

**Министерство образования и науки Российской Федерации**  
федеральное государственное автономное образовательное учреждение  
высшего образования  
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

---

Институт кибернетики  
Направление подготовки 09.03.01 Информатика и вычислительная техника  
Кафедра информатики и проектирования систем

**БАКАЛАВРСКАЯ РАБОТА**

Тема работы
<b>Разработка рекомендательной системы по подбору музыки</b>

УДК 004.415:78

Студент

Группа	ФИО	Подпись	Дата
8В2Б	Браневский Александр Ярославович		

Руководитель

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент	Наламвар Хитеш Санжай			

**КОНСУЛЬТАНТЫ:**

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент	Николаенко Валентин Сергеевич			

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент	Невский Егор Сергеевич			

**ДОПУСТИТЬ К ЗАЩИТЕ:**

Зав. кафедрой	ФИО	Ученая степень, звание	Подпись	Дата
ИПС	Демин Антон Юрьевич	К.Т.Н.		

## Запланированные результаты обучения по программе

Код результатов	Результат обучения (выпускник должен быть готов)	Требования ФГОС, критерии АИОР
P1	Применять базовые и специальные естественнонаучные и математические знания в области информатики и вычислительной техники, достаточные для комплексной инженерной деятельности.	Требования ФГОС (ОК-1, 10, ПК-4, 5, 6), критерий 5 АИОР (п. 1.1)
P2	Применять базовые и специальные знания в области современных информационных технологий для решения инженерных задач.	Требования ФГОС (ОК-11, 12, 13, ПК-1, 2, 11), критерий 5 АИОР (п.1.1, 1.2)
P3	Ставить и решать задачи комплексного анализа, связанные с созданием аппаратно-программных средств информационных и автоматизированных систем, с использованием базовых и специальных знаний, современных аналитических методов и моделей.	Требования ФГОС (ОК-1, 8, ПК-2, 4, 6), критерий 5 АИОР (п. 1.2)
P4	Разрабатывать программные и аппаратные средства (системы, устройства, блоки, программы, базы данных и т. п.) в соответствии с техническим заданием и с использованием средств автоматизации проектирования.	Требования ФГОС (ОК-2, 3, ПК-3, 4, 5), критерий 5 АИОР (п. 1.3)
P5	Проводить теоретические и экспериментальные исследования, включающие поиск и изучение необходимой научно-технической информации, математическое моделирование, проведение эксперимента, анализ и интерпретация полученных данных, в области создания аппаратных и программных средств информационных и автоматизированных систем.	Требования ФГОС (ОК-6, ПК-6, 7), критерий 5 АИОР (п.1.4)
P6	Внедрять, эксплуатировать и обслуживать современные программно-аппаратные комплексы, обеспечивать их высокую эффективность, соблюдать правила охраны здоровья, безопасность труда, выполнять требования по защите окружающей среды.	Требования ФГОС (ОК-4, 15, 16, ПК-9, 10, 11), критерий 5 АИОР (п. 1.5)
Универсальные компетенции		
P7	Использовать базовые и специальные знания в области проектного менеджмента для ведения комплексной инженерной деятельности.	Требования ФГОС (ОК-1, 4, ПК-1, 6, 7), критерий 5 АИОР (п. 2.1)
P8	Владеть иностранным языком на уровне, позволяющем работать в иноязычной среде, разрабатывать документацию, презентовать и защищать результаты комплексной инженерной деятельности.	Требования ФГОС (ОК-14, ПК-7), критерий 5 АИОР (п. 2.2)
P9	Эффективно работать индивидуально и в качестве члена группы, состоящей из специалистов различных направлений и квалификаций, демонстрировать ответственность за результаты работы и готовность следовать корпоративной культуре организации.	Требования ФГОС (ОК-2, 3, 4), критерий 5 АИОР (п. 2.3, 2.4)
P10	Демонстрировать знания правовых, социальных, экономических и культурных аспектов комплексной инженерной деятельности.	Требования ФГОС (ОК-1, 5, 9), критерий 5 АИОР (п. 2.5)
P11	Демонстрировать способность к самостоятельному обучению в течение всей жизни и непрерывному самосовершенствованию в инженерной профессии.	Требования ФГОС (ОК-6, 7), критерий 5 АИОР (п. 2.6)

**Министерство образования и науки Российской Федерации**  
 федеральное государственное автономное образовательное учреждение  
 высшего образования  
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
 ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Институт кибернетики  
 Направление подготовки 230100 Информатика и вычислительная техника  
 Кафедра информатики и проектирования систем

УТВЕРЖДАЮ:  
 Зав. кафедрой  
 \_\_\_\_\_  
 (Подпись)      \_\_\_\_\_      Демин А.Ю.  
 (Дата)                                      (Ф.И.О.)

