

Федеральное государственное автономное
образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Институт космических и информационных технологий
институт
Межинститутская базовая кафедра
«Прикладная физика и космические технологии»
кафедра

УТВЕРЖДАЮ

Заведующий кафедрой

_____ В.Е. Косенко
подпись инициалы, фамилия

« _____ » _____ 2019 г.

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

«Разработка метода сортировки музыкальных композиций по их звуковым
характеристикам»

тема

09.04.01 «Информатика и вычислительная техника»

код и наименование направления

09.04.01.03 «Информационные системы космических аппаратов и центров
управления полетами»

код и наименование магистерской программы

<u>Научный руководитель</u>	_____	_____	_____
	подпись, дата	канд. техн. наук, доцент кафедры МБК ПФиКТ	<u>М.В. Некрасов</u>
		должность, ученая степень	инициалы, фамилия
<u>Выпускник</u>	_____	_____	_____
	подпись, дата		<u>Р.Б. Величкин</u>
			инициалы, фамилия
<u>Рецензент</u>	_____	_____	_____
	подпись, дата	инженер-программист ДИТ ГХК	<u>Н.Ю. Зуева</u>
		должность, ученая степень	инициалы, фамилия
<u>Нормоконтролер</u>	_____	_____	_____
	подпись, дата		<u>Е.С. Сидорова</u>
			инициалы, фамилия

Красноярск 2019

Федеральное государственное автономное
образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Институт космических и информационных технологий
институт
Межинститутская базовая кафедра
«Прикладная физика и космические технологии»
кафедра

УТВЕРЖДАЮ

Заведующий кафедрой

_____ В.Е. Косенко

подпись

инициалы, фамилия

« _____ » _____ 20__ г.

ЗАДАНИЕ
НА ВЫПУСКНУЮ КВАЛИФИКАЦИОННУЮ РАБОТУ
в форме магистерской диссертации

Студенту Величкину Роману Борисовичу.

Группа КИ16-01-3М.

Направление 09.04.01 «Информатика и вычислительная техника».

Специализированная программа подготовки: 09.04.01.03 «Информационные системы космических аппаратов и центров управления полетами».

Тема магистерской диссертации (МД): «Разработка метода сортировки музыкальных композиций по их звуковым характеристикам».

Утверждено приказом по университету от « ____ » _____ 20__ г. № _____

Руководитель МД: М.В. Некрасов, канд. техн. наук, доцент кафедры МБК ПФиКТ.

Перечень разделов МД:

- 1 Обозначение проблемы и существующие решения.
- 2 Исследование и построение решения.
- 3 Реализация.

Перечень графического материала: представлен в виде слайдов презентации в количестве 18 штук.

Руководитель ВКР _____
подпись

М.В. Некрасов
инициалы и фамилия

Задание принял к исполнению _____
подпись

Р.Б. Величкин
инициалы и фамилия

« ____ » _____ 20__ г.

РЕФЕРАТ

Выпускная квалификационная работа по теме «Разработка метода сортировки музыкальных композиций по их звуковым характеристикам» содержит 67 страниц текстового документа и 20 использованных источников.

ЗВУК, МУЗЫКА, РЕКОМЕНДАТЕЛЬНАЯ СИСТЕМА, ЧАСТОТА, ТЕМП, ТОНАЛЬНОСТЬ, СРАВНЕНИЕ, АНАЛИЗ.

В данной работе рассматривается проблема, возникающая при необходимости ориентации в большом количестве музыки. Развитие технологий по работе со звуком сделало создание музыки доступным для многих людей, что привело к значительному увеличению ее количества и разнообразия.

В качестве решения было предложено разработать метод, который бы позволял находить схожие по звучанию песни, а также сортировать их по степени сходства, основываясь на анализе их звуковых характеристик.

Цель магистерской работы:

- выявить звуковые характеристики, на основе которых будут сравниваться музыкальные композиции;
- разработать автоматический способ определения значений звуковых характеристик музыкальной композиции;
- разработать методику сравнения и сортировки музыкальных композиций по степени сходства их звуковых характеристик;
- провести эксперимент и сравнить результаты работы созданных методик с результатами работы существующих музыкальных сервисов.

В результате работы были созданы алгоритмы для определения значений частотного баланса, темпа и тональности музыкальных композиций. Также был разработан метод поиска схожих по звучанию музыкальных композиций и прикладная программа, основанная на данном методе. По итогам эксперимента программа доказала действенность разработанного метода и имеющийся у него потенциал.

Предложена методика автоматизированной сортировки музыкальных композиций по степени сходства их звуковых характеристик: частотного баланса, темпа и тональности, полученной методом анализа саб-басовых частот.

Разработанную методику можно использовать для совершенствования существующих музыкальных сервисов, использующих рекомендательные системы с фильтрацией по содержанию. Также исходный код разработанной программы можно выложить в свободный доступ, что позволит другим разработчикам использовать его в свои программных продуктах.

СОДЕРЖАНИЕ

Введение	8
1 Обозначение проблемы и существующие решения	10
1.1 Обозначение проблемы	10
1.2 Существующие решения	11
1.2.1 Потокое мультимедиа	12
1.2.2 Появление музыкальных потоковых сервисов	13
1.2.3 Механизм работы музыкальных потоковых сервисов	20
1.2.3.1 Сеть дистрибуции содержимого	20
1.2.3.2 Виды протоколов, используемых при передаче потокового мультимедиа	25
1.2.4 Pandora	29
1.2.5 The Echo Nest и Spotify	31
1.2.6 Рекомендательные системы	34
1.2.6.1 Коллаборативная фильтрация	37
1.2.6.2 Фильтрация на основе содержания	38
1.2.6.3 Гибридные рекомендательные системы	39
1.2.6.4 Оценка эффективности рекомендательных систем	40
2 Исследование и построение решения	44
2.1 Выбор звуковых характеристик для анализа музыкальной композиции	44
2.2 Обоснование использования частотного баланса для поиска схожих по звучанию музыкальных композиций	47
2.3 Условия для применения	49
2.4 Определение частотного баланса музыкальных композиций	49
2.5 Определение темпа музыкальных композиций	51
2.6 Определение тональности музыкальных композиций	54
2.7 Сравнение музыкальных композиций по звуковым характеристикам	56
3 Реализация	59

3.1 Подготовка к эксперименту.....	59
3.2 План эксперимента.....	60
3.3 Результаты эксперимента.....	61
Заключение.....	63
Список использованных источников	65

ВВЕДЕНИЕ

В современном мире с небывалой скоростью растет количество информации хранимой на цифровых носителях, в том числе и музыки.

Совершенствование и удешевление музыкальных инструментов и звукозаписывающего оборудования, активное развитие технологий для работы со звуком, появление множества программ для создания и обработки музыки на компьютере, а также возможность дистанционного обучения через Интернет – в совокупности все это привело к значительному упрощению процесса создания музыки.

Благодаря этому появилось множество музыкантов и исполнителей, и значительно выросло количество музыки предлагаемой слушателям. Это привело к возникновению проблемы: как ориентироваться в таком количестве музыки.

В качестве решения этой проблемы появились потоковые музыкальные сервисы. Работают они по подписочной системе, предоставляя слушателям доступ ко всей своей музыкальной библиотеке за месячную абонентскую плату. Чтобы помочь слушателям ориентироваться в многообразии музыки сервисы используют рекомендательные системы – они анализируют музыкальные предпочтения слушателей, и подбирают песни, которые должны им понравиться.

При этом большинство музыкальных сервисов формирует рекомендации, основываясь на предпочтениях пользователей, либо используя метаданные, описывающие музыку: исполнитель, название альбома, жанр, энергичная или спокойная, инструментальная или вокальная, и т.д.

Существует возможность расширить список этих параметров, дополнив его параметрами, характеризующими звук музыкальной композиции, например: соотношение частот, тональность или число тактов в минуту. Это сделает рекомендации потоковых музыкальных сервисов более

точными, а значит, и упростить пользователям поиск необходимой им музыки.

1 Обозначение проблемы и существующие решения

1.1 Обозначение проблемы

С развитием человечеству приходится хранить и обрабатывать все большее количество информации. Стремительное увеличение объема хранимой информации, которое произошло с развитием информационных технологий, получило название «информационный взрыв».

Музыкальная сфера не стала исключением. Удешевление электроники позволило сделать музыкальную аппаратуру доступной. Появление мощных персональных компьютеров сделало возможным создание цифровых музыкальных библиотек, заменяющих живые инструменты. Теперь композиторы могут записывать оркестровые партии, не прибегая к помощи живого оркестра.

Также были разработаны компьютерные системы, предназначенные для работы со звуком - цифровые звуковые рабочие станции. Благодаря им были автоматизированы многие рутинные процессы, возникающие при звукозаписи и микшировании.

Широкополосный Интернет, видеохостинги, такие как YouTube, а также образовательные интернет-ресурсы, такие как Coursera и Udemy, способствовали появлению огромного количества учебных видеоматериалов, в том числе, и по профессии звукорежиссера. Многие профессиональные продюсеры и звукоинженеры публикуют в Интернете бесплатные видеоуроки, где подробно рассказывают о тонкостях своей работы, а также делятся опытом и разбирают конкретные примеры из своей практики.

В совокупности все эти факторы привели к значительному упрощению процесса создания музыки, а также ее записи и микширования. И как следствие, к росту объема доступной для прослушивания музыки.

Кроме того, развитие технологий для работы со звуком привело к появлению множества новых музыкальных жанров.

В настоящий момент в одной только библиотеке сервиса Spotify хранится более 45 млн. музыкальных композиций. Таким образом, у слушателей возникла проблема – как ориентироваться в столь значительном количестве разнообразной музыки.

1.2 Существующие решения

Долгое время после появления звукозаписи основными источниками информации о том, какую музыку стоит послушать были телевидение, радио, пресса и магазины с музыкальными записями. В прессе публиковались критические обзоры на исполнителей и их музыку. На радио и телевидении играли музыку, а также публиковали хит-парады с наиболее популярными песнями и исполнителями. В магазинах посетители могли спросить у консультантов совета о том, какие музыкальные диски стоит приобрести.

С распространением персональных компьютеров, появлением сети Интернет и аудио-формата mp3 слушателям стало проще обмениваться музыкой, хранить и коллекционировать ее.

Количество оцифрованной музыки росло, и возникла потребность в музыкальном каталоге с удобным поиском. Однако для создания такого каталога потребовался бы мощный сервер с высокоскоростным подключением к Интернету. Поэтому первыми каталогами стали сервисы и торрент-трекеры, работающие по технологии пиринговых сетей, которая избавила от необходимости хранить всю музыку в одном месте. Благодаря этой технологии пользователи могут скачивать музыку друг у друга, не нагружая центральный сервер, который нужен лишь для того, чтобы помочь пользователям «найти» друг друга. Первым сервисом, использующим технологию пиринговых сетей, стал Napster.

Поскольку такой обмен музыкой не подразумевал какой-либо платы, а каталоги оказались очень удобными, то продажи компакт-дисков начали падать. Правообладатели и авторы музыки не получали никакого денежного

вознаграждения от того, что пользователи скачивали их музыку, и звукозаписывающие компании начали судебный процесс, который привел к закрытию Napster. Такой обмен музыкой стали называть «пиратством» и признали незаконным. Однако, удобных сервисов, которые могли бы составить конкуренцию «пиратам», так и не появилось, поэтому музыкальная индустрия продолжала терпеть убытки.

В 2003 года компания Apple открыла музыкальный интернет-магазин iTunes Store, торгующий музыкой, который стал лидером в этой области на многие годы. Одной из ключевых особенностей стала возможность приобретать песни по отдельности, а не только целые альбомы. Таким образом, у слушателей появилась легальная возможность найти всю музыку, которая им необходима, в одном месте.

Появившиеся смартфоны постепенно заменили плееры, а доступный мобильный Интернет подстегнул развитие музыкальных потоковых сервисов, которые дали владельцам смартфонов возможность слушать музыку, не скачивая ее.

1.2.1 Потокое мультимедиа

Потоковое мультимедиа - мультимедиа, которое пользователь непрерывно получает от провайдера потокового вещания. Конечный пользователь клиента может использовать свой медиаплеер, чтобы начать воспроизведение цифрового видео- или аудиоконтента до того, как будет передан весь файл. Такой метод доставки мультимедиа является альтернативой загрузке файла, когда пользователю необходимо получить весь файл, чтобы воспроизвести его.

Определение «потоковое мультимедиа» применяется к доставке сигнала через телекоммуникационные сети.

Мультимедиа потоки бывают двух видов: по запросу или прямой трансляцией. Потоки информации, вызываемой по запросу пользователя,

хранятся на серверах продолжительный период времени. Прямой эфир доступен лишь короткий период времени, в течение которого идет трансляция. Например, при передаче видео со спортивных соревнований. Для потоковой передачи в реальном времени через Интернет требуются устройство для ввода исходного мультимедиа (видеокамера, микрофон, программное обеспечение для захвата экрана), кодер для оцифровки сигнала, и сервис для показа трансляции.

