

УДК 004.42

**ОБОСНОВАНИЕ АКТУАЛЬНОСТИ ПРОЕКТИРОВАНИЯ
ИНФОРМАЦИОННО-АНАЛИТИЧЕСКОЙ СИСТЕМЫ
УПРАВЛЕНИЯ ГОЛОСОМ В МЕДИАПРОЕКТАХ****¹Черепанов М.Д., ¹Безруких А.Д., ²Мельников В.А.,****²Пономарева А.О., ³Безруких Ю.А.**¹ФГАОУ ВО «Национальный исследовательский институт ИТМО», Санкт-Петербург;²ФГАОУ ВО «Сибирский федеральный университет», Красноярск;³ФГБОУ ВО «Сибирский государственный университет науки и технологий им. М.Ф. Решетнева»,
Красноярск, e-mail: expert-sib@yandex.ru

В статье представлено обоснование актуальности использования и проектирования приложения для управления файлами при помощи голоса с целью повышения эффективности проведения мероприятий компаний. Предлагаемое авторами решение основано на применении технологий нейронных сетей для голосового управления аудиофайлами при реализации проекта. Данная функция является преимуществом перед готовыми решениями, которые не ориентированы на ее реализацию, а также они не имеют возможности изменения конфигураций файлов. При этом существующие аналоги имеют высокую стоимость, что снижает их привлекательность при использовании в медиапроектах компаний. В работе были выявлены существующие на данный момент проблемы в этой области, а также проблемы, связанные с человеческим фактором, которые возникают в работе звукооператоров. Для более детального анализа и постановки проблемы был проведен опрос среди ведущих мероприятий, звукооператоров и радиоведущих. Результаты данного опроса помогли выявить проблемы, с которыми сталкиваются ведущие при работе со специализированным программным обеспечением, а также позволили определить необходимый функционал программного прототипа. Так, проектируемое приложение позволит избежать ошибок, которые возникают в результате искажения информации человеком при одновременном выполнении нескольких задач в проекте.

Ключевые слова: маркетинговый процесс, медиапроект, цифровой помощник, нейронная сеть, автоматизация процесса, распознавание речи

**SUBSTANTIATION OF THE RELEVANCE OF DESIGNING
AN INFORMATION AND ANALYTICAL VOICE
CONTROL SYSTEM IN MEDIA PROJECTS****¹Cherepanov M.D., ¹Bezrukikh A.D., ²Melnikov V.A.,****²Ponomareva A.O., ³Bezrukikh Y.A.**¹ITMO University, St. Petersburg;²Siberian Federal University, Krasnoyarsk;³Reshetnev Siberian State University of Science and Technology, Krasnoyarsk,
e-mail: expert-sib@yandex.ru

The article presents the rationale for the relevance of using and designing an application for file management using voice in order to increase the efficiency of media events of companies. The solution proposed by the authors is based on the use of neural network technologies for voice control of audio files during the implementation of the project. This function is an advantage over ready-made solutions that are not focused on its implementation, and they also do not have the ability to change file configurations. At the same time, existing analogues are expensive in cost, which reduces their attractiveness when used in media projects of companies. The work revealed the current problems in this area, as well as problems related to the human factor that arise in the work of sound engineers. For a more detailed analysis and formulation of the problem, a survey was conducted among the leading media events, sound engineers and radio hosts. The results of this survey helped to identify the problems faced by the presenters when working with specialized software, and also helped to determine the necessary functionality of the software prototype. So, the designed application will help to avoid errors that arise as a result of distortion of information by a person while simultaneously performing several tasks in the project.

Keywords: marketing process, media project, digital assistant, neural network, process automation, speech recognition

В современном бизнесе актуальной задачей является повышение эффективности современной организации за счет использования цифровых технологий. Так, они позволяют значительно сократить время реализации бизнес-задач, расстояние между потребителем и поставщиком товаров (услуг, работ), повысить привлекательность продукта за счет грамотного маркетинга

и работы с потребителями. Процесс маркетинга является первым в цепочке жизненного цикла продукции, что определяет стартовые характеристики производимого продукта и достижение высоких финансовых результатов для компании. Таким образом, совершенствование процесса маркетинга за счет новых и современных технологий является актуальной задачей. Популярные

сегодня маркетинговые технологии, такие как медиамероприятия, являются, по сути, событийными технологиями и направлены на продвижение продукта, бренда или услуг посредством личного взаимодействия, требуют применения современных и эффективных цифровых технологий для медиакоммуникаций. Рынок медиапроектов сегодня огромный и конкурентный. Эффективность таких проектов определяется использованием прогрессивных технологий, цифровых и технологических решений. Для проведения медиамероприятия на рынке существуют как крупные компании с десятками и сотнями человек, так и индивидуальные предприниматели, которые обычно работают сами на себя в качестве ведущих и часто нанимают звукооператоров для управления аудиоаппаратурой. Основными задачами этих людей являются аудиосопровождение, то есть запуск определенных музыкальных композиций в нужный момент сценария, контроль за микрофонами, которые использует ведущий, установка светового оборудования и иногда установка экранов для демонстрации видеоматериалов. Выполнять все эти задачи приходится почти одновременно, и часто при этом возникает сложность.

