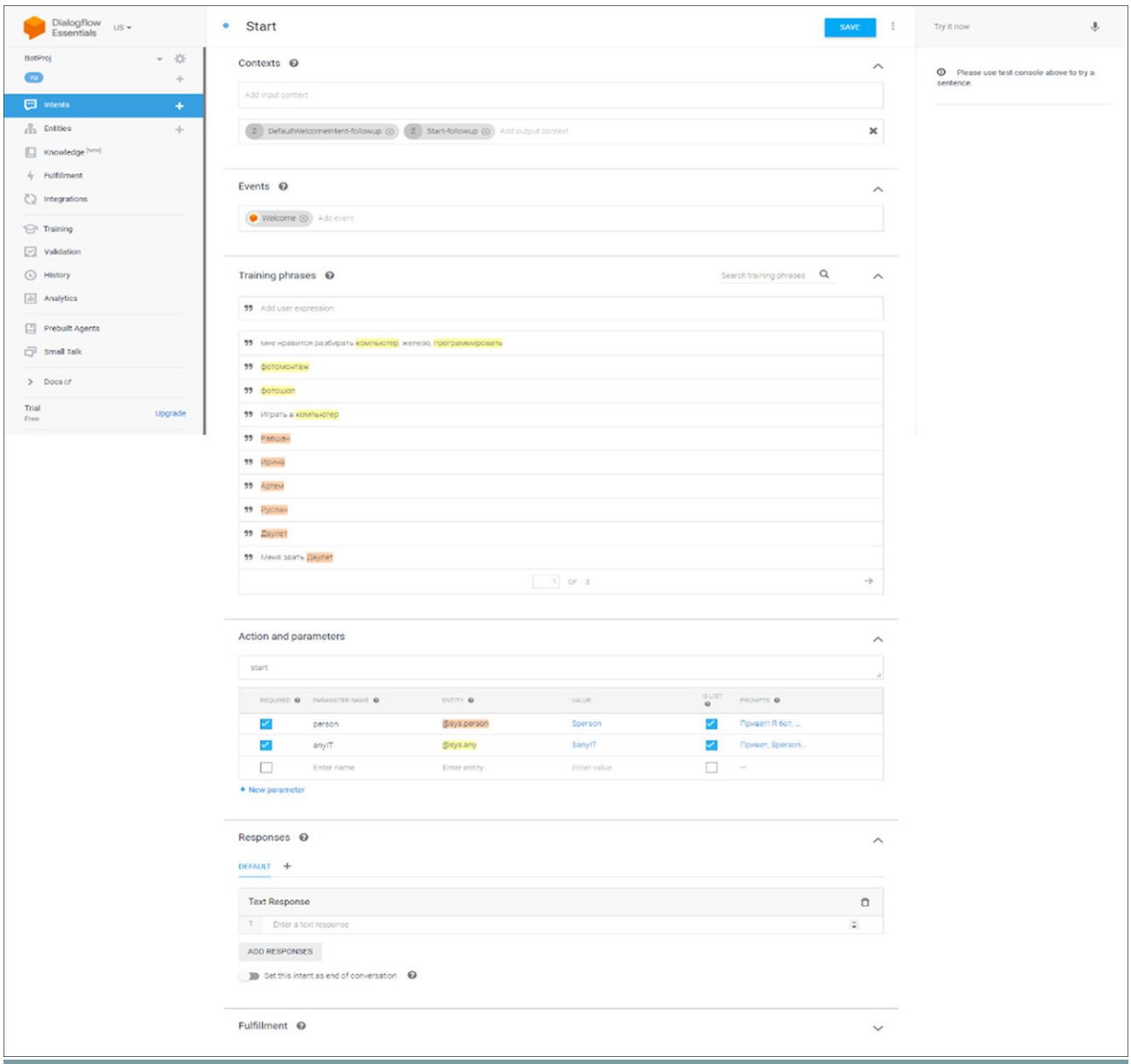

Рисунок 2 – Пример создания intent

При создании данного приложения использовался нативный Javascript.

Таким образом, интерфейс Web-приложения голосового помощника «IT-абитуриент» имеет поле для текстового ввода вопроса, кнопку для отправки голосового сообщения, а также окно для просмотра чата, в котором отображаются вопросы пользователя и текстовые ответы голосового ассистента. Голосовые ответы ассистента настроены для воспроизведения женским голосом.

Заключение

Исходя из проведенного анализа технологий распознавания естественного языка, можно сделать вывод, что для разработки прикладных голосовых ассистентов удобно использовать сервис распознавания естественного языка от Google DialogFlow, который функционирует на основе нейронных сетей с конструкцией self-attention с использованием технологии Word2Vec. В ходе исследования определен процесс разработки голосового ассистента «IT-абитуриент», выявлены возникшие проблемы и методы их решения. Голосовой помощник «IT-абитуриент» используется во время приемной кампании для помощи абитуриентам в выборе IT-специальностей Карагандинского университета им. академика Е.А. Букетова.