Для работы с потоковым контентом в Интернете пользователю необходимо иметь быстрое и стабильное интернет-соединение. Если у пользователя недостаточно пропускной способности при подключении к Интернету, у него могут возникать задержки и зависания контента, либо его буферизация будет происходить недостаточно быстро.

В настоящий момент наиболее популярными потоковым видеосервисами являются YouTube и Twitch. YouTube до недавнего времени работал только как видеохостинг, но в 2014 году на сайте появилась возможность создавать прямые трансляции. Twitch специализируется на прямых трансляциях, но на сайте также имеется возможность смотреть старые записи.

Netflix, Hulu и Amazon Prime Video являются наиболее популярными зарубежными потоковым видеосервисами, транслирующими фильмы и телепередачи. Самыми популярными российскими видеосервисами, транслирующими фильмы, являются ivi.ru, Амедиатека и Кинопоиск.

Крупнейшие зарубежные сервисы, транслирующие потоковую музыку: NetEase, SoundCloud, Spotify, российские: Яндекс.Музыка и BOOM.

1.2.2 Появление музыкальных потоковых сервисов

Первые попытки отображения мультимедийной информации на компьютерах начались еще на заре компьютерных технологий в середине 20-го века. Однако в течение нескольких последующих десятилетий в этом

направлении так и не было достигнуто значительного прогресса, в первую очередь из-за высокой стоимости и ограниченных возможностей компьютерного оборудования того времени. Основной технической проблемой потокового вещания было отсутствие достаточно производительного процессора, способного поддерживать требуемую скорость передачи данных. Также отсутствовала операционная система способная предотвращать переполнение памяти. С конца 1980-х до 1990-х годов начали появляться персональные компьютеры, обладающие достаточной производительностью для отображения мультимедийной информации. Однако в середине 1990-х годов мощности компьютерных сетей по-прежнему были ограничены, и аудио- и видеоматериалы обычно доставлялись путем загрузки цифрового файла с удаленного сервера и последующего хранения его на локальном диске, либо путем воспроизведения файла с компакт-диска.

В 1991 году был представлен первый коммерческий коммутатор Ethernet, который позволил создать более мощные компьютерные сети, что привело к появлению первых решений для потокового видео. В конце 1990-х и начале 2000-х годов пользователи получили значительно более широкий доступ к компьютерным сетям, особенно к Интернету. В начале 2000-х годов у пользователей появился доступ сетям с высокой пропускной способностью. Эти технологические усовершенствования упростили пользователям получение потоковых передач аудио- и видеоконтента. По мере того, как Интернет становился все более коммерциализированным, увеличивались объемы инвестиций в связанные с ним направления.

The Severe Tire Damage стала первой группой, которая вживую выступила в Интернете. 24 июня 1993 года группа отыграла концерт в научно-исследовательском центре Xerox PARC, где тогда разрабатывали технологию Mbone для вещания в Интернете с использованием многоадресной рассылки. В качестве демонстрации эффективности технологии выступление группы транслировалось в прямом эфире. Видео

имело разрешение 152 на 76 пикселей и восемь-двенадцать кадров в секунду, а качество звука, в лучшем случае, было сопоставимо с плохой телефонной связью. В мартовском интервью 2017 года один из участников этого музыкального коллектива Расс Хейнс заявил, что группа тогда для трансляции своего выступления использовала около половины всей пропускной способности Интернета.

Компания RealNetworks была пионером на рынках потокового мультимедиа, и провела в 1995 году одну из первых аудиопередач через Интернет, в ходе которой транслировалась бейсбольная игра между New York Yankees и Seattle Mariners. Первый симфонический концерт в Интернете был проведен 10 ноября 1995 года в Paramount Theatre в Сиэтле, штат Вашингтон. Концерт состоялся в сотрудничестве с The Seattle Symphony и различными приглашенными музыкантами из групп Guns 'n Roses, Velvet Revolver, Soundgarden, и Pearl Jam. Когда в 1995 году был запущен Word Magazine, они представили первые в мире потоковые саундтреки в Интернете.

Термин «потоковая передача» был впервые использован компанией Data Electronics Inc. для описания ленточных накопителей, предназначенных для запуска всей дорожки. Впервые видео стали называть «потоковой передачей» в компании Starlight Networks, а потом так стали характеризовать свои аудиопередачи и в компании RealNetworks. В начале 1990-х годов этот термин закрепился и стал общепринятым.

Первый коммерческий потоковый продукт появился в конце 1992 года и был назван StarWorks. StarWorks обеспечивал возможность свободного доступа к видео в формате MPEG-1 в корпоративных сетях Ethernet. StarWorks был разработан компанией Starlight Networks, которая также первой провела потоковую передачу видео в реальном времени и посредством Ethernet, и через Интернет. В конце 90-х после того, как Интернет стал доступен широкому кругу пользователей, потоковое видео в

сети стало развиваться благодаря стартапам VDOnet, приобретенному RealNetworks, и Precept, приобретенному Cisco.

В 1995 году Microsoft разработала медиапроигрыватель, известный как ActiveMovie, который позволял осуществлять потоковую передачу мультимедиа и имел собственный потоковый формат. Впоследствии данные разработки получили развитие в виде функции потоковой передачи в Windows Media Player 6.4 в 1999 году. В июне 1999 года Apple также представила формат потоковой передачи мультимедиа - приложение QuickTime 4. Позднее он получил широкое распространение на веб-сайтах вместе с другими потоковыми форматами RealPlayer и Windows Media. Конкурирующие форматы на веб-сайтах требовали, чтобы каждый пользователь загружал соответствующие приложения для потоковой передачи, и в результате многим пользователям приходилось иметь все три приложения на своем компьютере.

В 2000 году Industryview.com запустил свой «крупнейший в мире веб-сайт потокового видеоархива», чтобы помочь компаниям в продвижении. Веб-трансляции стали новым инструментом для маркетинга и рекламы, который сочетал в себе захватывающую природу телевидения и интерактивность Интернета. Способность собирать данные и отзывы от потенциальных клиентов привела к быстрому развитию этой технологии.

В 2002 году интерес к единому унифицированному потоковому формату и широкое распространение Adobe Flash побудили к разработке формата потокового видео HTML5, который теперь используется в проигрывателях на многих популярных видеохостингах, таких как YouTube. Растущий потребительский спрос на показ прямых трансляций привел к появлению Twitch в 2011 году, а также его мобильного аналога Periscope в 2015 году. В настоящий момент многие социальные сети имеют возможность проводить прямые трансляции.

В начале 1999 года в Халле, штат Массачусетс был запущен сервис Napster, который построен на технологии пиринговых сетей, где пользователи могли свободно загружать и скачивать mp3-файлы.

На рисунке 1 изображена структура пиринговой сети.

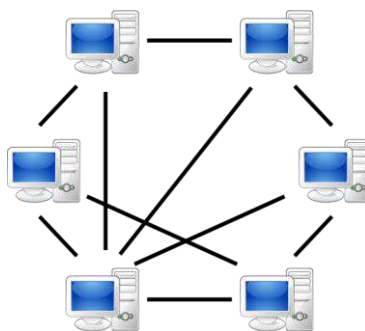


Рисунок 1 – Структура пиринговой сети

Платформа была разработана Шоном и Джоном Фаннингом, а также Шоном Паркером. В интервью 2009 года Шон Фаннинг объяснил, что Napster «был создан с целью удовлетворить потребность людей в том, чтобы найти всю эту музыку, особенно какие-то редкие вещи, которые невозможно найти в магазине звукозаписей». Это событие подорвало музыкальную индустрию, сделав песни, которые ранее приходилось покупать, доступными для бесплатного прослушивания любому пользователю Napster.

Такой подход произвел революцию в восприятии прав потребителей на цифровые товары - он сделала музыку свободно распространяемой. Napster быстро собрал миллионы пользователей и рос быстрее, чем любой другой бизнес в истории. На пике своей популярности Napster насчитывал около 80 миллионов пользователей по всему миру. Сайт получал такой большой трафик, что многим учебным заведения приходилось блокировать доступ к Napster, потому что студенты, обменивающиеся музыкальными файлами, перегружали сети.

Появление Napster привело к созданию множества других сервисов, работающих на такой же технологии: LimeWire, BitTorrent и Pirate Bay.

Впоследствии эти сервисы столкнулись с проблемами. Сначала это случилось с сервисом Napster осенью в 2001 году, против которого были поданы многочисленные судебные иски за несанкционированное распространение материалов, защищенных авторским правом. Иски были поданы различными звукозаписывающими лейблами, все из которых были дочерними компаниями Universal Music Group, Sony Music Entertainment, Warner Music Group или EMI. В дополнение к этому Ассоциация звукозаписывающей индустрии Америки также подала иск против Napster, что в конечном итоге привело к закрытию Napster в 2001 году.

Предписание, вынесенное в отношении Napster, положило конец короткому периоду, когда музыка была общедоступной. Другие пиринговые сети также имели успех в обмене mp3-файлами, но позже все они встретили похожую участь в суде. Постановление создало прецедент, согласно которому защищенный авторским правом цифровой контент не может свободно тиражироваться и распространяться без согласия владельца, тем самым укрепляя права собственности, как исполнителей, так и звукозаписывающих компаний.

В 1999 году был разработан новый сервис потокового аудио для использования на сайте TuneTo.com, настраиваемой интернет-радио платформе. В апреле 2001 года сервис TuneTo.com был приобретен стартапом Listen.com, который имел большую музыкальную онлайн-библиотеку. На основе разработок этих двух компаний 3 декабря 2001 года был запущен первый в мире музыкальный потоковый сервис Rhapsody. Rhapsody работал по модели платной подписки и позволял слушать музыку по запросу. К 2002 году Rhapsody подписал контракты со всеми пятью крупными звукозаписывающими компаниями того времени: BMG Entertainment, EMI Recorded Music, Universal Music Group, Warner Music Group и Sony Music Entertainment.

Реакция музыкальной индустрии на потоковую передачу музыки изначально была негативной. Наряду с музыкальным пиратством потоковые

сервисы разрушили рынок и способствовали значительному падению выручки с 14,6 млрд. долларов США в 1999 году до 6,3 млрд. долларов США в 2009 году. Из-за того, что музыка стала доступной бесплатно в Интернете, количество продаваемых физических копий альбомов стало падать, а продажи цифровых копий росли недостаточно быстро. Результатом стало то, что звукозаписывающие лейблы вкладывали больше средств в «безопасных» исполнителей - чартовая музыка стала более привлекательной для продюсеров, чем исполнители с уникальным звуком. Однако в 2018 году доходы от потоковой передачи музыки превысили прочие доходы, более традиционные для индустрии: продажи записей, продажи альбомов, цифровые загрузки. Только в 2017 году рост доходов от потокового прослушивания музыки составил 41,1%, когда общие доходы отрасли выросли всего на 8,1%. Доходы от потокового прослушивания музыки являются одной из крупнейших движущих сил роста музыкальной индустрии.

Современные музыкальные потоковые платформы, такие как Deezer, Apple Music, SoundCloud и Amazon Prime Music, предоставляют доступ к своей музыке только по премиум-подписке. Хотя, некоторые платформы, например, Яндекс.Музыка, BOOM и Spotify, кроме платной подписки также дают пользователям возможность слушать музыку бесплатно, но с некоторыми ограничениями по качеству звука и доступным функциям, а также с показом рекламы.

В последние годы популярность музыкальных потоковых платформ быстро растет. Spotify по состоянию на 1 января 2019 год имеет более 207 миллионов пользователей в 78 странах, SoundCloud имеет 175 миллионов, а Apple Music имеет около 60 миллионов. По мере того, как увеличилось число поставщиков потоковой музыки, конкуренция вынудила их снижать стоимость премиум-подписок, что также привело к снижению уровня музыкального пиратства.

1.2.3 Механизм работы музыкальных потоковых сервисов

Каждый из потоковых музыкальных сервисов имеет в своей библиотеке десятки миллионов песен. К библиотеке постоянно подключены миллионы пользователей со всего мира. При таком объеме передаваемой информации у пользователей при воспроизведении музыки могут возникать задержки или даже прерывания связи. Чтобы избежать таких проблем, сервера, откуда идет потоковое вещание, располагают территориально ближе к потребителю. Такой способ организации инфраструктуры называется сетью дистрибуции содержимого (CDN).

Перед отправкой файла в потоке, его предварительно сжимают. Для аудиопотока используют следующие форматы кодирования: MP3, Vorbis, AAC или Opus. Для сжатия видеопотока используют форматы H.264, HEVC, VP8 или VP9. Кодированные аудио- и видеопотоки собираются в контейнерный «битовый поток», такие как MP4, FLV, WebM, ASF или ISMA.