**ЗАДАНИЕ**  
**на выполнение выпускной квалификационной работы**

В форме:

Бакалаврской работы
---------------------

Студенту:

Группа	ФИО
8В2Б	Браневский Александр Ярославович

Тема работы:

Разработка рекомендательной системы по подбору музыки	
Утверждена приказом директора (дата, номер)	

Срок сдачи студентом выполненной работы:	
--	--

**ТЕХНИЧЕСКОЕ ЗАДАНИЕ:**

<p><b>Исходные данные к работе</b></p> <p><i>(наименование объекта исследования или проектирования; производительность или нагрузка; режим работы (непрерывный, периодический, циклический и т. д.); вид сырья или материал изделия; требования к продукту, изделию или процессу; особые требования к особенностям функционирования (эксплуатации) объекта или изделия в плане безопасности эксплуатации, влияния на окружающую среду, энергозатратам; экономический анализ и т. д.).</i></p>	<ul style="list-style-type: none"> <li>- научная литература по заданной тематике</li> <li>- экономическая и юридическая литература</li> </ul>
<p><b>Перечень подлежащих исследованию, проектированию и разработке вопросов</b></p> <p><i>(аналитический обзор по литературным источникам с целью выяснения достижений мировой науки техники в рассматриваемой области; постановка задачи исследования, проектирования, конструирования; содержание процедуры исследования, проектирования, конструирования; обсуждение результатов выполненной работы; наименование дополнительных разделов, подлежащих разработке; заключение по работе).</i></p>	<ul style="list-style-type: none"> <li>- анализ возможных подходов к решению данной задачи</li> <li>- исследование уже имеющихся аналогов на рынке</li> <li>- изучение технологии SPARK для обеспечения выполнения приложения распределено, изучение языка SCALA для разработки системы</li> </ul>

<b>Перечень графического материала</b> <i>(с точным указанием обязательных чертежей)</i>	- скриншоты результатов работы приложения - презентация выполненной работы
---	---

**Консультанты по разделам выпускной квалификационной работы**

Раздел	Консультант
Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	Ассистент кафедры Менеджмента, к.э.н. Николаенко Валентин Сергеевич
Социальная ответственность	Ассистент кафедры Экологии и безопасности жизнедеятельности, к.б.н. Невский Егор Сергеевич

<b>Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику</b>	
---	--

**Задание выдал руководитель:**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент	Наламвар Хитеш Санжай			

**Задание принял к исполнению студент:ы**

Группа	ФИО	Подпись	Дата
8В2Б	Браневский Александр Ярославович		

**Министерство образования и науки Российской Федерации**  
федеральное государственное автономное образовательное учреждение  
высшего образования  
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

---

Институт кибернетики  
Направление подготовки 09.03.01 Информатика и вычислительная техника  
Уровень образования бакалавр  
Кафедра информатики и проектирования систем  
Период выполнения \_\_\_\_ (осенний / весенний семестр 2015/2016 учебного года) \_\_\_\_

Форма представления работы:

Бакалаврская работа
---------------------

(бакалаврская работа, дипломный проект/работа, магистерская диссертация)

**КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН**  
**выполнения выпускной квалификационной работы**

Срок сдачи студентом выполненной работы:	17.06.2016 г.
--	---------------

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
	Основная часть	70
	Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	15
	Социальная ответственность	15

Составил преподаватель:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент	Наламвар Хитеш Санжай			

**СОГЛАСОВАНО:**

Зав. кафедрой	ФИО	Ученая степень, звание	Подпись	Дата
ИПС	Демин Антон Юрьевич	к.т.н.		

