

Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 621.395.44:004.93

Полозок
Диана Сергеевна

Кроссплатформенное мобильное приложение для обмена информацией с функцией распознавания голоса

АВТОРЕФЕРАТ

на соискание степени магистра информатики и вычислительной техники по специальности 1-40 80 04 «Математическое моделирование, численные методы и комплексы программ»

_____ Д.С.Полозок
Научный руководитель
Ганжа Виктор Александрович
кандидат физико-математических наук, доцент
_____ В.А.Ганжа

Минск 2015

ВВЕДЕНИЕ

Разработка мобильных приложений заняла важное место в IT сфере. Каждая компания стремится создать мобильную версию своего сервиса. В некоторых сегментах рынка мобильное продвижение – это, фактически, обязательное условие для существования и развития бизнеса. Мобильные приложения несут в себе определённый эмоциональный заряд. Большинству людей нравится пользоваться такими современными программами. Мобильное приложение – это современный высокоэффективный бизнес-инструмент, с помощью которого оперативно доносится коммерческое предложение для пользователя. Стоимость разработки мобильных приложений зависит от сложности и сроков изготовления.

Основная сложность в разработке мобильных приложений заключается в многообразии операционных систем. Для каждой мобильной платформы требуется знание разных технологий и языков программирования, за счет чего увеличивается стоимость и время разработки. Однако в феврале 2013 года американская компания Xamarin выпустила продукт, позволяющий разрабатывать кроссплатформенные мобильные приложения. С использованием их продукта для разработки мобильных приложений для основных ОС будет достаточно специалистов со знаниями одного языка программирования, что существенно сокращает стоимость и время разработки.

Речевое общение является естественным и удобным для человека. Задача распознавания речи состоит в том, что бы убрать посредника в общении человека и компьютера. Управление машиной голосом в реальном времени, а также ввод информации посредством человеческой речи намного упростит жизнь современного человека. Научить машину понимать без посредника тот язык, на котором говорят между собой люди – задачи распознавания речи.

Ученые и инженеры уже много лет решают проблему речевого общения человека и машины. Первое устройство для распознавания речи появилось в 1952 году, оно могло распознавать произнесённые человеком цифры. Коммерческие программы по распознаванию речи появились в начале девяностых годов. Стоит отметить, что распознавание речи – очень сложная и актуальная задача, которой в настоящее время занимаются специалисты крупнейших мировых компаний.

ОБЩАЯ ХАРАКТЕРИСТИКА

Основной целью данной магистерской диссертации является написание кроссплатформенного мобильного приложения для обмена информацией (текстовой и графической), которое будет работать на операционных системах iOS, Android и Windows Phone. Кроссплатформенность приложения позволит снизить стоимость и время разработки. Дополнительными возможностями являются возможность отправки медиа-сообщений (картинок), текущих GPS-координат пользователя, а также создание групповых чатов с другими пользователями. В рамках данной работы также рассмотрены алгоритмы распознавания голоса и их использование в различных системах. В приложение добавлена функция распознавания голоса, на основе одного из простейших алгоритмов.

В первой главе рассмотрена предметная область по теме магистерской диссертации, проанализирована роль мессенджеров в современной жизни человека, а также рассмотрено использование технологий распознавания речи.

Для разработки кроссплатформенного приложения обмена информацией потребовалось изучение целого ряда информационных технологий. Во второй главе подробно описаны использованные технологии, рассмотрены их возможности, подробно оценены плюсы и минусы. Также описан язык программирования и платформа, при помощи которых было реализовано мобильное приложение.

Третья и четвертая главы посвящены проектированию разработанного приложения. В них описана архитектура приложения, модуль графического интерфейса, модуль бизнес-логики, модуль доступа к данным. Также рассмотрены использованные шаблоны проектирования. Согласно основным принципам проектирования мобильных приложений, разработанное приложение делится на 3 модуля. В данном контексте модуль – это полноценное структурное составляющие приложения, которое характеризуется низкой внешней зависимостью при высокой внутренней связанности. В четвертой главе подробно описаны каждый из модулей, а также приведены примеры использованных алгоритмов.