Полученный таким образом поток доставляется с сервера клиенту, например, пользователю с компьютером, подключенным к Интернету. У потоковых мультимедийных сервисов есть широкий выбор протоколов для осуществления такой передачи.

Музыка, которую сервис предлагает своим пользователям, подбирается для каждого слушателя индивидуально - с учетом его музыкальных предпочтений. За это отвечают рекомендательные системы.

1.2.3.1 Сеть дистрибуции содержимого

Сеть дистрибуции содержимого (CDN) - это географически распределенная сеть прокси-серверов и их центров обработки данных. Цель состоит в том, чтобы обеспечить высокую доступность и высокую производительность за счет пространственного распределения услуг по отношению к конечным пользователям.

На рисунке 2 изображена структура сети дистрибуции содержимого.

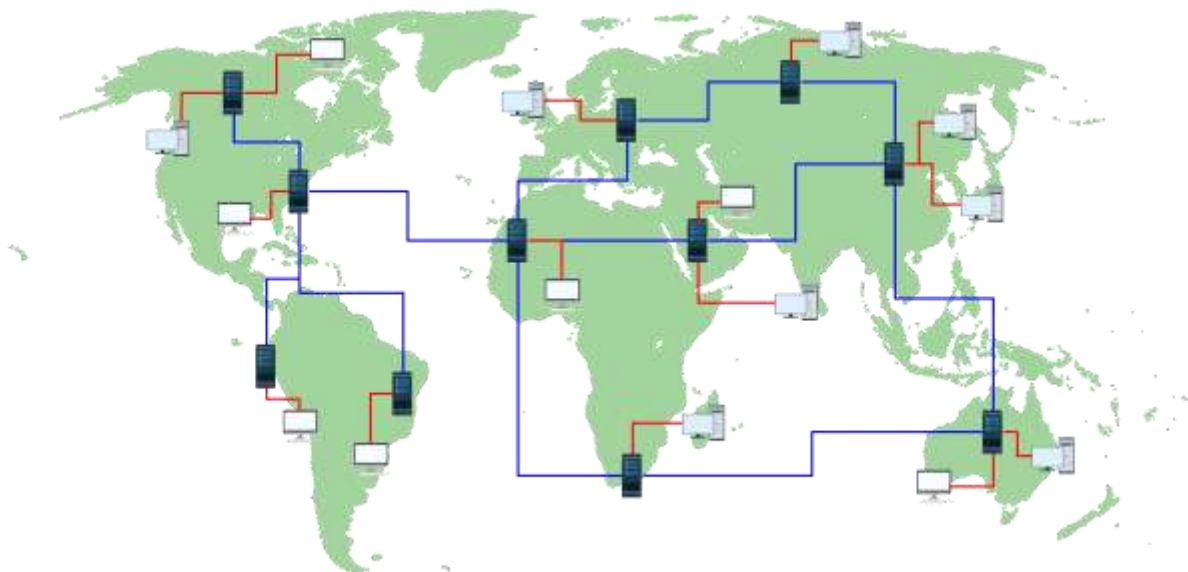


Рисунок 2 – Структура сети дистрибуции содержимого

Сегодня CDN обслуживают большую часть интернет-контента, включая веб-объекты (текст, графику и скрипты), загружаемые объекты (медиафайлы, программное обеспечение, документы), приложения (электронная коммерция, порталы), потоковую передачу мультимедиа, СМИ и сайты социальных сетей.

CDN - это слой в интернет-экосистеме. Владельцы контента, такие как медиакомпании и поставщики услуг электронной коммерции, платят операторам CDN за доставку контента конечным пользователям. В свою очередь, CDN платит провайдерам, операторам связи и операторам сетей за размещение своих серверов в их центрах обработки данных.

CDN - это общий термин, охватывающий различные типы услуг доставки контента: потоковое мультимедиа, загрузка программного обеспечения, ускорение веб- и мобильного контента, лицензированное / управляемое CDN, прозрачное кэширование и сервисы для измерения производительности CDN, балансировки нагрузки, коммутации нескольких CDN, аналитика, а также «облачный интеллект». Поставщики CDN также

предоставляют дополнительные услуги, такие как безопасность, использование брандмауэров и защита от DDoS.

Узлы CDN обычно развертываются в нескольких местах, часто по нескольким магистральным линиям. Такое расположение позволяет увеличить пропускную способность для пользователей сервиса, и делает контент доступнее по всему миру. Количество узлов и серверов, составляющих CDN, варьируется в зависимости от архитектуры. Некоторые достигают тысяч узлов с десятками тысяч серверов, расположенных по всему миру. Другие строят глобальную сеть и при этом имеют небольшое количество географических точек доступа.

Для оптимизации производительности можно выбрать места, которые лучше всего подходят для предоставления контента пользователю. Это можно определить, выбрав места с наименьшим количеством переходов, наименьшим количеством сетевых секунд от запрашивающего клиента или с максимальной доступностью с точки зрения производительности сервера. При оптимизации стоимости можно выбрать места, которые являются наименее дорогими.

Большинство провайдеров CDN будут предоставлять свои услуги через различные, определенные наборы точек доступа, в зависимости от покрытия желаемого региона, например: США, Европа, Азиатско-Тихоокеанский регион и т. д.

Сети CDN растут за счет приобретения новых серверов и создания новых точек доступа, либо за счет средств совместного размещения своего оборудования у других провайдеров.

Интернет был разработан в соответствии со сквозным принципом. По этому принципу базовая сеть остается относительно простой, а все сложные задачи решаются конечным точкам сети: хостами и клиентами. В результате базовая сеть решает только задачу отправки пакетов данных.

Сети дистрибуции содержимого дополняют сквозную транспортную сеть, распространяя в ней различные интеллектуальные приложения,

использующие методы, предназначенные для оптимизации доставки контента. Получившееся тесно интегрированное наложение использует веб-кэширование, балансирует нагрузки на сервера и настраивает маршрутизацию запросов.

Веб-кэши хранят данные, наиболее востребованные пользователями, на серверах, к которым пользователи обращаются чаще всего. Благодаря этому увеличивается пропускная способность сети, снижается нагрузка на сервер, и для пользователя уменьшается время отклика при обращении к данным, хранящимся в кэше.

Для балансировки нагрузки на сервера используются сервисные и аппаратные методы. Поступающие клиентские запросы распределяются между серверами и веб-кэшами. Таким образом, достигается оптимальный режим работы серверов, а также повышается общая надежность сети – сервис сохраняет работоспособность даже при отказе одного или нескольких серверов.

Маршрутизация запросов организуется так, чтобы клиентские запросы направлялись к серверу с необходимыми данными, который мог бы наилучшим образом обработать эти запросы. Это может быть перенаправление запроса на сервер, расположенный к клиенту ближе всего, или к серверу, наименее загруженному в настоящий момент.

Некоторые сети дистрибуции содержимого используют пиринговые сети. В пиринговых сетях клиенты не только получают контент, но и предоставляют его другим пользователям – благодаря этому все больше пользователей получает доступ к контенту. Таким образом, сервисы, ориентированные на доставку контента и использующие пиринговые сети, могут работать лучше, чем при использовании клиент-серверной архитектуры. Это свойство является одним из основных преимуществ использования пиринговых сетей, поскольку оно позволяет снизить затраты на настройку и эксплуатацию сети. Так, например, музыкальный потоковый

сервис Spotify до 2014 года в дополнение к прочим сетевым решениям использовал пиринговые сети.

Если владельцы потокового сервиса не удовлетворены вариантами или стоимостью услуг CDN, они могут создать свою собственную CDN. Это называется частной сетью дистрибуции содержимого. Частная CDN состоит из сетевых узлов, которые обслуживают контент только своего владельца.

Быстрый рост потокового трафика требует больших капиталовложений от интернет-провайдеров, чтобы удовлетворить этот спрос и удержать абонентов, предоставляя им качественные услуги.

Чтобы решить эту проблему, поставщики телекоммуникационных услуг (TSP) начали запускать свои собственные сети дистрибуции содержимого, чтобы снизить требования к магистральной сети и снизить объем необходимых инвестиций в инфраструктуру.

Поскольку телекоммуникационным компаниям принадлежат сети, по которым передается мультимедийный контент, то телекоммуникационные CDN имеют преимущества перед прочими CDN.

Эти компании владеют «последней милей» - каналом связи, соединяющий компьютер клиента и оборудование провайдера. Поэтому они могут располагать сервера с контентом ближе к конечному пользователю. Такая инфраструктура позволяет мультимедийным данным проходить меньшее расстояние по сети Интернет, а также доставлять их быстрее и надежнее.

У телекоммуникационных CDN также имеется преимущество в стоимости услуг, поскольку традиционным CDN приходится приобретать у них пропускную способность и включать маржу оператора в собственные затраты.

Кроме того, используя собственную инфраструктуру для дистрибуции контента, операторы связи могут лучше контролировать использование своих ресурсов. Операции управления контентом, выполняемые CDN, обычно применяются без (или с очень ограниченными) знаниями о структуре сети

операторов связи, с которыми они взаимодействуют или имеют деловые отношения. Это создает ряд проблем для операторов связи, которые имеют ограниченные возможности для устранения проблем, которые могут быть вызваны операциями CDN.

Развертывание телекоммуникационных сетей дистрибуции содержимого позволяет операторам осуществлять собственные операции по управлению контентом, что позволяет им лучше контролировать использование своих ресурсов и сохранять уровень качества услуг, предоставляемых пользователям [16].

1.2.3.2 Виды протоколов, используемых при передаче потокового мультимедиа

Чаще всего для отправки потокового мультимедиа используются протокол управления передачей (TCP) и протокол пользовательских датаграмм (UDP). Основная цель TCP - обеспечить надежный процесс передачи. Механизм работы TCP таков, что он не будет передавать другие пакеты, если предыдущие не достигли своего конечного пункта назначения. Однако это может занять много времени, потому что потери пакетов периодически случаются. UDP, в отличие от TCP, не будет тратить дополнительные усилия на исправление ошибок при доставке, он будет продолжать поддерживать непрерывный поток информации. Эта функция делает UDP более подходящим для потоковой передачи мультимедиа в реальном времени. Однако из-за того, что обычно TCP используется для различных действий в Интернете, транспортный протокол UDP может быть заблокирован некоторыми брандмауэрами. Кроме того, TCP является предпочтительным вариантом для потоковой передачи, когда небольшие задержки не имеют большого значения.

Протокол потоковой передачи в реальном времени (RTSP), транспортный протокол в реальном времени (RTP) и протокол управления

транспортировкой в реальном времени (RTSP) были специально разработаны для потоковой передачи мультимедиа по сетям. Основной целью протокола RTSP является установление и мониторинг медиа-сессий между сервером и клиентом. Кроме того, в качестве протокола управления сетью его основная функция заключается в контроле производительности потоковых серверов. Одна из замечательных особенностей RTSP - это его гибкость. Протокол позволяет переключаться с TCP на UDP с целью обеспечения лучшего качества потоковой передачи. Кроме того, RTSP дает пользователям возможность управлять процессом потоковой передачи посредством команд: воспроизведение, пауза, остановка и запись. Для обеспечения передачи медиафайлов серверы RTSP используют протоколы RTP и RTCP. Основной функцией RTP является создание и поддержка потоковой передачи мультимедиа. RTCP, в свою очередь, отслеживает информацию о передаче и ее качестве, также позволяет синхронизировать несколько одновременных потоков. Если потоковая передача в реальном времени требует своевременной доставки данных и допускает небольшое количество пропущенных пакетов, реализации RTP в основном основаны на UDP. Тем не менее, TCP также подходит для использования RTP, несмотря на то, что он не соответствует требованиям потока RTP.

Очень популярен протокол обмена сообщениями в реальном времени (RTMP). Этот протокол потокового мультимедиа, разработанный компанией Adobe, предназначен для потоковой передачи мультимедиа в реальном времени и по требованию через Adobe Flash Player. RTMP является протоколом на основе TCP и поэтому обеспечивает надежную передачу данных и стабильное соединение. Кроме того, этот протокол является гибким и позволяет осуществлять потоковую передачу видео, аудио и даже текстовых данных в различных форматах на разные устройства. Отличительной особенностью RTMP является низкая задержка. По этому показателю RTMP превосходит все прочие протоколы, даже такие как HTTP и HLS. Еще одна особенность этого протокола - это мультиплатформенность,

благодаря которой его можно использовать для устройств Mac, iOS, Android, Windows и многих других. Кроме того, как и RTSP, RTMP позволяет пользователям управлять потоком с помощью аналогичных команд и отслеживает действия пользователей в течение всего процесса потоковой передачи. Тем не менее, протокол имеет определенные недостатки. RTMP хоть и работает на широком спектре устройств, но на некоторых устройствах, например от компании Apple, он показывает меньшую стабильность и производительность чем проприетарные протоколы. Кроме того, могут возникнуть некоторые проблемы с потоковой передачей, если у полосы пропускания нестабильная скорость, поскольку RTMP не приспособлен к перепадам скорости.