СПИСОК ЛИТЕРАТУРЫ

1. Поначугин А.В., Пичужкина Д.Ю., Смекалова Е.С. Голосовой помощник как технология обработки данных // Наука без границ. – 2020. – № 6 (46) – С. 96-100. [Электронный ресурс]. URL: https://cyberleninka.ru/article/n/golosovoy-pomoschnik-kak-tehnologiya-obrabotki-dannyh
2. Артозей Е.А. Разработка алгоритмов идентификации голосовых команд для управления бортовой системой автомобиля // Современная техника и технологии, 2015. – № 9. [Электронный ресурс]. URL: http://technology.snauka.ru/2015/09/7903.
3. Яворский В.В. Оптимизация и математические методы принятия решений: учеб. пособие / В.В. Яворский; Федеральное агентство по образованию, Томский гос. ун-т систем упр. и радиоэлектроники. – Томск: Том. гос. ун-т систем упр. и радиоэлектроники, 2006. – 215 с.
4. Богомолов Ю.А. Обзор моделей нейронных сетей для обработки естественного языка // Научно-образовательный журнал для студентов и преподавателей «StudNet» № 4 / 2020. [Электронный ресурс]. URL: https://cyberleninka.ru/article/n/obzor-modeley-neyronnyh-setey-dlya-obrabotki-estestvennogo-yazyka.
5. Transformer: A Novel Neural Network Architecture for Language Understanding. [Электронный ресурс]. URL: https://ai.googleblog.com/2017/08/transformer-novel-neural-network.html
6. Харламов А.А., Гордеев Д.И. Дистрибутивная vs сетевая семантика в диалоговых системах // Проблемы искусственного интеллекта. – 2019. – № 2 (13). – С. 93-126. [Электронный ресурс]. URL: https://cyberleninka.ru/article/n/distributivnaya-vs-setevaya-semantika-v-dialogovyh-sistemah
7. Хлопенкова А.Ю., Белов Ю.С. Методы обработки естественного языка в виртуальных голосовых помощниках // E-Scio. – 2019. [Электронный ресурс]. URL: https://cyberleninka.ru/article/n/metody-obrabotki-estestvennogo-yazyka-v-virtualnyh-golosovyh-pomoschnikah
8. Бобровский И.А., Булатов А.Д., Суслов И.Ю. Развитие интерфейсов для взаимодействия с ЭВМ // Наука. Технологии. Инновации // Сборник научных трудов в 9 ч. / Под ред. Гадюкиной А.В. – Новосибирск: Изд-во НГТУ, 2020. – С. 97-101.
9. Dialogflow: описание. [Электронный ресурс]. URL: https://roi4cio.com/catalog/product/dialogflow
10. Что умеет Dialogflow? [Электронный ресурс]. URL: https://habr.com/ru/post/502688/

Дауыстық ассистентті жасауға арналған табиғи тілді тану технологиялары

[1]***СПИРИНА Елена Александровна,** п.ғ.к., қаумдастырылған профессор, sea_spirina@mail.ru,*
[1]***ЖҰМАБЕКОВ Әділхан Ералыұлы,** магистрант, sk1llahh2@gmail.com,*
[1]***ГОРБУНОВА Надежда Александровна,** п.ғ.к., профессордың ассистенті, ant_nadezhda@mail.ru,*
[1]***КАЗИМОВА Динара Ашубасаровна,** п.ғ.к., профессор, dinkaz73@mail.ru,*
[2]***СЕЛЕЗНЕВА Руслана,** т.ғ.к., профессор, seleznevaruslana@gmail.com,*
[1]*«Е.А. Бөкетов атындағы Қарағанды университеті» КеАҚ, Қазақстан, Қарағанды, Университет көшесі, 28,*
[2]*Хуманитас жоғары мектебі, Польша, Сосновец, 41-200, Ян Килиньский көшесі, 43,*

**автор-корреспондент.*

***Аңдатпа.** Жұмыста табиғи тілді танудың негізгі технологиялары және DialogFlow бұлтты қызметі негізінде дауыстық ассистентті (көмекшіні) құру ерекшеліктері қарастырылады. Мысал ретінде академик Е.А. Бөкетов атындағы Қарағанды университетінің IT саласындағы білім беру бағдарламалары туралы талапкерлердің сұрақтарына жауап беретін «IT-талапкер» деп аталатын дауыстық көмекші қосымшасы әзірленді. Мақалада танымал Google Assistant, Amazon Echo (Alexa), Microsoft Cortana, Yandex Alisa сияқты дауыстық ассистенттерінің жұмыс принциптері талданған. Google DialogFlow табиғи тілді тану платформасы мобильді қосымшаларға диалогтық қолданушы интерфейсін жасау және біріктіру үшін үлкен мүмкіндіктер беретіні анықталды. DialogFlow бұлтты қызметі негізінде дауыстық ассистентті әзірлеу алгоритмі қарастырылған.*