**ЗАДАНИЕ ДЛЯ РАЗДЕЛА  
«ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И  
РЕСУРСОСБЕРЕЖЕНИЕ»**

Студенту:

<b>Группа</b>	<b>ФИО</b>
8В2Б	Браневский Александр Ярославович

<b>Институт</b>	Кибернетики	<b>Кафедра</b>	Информатики и проектирования систем
<b>Уровень образования</b>	Бакалавр	<b>Направление/специальность</b>	09.03.01

**Исходные данные к разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:**

1. Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих	На основании информации, представленной в научных статьях и публикациях, аналитических материалах, статистических бюллетенях и изданиях, нормативно-правовых документах, определить методику расчета экономической эффективности.
2. Нормы и нормативы расходования ресурсов	
3. Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования	

**Перечень вопросов, подлежащих исследованию, проектированию и разработке:**

1. Оценка коммерческого потенциала, перспективности и альтернатив проведения НИ с позиции ресурсоэффективности и ресурсосбережения	Оценка ресурсной, социальной эффективности НИ и потенциальных рисков.
2. Планирование и формирование бюджета научных исследований	
3. Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования	

**Перечень графического материала (с точным указанием обязательных чертежей):**

1. Оценка плана проведения НИ
2. QuaD-технология
3. Матрица SWOT
4. Морфологический анализ

<b>Дата выдачи задания для раздела по линейному графику</b>	
---	--

**Задание выдал консультант:**

<b>Должность</b>	<b>ФИО</b>	<b>Ученая степень, звание</b>	<b>Подпись</b>	<b>Дата</b>
Ассистент	Николаенко Валентин Сергеевич			

**Задание принял к исполнению студент:**

<b>Группа</b>	<b>ФИО</b>	<b>Подпись</b>	<b>Дата</b>
8В2Б	Браневский Александр Ярославович		

## РЕФЕРАТ

Выпускная квалификационная работа содержит с., 73 рис., 5 табл., 14 источников, 4 прил.

Ключевые слова: machine learning, big data, data mining, SPARK, большие данные, машинное обучение, рекомендательная система, рекомендация музыки, SCALA, MLLib, MAPREDUCE.

**Объект исследования** – сервис позволяющий делать рекомендации музыки целевому клиенту, а также способного работать распределено на нескольких устройствах для обеспечения большей скорости обработки исходных данных.

**Цель работы** – разработка сервиса позволяющего рекомендовать музыкальный продукт целевому клиенту.

В процессе разработки был проведен анализ уже существующих решений для обеспечения подобного функционала, были изучены технологии позволяющие организовать приложение способное обрабатывать большие данные распределено, на разных узлах.

В результате было разработано приложение, позволяющее делать рекомендации на основе некоторой имеющейся базы данных оценок пользователей музыкальных композиций.

**Область применения** – веб-сервисы, имеющие задачу делать рекомендации товара целевому клиенту, самостоятельное приложение.

В дальнейшем планируется продолжить работу по улучшению качества рекомендаций и скорости обработки данных, также возможно интегрировать данную систему в некоторый веб-сервис. Также планируется попробовать данную систему для рекомендации других товаров.

## **ОПРЕДЕЛЕНИЯ, ОБОЗНАЧЕНИЯ, СОКРАЩЕНИЯ, НОРМАТИВНЫЕ ССЫЛКИ**

В данной работе применены следующие термины с соответствующими определениями:

1. ALS (Alternating least squares) – метод наименьших квадратов,
2. API – интерфейс прикладного программирования, набор готовых объектно-ориентированных процедур и функций, предоставляемых операционной системой, приложением, библиотекой, сервисом для использования в разрабатываемых программных продуктах,
3. Big data – «большие данные», обобщенная методология обработки неструктурированных данных огромных объемов, получаемых в результате жизнедеятельности во всех сферах активности человечества,
4. Framework (фреймворк) – библиотека предоставляющая высокоуровневый интерфейс к некоторому программному функционалу,
5. JAVA – язык программирования,
6. JVM – Java Virtual Machine среда для исполнения Java/Scala байт кода,
7. SCALA – JVM совместимый язык, возможно использовать совместно с Java (Scala больше ориентирован на функциональное программирование),
8. SPARK – библиотека для обработки больших данных,
9. SPARK MLlib (MLlib) – библиотека, содержащая алгоритмы машинного обучения для больших данных.