HTTP регулирует взаимодействие между браузерами и серверами, а также позволяет передавать различные типы мультимедийного контента в Интернете. Кроме того, протокол HTTP может передавать текст и двоичные данные. HTTP очень хорошо работает с различными межсетевыми экранами, так как он основан на протоколе TCP. Однако он также адаптирован для использования UDP, например, в протоколах SSDP и HTTPU. Отличительной особенностью HTTP от перечисленных выше RTSP и RTMP является адаптивный битрейт - протокол позволяет медиаплееру оценивать пропускную способность сети и транслирует ему потоковое мультимедиа в подходящем для него качестве. Еще одной отличительной особенностью HTTP является то, что это протокол без сохранения состояния. Это означает, что все команды выполняются отдельно, без учета предыдущих. Эта особенность иногда не позволяет создавать интерактивные веб-сайты, которые интеллектуально реагируют на запросы пользователей. Несмотря на некоторые недостатки протокола HTTP, в последнее время очень распространенной стала тенденция отказываться от классических потоковых протоколов RTSP и RTMP и возвращаться к принципу простой загрузки по протоколу HTTP. Это стимулирует появление новых протоколов потокового мультимедиа, основанных на HTTP. Эти протоколы привлекательны тем, что

они имеют адаптивную потоковую передачу, одновременную поддержку большего количества потоков, а также обеспечивают легкое переключение с одного потока на другой.

Протокол потоковой передачи HTTP Live Streaming (HLS) был разработан компанией Apple для поддержки потокового мультимедиа с адаптивным битрейтом на своих устройствах. Однако в настоящее время этот протокол стал популярным и теперь используется, в том числе, на устройствах других производителей. В основе работы HLS алгоритм, который делит поток на ряд загружаемых файлов и кодирует их в разном качестве. Для каждого варианта качества генерируется индекс с расположением его файлов. Это позволяет медиаплееру получателя прямо во время воспроизведения подбирать качество потокового мультимедиа соответствующее текущей пропускной способности сети. Для прямых трансляций протокол HLS используется не так часто из-за высокой задержки передачи данных.

Существуют протоколы схожие с HLS - HTTP Dynamic Streaming (HDS) от Adobe и Smooth Streaming от Microsoft. Они также обеспечивают передачу потокового мультимедиа с адаптивным битрейтом через обычные HTTP.

MPEG-DASH - это технология потоковой передачи с адаптивным битрейтом, основанная на протоколе HTTP. MPEG-DASH является наиболее популярным решением среди протоколов для потоковой передачи с адаптивным битрейтом, поскольку имеет статус международного стандарта. Он использует TCP в качестве транспортного протокола и, как и HLS, обеспечивает потоковую передачу мультимедийных файлов, разделенных на несколько фрагментов. Кроме того, с помощью MPEG-DASH медиаплеер пользователя воспроизводит мультимедиа надлежащего качества в соответствии с несколькими факторами: скоростью соединения, возможностями устройства и предпочтениями пользователя. Основная цель этого протокола - обеспечить легкий процесс потоковой передачи любых

медиафайлов на любые устройства. MPEG-DASH не зависит от кодеков, что означает, что он подходит для медиаконтента, закодированного с помощью любых кодеков. Это позволяет уменьшить число технических проблем, возникающих во время вещания, и затраты на транскодирование. Таким образом, с MPEG-DASH можно использовать контент, закодированный проприетарными кодеками Apple, Adobe или Google, на любых устройствах других производителей. Благодаря этому, пользователям не нужно беспокоиться о том, поддерживается ли медиаконтент, который они хотят воспроизводить, их устройствами [13].

1.2.4 Pandora

Pandora Radio (также известно как Pandora Internet Radio или просто Pandora) - это музыкальный сервис, предназначенный для потоковой передачи музыки и автоматического формирования музыкальных рекомендаций, созданный в рамках реализации Music Genome Project. Сервис, управляемый компанией Sirius XM Satellite Radio, доступен только в США. Pandora предлагает пользователям песни, имеющие сходные музыкальные черты с той музыкой, которую они предпочитают. Затем пользователи отмечают – нравится им предложенная песня или нет, и эта оценка учитывается при последующем подборе песен для пользователей. Доступ к сервису можно получить либо через веб-браузер, либо через мобильное приложение. Pandora - это сервис freemium, основные функции бесплатны, но при прослушивании музыки присутствует реклама, а также отсутствуют дополнительные функции, такие как улучшенное качество потоковой передачи, загрузка музыки и скачивание музыки, которые предлагаются через платную подписку. Формируемые Pandora плейлисты строятся на основании алгоритма Music Genome Project.

Компания была основана в 2000 году Уиллом Глейзером, Джон Крафтом и Тим Вестергреном, и называлась тогда Savage Beast Technologies.

Их идея заключалась в том, чтобы создать радио, где каждый слушатель мог бы слушать собственную радиостанцию с музыкой, которой нравится именно ему. К 2001 году компания добилась первоначального финансирования в размере 2 миллионов долларов и запустила свой сервис. Сначала он был исключительно платным, но с привлечением рекламодателей, у слушателей появилась возможность слушать музыку бесплатно.

В 2014 году у Pandora среднемесячное число пользователей составляло около 76 миллионов пользователей в месяц, а на сервис приходилось около 70% всех прослушиваний среди музыкальных потоковых сервисов в США.

В феврале 2019 года компания Sirius XM Satellite Radio приобрела сервис Pandora за 3,5 миллиарда долларов.

Music Genome Project был впервые задуман Уиллом Глейзером и Тимом Вестергреном в конце 1999 года. Music Genome Project - это попытка описать музыку с использованием более 450 атрибутов и сложного математического алгоритма для их организации. Music Genome Project в настоящее время состоит из 5 подгеномов, характеризующих музыкальные жанры: поп, рок, хип-хоп, электроника, джаз, этническая музыка и классика. Под руководством Нолана Гассера и команды музыковедов первоначальные атрибуты были позже уточнены и расширены.

Каждая песня представлена вектором, содержащим значения для приблизительно 450 «генов» (аналогично генам, определяющим признаки для организмов в области генетики). Каждый ген соответствует характеристике музыки, например, пол ведущего вокалиста, уровень дисторшена на электрогитаре, тип бэк-вокала и т. д. Но точного описания генов в публичном доступе нет – это коммерческая тайна. Для определения значения параметров каждая песня анализируется музыкантом, этот процесс занимает от 20 до 30 минут на песню. Десять процентов песен повторно анализируются другим музыкантом, чтобы поддерживать уровень качества проводимого анализа.

Рок- и поп-песни имеют 150 генов, рэп-песни имеют 350, и джазовые песни имеют приблизительно 400. Другие жанры музыки, такие как этническая и классическая музыка, имеют около 450 генов. Система зависит от достаточного количества генов для получения полезных результатов. Каждому гену присваивается значение от 0 до 5 с шагом в половину целого.

Рекомендательная система сравнивает вектора различных песен и находит среди них похожие, чтобы на основании этих данных строить рекомендации для пользователей.

Music Genome Project был полностью разработан Pandora Media и остается основной технологией, используемой для Pandora Radio, его потокового музыкального сервиса. Ранее сервис лицензировал эту технологию для использования другими компаниями, но сегодня она используется только в Pandora Radio [20].

1.2.5 The Echo Nest и Spotify

The Echo Nest - это компания, выпускавшая программные инструменты для интеллектуального анализа музыки, которые могли использовать сторонние разработчики и медиакомпании.

Компания The Echo Nest появилась как исследовательское подразделение MIT Media Lab, основной задачей которого было анализ аудио и текстового содержания музыки.

Его создатели ставили перед собой ряд задач, связанных с музыкой: создание акустических отпечатков для идентификация музыкальных композиций по отдельным фрагментам, составление музыкальных рекомендаций и плейлистов для слушателей, анализ данных для потребителей и разработчиков.

В основу технологий компании The Echo Nest, основанной в 2005 году, легли диссертационные работы основателей компании Тристана Джехана и Брайана Уитмена [8, 18].

С 2010 по 2012 годы The Echo Nest получила финансирование в размере 24,3 млн. долларов США от нескольких венчурных фондов. Линейка продуктов The Echo Nest была основана на базе данных, содержащей информацию о 30 миллионах песен. Данные были получены посредством интеллектуального анализа имеющихся музыкальных композиций и сканирования сети Интернет для получения пользовательских отзывов об этих песнях. База данных была доступна для сторонних разработчиков через API. За все время существования компании данным API воспользовались более 7000 разработчиков для создания собственных музыкальных приложений.

Создатели The Echo Nest противопоставляли свои разработки музыкальному потоковому сервису Pandora – они также выделяли параметры, характеризующие музыкальные композиции, но в отличие от Pandora они делали это автоматически без участия человека и гораздо быстрее. На анализ каждой песни у Pandora уходит до 20-30 минут, а технологии The Echo Nest позволяют сделать это за несколько секунд.

The Echo Nest опубликовала базу данных об 1 млн. песен, которую сторонние разработчики и ученые могли использовать для своих исследовательских целей. В июне 2011 года компания выпустила Echoprint - библиотеку акустических отпечатков с открытым исходным кодом.

В марте 2014 года компания The Echo Nest была приобретена Spotify за 49,7 млн. евро. API для сторонних разработчиков был закрыт 31 мая 2016 года.

Spotify – музыкальный потоковый сервис, разработка которого началась в 2006 году в Стокгольме в Швеции. В то время в Швеции уже появились торрент-трекеры Kazaa и The Pirate Bay, работающие на технологии пиринговых сетей и способствовавшие незаконному обмену файлами. Первоначально, Spotify также работал на технологии пиринговых сетей и не хранил музыку на своих серверах.

Потребовалось несколько лет, чтобы убедить правообладателей стать участниками платформы – компания договорилась с Sony, Warner Music и Universal о размещении лицензионных треков, и в 2008 году сервис запустился в Швеции, Норвегии, Франция, Испания и Великобритания.

Сервис начал показывать рекламу и это позволило выплачивать вознаграждение артистам, которые согласились, чтобы их музыка присутствовала на Spotify. Шведская компания создала прибыльную платформу для правообладателей и предложила потребителям правовую альтернативу, которая позволяла им слушать музыку без необходимости ее приобретать или скачивать, не нарушая при этом авторских прав. Позднее была создана премиум-подписка, благодаря которой пользователям стала доступна музыка в лучшем качестве и без рекламы. Появление платной подписки позволило увеличить объем отчислений, выплачиваемых владельцам авторских прав.

Стриминговый сервис Spotify имеет два тарифа: бесплатное прослушивание музыки с наличием рекламы и премиум-подписку за \$9.99 в месяц без рекламы и с быстрой загрузкой. Компания в качестве отчислений выплачивает 70% от всех своих доходов владельцам авторских прав, которые затем в зависимости от договоренностей, распределяют доходы между исполнителями, студиями и другими участниками музыкальной индустрии. Указанные 70%, которые достаются правообладателям, распределяются между песнями соответственно количеству их прослушиваний пользователями Spotify.

Важным вкладом в успех Spotify стала его возможность взаимодействия с социальными медиа. Пользователи могут создавать плейлисты с песнями, соответствующими их интересам, настроению и предпочтениям. Созданные плейлисты можно отправлять другим пользователям, а также публиковать их в социальных сетях, например в Facebook. Такое социальное взаимодействие помогает открывать новых исполнителей, которые в противном случае остались бы незамеченными.

Согласно финансовому отчету 2014 года сервис Spotify работает в 58 странах и у него более 60 миллионов пользователей. 15 миллионов клиентов, имеющих премиум-подписку, принесли компании 90% выручки. Для продвижения премиум-подписки Spotify договорился с операторами мобильной связи и интернет-провайдерами, чтобы те включали подписку Spotify в пакет своих услуг. В результате, средний доход Spotify на одного пользователя премиум-подписки вырос до 73 долларов в 2014 в год. Несмотря на это, Spotify до сих пор не удается получать доход - при выручке в 1,37 млрд. долларов США в 2014 году операционные убытки по-прежнему составляют 226,14 млн. долларов США.

Spotify и другие компании были раскритикованы некоторыми исполнителями за низкие гонорары, выплачиваемые потоковыми музыкальными сервисами. Иногда известные исполнители откладывают или вообще воздерживаются от выпуска своих песен в этих сервисах, а вместо этого предпочитают продавать свою музыку в цифровых магазинах.

В конце 2016 года компания Spotify купила The Echo Nest и все ее технологии, после чего был закрыт доступ к API и всем прочим инструментам The Echo Nest.

В настоящий момент разработки The Echo Nest используются в музыкальных рекомендациях Spotify.

