

УДК 657

РАЗРАБОТКА ГОЛОСОВОГО ЕСТЕСТВЕННО-ЯЗЫКОВОГО ПОЛЬЗОВАТЕЛЬСКОГО ИНТЕРФЕЙСА С ПРИМЕНЕНИЕМ ТЕХНОЛОГИЙ СИНТЕЗА И РАСПОЗНАВАНИЯ РЕЧИ ДЛЯ ВЗАИМОДЕЙСТВИЯ С МОБИЛЬНЫМ ANDROID-ПРИЛОЖЕНИЕМ «STUDY PLANNER».

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

Научный руководитель: О.Н. Ванеев, Доцент (к.н.)

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

Актуальность

Сегодня много кто решает повседневные задачи на ходу — с телефона. С его помощью можно проверить почту, отправить документы и фотографии, найти ближайший банкомат или построить автомобильный маршрут. Не для всех подобных задач удобно пользоваться клавиатурой, поэтому сейчас одно из самых актуальных направлений мобильной разработки — это управление голосом.

Использование привычного для пользователя естественного языка для организации диалога с компьютерными системами реализуется средствами естественно-языкового и речевого пользовательского интерфейса, обладающего рядом преимуществ: минимальной подготовкой пользователя для работы с системой, простотой и высокой скоростью задания произвольных запросов к системе. Использование речевого синтеза при этом позволяет снизить нагрузку на пользователя по восприятию результатов обработки, используя не графический интерфейс (зрительную систему), а посредством речи, а так же являясь, при использовании синтеза речи, более привычной формой человеческого взаимодействия — диалогом.

Голосовой интерфейс качественным образом изменяет способ, а следовательно и эффективность взаимодействия пользователя с системой. Голосовой поиск от компании Google и голосовой ассистент Siri от компании Apple являются этому яркими примерами, подтверждая насущную необходимость внедрения речевых технологий, в частности распознавания речи и голосовых интерфейсов.

С помощью мобильного Android-приложения “Study Planner” пользователи могут планировать свое учебное расписание занятий, записывать домашние задания и ставить напоминания о каких-либо важных событиях. Вся информация вводится вручную, а для того чтобы узнать, например, «какие занятия будут завтра», необходимо открывать приложение.

Исходя из этого, было решено разработать голосовой естественно-языковой интерфейс для взаимодействия с приложением “Study Planner”.

Цели разработки:

- Обеспечить голосовой ввод информации и выполнение функций внутри приложения с помощью голосового управления.
- Обеспечить голосовой вывод информации
- Ускорить процесс получения информации
- Упростить взаимодействие с мобильным приложением Study Planner”.

Задачи:

- Построить модель естественно-языкового интерфейса
- Разработать голосовой естественно-языковой пользовательский интерфейс

Используемые технологии

Для решения задач будут применяться следующие технологии:
Google Speech

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

3. 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. Вот 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) равно «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 обучится и будет способен дать ответы на многие вопросы.

Подведение итогов

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. Естественно-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/>