Цель исследования – обоснование актуальности проектирования автоматизированной информационно-аналитической системы для управления файлами при помощи голоса с целью частичной передачи действий по управлению аудиофайлами в медиапроектах. Для достижения поставленной цели были выполнены следующие задачи: исследование актуальности разрабатываемой системы; моделирование сопровождаемых системой бизнес-процессов; разработка ключевых элементов проектируемой системы.

Материалы и методы исследования

В ходе исследования были использованы теоретические методы исследования – изучение и анализ научной литературы, обобщение, существующий опыт разработки подобных продуктов, а также апробация результатов исследования в рамках выпускной квалификационной работы при проектировании программного продукта.

Результаты исследования и их обсуждение

Технологии искусственного интеллекта сильно упрощают рутинные процессы, которые реализуются в рамках медиапроектов. Такие методы, как статистический анализ, распознавание силуэтов или символов

на фото, а также распознавание человеческой речи в аудиофайлах, – тому подтверждение. Автоматизация процессов с использованием цифровых технологий значительно повышает результативность и эффективность процессов. Спрос на автоматизацию технологических процессов при помощи голоса значительно вырос за последнее время [1, 2]. Проблемы, которые можно решить с помощью автоматизации, связаны также с человеческим фактором. Люди совершают ошибки из-за невнимательности, отвлеченности или утомляемости, в отличие от компьютера, который выполняет команды точно по плану. Успех проведения мероприятия зависит от проектной команды, которая проектирует и реализует подобные мероприятия. Создание голосового помощника для ведущего мероприятия, который, получая команды через микрофон, мог бы переключать аудиофайлы, может существенно сократить список задач звукооператора, позволив отказаться от аудиосопровождения либо полностью заменить его, учитывая, что остальные задачи можно решить до начала мероприятия.

Речевые сигналы изначально являлись одним из основных средств передачи информации человеком. Люди ежедневно используют речь для взаимодействия с миром и в ближайшем времени не откажутся от нее. Этим можно объяснить желание взаимодействовать с техникой при помощи голоса. Такие технологии, как голосовой ввод клавиатуры или голосовые помощники в телефонах, уже давно вошли в обиход. Цифровые помощники: «Алиса» от Яндекса, «Siri» от «Apple» и «Google Ассистент» – являются самыми популярными голосовыми помощниками в современных умных устройствах, операционных системах и браузерах.

В основе любого голосового помощника лежит искусственный интеллект, ответственный за распознавание человеческой речи. На сегодняшний момент существует большое количество систем для распознавания речи, все они имеют свои особенности. На данный момент существует два основных вида технологий распознавания речи [3]:

- зависимое от диктора, которое предполагает первоначальное обучение системы распознавания конкретного голоса при получении для обучения эталонов слов и фраз;
- не зависимое от диктора распознавание речи. В этом случае применяются уже обученные системы, которые учатся не словами, а звуками, взятыми из миллионов аудиозаписей речей разных людей.

Также у систем распознавания существует возможность непрерывной обработки речевого сигнала, при этом необходимо учитывать скорость, с которой говорит пользователь. В потоке речи многие звуки могут сливаться, и системе сложнее распознать, где заканчивается одно слово и начинается другое. В разговорной речи также начало и конец слов влияют на последующие. Анализ существующих специальных программ, воспроизводящих аудиофайлы, для диджеев, а также систем автоматизированного голосового управления показал, что в каждом из сравниваемых объектов обнаруживаются общие с другими функциональные возможности, а также выделяются те критерии, которые отличают их друг от друга. Далее нами было проведено сравнение существующих программных решений для воспроизведения аудиофайлов. Были взяты три программы для воспроизведения аудиофайлов: «AIMP», «VirtualDJ», «RadioPlayerPro». Данные программы выбраны на основании того, что они чаще всего используются звукооператорами и радиоведущими мероприятиями. Программные продукты сравнивались по критериям, которые были определены экспертным способом. По итогам сравнения была составлена таблица 1.