1.2.6 Рекомендательные системы

Рекомендательная система представляет собой подвид системы фильтрации информации, которая старается предсказать «рейтинг», который пользователь может дать объекту. Рекомендательные системы используются в самых разных областях, но более всего известны формированием рекомендаций для видео- и музыкальных сервисов, таких как Netflix, YouTube и Spotify, рекомендаций товаров для интернет-магазинов, таких как Amazon, или рекомендаций контента для социальных сетей, таких как

Facebook и Twitter. В рекомендательных системах чаще всего используются коллаборативная фильтрация и фильтрация на основе содержания.

На рисунке 3 изображены виды рекомендательных систем, а также данные, на основании которых строятся рекомендации для пользователей.

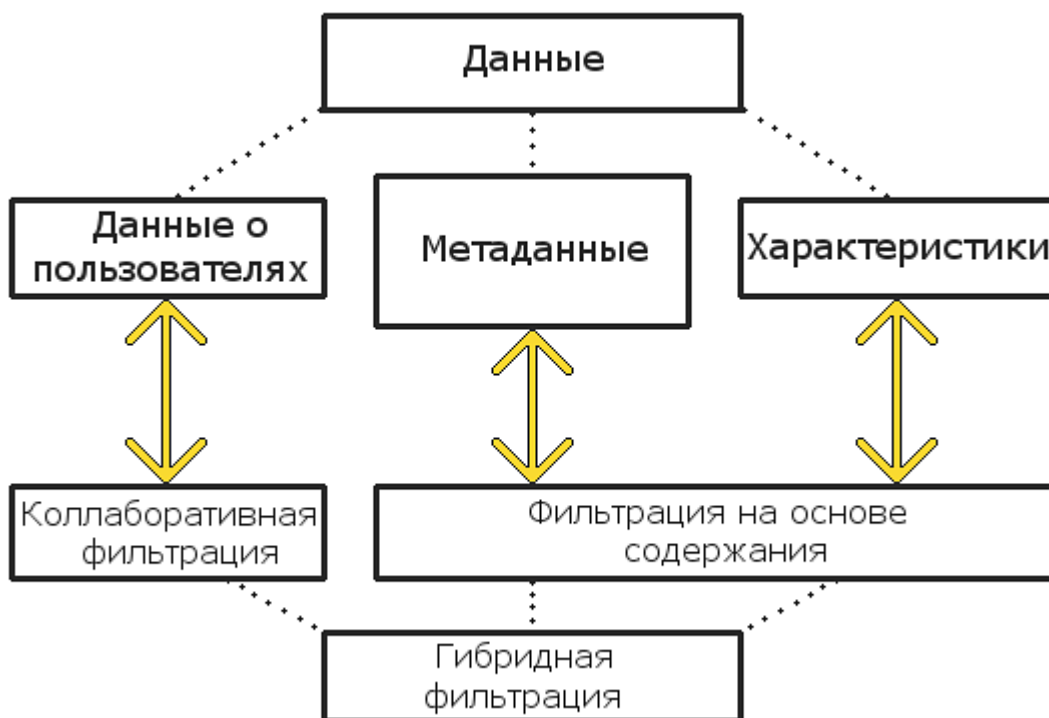


Рисунок 3 – Виды рекомендательных систем

Коллаборативная фильтрация создает модель на основе прошлого поведения пользователя (ранее приобретенные или выбранные товары, а также количественные оценки этих товаров), а также аналогичных решений, принятых другими пользователями. Затем эта модель используется для прогнозирования того, какие товары могут заинтересовать пользователя. Фильтрация на основе содержания использует ряд характеристик объекта с предварительно обозначенными количественными значениями, чтобы рекомендовать пользователям объекты со схожими характеристиками. Современные рекомендательные системы обычно используют сразу оба способа фильтрации.

Различия между коллаборативной фильтрацией и на основе содержания можно продемонстрировать, сравнив механизм построения рекомендаций в двух музыкальных сервисов - Last.fm и Pandora.

Last.fm сравнивает, какие группы или песни чаще всего слушает пользователь с тем, что слушают другие пользователи. На основании данного анализа создается плейлист с рекомендуемыми песнями, которых нет в библиотеке пользователя, но которые часто слушают другие пользователи со схожим музыкальным вкусом. Это пример коллаборативной фильтрации.

Pandora использует характеристики песни или исполнителя (множество из 400 атрибутов, определенных посредством Project Music Genome), чтобы создать плейлист, который будет содержать музыку со схожими характеристиками. Оценки пользователей, выставляемые музыке, используются, чтобы понять какие характеристики являются наиболее важными для каждого конкретного пользователя. Это пример фильтрации на основе содержания.

У каждого типа фильтрации есть свои сильные и слабые стороны. В приведенном выше примере Last.fm требует большого количества информации о пользователе, чтобы давать точные рекомендации. Это проблема холодного запуска, которая часто встречается в системах коллаборативной фильтрации. В то время как Pandora требуется немного меньше информации для того, чтобы начать работу, получаемые ею плейлисты гораздо менее разнообразны по жанровому составу и составу исполнителей.

Рекомендательные системы являются альтернативой поисковым алгоритмам, поскольку они помогают пользователям обнаруживать объекты, которые они не смогли бы найти иным способом.

1.2.6.1 Коллаборативная фильтрация

Одним из наиболее широко используемых подходов к созданию рекомендательных систем является коллаборативная фильтрация. Коллаборативная фильтрация основана на предположении, что в будущем людям будут нравиться те же вещи, которые им нравились в прошлом. Система делает рекомендации, используя только информацию о предпочтениях пользователей и оценках, которые они ставили объектам. Система рекомендует пользователю объекты, которые понравились другим людям со схожими предпочтениями.

Ключевое преимущество коллаборативной фильтрации заключается в том, что этот метод не нуждается в изучении свойств объекта, и, следовательно, способен точно рекомендовать такие сложные вещи, как фильмы или музыка. Многие социальные сети используют коллаборативную фильтрацию, чтобы изучив связи между пользователями, рекомендовать им новых друзей и группы.

Для построения модели поведения пользователя часто проводятся явные и неявные сборы данных.

Примеры явного сбора данных включают следующее:

- предложить пользователю оценить объект по какой-нибудь шкале;
- предложить пользователю расположить объекты в порядке от лучшего до худшего;
- показать пользователю два объекта и предложить выбрать из них лучший;
- предложить пользователю создать список объектов, которые ему нравятся.

Примеры неявного сбора данных включают в себя следующее:

- учет объектов, которые пользователь просматривал;
- учет количества времени, которое пользователь уделил просмотру каждого объекта;

- анализ аккаунтов пользователя в социальных сетях.

Рекомендательные системы, построенные на коллаборативной фильтрации, имеют несколько недостатков.

1 Холодный запуск.

Для того, чтобы давать пользователю точные рекомендации, необходимо собрать о нем большой объем информации. По этой причине тяжело создавать рекомендации для новых пользователей. Схожая проблема возникает и с появлением новых объектов – им тяжело попасть в рекомендации, пока никто из пользователей их не оценил.

2 Масштабируемость.

Многие сервисы, использующие рекомендательные системы с коллаборативной фильтрацией, имеют миллионы пользователей и миллионы товаров. Поэтому для построения рекомендаций, учитывающих отношения такого количества пользователей и товаров, требуются значительные вычислительные мощности.

3 Разреженность.

Для построения эффективных рекомендаций необходимо большое количество данных, но зачастую их не хватает, поскольку лишь небольшое число наиболее активных пользователей ставит оценки объектам [4].

1.2.6.2 Фильтрация на основе содержания

Другим распространенным подходом при разработке рекомендательных систем является фильтрация на основе содержания. Этот подход для построения рекомендаций использует характеристики объектов и предпочтения пользователей.

В рамках данного подхода каждый объект описывается набором характеристик, текстовых и числовых. По мере взаимодействия пользователя с рекомендательной системой, для пользователя формируется профиль предпочтений, где содержатся данные по заинтересовавшим его объектам.

Затем, на основании предпочтений пользователя, для него строятся рекомендации. Таким образом, система предлагает пользователю объекты, схожие с теми, что ему понравились ранее.

Пользовательский профиль постоянно обновляется – рекомендательная система анализирует, какие характеристики объектов являются наиболее важными для пользователя, и корректирует в его профиле соответствующие значения.

Ключевая проблема с фильтрацией на основе содержания заключается в том, не всегда возможно использовать пользовательские предпочтения в отношении одного типа объектов для других типов объектов. Когда система ограничена рекомендацией объектов того же типа, который уже используется пользователем, ценность такой системы рекомендаций значительно меньше, чем когда система может рекомендовать другие типы объектов из других сервисов. Например, гораздо эффективнее, когда на основе просмотра новостей можно рекомендовать музыку, видео, продукты и обсуждения из разных сервисов.

Рекомендательные системы на основе содержимого также могут включать системы рекомендаций на основе мнений. В некоторых сервисах пользователи имеют возможность оставлять текстовые отзывы о товарах. Эти отзывы являются неявными данными и также могут использоваться рекомендательной системой для описания объекта и оценки отношения пользователя к этому объекту [9, 11].

1.2.6.3 Гибридные рекомендательные системы

Большинство рекомендательных систем в настоящее время является гибридными, сочетая коллаборативную фильтрацию и фильтрацию по содержанию. Исследования показали, что гибридные рекомендательные системы в большинстве случаев дают более точные рекомендации, чем рекомендательные системы, использующие лишь один вид фильтрации [3].

Благодаря сочетанию нескольких видов фильтрации рекомендательные системы лишены различных проблем, которые свойственны отдельным видам фильтрации.

Два главных преимущества объединения моделей — это увеличение точности рекомендаций и возможность более гибкой настройки на разные группы клиентов. Недостатки — меньшая интерпретируемость и большая сложность реализации и поддержки.

Существует несколько стратегий для формирования рекомендаций в гибридных рекомендательных системах:

1. Взвешенный — считать средневзвешенный прогноз по нескольким оценкам;
2. Комбинация — предсказания отдельных моделей объединяются и передаются в единый алгоритм рекомендаций;
3. Переключение — для разных продуктов/пользователей применять различные алгоритмы;
4. Смешивание — вычисляются рекомендации по разным алгоритмам, а потом просто объединяются в один список.

1.2.6.4 Оценка эффективности рекомендательных систем

Основной характеристикой для оценки рекомендательных систем является точность создаваемых ими рекомендаций. Чаще всего для измерения точности используется среднеквадратическое отклонение. Иногда вместо измерения точности проводятся пользовательские исследования. Для этого несколько десятков или сотен пользователей получают рекомендации, созданные с использованием различных методик. После чего пользователи оценивают, какие из рекомендаций являются лучшими.

Кроме точности существует ряд других характеристик, которые также важно учитывать при оценке рекомендательных систем.

1 Разнообразие.

Пользователи, как правило, более удовлетворены рекомендациями, которые предлагают им разнообразные товары, например музыку от разных исполнителей.

2 Постоянство рекомендации.

У пользователей иногда может не быть времени, чтобы внимательно изучить рекомендации. Поэтому иногда необходимо повторно показывать те же самые рекомендации или давать возможность пользователям еще раз оценить объекты.

3 Конфиденциальность.

Рекомендательные системы часто сталкиваются с проблемами конфиденциальности, потому что многие пользователи предпочли бы скрыть свои вкусы от других людей. Создание профилей пользователей с использованием коллаборативной фильтрации также может быть проблематичным с точки зрения конфиденциальности. Связи, построенные такой рекомендательной системой между пользователями и объектами, позволяют идентифицировать людей в обезличенном наборе данных.

4 Новизна.

Мера того, «насколько удивительны рекомендации». Например, рекомендательная система, которая рекомендует молоко покупателю в продуктовом магазине, может быть совершенно точной, но это не очень хорошая рекомендация, потому что это предлагает пользователю очевидный факт [12, 14].

1.2.7 Использование рекомендательных систем в музыкальных потоковых сервисах

Музыкальные потоковые сервисы также используют рекомендательные системы. С их помощью сервисы пытаются понять музыкальный вкус пользователя и предлагать ему соответствующую музыку.

Чаще всего в музыкальных потоковых сервисах используются гибридные рекомендательные системы, созданные на основе сразу двух подходов. При этом большинство музыкальных потоковых сервисов для построения рекомендаций использует для этого описательные метаданные музыкальных композиций. Такими метаданными являются: название музыкальной композиции, имя исполнителя, жанр, темп и прочее. Чаще всего эта информация предоставляется автором или издателем музыки. Вместо метаданных можно использовать характеристики, относящиеся непосредственно к музыке, но так делают лишь два музыкальных сервиса: Pandora и Spotify. Этот подход является затратным, из-за необходимости проводить анализ значительного количества музыки. Но в тоже время, такой подход позволяет этим сервисам делать более точные музыкальные рекомендации.

- Pandora – этот музыкальный потоковый сервис для классификации использует не только описательные метаданные, но и параметры, характеризующие звуковую составляющую в песнях. Ключевой особенностью является то, что значения параметров для музыкальных композиций устанавливают эксперты – специалисты компании. В среднем на работу с одной песней уходит до 30 минут [20].