В пятой главе описаны подробности реализации мобильного приложения. Также в ней можно найти различные диаграммы, при помощи которых можно понять функциональное назначение системы. Для описания того, что будет делать система в процессе своего функционирования, приведен пример диаграммы

вариантов использования. Диаграмма вариантов использования является исходным концептуальным представлением или концептуальной моделью системы в процессе ее проектирования и разработки.

Одной из целей магистерской диссертации являлось изучение существующих алгоритмов распознавания речи, а также рассмотрение вопроса, есть ли возможность сделать свою систему распознавания речи, пусть даже на довольно ограниченном по размеру словаре, без использования «внешних» сервисов, при этом, чтобы она работала быстро и с приемлемым качеством. В шестой главе подробно описан алгоритм динамического трансформирования времени (DTW), реализующего распознавание речи. DTW алгоритмы являются очень полезными для распознавания отдельных слов в ограниченном словаре, однако их использование для распознавания беглой речи не представляется возможным.

В рамках данной работы реализован простой алгоритм системы распознавания речи по короткому словарю, основанный на анализе статистического распределения мел-кепстральных коэффициентов (Mel-frequency cepstrum coefficients, MFCC). Его описание, подробности реализации и результаты тестирования приведены в главе номер семь.

В реализованном приложении основной акцент был сделан на функционал приложения, поэтому в качестве дизайна был реализован минималистический интерфейс, не перегруженный излишними деталями. В восьмой главе описан пользовательский интерфейс приложения, приведены примеры скриншотов, а также даны пользовательские инструкции по его использованию.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Мобильное приложение для обмена текстовой и графической информацией было реализовано при помощи фреймворка Xamarin.

Xamarin — это фреймворк для кроссплатформенной разработки мобильных приложений (iOS, Android, Windows Phone) с использованием языка C#. Его идея заключается в том, программный код реализуется на языке C#, с применением всех возможностей данного языка. При этом имеется полный доступ ко всем возможностям SDK платформы и родному механизму создания UI, получая на выходе приложение, которое, строго говоря, ничем не отличается от приложения, разработанного на языке, предназначенном для определенной платформы, и не уступает им в производительности. Xamarin основан на open-source реализации платформы .NET — Mono. Эта реализация включает в себя собственный компилятор C#, среду выполнения, а так же основные .NET библиотеки. Цель проекта — позволить запускать программы, написанные на C#, на операционных системах, отличных от Windows — Unix-системах, Mac OS и других.

В процессе разработки программного обеспечения были использованы шаблоны проектирования (паттерны). Паттерн – повторяемая архитектурная конструкция, представляющая собой решение проблемы проектирования в рамках некоторого часто возникающего контекста.

Проектирование приложения основано на известном подходе «проектирования на основе предметной области» (DDD: Domain Driven Design). Данный подход подразумевает использование:

- Объектно-ориентированного проектирования
- Шаблоны проектирования.
- Разработку по средствам тестирования.

Основываясь на данном подходе, была определена трехуровневая модель приложения.

Согласно данной модели приложение делится на 3 модуля:

1. Модуль графического интерфейса, который будет отображаться на устройстве и взаимодействовать с пользователем.
2. Модуль бизнес-логики, которая представляет объектную модель системы и осуществляет основные функциональные обязанности системы.
3. Модуль доступа к данным, который реализует доступ к различным данным, независимо от того хранится в локальной базе данных или поставляются от удалённых сетевых ресурсов.

Преимущества такой модели:

- Масштабируемость.
- Конфигурируемость – изолированность уровней друг от друга, позволяет (при правильном развертывании архитектуры) быстро и простыми средствами переконфигурировать систему при возникновении сбоев или при плановом обслуживании на одном из уровней.

В качестве концепции разработки части графического интерфейса была выбрана концепция MVVM (model-view-viewmodel). Основная идея этой концепции состоит в разделении бизнес-логики (модели) от её визуализации (представления, вида).

В данную модель приложения был заложен паттерн «Мост». Для чего интерфейс каждого модуля был определён и в модуле более высокого уровня. Таким образом, в реализации классов использовалась не конкретная реализация подмодуля, а интерфейс. Плюсы данного подхода:

- Позволяет отделить реализацию от интерфейса. Тем самым реализацию абстракции можно конфигурировать во время выполнения, и динамически изменять реализацию.

- Облегчает разбиение системы на слои и тем самым позволяет улучшить её структуру.