## ОГЛАВЛЕНИЕ

РЕФЕРАТ .....	7
ВВЕДЕНИЕ .....	10
1. ОБЗОР ПРЕДСТАВЛЕННЫХ РЕШЕНИЙ НА РЫНКЕ.....	13

## ВВЕДЕНИЕ

С ростом мировой экономики и соответствующего ей роста товарооборота, появляется потребность в совершенствовании «предложения» за счет алгоритмов, обеспечивающих подбор и рекомендацию товара целевому клиенту, который в нем заинтересован. Такой подход повышает конверсию и эффективность рекламы и интернет-магазинов, то есть увеличивает показатель перехода посетителей в покупателей.

Алгоритмы основанные на коллаборативной фильтрации, используя статистику пользователя на веб-ресурсе (его оценки определенной продукции, покупки и т.д.) определяют на ее основе его вкусы и предпочтения. Все это дает возможность в подборе и рекомендации наилучшего предложения.

В наше время такие системы являются частью любого уважающего себя серьезного веб-сервиса (Amazon, Google, Yahoo и т.д.)

Также подобная рекомендательная система в большинстве случаев должна будет работать с большими объемами данных (например количество пользователей сервиса last.fm составляет свыше 20 миллионов человек, а количество прослушанных композиций с момента основания сервиса почти 100 миллиардов). Очевидным является то, что для создания рекомендательной системы, позволяющей анализировать такой поток данных, необходима возможность обеспечения обработки большого потока данных. Для этого существует концепция MapReduce.

MapReduce использована компанией Google и представляет из себя распределенную технологию вычислений, обеспечивающих параллелизм.

Опишем более подробно технологию MapReduce. Она выделяет две процедуры: Map и Reduce.

Процедура Map проводит предварительную обработку входных данных на отдельных узлах кластера. Процедура Reduce проводит итоговую свёртку данных.

Достоинством MapReduce является то, что он позволяет распределено производить процедуры начальной предобработки и свертки данных. Операции начальной предобработки работают независимо друг от друга и могут быть распараллелены. Хотя эта технология может быть менее эффективной по сравнению с другими алгоритмами последовательной обработки, MapReduce может быть использован с массивами big data, обрабатываемыми на распределенной сети серверов. MapReduce может произвести сортировку петабайт неструктурированной информации, и на это уйдет всего лишь несколько часов. Параллелизация дает восстановления после частичных сбоев серверов: если в рабочем узле, производящем операцию предварительной обработки или свертки, возникает сбой, то его работа может быть передана другому рабочему узлу (при условии, что входные данные для проводимой операции доступны).

В результате было реализовано приложение позволяющее строить рекомендательные модели на основе больших объемов данных (в работе использовалась база данных с last.fm объемом 2 ГБ.) и на основе них выполнять рекомендации. Данное приложение было протестировано в среде, состоящей из 4 кластеров с конфигурацией представленной ниже:

```
root@ts-datastor:~# lscpu
Architecture:          x86_64
CPU op-mode(s):        32-bit, 64-bit
Byte Order:            Little Endian
CPU(s):                 4
On-line CPU(s) list:   0-3
Thread(s) per core:    1
Core(s) per socket:    4
Socket(s):              1
NUMA node(s):          1
Vendor ID:              GenuineIntel
CPU family:             6
Model:                  44
Model name:             Westmere E56xx/L56xx/X56xx (Nehalem-C)
Stepping:               1
CPU MHz:                2599.990
BogoMIPS:               5199.98
Hypervisor vendor:     KVM
Virtualization type:   full
L1d cache:              32K
L1i cache:              32K
L2 cache:               4096K
NUMA node0 CPU(s):     0-3
```

*Рисунок 1.1. Конфигурация системы с развернутым SPARK*

## **1. ОБЗОР ПРЕДСТАВЛЕННЫХ РЕШЕНИЙ НА РЫНКЕ**

Как уже упоминалось ранее в настоящее время на рынке существуют аналоги обеспечивающие подобную функциональность. Однако данная область знаний имеет высокий входной порог входа и количество сервисов, предоставляющих качественный сервис, не так много. Таким образом, разработка рекомендательной системы является перспективным направлением с экономической точки зрения, однако требует большого количества ресурсов и времени, так как практически все аналоги разрабатываются крупными мировыми компаниями, с привлечением большого числа разработчиков.