- The Echo Nest/Spotify – компания The Echo Nest так же, как и Pandora выделила большой набор характеристик, но описывающих не только особенности звучания музыкальных композиций, но и особенности их аудиосигнала. Это стало возможным благодаря тому, что в The Echo Nest смогли автоматизировать процесс определения значений характеристик.

До недавнего времени компания предоставляла сторонним разработчикам доступ к своим инструментам. Благодаря этому сторонние разработчики могли создавать различные музыкальные сервисы. Однако, после приобретения компании сервисом Spotify, доступ ко всем этим инструментам был закрыт.

Компании The Echo Nest удалось решить главную проблему фильтрации на основе содержания. У нее получилось автоматизировать процесс определения значений звуковых характеристик, и на основании этих данных провести классификацию музыкальных композиций. Благодаря тому, что используются не метаданные, а звуковые параметры, характеризующие музыку, рекомендательная система Spotify показывает лучшие результаты по сравнению с другими музыкальными сервисами [19].

По причине того, что доступа к технологиям The Echo Nest больше нет, возникает потребность в автоматизированном методе, позволяющем определять значение звуковых характеристик музыкальных композиций, чтобы на основании этих данных находить похожие по звучанию композиции. В дальнейшем такой метод может быть использован в рекомендательных системах.

Для создания такого метода необходимо решить следующие задачи:

- сформировать список звуковых характеристик, на основе которых будут сравниваться музыкальные композиции;
- разработать алгоритм определения значений выбранных звуковых характеристик, а также метод для сравнения и сортировки музыкальных композиций по этим характеристикам;
- провести эксперимент и сравнить результаты работы разработанного метода с результатами существующих музыкальных сервисов.

2 Исследование и построение решения

В первую очередь необходимо определить звуковые характеристики, по которым можно было бы сравнивать музыкальные композиции. Выбирая характеристики, следует учитывать следующее:

- какую информацию содержит характеристика, насколько существенно ее влияние на звучание музыкальной композиции;
- возможность реализации автоматизированного определения значения характеристики.

2.1 Выбор звуковых характеристик для анализа музыкальной композиции

Перечислим характеристики, свойственные музыке.

1 Ноты.

Содержащаяся информация:

- Поскольку практически любая музыкальная композиция представляет собой мелодии, составленные из нот, то знание нотации позволяет нам получить исчерпывающее представление о композиции.

Возможность реализации:

- Крайне сложно реализовать на практике из-за того, что в музыке смешивается звучание множества инструментов, каждый из которых имеет свои ноты и мелодии.

- Некоторые музыкальные композиции атональны или состоят только из шумов, а значит определить звучащие ноты невозможно.

2 Тональность.

Содержащаяся информация:

- По тональности можно многое узнать о музыкальном произведении, например, его настроение - веселое оно или грустное. Также

можно определить его культурное происхождение, например, европейская это музыка или арабская.

Возможность реализации:

- Поскольку тональность – это лишь набор рекомендательных правил, чтобы подобрать ноты для звучания, которое задумал композитор, то они соблюдаются не всегда. Это снижает возможность точно определить тональность музыкальной композиции.

3 Темп.

Содержащаяся информация:

- По темпу также можно оценить характер музыки: танцевальная (и вид танца, для которого создавалась музыка), медленная (а значит чаще всего и грустная), быстрая (подходит для какой-то активности, например, спортивной).

Возможность реализации:

- Можно определить количество тактов в минуту. Такт – это расстояние между сильными долями музыкальной композиции.

Сложность заключается в том, что, даже узнав число тактов в минуту, нельзя с уверенностью утверждать - быстрая это композиция или медленная. Связано это с тем, что такт обозначает расстояние между частями музыкальной композиции, но не определяет скорость ее исполнения. На практике это приводит к тому, что одну и ту же мелодию можно записать в два такта, и исполнить ее в быстром темпе, а можно записать в один такт, и исполнить в два раза медленнее. Для слушателя же не будет никакой разницы – в обоих случаях музыка будет звучать одинаково.

4 Инструменты, звучащие в музыкальной композиции.

Содержащаяся информация:

- Некоторым жанрам характерно использование определенных музыкальных инструментов. А значит, если определить наличие тех или иных инструментов в музыкальной композиции, то можно определить и ее

жанр. Например: гитара свойственна року, обилие скрипичных инструментов означает, что играет барочная музыка.

Возможность реализации:

- Определить наличие инструментов, звучащих в музыкальной композиции, сложно по многим причинам, например: инструменты в музыке могут звучать одновременно, их звучание может сильно варьироваться, а также одни инструменты могут подражать другим.

5 Характер и техника исполнения.

Содержащаяся информация:

- Позволяет еще более точно определить нюансы звучания музыки: спешный фортепианный этюд, гитарный тэппинг, горловое пение, нестандартное исполнение классического оркестрового произведения и т.д.

Возможность реализации:

- Это сложная для решения задача. Единственным выходом видится создание интеллектуальной системы, которая сама будет находить особенности в одних музыкальных композициях и искать их в других музыкальных композициях.

6 Частотный баланс.

Содержащаяся информация:

- Поскольку музыка – это набор волн различной частоты и величины, то анализ этих волн позволяет нам получить совокупное представление о музыкальной композиции.

Возможность реализации:

- Существующие алгоритмы по обработке звука позволяют реализовать анализ данной характеристики в полной мере.

Для поиска схожих музыкальных композиций лучше всего подойдут частотный баланс, темп и тональность. Их достаточно, для того чтобы в полной мере описать песню, а также их анализ возможно автоматизировать. Однако, если темп и тональность – это известные характеристики и не

нуждаются в дополнительном пояснении, то сравнение частотных балансов необходимо обосновать.

2.2 Обоснование использования частотного баланса для поиска схожих по звучанию музыкальных композиций

Для демонстрации механизма анализа частотного баланса сравним два жанра: танцевальную и рок-музыку. Для этого проанализируем несколько музыкальных композиций, относящихся к каждому из жанров.

На рисунке 4 изображены частотные балансы в жанрах рок-музыка и танцевальная музыка.

Узкой линией обозначен график танцевальной музыка, широкой линией – график рок-музыки. Прямоугольниками выделены наиболее различающиеся диапазоны частот.

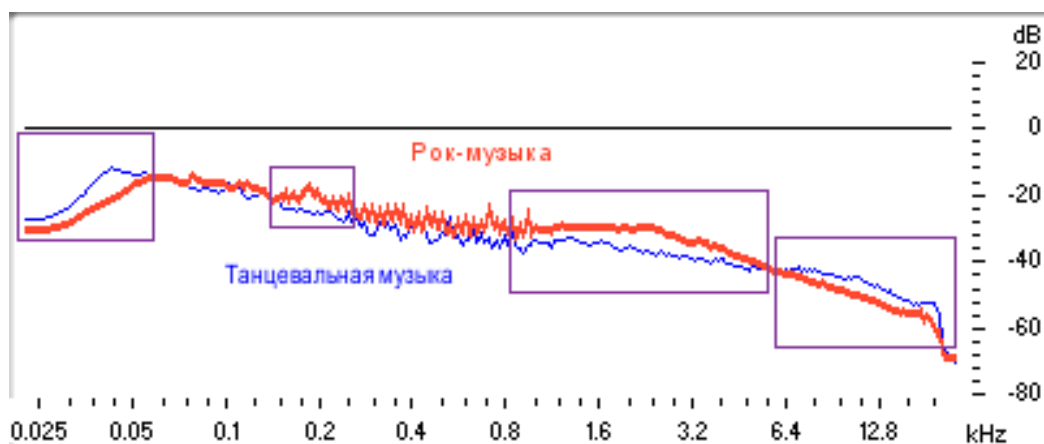


Рисунок 4 – Сравнение танцевальной и рок-музыки

Графики показывают, что наибольшее расхождение у жанров присутствует в суб-басах, на небольшом участке у средних частот, а также на всем протяжении верхних частот. Из этого можно сделать вывод, что между жанрами существуют не только различия в звучании, но и в частотном балансе.

Однако, трудно сказать насколько существенны эти различия без результатов других сравнений. Поэтому для представления противоположной ситуации найдем схожие по звучанию песни. Лучший вариант в такой ситуации – рассмотреть какой-нибудь отдельный музыкальный альбом. Дело в том, что перед выпуском на альбоме все песни на нем приводятся к звуковому единообразию.

Для исследования возьмем альбом «Трава у дома» группы «Земляне», которая исполняет музыку в жанре рок.

Проанализируем частоты всего альбома и отдельной песни с этого альбома – «Поверь в мечту», чтобы сравнить.

На рисунке 5 изображены частотные балансы песни «Поверь в мечту» и всего альбома «Трава у дома».

Узкой линией обозначен график музыкального альбом «Трава у дома», а широкой линией - график песни «Поверь в мечту».

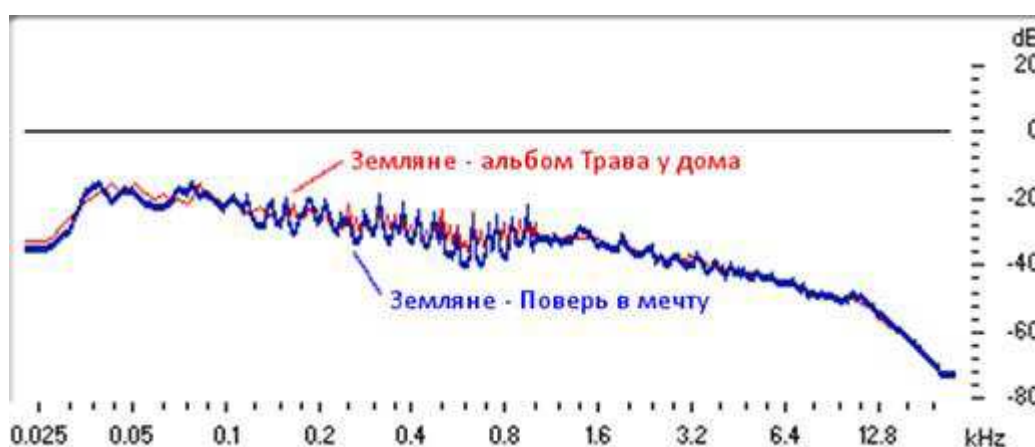


Рисунок 5 – Сравнение песни «Поверь в мечту» с альбомом «Трава у дома» группы «Земляне»

Как видно, частотные кривые альбома и песни практически совпадают. На этом фоне расхождения между частотами в предыдущем сравнении выглядят значительными.

На основе этого следует вывод, что схожие по звучанию музыкальные композиции также схожи по частотному балансу, а отличающиеся по звучанию имеют расхождения в частотном балансе.

2.3 Условия для применения

Для анализа характеристик музыкальных композиций необходимо, чтобы вся используемая музыка имела один формат. В предложенном методе будет использоваться формат MPEG-3 Layer 3 с битрейтом 320 кбит/с и частотой дискретизации 44100 Гц, поскольку это наиболее распространенный формат для хранения музыки.

2.4 Определение частотного баланса музыкальных композиций

Баланс частот музыкальной композиции будет основной характеристикой метода для поиска и сравнения песен между собой.

Для работы будут использоваться диапазон частот, в котором слышит человек - от 25 Гц до 20 кГц. При работе с частотами в музыкальной индустрии следуют стандарту ISO 266 и используют логарифмическую шкалу. Для составления частотного баланса музыкальной композиции также будет использована логарифмическая шкала. Разместим на ней через равное расстояние 500 точек, каждая из которых будет обозначать соответствующую частоту. Чем больше точек-частот будет использоваться для анализа, тем выше будет его точность. В данной работе было решено использовать 500 точек, потому что увеличение числа точек больше этого количества показывает незначительный прирост точности. Отсчет точек будет вестись от нижней границы в 25 Гц ($n = 1$) до верхней границы в 20 кГц ($n = 500$).

Для равномерного распределения частот по логарифмической шкале была использована следующая формула (1):

$$F = 25 * 1.0135^{n-1}, \quad (1)$$

где F – это значение частоты n ;

n – это порядковый номер частоты, имеет диапазон от 1 до 500.

На каждый отсчет дискретизации песни для каждой частоты мы будем определять уровень энергии звука. После этого мы посчитаем среднее арифметическое для уровня энергии каждой из частот. Полученная совокупность значений и будет представлять собой частотный баланс.

Формула для вычисления частотного баланса музыкальной композиции (2):

$$M_n = \frac{\sum_{s=1}^s E_n^s}{s}, \quad (2)$$

где M_n – это среднее арифметическое значение энергии звука частоты n ;

n – это порядковый номер частоты, имеет диапазон от 1 до 500;

s – это номер отсчета дискретизации;

E_n^s – это уровень громкости частоты n в отсчете s .

На рисунке 6 изображено как определяется уровень энергии звука для частоты 25 Гц в одном отсчете дискретизации.

