

УДК 004

РАЗРАБОТКА ДИАЛОГОВОГО ИНТЕРФЕЙСА ДЛЯ ANDROID – ПРИЛОЖЕНИЯ STUDY PLANNER С ПОМОЩЬЮ NLU-ПЛАТФОРМЫ API.AI

Вагайцев А.В., студент гр. ИТм-151, 2 курс

Научный руководитель: Ванеев О.Н., к.т.н., доцент

Кузбасский государственный технический университет имени Т.Ф. Горбачева
г. Кемерово

Множество людей ежедневно взаимодействуют с различными программными системами. В результате внедрения компьютерных систем во все сферы человеческой жизни все больше проявляется необходимость перехода от визуальных и командных интерфейсов к естественно-языковым. Использование привычного для пользователя естественного языка для организации диалога с компьютерными системами реализуется средствами естественно-языкового и речевого пользовательского интерфейса.

Программы, способные общаться с человеком на естественном языке уже активно используются в медицинской отрасли, беспилотных автомобилях, в виртуальных цифровых помощниках, и других областях. Мы становимся свидетелями того, как компьютеры постепенно начинают общаться с нами на естественном языке в различных сферах общественной жизни.

По мнению многих экспертов, сегодня наиболее перспективным направлением применения технологий осмысленного анализа естественного языка являются помощники, консультирующие пользователя по различным вопросам и выполняющие различные задания. Достаточно известными сейчас являются голосовые помощники Apple Siri, Microsoft Cortana и Google Now [41].

Во времена, где все вбивается и печатается, голосовой помощник может стать не только новым способом ввода данных, но и новым способом работы с информацией. Многие разработчики смогут переделать свои продукты, чтобы пользователи общались голосом. Безусловно, это совершенно новый способ взаимодействия, который применим к новым задачам. В ближайшем будущем голосовые ассистенты смогут в полной мере удовлетворять всем потребительским запросам пользователей и станут настоящими посредниками между клиентами и представителями бизнеса.

Согласно прогнозам, к 2020 году большая часть поисковых запросов будет создаваться голосом, а не текстом. По данным Kleiner Perkins Caufield & Byers, более 25% поисковых сессий пользователей в панели Windows 10 осуществлялось при помощи голосового взаимодействия с интерфейсом. 40% пользователей начали пользоваться мобильными ассистентами в течение последних 6 месяцев, что демонстрирует резкий рост в популярности и востре-

бованности. 20% запросов, получаемых в Google, голосовые. Чаще всего голосовым поиском пользуются подростки. [5]

Столь ощутимый рост популярности голосовых помощников можно объяснить быстрым развитием технологий. Например, в 2013 году степень точности распознавания голосовых запросов для поисковой платформы Google была ниже 80%. Всего за 2 года она достигла 90%. Машинное обучение продвинулось резко вперед за счет “глубинного обучения” — одной из форм адаптивных искусственных нейронных сетей, обученных методом “обратного распространения ошибки”. Этот скачок в производительности алгоритмов обработки информации сопровождался значительным прогрессом в аппаратной технологии для основных операций, таких как зондирование, восприятие и распознавание объектов. Новые платформы и рынки для продуктов, управляемых данными, а также экономические стимулы, способствовали к появлению технологий с ИИ.

В последнее время появилось большое количество сервисов, с помощью которых можно разрабатывать диалоговые интерфейсы для своих собственных приложений, например, “API.AI”, “WIT.AI”, “Microsoft Bot Framework”. Такие сервисы способны не только «понимать» естественную речь, но и давать пользователю осмысленные ответы. С учетом возможности обучения подобных сервисов на выборке данных открываются огромные возможности для создания «умных» виртуальных помощников. Именно поэтому построение диалоговых интерфейсов является актуальной задачей.

В рамках исследовательской работы разрабатывается голосовой естественно-языковой пользовательский интерфейс для android – приложения Study Planner. С помощью android-приложения “Study Planner” пользователи могут планировать свое учебное расписание занятий, записывать домашние задания и ставить напоминания о каких-либо важных событиях. С помощью голосового интерфейса можно будет выполнить большинство задач, не прикасаясь к телефону – достаточно будет произнести определенную фразу, и ассистент даст нужный ответ. Такой способ взаимодействия очень часто является более естественным и удобным, чем непосредственное взаимодействие с приложением с помощью графического интерфейса.

Для решения задач будут применяться следующие технологии:

1. Android Studio 1.5

Android Studio — это бесплатная среда разработки, предоставляющая интегрированные инструменты для разработки и отладки приложений для ОС Android. Android Studio включает в себя: Android SDK; инструменты для разработки дизайна, тестирования и отладки; последнюю версию платформы Android для компиляции; последнюю версию образа Android для запуска ваших приложений.

2. Google Speech Recognition API

Технология распознавания речи интегрирована в мобильные телефоны и компьютеры, где можно ввести информацию с помощью голоса. Сам процесс распознавания речи осуществляется на серверах Google. В Android-

устройствах версии 4.1 и выше позволяет выполнить распознавание речи offline.

3. Синтезатор речи Google TTS (Text To Speech)

Google TTS – бесплатный синтезатор речи, доступный на всех Android-устройствах версии 4.0 и выше

4. CMU Pocketsphinx

CMU Sphinx сейчас является крупнейшим проектом по распознаванию человеческой речи. В данной работе, **Pocketsphinx** используется для голосовой активации. С помощью голосовой активации можно запустить определенные функции внутри приложения, произнеся определенную команду. Это система распознавания речи, которая работает локально на устройстве без использования сети интернет и умеет распознавать одну или несколько активационных команд во входящем потоке речи.