- Скрывает детали реализации используемого модуля от модуля его использующего.

Для реализации доступа из модуля более высокого уровня к модулям более низкого уровня без обращения к конкретной реализации (работая только с интерфейсом), был спроектирован и реализован собственный контейнер ServiceLocator.

В разработанном мобильном приложении был реализован простейший алгоритм распознавания речи.

В системах распознавания речи, содержащих слова, распознавание требует сравнения между входным словом и различными словами в словаре. Эффективное решение проблемы лежит в динамических алгоритмах сравнения, целью которого является введение временных масштабов двух слов в оптимальное соответствие. Алгоритмы такого типа являются динамическими алгоритмами трансформации временной шкалы. Таким вот образом механические колебания превращаются в набор чисел, пригодный для обработки на современных ЭВМ.

Отсюда следует, что задача распознавания речи сводится к «сопоставлению» множества численных значений (цифрового сигнала) и слов из некоторого словаря (русского языка, например).

Сегодня системы распознавания звука строятся на основе принципов признания форм распознавания. Методы и алгоритмы, которые использовались до сих пор, могут быть разделены на четыре больших класса:

Методы дискриминантного анализа, основанные на Байесовской дискриминации;

Скрытые модели Маркова;

Динамическое программирование – временные динамические алгоритмы (DTW);

Нейронные сети;

В данной работе приводится пример и альтернатива алгоритму динамического программирования DTW, реализующего распознавание речи. В приложении был реализован простой алгоритм системы распознавания речи по короткому словарю, основанный на анализе статистического распределения мел-кепстральных коэффициентов (Mel-frequency cepstrum coefficients, MFCC). Основная идея заключается в использовании линейного дискриминантного анализа для идентификации слова. Т.к. слова могут быть различной длины, возникает вопрос: каким образом преобразовать последовательность произвольного числа MFCC-векторов в вектор фиксированной размерности. Было решено находить места «сгущения» распределения этих векторов и в качестве результирующего вектора брать конкатенацию векторов, являющихся центрами «сгущений».

В качестве эксперимента была создана система, которая умеет распознавать 14 слов, записанных одним голосом. Для обучения каждое слово было записано 4-5 раз, а для тестирования — 7 раз. Итого база обучения содержит 63 файла, а база тестирования — 98.

Результат тестирования на базе обучения показал уровень ошибки распознавания слов (WER) 1,6%, а на базе тестирования 5,1%. Для того, чтобы любая система (включая описанную здесь) могла качественно распознавать речь любого человека, необходимо иметь огромную базу обучения с записью всех слов, произнесенных разными людьми в разном эмоциональном состоянии с использованием различных записывающих устройств (телефон, микрофон, подслушивающее устройство и т.п.).

ЗАКЛЮЧЕНИЕ

Во время работы над проектом были изучены способы построения сложных систем передачи данных, методы оптимизации работы с базой данных, способы организации взаимодействий между модулями сложного приложения.

Созданное приложение позволяет пользователям обмениваться текстовой и графической информацией, создавать групповые разговоры, пересылать координаты своего местонахождения. Главное достоинство данного приложения заключается в том, что оно кроссплатформенное, и написанное единожды, оно может работать на принципиально различных мобильных платформах, таких как iPhone, Android и Windows Phone. Это существенно сокращает стоимость и время разработки. В логику кроссплатформенного приложения можно без труда вносить любые изменения, которые будут отображаться одновременно для всех платформ. В процессе разработки приложения были подробно изучены современные возможности для распознавания голоса. Также был реализован простой алгоритм системы распознавания речи по короткому словарю, основанный на анализе статистического распределения мел-кепстральных коэффициентов. Реализованная система распознавания речи работает на довольно ограниченном по размеру словаре, однако, без использования «внешних» сервисов, при этом быстро и с приемлемым качеством.

Применение всех возможностей фреймворка Xamarin и платформы .NET, а также соблюдение основных правил эргономики с учетом особенностей и требований для мобильных устройств, позволило создать решение на базе многоуровневой системы с минимальными зависимостями между уровнями, максимально ускорив быстродействие приложения и уменьшив потребление энергии во время работы приложения.

Результатом проекта является приложение с богатым функционалом, удобным для пользователя интерфейсом, а так же легкое в сопровождении и понимании.