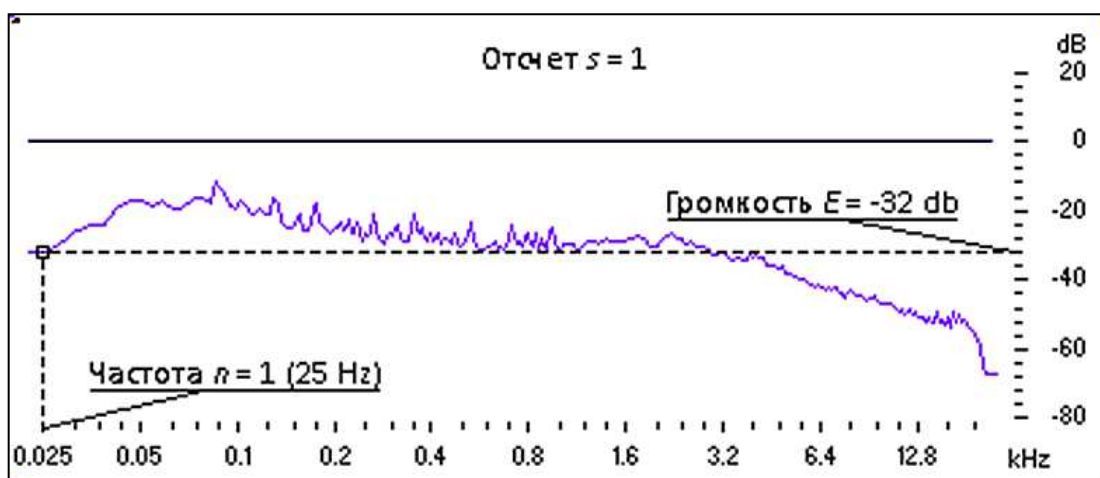


Рисунок 6 – Определение уровня энергии звука для частоты 25 Гц

2.5 Определение темпа музыкальных композиций

Человеческий слух определяет ритм музыки как периодическую последовательность ударов. Сигнал, который перехватывается ухом, содержит определенную энергию. Чем больше энергии содержит звук, тем громче он нам кажется. Но звук будет слышен как удар, только если его энергия в значительной степени превосходит энергетическую историю звука, то есть если мозг обнаруживает жесткую вариацию звуковой энергии. Таким образом, ритм определяется вариацией звуковой энергии. Поэтому чтобы найти всплеск звуковой энергии – необходимо сравнивать его со средней энергией окружающего звука. При этом нет смысла сравнивать со всей музыкальной композицией, потому что в ней могут присутствовать как громкие части, так и тихие [17].

Необходимо определить какие по продолжительности отрезки музыки следует использовать для определения всплесков звуковой энергии.

Для этого необходимо взять минимальный размер ноты - 1/64 такта, и определить ее длительность. Известно, что чем выше темп музыкальной композиции, тем короче будет звучать эта нота.

Согласно исследованию Д. Оливета, посвященному поиску закономерностей в музыке, самый быстрый темп, используемый в музыке, составляет 238 BPM [10]. В таком темпе 1/64 такта будет иметь продолжительность 15 мс.

При частоте дискретизации в 44100 Гц, длительность в 15 мс составит 661,5 отсчета. Поскольку отсчеты не могут быть разделены на части, то округлим дробную величину и получим 662 отсчета. Для дальнейших расчетов она будет обозначена как c .

Посчитаем количество звуковой энергии в выделенном отрезке музыкальной композиции по формуле (3):

$$E_c = \sum_{s=1}^c E^s, \quad (3)$$

где E_c – это количество звуковой энергии, содержащейся в c отсчетов;

c – это количество отсчетов содержащееся в 1/64 такта;

s – это порядковый номер отсчета дискретизации;

E^s – это количество энергии звука, содержащейся в отсчете s .

Для определения ближайшей средней звуковой энергии достаточно использовать фрагмент музыки длительностью в 1 с. Если взять фрагмент короче, то всплески будут восприниматься как шум, а не как ритмичные удары. Если фрагмент будет длиннее, то это негативно скажется на точности расчетов. Для анализа музыкальных композиций используется частота дискретизации 44100 Гц, следовательно, длительность 1 секунды составляет 44100 отсчетов.

Поскольку ближайшая средняя энергия звука будет сравниваться с количеством энергии для выделенного отрезка, то для этого необходимо

определить, сколько таких же отрезков звука длительностью в 662 отсчета содержится в 1 с.

Определив общий уровень энергии звука для фрагмента в 1 с и поделив его на количество отрезков, мы найдем ближайшую среднюю энергию звука по формуле (4):

$$E_{avg} = \frac{1}{R/c} \sum_{s=1}^R E^s, \quad (4)$$

где E_{avg} – это ближайшая средняя энергия звука, содержащаяся в R отсчетов;

R – это частота дискретизации;

c – это количество отсчетов содержащееся в 1/64 такта;

s – это порядковый номер отсчета дискретизации;

E^s – это количество энергии звука, содержащейся в отсчете s .

Далее следует определить насколько большой должна быть разница между количеством звуковой энергии выделенного участка и ближайшей средней энергией звука, чтобы можно было сделать вывод, что в выделенном отрезке звучит удар. Экспериментальным путем удалось выяснить, что разница в 1,3 раза является оптимальной. Обозначим ее как K .

Следовательно, удар будет слышен, если будет выполняться следующее условие по формуле (5):

$$E_c \geq K * E_{avg}, \quad (5)$$

где E_c – это количество звуковой энергии, содержащейся в выделенном отрезке звука;

E_{avg} – это ближайшая средняя энергия звука;

K – это коэффициент разницы между этими энергиями.

Обнаружив удары, мы можем вычислить время между ними, а из него рассчитать темп музыкальной композиции. Темп считается как количество четвертей такта, звучащих за минуту.

Формула для вычисления темпа музыкальной композиции (6):

$$BPM = \frac{60}{t_2 - t_1}, \quad (6)$$

где BPM – это темп музыкальной композиции;

t_1 – это время первого удара;

t_2 – это время второго удара.

Поскольку данный алгоритм может обнаруживать не только удары, звучащие раз в четверть, но и в половину, и в 1/8 такта, то BPM может получиться слишком большим или маленьким. Для BPM оптимальным будет значение между 100 и 200, и чтобы темп оказался в этом диапазоне, его можно умножить, либо разделить на 2 или 4.

2.6 Определение тональности музыкальных композиций

Тональность музыкальной композиции можно определить, если понять, какие ноты звучат в композиции. Как было уже сказано в главе 2.1, определение нот, звучащих в музыкальной композиции – это сложная задача, которая не решена в полной мере до сих пор из-за того, что в музыке часто звучат несколько нот и инструментов одновременно.

Есть ряд проблем, которые мешают разработать совершенный механизм по определению звучащих нот. Но, поскольку стоит более общая задача – определить тональность, то для ее решения нет необходимости знать все ноты, звучащие в композиции – для этого будет достаточно некоторого количества.

В музыке существует частотный диапазон, где инструменты не смешиваются, потому что исполнители стараются сохранить в нем чистое звучание - это саб-бас. Также задачу упрощает то, что в нижних частотах очень часто звучат основные ноты тонального ряда.

Для определения нот будет использоваться диапазон частот от 25 до 63 Гц, в котором располагаются саб-басы. Частоты ниже 25 Гц человек не слышит, поэтому при микшировании музыки их обрезают для сохранения четкой слышимости басов. В этих частотах располагаются ноты контроктавы и субконтроктавы.

В таблице 1 перечислены ноты контроктавы и субконтроктавы, а также их частоты.

Таблица 1 – Контроктава и субконтроктава

Название ноты	Латинское обозначение	Частота, Гц
Си ₁	B1	61,7
Ля ₁ # (Си ₁ b)	A1# (B1 b)	58,3
Ля ₁	A1	55
Соль ₁ # (Ля ₁ b)	G1# (A1 b)	51,9
Соль ₁	G1	49
Фа ₁ # (Соль ₁ b)	F1# (G1 b)	46,2
Фа ₁	F1	43,7
Ми ₁	E1	41,2
Ре ₁ # (Ми ₁ b)	D1# (E1 b)	38,9
Ре ₁	D1	36,7
До ₀ # (Ре ₁ b)	C0# (D1 b)	34,6
До ₀	C0	32,7
Си ₀	B0	30,9
Ля ₀ # (Си ₀ b)	A0# (B0 b)	29,1
Ля ₀	A0	27,5
Соль ₀ # (Ля ₀ b)	G0# (A b)	26

Каждый отсчет дискретизации определяется частота с наибольшей энергией звука, после чего определяется ближайшая к ней по частоте нота по формуле (7):

$$N^s \approx \max\{F_1^s, \dots, F_x^s\}, \quad (7)$$

где N^s – это музыкальная нота, звучащая в отсчете s ;

$\{F_1^s, \dots, F_x^s\}$ – это диапазон частот от 25 до 63 Гц;

s – это порядковый номер отсчета дискретизации.

На основании нот, прозвучавших в саб-басу, определяется тональность музыкальной композиции и ее тональный лад – мажорный или минорный.

2.7 Сравнение музыкальных композиций по звуковым характеристикам

Сначала необходимо сравнить музыкальные композиции отдельно по каждой из звуковых характеристик. Из полученных значений будет вычислена итоговая разница между композициями.

Для того чтобы понять насколько две музыкальные композиции различаются по частотному балансу, следует найти у них разницу между соответствующими частотами. Поскольку важна лишь величина разницы, то она считается без учета знака.

Величина разницы частотных балансов будет равна отношению суммы полученных значений к числу используемых частот и находится по формуле (8):

$$D = \frac{\sum_{n=1}^{500} |M_n^a - M_n^b|}{n}, \quad (8)$$

где D – это относительная разница между частотными балансами музыкальных композиций a и b ;

M_n – это среднее арифметическое значение энергии звука частоты n ;

a – это выбранная музыкальная композиция;

b – это музыкальная композиция, которая сравнивается с выбранной музыкальной композицией;

n – это порядковый номер частоты, имеет диапазон от 1 до 500.

Величина разницы темпов у двух музыкальных композиций будет равна отношению разницы в темпах к значению темпа выбранной композиции. Разница между темпами считается без учета знака и рассчитывается по формуле (9):

$$BPM_{\Delta} = \frac{|BPM^a - BPM^b|}{BPM^a}, \quad (9)$$

где BPM_{Δ} – это относительная разница в темпе между музыкальными композициями a и b . BPM_{Δ} принимает значение от 0 до 1, где 0 означает, что у композиций одинаковый темп, а 1 означает, что темпы композиций различаются в два раза.

Далее музыкальные композиции сравниваются по тональностям. Опытным путем было установлено, что если у композиций различаются

тональные лады, то итоговое значение разницы следует увеличить на 10%. Обозначим этот параметр как L .

Итоговая относительная разница между двумя музыкальными композициями по результатам сравнения звуковых характеристик будет считаться следующим образом по формуле (10):

$$G = D * (1 + BPM_{\Delta} + L), \quad (10)$$

где G – итоговое значение разницы между двумя музыкальными композициями, принимает положительное значение, где, чем больше число, тем сильнее разница в звучании между композициями, а 0 означает, что композиции идентичны по звучанию;

L – параметр, принимающий значение 0.1, если сравниваемые композиции имеют различные тональные лады, и 0 если одинаковые.

3 Реализация

Методика была реализована в виде компьютерного приложения для Windows, работающего с базой данных. Приложение было разработано на Python версии 3.6.5, а для хранения информации о музыке была использована база данных PostgreSQL версии 10.3.

На рисунке 7 изображен внешний вид разработанного приложения.



Рисунок 7 – Внешний вид приложения

3.1 Подготовка к эксперименту

Программой по разработанному методу было проанализировано 25 тысяч музыкальных композиций различных жанров. Музыкальные композиции имели формат MPEG-3 Layer 3 с битрейтом 320 кбит/с и частотой дискретизации 44100 Гц.

Значения звуковых характеристик, полученные по результатам анализа, были загружены в базу данных.

На рисунке 8 изображен фрагмент базы данных с полученной информацией.

SONG_ID	F_1	F_2	F_3	F_4	F_5	F_6	F_7
12590	-32	-32	-32	-32	-32	-32	-32
12589	-33,9	-33,9	-33,9	-33,9	-33,9	-33,9	-33,9
12588	-32,3	-32,2	-31,8	-30,7	-31	-32	-32
12587	-31	-31	-31	-32	-31	-31	-31
12586	-32,1	-32	-31,9	-31	-32	-31	-32
12585	-33	-32	-31	-32,1	-31	-30	-32
12584	-33,9	-33,8	-33,9	-33,7	-33,7	-33,7	-33,9
12583	-33	-32,9	-32,9	-32,9	-32,9	-33,9	-32,9
12582	-32,2	-32	-31	-31	-31	-32	-32
12581	-32,4	-32	-31	-31	-32	-31,5	-32
12580	-32	-32,1	-31,1	-30	-31	-31	-32
12579	-34	-33,9	-33,2	-33	-32,5	-32	-32
12578	-32,3	-32	-31	-32,2	-31	-32	-32

Рисунок 8 – Фрагмент базы данных

3.2 План эксперимента

Для проведения эксперимента было выбрано 5 музыкальных композиций следующих жанров:

- хип-хоп;
- электронная танцевальная музыка;
- метал;
- брейкор;
- оркестровая музыка.