Все эти программы под управлением человека хорошо выполняют свои функции

по воспроизведению аудиофайлов, а также имеют большое количество функций как для сведения, так и для конфигурации аудиодорожек. Некоторые из этих программ, такие как AIMP и RADIOPlayerPRO, подходят для проведения простых мероприятий, где все действия расписаны по секундам, и эти программы могут работать без постоянного контроля, но во время сложных мероприятий все действия носят импровизационный характер и рассчитать время запуска нужной композиции невозможно. Данные программы не являются решением поставленной проблемы, но могли бы помочь в этом, если бы имели в своем функционале технологии управления при помощи голоса – так называемые голосовые помощники.

На следующем этапе анализа были выбраны голосовые помощники, которые активно используются в медиаиндустрии, некоторые из них интегрированы в операционные системы персональных компьютеров и мобильных устройств, некоторые – в браузеры, а некоторые интегрируются в фирменные устройства и приложения. Для сравнения голосовых помощников были выбраны: «Алиса» от Яндекса, «Siri» от «Apple» и «Google Ассистент». Это самые популярные голосовые помощники в современных умных устройствах, операционных системах и браузерах [4].

Таблица 1

Результаты сравнения программ для запуска аудиофайлов

Критерий	«AIMP»	«Virtual DJ»	«RADIO Player Pro»
Количество одновременно запущенных аудиодорожек	1	До 4	1
Возможность наложения аудиодорожек друг на друга	Нет	Есть	Нет
Удобство доступа к файлам на компьютере	2	3	1
Наличие полезных функций для сведения аудиофайлов	Есть	Есть	Есть
Возможность изменения конфигураций аудиодорожки	Есть	Есть	Нет

Таблица 2

Результаты сравнения голосовых помощников

Критерий	«Алиса»	«Siri»	«Google Ассистент»
Поддержка русского языка	Есть	Есть (не для всех команд)	Есть (не для всех команд)
Возможность интеграции в сторонние программы	Нет	Нет	Нет
Возможность постоянного распознавания речи	Нет	Нет	Нет
Необходимо ли заранее обучать систему	Нет	Да	Нет
Необходимо ли наличие интернет-соединения	Да	Да	Да

Таблица 3

Результаты сравнения библиотек для распознавания

Критерий	«Vosk»	«Artyom.js»	«Speech Recognition»
Поддержка русского языка	Есть	Есть	Есть
Возможность постоянного распознавания речи	Есть	Есть	Есть
Необходимо ли заранее обучать систему	Нет	Нет	Нет
Возможность автономной работы	Есть	Есть	Есть

Анализ показал, что выбранные голосовые помощники не удовлетворяют решению поставленной задачи, потому что они не имеют возможности интеграции в сторонние программные обеспечения и требуют для своей работы постоянное интернет-соединение. Так как ни один из рассмотренных голосовых помощников не удовлетворяет требованиям для решения поставленной проблемы, принято решение разработать приложения для воспроизведения аудиофайлов со встроенной системой голосового управления. Основным компонентом в данной системе служит распознавание речи. Как показал обзор голосовых помощников, главной их проблемой является невозможность интеграции в сторонние программы. Итоги анализа представлены в таблице 2.

Для разрабатываемой программы принято решение использовать API и библиотеки с уже готовыми методами, классами, функциями для распознавания речи, потому что они имеют открытый исходный код в общем доступе, соответственно, их можно интегрировать в разрабатываемую систему. Для сравнения и последующего выбора были взяты 3 самые распространенные на сегодняшний день API и библиотеки для распознавания: «Vosk», «Artyom.js», «SpeechRecognition» [5]. Так, рассмотренные библиотеки показали одинаковые результаты при сравнении, и все они удовлетворяют поставленным требованиям. Выбор определенной библиотеки будет зависеть от языка программирования, на котором будет написан программный прототип разрабатываемой системы. Прямым аналогом приложения для управления аудиофайлами при помощи голоса является человек. Получая от ведущего определенную команду, диджей находит нужный файл и запускает его в плеере, используя различные техники сведения (табл. 3).

Изучив существующие возможности рынка цифровых технологий, авторы пришли к выводу, что нет универсальных цифровых инструментов, которые бы одинаково эффективно работали в разных проектах.

В результате анализа сделан вывод, что влияние человеческого фактора на результат, который мы получаем от цифрового помощника, что приводит иногда к существенным ошибкам. С целью устранения недостатков, выявленных в существующих аналогах, авторы разработали концепцию усовершенствованного голосового помощника со специализацией в медиапроектах.