Архитектура естественно-языкового интерфейса

Разрабатываемый голосовой интерфейс будет выполнять простую и рутинную работу за пользователя, например, добавлять новое занятие в расписание, записывать домашнее задание на следующий урок, отвечать на вопросы, связанные с учебой, например, «Какие завтра занятия?». При этом взаимодействие должно быть простым, интуитивно понятным, а диалог должен быть естественным. Это позволит пользователю достаточно быстро обучиться основным правилам и принципам работы с интерфейсом, а также в диалоге с ним использовать естественную речь. Эти задачи решает естественно-языковой интерфейс.

Естественно-языковой интерфейс состоит из нескольких компонентов:

1. **Диалоговый компонент.** В его задачи входит непосредственное взаимодействие с пользователем, ведение диалога.
2. **Компонент понимания высказываний.** Данный компонент предназначен для семантического анализа и интерпретации высказывания. На этапе анализа выделяются описания сущностей, их свойства и отношения между ними. В задачи интерпретации входит построение списка команд из исходных гипотез.
3. **Компонент генерации высказываний.** Компонент решает в соответствии с результатами, полученными остальными компонентами системы, одну основную задачу: синтез высказывания, т.е. преобразование смысла в высказывание на естественном языке.
4. **База знаний.** Содержит весь используемый лексикон, команды, описание сущностей, их свойств и отношений между ними.
5. **Модуль обучения.** Данный компонент будет автоматически наполнять базу знаний на основе запросов пользователя. Это могут быть как новые правила, так и уточнения к ним.

Архитектура голосового интерфейса

В задачи голосового интерфейса входит синтез ответа голосом и распознавание речи. Поэтому выделяется:

1. **Компонент синтеза речи.** На вход данному компоненту подается текстовая строка, а на выходе ответ, синтезированный с помощью голосового движка.
2. **Компонент распознавания речи.** Преобразует исходный сигнал (речь) в текст.

Взаимодействие пользователя с интерфейсом

Для того, чтобы начать диалог с интерфейсом, пользователь должен выполнить несколько простых действий. Во т примерная схема взаимодействия пользователя с приложением с помощью голосового интерфейса:

- Произнести определенную активационную фразу, которая запустит распознавание речи. Например, это может быть «Окей, ассистент».
- Далее произнести команду, например, «Какие завтра занятия?»
- Модуль распознавания речи преобразует исходный сигнал в текст и передаст его в модуль семантического и синтаксического анализа.
- Модуль анализа разберет исходную фразу и выделит сущности, их описание и свойства, а также отношения между ними. Например, у нас будет сущность «Занятие» и «Дата»
- Далее модуль интерпретации сопоставит синтаксическую фразу с формальной и найдет соответствующие правила. Это будет правило Завтра(Занятие) равное «Найти(\$Занятие, \$Дата)»
- После этого передается запрос внутрь приложения для поиска занятий на завтра. Для этого будет выполнен SQL запрос к БД.
- Модуль генерации высказываний принимает результат выполненного SQL запроса. Если данных нет, то модуль генерирует высказывание «Завтра занятий нет, можете отдохнуть». Если данные есть, то модуль генерирует высказывание «У вас завтра \$количество\$ занятий. Показать какие?».

При этом если пользователь задаст вопрос, на который интерфейс не сможет ответить, то пользователю будет предложено уточнить вопрос, и если на него будет дан положительный ответ, то данная цепочка вопросов будет записана в базу знаний. Например, пользователь может спросить «Завтра будет занятие по Иванову?». Интерфейс спросит «Что значит Иванову». Пользователь скажет «Это преподаватель». Тогда интерфейс применит правило «Найти(Найти(\$Занятие, Преподаватель(\$Фамилия)), \$Дата).

Благодаря такому подходу, отпадает необходимость описывать дословно все возможные фразы, достаточно описать правила для выполнения команд и разработать механизм анализа и интерпретации высказываний. А если сделать базу знаний общей для всех пользователей, то интерфейс быстрее обучиться и будет способен дать ответы на многие вопросы.

Используя описанные выше технологии и архитектуры интерфейсов, будет разработан голосовой естественно-языковой интерфейс, который упростит работу пользователя с приложением.

Список литературы:

1. Практика для студентов. Кафедра компьютерных систем и сетей, Хмельницкий национальный университет. Мясищев А.А. [Электронный ресурс] – Режим доступа: http://khnu.km.ua/root/kaf/ksm/my_syte_g/
2. А.В. Фролов, Г.В. Фролов, Синтез и распознавание речи. Современные решения, "Москва" 2008г
3. И.С. Евдокимова. Естественнo-языковые системы: курс лекций. - Улан-Удэ: Изд-во ВСГГУ, 2006. - 92 с.: илл.
4. "Voice User Interface Design" Michael H. Cohen, James P. Giangola, Jennifer Balogh 2004. — 336 с:
5. Официальный сайт CMU Pocketsphinx [Электронный ресурс]. – Режим доступа: <http://cmusphinx.sourceforge.net/>
6. Материалы с сайта Википедия. [Электронный ресурс]. URL: <http://ru.wikipedia.org/>