Практически всегда основной задачей такой системы является популяризация ресурса и привлечения новых клиентов, также такая система может выступать, например, как часть рекламного сервиса.

## 1.1. Обзор существующих аналогов рекомендательных систем

Ниже будут рассмотрены популярные аналоги и их возможности. Сразу стоит заметить, что всю информацию о системы собрать не получится так как компании заботятся о сохранности своей интеллектуальной собственности. Подходы к решению данной проблемы могут сильно отличаться как по составу программной части, так и по аппаратной. Анализ будет проводиться с помощью сводки в виде таблицы, сформированной по определенным параметрам.

1. **Last.fm** – одна из самых популярных сервисов по подбору музыки на рынке. Огромная музыкальная база, огромное количество клиентов, креативный дизайн, большие возможности в плане рекомендаций музыки. Также данный сервис предоставляет API-интерфейс для разработчиков, давая возможность, например, собрать информацию и предпочтения по определенным пользователям (что слушали, сколько раз и т.д.)

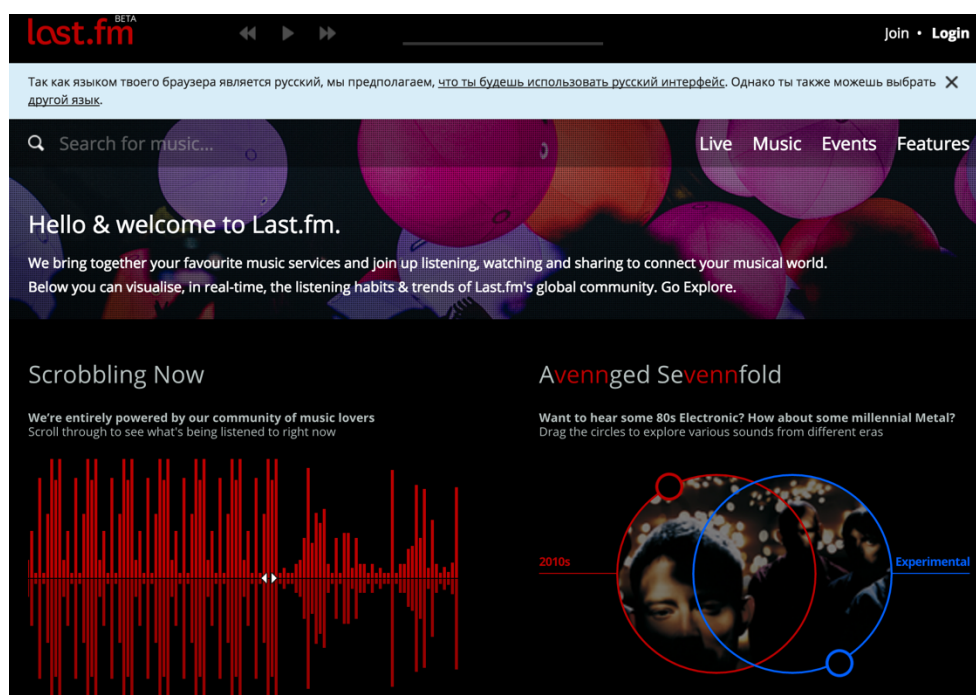


Рисунок 1.2. Заглавная страница Last.fm

Таблица 1.

## Обзорная таблица веб-сервиса last.fm

Приблизительное количество клиентов:	40 000 000 регистраций
Объем оперируемых данных:	10 000 000+ саундтреков
Момент открытия веб-сервиса:	2002г.
Приблизительная стоимость ресурса:	30 мая 2007 года американская компания «CBS Corp.» приобрела сайт «Last.fm» за \$280 млн (£140 млн). Эта сделка стала крупнейшим в истории приобретением британского веб-сервиса
Количество разработчиков проекта	Нет данных

**Преимущества:**

- Интуитивный интерфейс;
- Огромная клиентская база;
- Рекомендации аудио;
- API для разработчиков

**Недостатки:**

- Подписка на прослушивание радио стоит 5\$;
- Нельзя удалить свой аккаунт с сайта;
- Изменение политики сайта с целью монетизации проводится не путём добавления новых возможностей, а путём отключения бесплатной функциональности.