На первом шаге авторами были сформулированы требования к проектируемому результату. Выделены две категории требований к проектируемому цифровому помощнику:

1) нефункциональные требования: результатом программного проекта в данном случае должен быть сайт, так как, по мнению экспертов, это позволяет иметь доступ к системе в удаленном формате. Так как это сайт, то необходимо учитывать возможность его использования на мобильных устройствах, потому что не многие пользователи имеют в своем распоряжении компьютер или ноутбук. По мнению экспертов, в дизайне сайта нужно использовать темносиний и графитовый цвета, а основные блоки должны быть с белым фоном;

2) функциональные требования: программный прототип должен состоять из следующих компонентов: компонент управления информационными страницами; компонент управления сценариями; компонент «Модель автоматического преобразования сигнала»; компонент «Автоматическое распознавание»; компонент «Выборка».

Для пользователя должны быть доступны модули: регистрация; авторизация; личный кабинет с созданными сценариями; создание, редактирование и удаление сценариев; запуск сценария. Для корректной работы данных модулей необходимо получение доступа к микрофону пользователя, так как он требуется для передачи голосовых команд в систему. Администратору должны быть доступны все модули, доступные пользователю, а также возможность доступа в панель администратора, при помощи которой он может взаимодействовать с аккаунтами пользователей и баз данных.



Диаграмма прецедентов разрабатываемой системы

Для моделирования основных процессов и архитектуры проекта информационной системы необходима разработка UML-диаграммы для описания объектов в едином синтаксисе, понятном всем. Для написания кода важно разработать проект приложения, декомпозировать основные процессы. Для этого в первую очередь необходимо определить модели акторов системы – действующих субъектов. Так как системой будут пользоваться ведущие и у них будет единственная роль – обычный пользователь системы. Также для контроля и поддержки работоспособности системы требуется еще один вид акторов – это администратор.

На основе проведенного анализа была разработана диаграмма прецедентов (Use-case diagram), которая отражает взаимосвязь участников процесса, их роли с прецедентами (рис.). В системе два участника: пользователь и администратор. Прецеденты для пользователя: регистрация в системе, авторизация, просмотр страниц, создание и удаление сценариев и т.д. Роль администратора: администрирование системы – создание и удаление пользователей.

Предложенная логика является основой для дальнейшего проектирования информационно-аналитической системы для управления файлами с помощью голоса. Алгоритм дальнейших мероприятий по проектированию должен включать: логическую модель базы данных, дающую пред-

ставление о связях в базе данных; диаграмму последовательности взаимодействия; диаграмму активности процесса запуска сценария; контекстную схему по стандарту IDEF0; декомпозицию схемы IDEF0; структурную схему сайта; выбор средств реализации для серверной и клиентской частей.

Предполагается, что для написания серверной части будет применен фреймворк Django и языке Python. Для клиентской части необходимо использовать язык разметки документов HTML; объектно-ориентированный язык JavaScript, фреймворк jQuery; для описания внешнего вида веб-страницы – язык CSS, фреймворк Bootstrap.

Заключение

Таким образом, был осуществлен анализ сферы проведения медиамероприятий. Были выявлены существующие на данный момент проблемы в этой области, такие как проблема нехватки кадров, а также проблемы, связанные с человеческим фактором, которые возникают в работе звукооператоров. Для более детального анализа и постановки проблемы был проведен опрос среди ведущих медиамероприятий, звукооператоров и радиоведущих. Результаты данного опроса помогли выявить проблемы, с которыми сталкиваются ведущие при работе со специализированным программным обеспечением, а также помогли определить необходимый функционал про-

граммного прототипа. Также был проведен сравнительный анализ существующих программных решений и технологий, в результате был сделан вывод, что существующие готовые решения не подходят для решения поставленной проблемы.

В итоге можно сказать, что возникает необходимость разработки нового программного продукта, который будет направлен на достижение целей мероприятия с учетом выявленных проблем и предложенного алгоритма.

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.

2. Безруких А.Д., Черепанов М.Д., Сизова Т.Н., Безруких Ю.А. Сверточная нейронная сеть в решении задачи: определение пола и возраста // Инновации в химико-лесном комплексе: тенденции и перспективы развития: сборник материалов Всероссийской научно-практической конференции. Красноярск, 2021. С. 108-111.

3. Тампель И.Б., Карпов А.А. Автоматическое распознавание речи: учебное пособие. СПб.: Университет ИТМО, 2016. 138 с.

4. Морозова А.А. Голосовой помощник «Алиса»: достоинства и недостатки коммуникации с «умным собеседником» // Художественное произведение в современной культуре: творчество – исполнительство – гуманитарное знание: сборник статей и материалов / сост. А.С. Макурина. Челябинск: Издательство: Южно-Уральский государственный институт искусств им. П.И. Чайковского, 2021. С. 240-242.

5. Библиотека распознавания речи: что это и какие существуют [Электронный ресурс]. URL: https://codernet.ru/articles/drugoe/biblioteka_raspoznavaniya_rechi_chno_eto_i_kakie_sushhestvuyut/ (дата обращения: 27.02.2023).