***Кілт сөздер:** жасанды интеллект, Web-қосымша, дауыстық ассистент, дауыстық көмекші, Google Alisa, DialogFlow бұлтты қызметі, дауыстық көмекшіні дамыту алгоритмі.*

Natural Language Recognition Technologies for the Development of a Voice Assistant

[1]***SPIRINA Yelena,** Cand. of Ped. Sci., Associate Professor, sea_spirina@mail.ru,*
[1]***ZHUMABEKOV Adilkhan,** master student, sk1llahh2@gmail.com,*
[1]***GORBUNOVA Nadezhda,** Cand. of Ped. Sci., Assistant Professor, ant_nadezhda@mail.ru,*
[1]***KAZIMOVA Dinara,** Cand. of Ped. Sci., Professor, dinkaz73@mail.ru,*
[2]***SELEZNEVA Ruslana,** Cand. of Tech. Sci., Professor, seleznevaruslana@gmail.com,*
[1]*NCJSC «Karaganda Buketov University», Kazakhstan, Karaganda, University Street, 28,*
[2]*Wyższa Szkoła Humanitas, Poland, Sosnowiec, 41-200, Jan Kiliński Street, 43,*

**corresponding author.*

***Abstract.** The paper discusses the basic technologies of natural language recognition and the features of creating a voice assistant based on the DialogFlow cloud service. As an example, an application has been developed a voice assistant*

«IT-entrant», which answers entrants' questions about educational programs in the field of IT at Karaganda University named after E.A. Buketov. The article analyzes the principles of the popular voice assistants such as Google Assistant, Amazon Echo (Alexa), Microsoft Cortana, Yandex Alisa. It is revealed that the natural language recognition platform Google DialogFlow is a convenient tool for developing a dialog user interface in mobile applications. The algorithm for developing a voice assistant based on the DialogFlow cloud service is considered.

Keywords: *artificial intelligence, natural language recognition, Web application, voice assistant, Google Alisa, DialogFlow cloud service, voice assistant development algorithm.*

REFERENCES

1. Ponachugin A.V., Pichuzhkina D.Yu., Smekalova E.S. Golosovoy pomoschnik kak tehnologiya obrabotki dannyh // Nauka bez granits. – 2020. – No. 6 (46) – pp. 96-100. [Elektronniy resurs]. URL: https://cyberleninka.ru/article/n/golosovoy-pomoschnik-kak-tehnologiya-obrabotki-dannyh
2. Artozey E.A. Razrabotka algoritmov identifikatsiya golosovikh komand dlya upravleniya bortovoy sistemoy avtomobilya // Sovremennaya tekhnika i tekhnologii, 2015. – No. 9. [Elektronniy resurs]. URL: http://technology.snauka.ru/2015/09/7903.
3. Yavorskiy V.V. Optimizatsya i matematicheckie metody prinyatiya resheniy: ucheb. posobie / Yavorskiy V.V; Federalnoe agentstvo po obrazovaniyu, Tomskiy gos. un-t sistem upr. i radioelektroniki, 2006. – 215 p.
4. Bogomolov Yu.A. Obzor modeley neyronnih setey dlya obrabotki estestvennogo yazyka // Nauchno-obrazovatelnyi zhyrnal dlya studentov i prepodavateley «StudNet» no. 4 / 2020. [Elektronniy resurs]. URL: https://cyberleninka.ru/article/n/obzor-modeley-neyronnyh-setey-dlya-obrabotki-estestvennogo-yazyka.
5. Transformer: A Novel Neural Network Architecture for Language Understanding. – [Elektronniy resurs]. URL: https://ai.googleblog.com/2017/08/transformer-novel-neural-network.html
6. Kharlamov A.A., Gordeev D.I. Distributivnaya vs setevaya semantika v dialogovyh sistemah // Problemy iskusstvennogo intellekta. – 2019. – No. 2 (13). – pp. 93-126. [Elektronniy resurs]. URL: https://cyberleninka.ru/article/n/distributivnaya-vs-setevaya-semantika-v-dialogovyh-sistemah
7. Khlopenkova A.Yu., Belov Yu.S. Metody obrabotki estestvennogo yazyka v virtualnyh golosovyh pomoschnikah // E-Scio. – 2019. [Elektronniy resurs]. URL: https://cyberleninka.ru/article/n/metody-obrabotki-estestvennogo-yazyka-v-virtualnyh-golosovyh-pomoschnikah
8. Bobrovskiy I.A., Bulatov A.D., Suslov I.Yu. Razvitie interfeysov dlya vzaimodeystviya s EVM // Nauka. Technologii. Innovatsii // Sbornik nauchnykh trudov v 9 ch. / Pod red. Gadyukinoy A.V. – Novosibirsk: Publ. NGTU, 2020. – pp. 97-101.
9. Dialogflow: opisanie. [Elektronniy resurs]. URL: https://roi4cio.com/catalog/product/dialogflow
10. Chto umeet Dialogflow? [Elektronniy resurs]. URL: https://habr.com/ru/post/502688/