**2. Youtube** – видеохостинг, предоставляющий пользователям услуги хранения, доставки, показа и монетизации видео. Пользователи могут загружать, просматривать, оценивать, комментировать и делиться теми или иными видеозаписями. Благодаря простоте и удобству использования YouTube стал популярнейшим видеохостингом и третьим сайтом в мире по количеству посетителей. В январе 2012 ежедневное количество просмотров видео на сайте достигло 4 млрд. На сайте представлены фильмы, музыкальные клипы, трейлеры, новости, образовательные передачи, а также любительские видеозаписи, включая видеоблоги, слайд-шоу, юмористические видеоролики и прочее.



*Рисунок 1.3. Логотип Youtube*



## Основные характеристики:

Таблица 2.

Обзорная таблица веб-сервиса Youtube

Приблизительное количество клиентов:	1 000 000 000 + регистраций
Объем оперируемых данных:	21 Терабайт / день
Момент открытия веб-сервиса:	2005 г.
Приблизительная стоимость ресурса:	816 723 687.00 USD
Количество разработчиков проекта:	До того, как стали дочерней компанией Google, разработчики были небольшим стартапом, около 20 человек.

## Преимущества:

- Оригинальный дизайн;
- Высокий доход с рекламы (по неофициальным данным более \$2млн в день);
- Передовые технологии в плане рекомендаций видео/музыки, не имеющие аналогов;
- Высокая скорость работы при огромном потоке входных данных;

## Недостатки:

- Частые иски в связи с нарушением авторских прав;

### 3. 8tracks

Достаточно популярный ресурс за рубежом, предоставляющий возможность загружать свой плейлист, и, как следствие, получать рекомендации наиболее популярных исполнителей максимально соответствующих уже имеющимся. Также, в дополнение, сайт имеет приятный дизайн.