Выбор жанров обоснован тем, что некоторые из них имеют хорошо узнаваемое звучание, в то время, как элементы других жанров присутствуют в большом количестве музыки, а потому их не так легко определить. Было отобрано 10 испытуемых, среди которых 5 имеют оконченное начальное музыкальное образование. Один из испытуемых несколько лет занимается

написанием электронной музыки, еще один из испытуемых несколько лет играет на электрогитаре. Также был приглашен эксперт, работающий звукорежиссером, для проведения дополнительной оценки.

Сначала испытуемые должны были прослушать предложенные музыкальные композиции. После этого требовалось найти в сети Интернет по 10 схожих по звучанию песен для каждой из композиций. Чтобы показать испытуемым насколько трудоемким может быть поиск необходимых композиций, им было запрещено пользоваться потоковыми музыкальными.

На следующем этапе эксперимента испытуемые должны были загрузить 5 выбранных композиций в разработанное приложение, а также в Google Play Music и Spotify. Затем они должны были оценить первые 10 песен, предложенных каждым из сервисов на каждую из загруженных композиций. Оценивалось сходство по звучанию между загруженной музыкальной композицией и предложенной песней. Оценка производилась по шкале от 1 до 10, где 1 означает, что композиции совершенно не похожи по звучанию, а 10 означает, что композиции имеют идентичное звучание. Всего каждый испытуемый выставил по 50 оценок каждому из приложений.

Каждый испытуемый выполнял задания самостоятельно, ни один из них не видел, как работали другие испытуемые.

В заключительном этапе принимал участие эксперт, который провел дополнительную оценку всех композиций предложенных сервисами в ходе эксперимента.

3.3 Результаты эксперимента

Поиск схожих по звучанию музыкальных композиций вручную даже для одной композиции у некоторых пользователей занял до 30 минут. Это демонстрирует тот факт, насколько сложно найти необходимую музыку, даже имея доступ к сети Интернет.

Из выставленных оценок для каждого испытуемого методом нахождения среднего арифметического была рассчитана обобщенная оценка для каждого из приложений. После чего таким же способом для каждого приложения были получены оценки эксперта и итоговые оценки.

В таблице 2 приведены обобщенные оценки, данные испытуемыми и экспертом, а также итоговые оценки для каждого из сервисов.

Таблица 2 – Результаты эксперимента

Номер испытуемого	Разработанное приложение		Google Play Music		Spotify	
	Оценка испытуемого	Оценка эксперта	Оценка испытуемого	Оценка эксперта	Оценка испытуемого	Оценка эксперта
1	7,1	6,8	8,2	8,5	9	9,5
2	4,9	6,8	8,2	8	9,5	9,2
3	7,8	6,8	5,8	8,2	9,5	9,4
4	4	6,8	8	8,2	10	9,4
5	5,5	6,8	9,1	8,4	9	9,4
6	6	6,8	8	8,5	10	9,5
7	7	6,8	7,4	8,4	10	9,5
8	7,5	6,8	8	8,3	9,5	9,3
9	5,2	6,8	8,5	8	9	9,2
10	7	6,8	9,8	8,5	9,5	9,6
Итоговый результат	6,2	(+9,7%) 6,8	8,1	(+2,5%) 8,3	9,5	(-1,1%) 9,4

Расхождения в оценках испытуемых и эксперта составили менее 10%, что говорит о достоверности данных оценок. Эксперимент показал, что предложенный в работе метод позволяет решить проблему ориентации в больших объемах музыки. Согласно оценкам эксперта программа, разработанная на основе предложенной методики, дает лишь на 28% менее точные рекомендации, чем Spotify. Эта разница не столько значительна, и при увеличении объемов музыкальной библиотеки, результат работы метода может быть улучшен.

ЗАКЛЮЧЕНИЕ

В данной работе рассматривается проблема, возникающая при необходимости ориентации в большом количестве музыки. Развитие технологий по работе со звуком сделало создание музыки доступным для многих людей, что привело к значительному увеличению ее количества и разнообразия.

В настоящее время эта проблема решается с помощью рекомендательных систем, которые могут оценивать музыкальные предпочтения пользователя и предлагать ему подходящую музыку. Однако наиболее точный вид рекомендательных систем – с фильтрацией на основе содержания, в большинстве случаев ищет схожие музыкальные композиции по описательным данным.

В качестве альтернативы было предложено разработать метод, который бы позволял находить схожие по звучанию песни, основываясь на анализе их звуковых характеристик.

В результате работы был создан алгоритм определения значений частотного баланса, темпа и тональности музыкальных композиций. Также был разработан метод поиска и сортировки музыкальных композиций по степени сходства их звуковых характеристик.

На основе метода была реализована прикладная программа, работающая вместе с базой данных. Программа позволила провести эксперимент и апробировать разработанный метод. Испытуемые получили песни и попробовали найти схожие по звучанию песни, используя предложенную программу и популярные потоковые сервисы. По итогам эксперимента программа доказала действенность разработанного метода и имеющийся у него потенциал – точность ее рекомендаций всего на 28% меньше чем у Spotify.

Предложена методика автоматизированной сортировки музыкальных композиций по степени сходства их звуковых характеристик: частотного

баланса, темпа и тональности, полученной методом анализа саб-басовых частот.

Разработанную методику можно использовать для совершенствования существующих музыкальных сервисов, использующих рекомендательные системы с фильтрацией по содержанию. Также исходный код разработанной программы можно выложить в свободный доступ, что позволит другим разработчикам использовать его в свои программных продуктах.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1 Величкин, Р. Б. Разработка метода поиска музыкальных композиций схожих по звуковым характеристикам / Р.Б. Величкин // Молодой ученый. — 2019. — №27. — Режим доступа: <https://moluch.ru/archive/265/61328>.

2 Тюлин, Ю. Н. Октавная система / Ю. Н. Тюлин // Краткий теоретический курс гармонии. : Издательство «Музыка» – Москва, 1978. – Гл. 5. – С. 93–115.

3 Bouneffouf, Djallel (2012), “Following the User’s Interests in Mobile Contex-Aware Recommender Systems: The Hybrid-e-greedy Algorithm”, Proceedings of the 2012 26th International Conference on Advanced Inforamtion Networking and Applications Workshops, Lecture Notes in Computer Science, IEEE Computer Society, pp. 657-662, on 2014-05-14.

4 Breese, John S. (1998). Empirical Analysis of Predictive Algorithms for Collaborative Filtering. [Электронный ресурс] / John S. Breese, David Heckerman, Carl Kadie // Microsoft Corporation. – Режим доступа: <https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/tr-98-12.pdf>.

5 Coffey, Aoife. (2016). The impact that music streaming services such as Spotify, Tidal and Apple Music have had on consumers, artists and the music industry itself. [Электронный ресурс] / Aoife, Coffey // University of Dublin. – Режим доступа: <https://scss.tcd.ie/publications/theses/diss/2016/TCD-SCSS-DISSERTATION-2016-027.pdf>.

6 Eriksson, Maria. Close Reading Big Data: The Echo Nest and the Production of (Rotten) Music Metadata. [Электронный ресурс] / Maria, Eriksson // First Monday. – 2016 – №21-7. – Режим доступа: <https://firstmonday.org/ojs/index.php/fm/article/view/6303/5530>.

7 Hilbert, M. (2012). How to Measure “How Much Information”? Theoretical, Methodological, and Statistical Challenges for the Social Sciences (pp. 1042-1055). Internatonal Journal of Communication.

8 Jehan, Tristan. (2005). Creating Music by Listening. [Электронный ресурс] / Tristan, Jehan // MIT Media Lab. – Режим доступа: <http://web.media.mit.edu/~tristan/phd/dissertation>.

9 Kaitila, Juuso. (2017). A content-based music recommender system. [Электронный ресурс] / Juuso, Kaitila // University of Tampere. – Режим доступа: <https://trepo.tuni.fi/bitstream/handle/10024/101198/GRADU-1495623946.pdf>.

10 Olivet, D. BMPs Musicians Use Most. [Электронный ресурс] / D. Olivet // Tonal Trends Pop Music Theory for Songwriters. – Режим доступа: <http://www.tonaltrends.com/bpms-musicians-use-most/4580116161>.

11 Prem Melville, Raymond J. Mooney, and Ramadass Nagarajan Agus. (2002) Content-Boosted Collaborative Filtering for Improved Recommendations. Proceedings of Eighteen National Conference on Artificial Intelligence (AAAI-2002), pp. 187-192, Endmonto, Canada, July 2002.

12 Ricci, Francesco. Recommender Systems Handbook. Second edition / Francesco Ricci, Lior Rokach, Bracha Shapira. – Switzerland AG : Springer, 2015. – 1003 с.

13 Santos-González, Iván. (2017). Implementation and Analysis of Real-Time Streaming Protocols. [Электронный ресурс] / Iván Santos-González, Alexandra Rivero-García, Jezabel Molina-Gil, Pino Caballero-Gil // University of La Laguna. – Режим доступа: <https://pdfs.semanticscholar.org/413d/09cb44a8c8c3dfab7f4f92f18a878b6d7855.pdf>.

14 Shuyang, Zhao. (2014). A personalized hybrid music recommender based on empirical estimation of user-timbre preference. [Электронный ресурс] / Zhao, Shuyang // Tampere University of Technology. – Режим доступа: <https://pdfs.semanticscholar.org/42b7/4109074d6a8665c001dbec3410e8698eff50.pdf>.

15 Suied, C., Agus, T. R., Thorpe, S. J., & Pressnitzer, D. (2013). Processing of short auditory stimuli: the rapid audio sequential presentation paradigm (RASP). In *Basic Aspects of Hearing* (pp. 443-451). Springer New York.

16 Zakas, Nicholas C. (2011). How content delivery networks (CDNs) work. [Электронный ресурс] / Nicholas C., Zakas. – Режим доступа: <https://humanwhocodes.com/blog/2011/11/29/how-content-delivery-networks-cdns-work/>.

17 Ziccardi, Marco. (2015). Beats detection algorithms. [Электронный ресурс] / Marco, Ziccardi. – Режим доступа: <http://mziccard.me/2015/05/28/beats-detection-algorithms-1>.

18 Whitman, Brian A. (2005). Learning the meaning of music. [Электронный ресурс] / Brian A., Whitman // MIT Libraries. – Режим доступа: <https://dspace.mit.edu/handle/1721.1/32500>.

19 Pandora vs. Spotify - Satisfaction Study. [Электронный ресурс] // Bridge Ratings LLC. – Режим доступа: <http://www.bridgeratings.com/pandora-vs-spotify>.

20 Пат. US7003515B1 США. Consumer item matching method and system / William T. Glaser, Timothy B. Westergren, Jeffrey P. Stearns, Jonathan M. Kraft; заявитель и патентообладатель Pandora Media LLC. – № 60/291821 ; заявл. 16.05.02 ; опубл. 21.02.06. – 4 с.

Федеральное государственное автономное
образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Институт космических и информационных технологий
институт

Межинститутская базовая кафедра
«Прикладная физика и космические технологии»
кафедра

УТВЕРЖДАЮ

Заведующий кафедрой

подпись

« 01 » 07

В.Е. Косенко

инициалы, фамилия

2019 г.

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

«Разработка метода сортировки музыкальных композиций по их звуковым характеристикам»

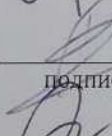
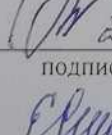
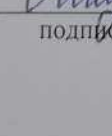
тема

09.04.01 «Информатика и вычислительная техника»

код и наименование направления

09.04.01.03 «Информационные системы космических аппаратов и центров управления полетами»

код и наименование магистерской программы

<u>Научный руководитель</u>	 подпись, дата 25.06.2019	канд. техн. наук, доцент кафедры МБК ПФиКТ должность, ученая степень	<u>М.В. Некрасов</u> инициалы, фамилия
<u>Выпускник</u>	 подпись, дата 25.06.19		<u>Р.Б. Величкин</u> инициалы, фамилия
<u>Рецензент</u>	 подпись, дата 25.06.19	инженер-программист ДИТ ГХК должность, ученая степень	<u>Н.Ю. Зуева</u> инициалы, фамилия
<u>Нормоконтролер</u>	 подпись, дата 25.06.19г.		<u>Е.С. Сидорова</u> инициалы, фамилия

Красноярск 2019