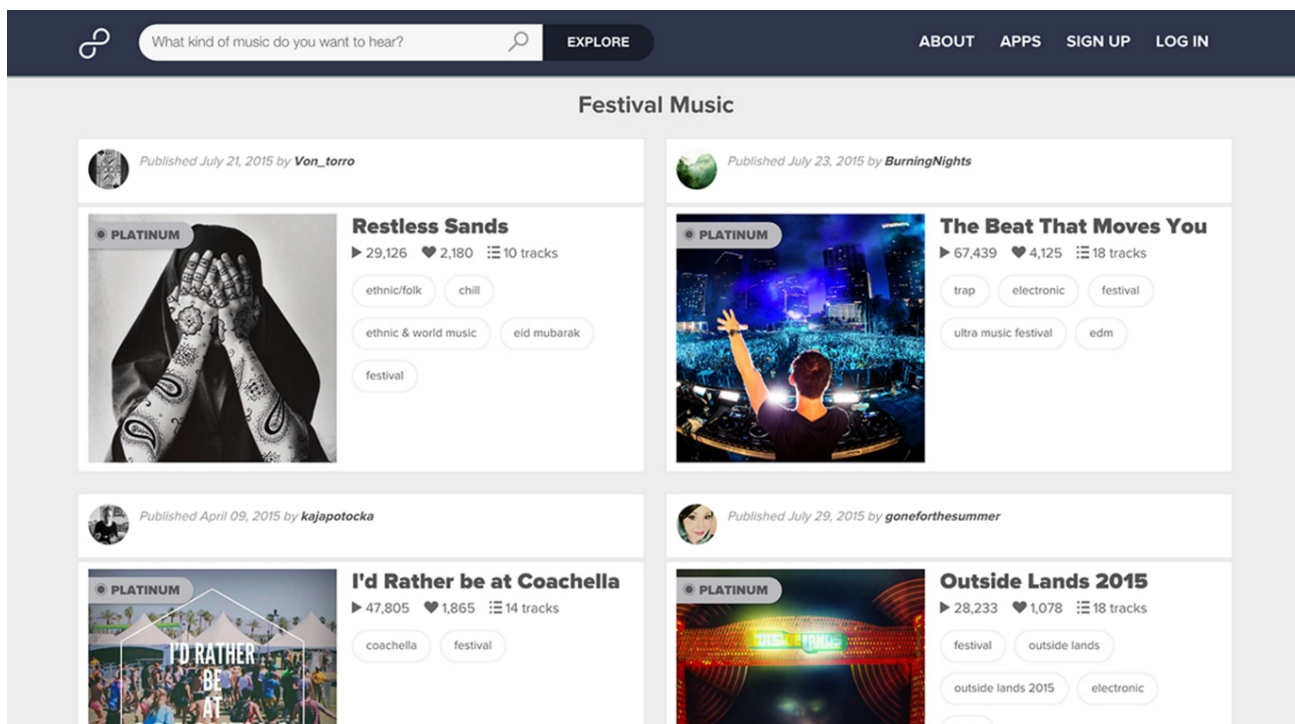


Рисунок 1.4. Заглавная страница 8tracks

## Основные характеристики:

Таблица 3.

Обзорная таблица веб-сервиса 8 tracks

Приблизительное количество клиентов:	8 000 000 пользователей
Объем оперируемых данных:	3 000 000 саундтреков
Момент открытия веб-сервиса:	8 августа 2008
Приблизительная стоимость ресурса:	26 000 000\$
Количество разработчиков проекта:	Нет данных

## Достоинства:

- Оригинальный дизайн;
- Большая музыкальная база, на основе которой делаются рекомендации;
- Рекомендации аудио на основе пользовательских отзывов;

## Недостатки:

- Нет русской локализации;
- Большинство функций платные;

## 4. Grooveshark

На данный момент этот веб-ресурс закрыт из-за исков правообладателей, однако в свое время был очень популярен за рубежом. У сервиса была уникальная функция, позволяющая подбирать саундтреки похожие на те, которые находились в очереди на прослушивание у пользователя. Также у пользователя была возможность оценить качество каждой отдельной рекомендации (данный отзыв позволял улучшать качество рекомендаций в дальнейшем).

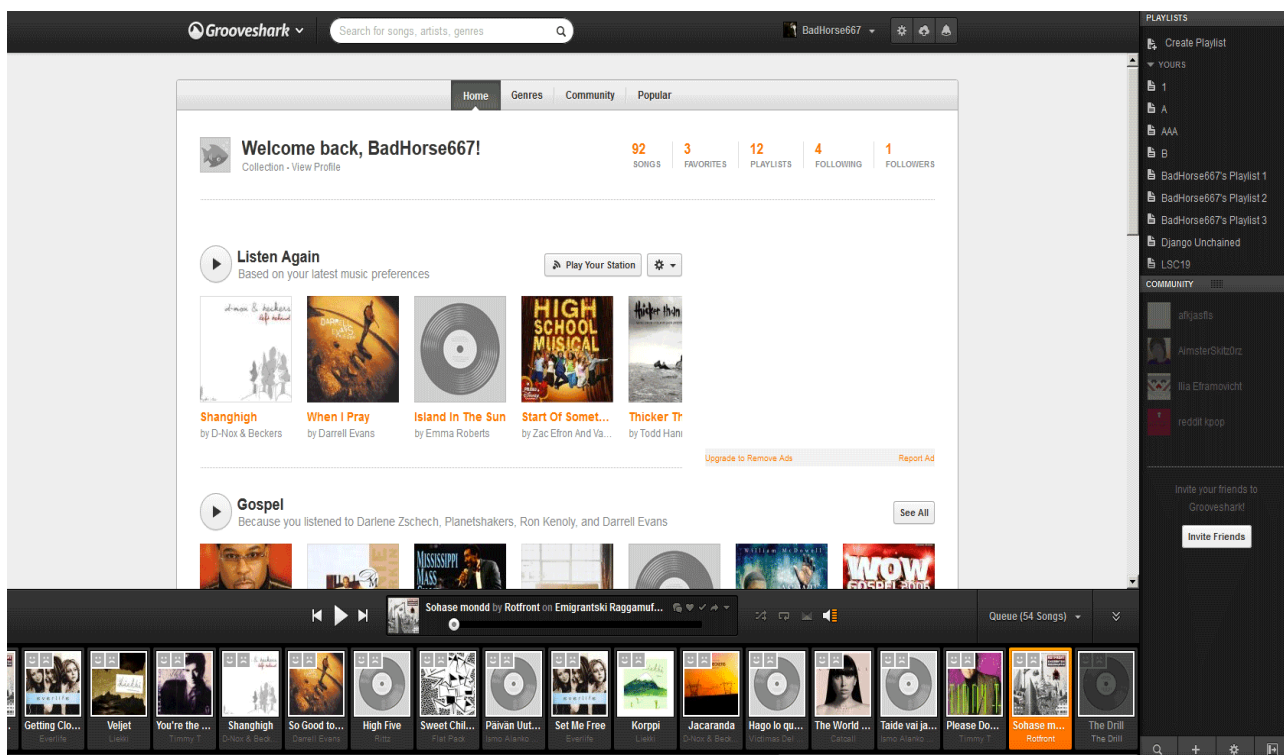


Рисунок 1.5. Заглавная страница GrooveShark

## Основные характеристики:

Таблица 4.

Обзорная таблица веб-сервиса GrooveShark

Приблизительное количество клиентов:	3 000 000 пользователей
Объем оперируемых данных:	–
Момент открытия веб-сервиса:	2007 – 2015
Приблизительная стоимость ресурса:	7 900 000\$
Количество разработчиков проекта:	Разрабатывался командой из 5 разработчиков, принадлежавшей фирме «Escape Media Groop»

### Достоинства:

1. Оригинальный дизайн;
2. Постоянное улучшение качества системы рекомендаций, основанной на отзывах пользователей;
3. API для разработчиков.

### Недостатки:

1. Проблемы с правообладателями (было размещено много нелегального медиаконтента), что послужило причиной закрытия ресурса;
2. Платная подписка.

## **ПРИЛОЖЕНИЕ М**

# **Менеджмент, ресурсоэффективность и ресурсосбережение**

### **1. Введение**

Результатом разработки рассматриваемого проекта является приложение, представляющее собой систему способную осуществлять рекомендацию музыки. Данное приложение может использоваться как самостоятельный проект или же составлять часть некоторого веб-сервиса. Сейчас практически каждый популярный крупный веб-сервис имеет в своем распоряжении рекомендательную систему и улучшение качества рекомендаций для таких сервисов является одной из приоритетных задач. Так как увеличивая качество рекомендаций с большой вероятностью мы также увеличим коэффициент конверсии посетителей в покупателей.

Таким образом целевой аудиторией данного проекта могут быть как компании заинтересованные в продвижении некоторого продукта, так и люди занимающиеся научной работой в области машинного обучения. В настоящее время машинное обучение является активно развивающейся областью и все больше число компаний стараются использовать его элементы в крупных проектах.

В данном разделе будут рассмотрены методы, которые помогут обозначить актуальность и необходимость разработки. Также будут рассмотрены возможные пути улучшения результатов и дальнейших исследований в данной отрасли.

## 2. Цели и задачи

Данный раздел поможет определить коммерческий потенциал рассматриваемого проекта, поможет определить его возможные перспективы на рынке, а также целесообразность вложения денежных средств в разработку.

Решение поставленных задач будет осуществляться с помощью QuaD-анализа, который представляет собой инструмент для измерения характеристик, описывающих качество разработки. Также для улучшения качества анализа будут применяться SWOT и морфологический анализ.

Для проведения QuaD анализа необходимо проведение следующих задач:

1. Определение показателей отвечающих за коммерческий потенциал разрабатываемого продукта и показателей отвечающих за качество разработки
2. Нахождение весов показателей, основываясь на их значимости. Вычисление средневзвешенного значения показателя качества и перспективности
3. Анализ полученных результатов

Для проведения анализа типа SWOT осуществляется выполнение следующих задач:

1. Выявление сильных и слабых сторон проекта, определение его потенциала, определение возможных угроз
2. Установление взаимосвязи между ними
3. Анализ полученных результатов

### 3. Технология QuaD

Технология QuaD применяется для количественной оценки характеристик, которые описывают перспективность разработанного продукта на рынке и его качество, используя такие характеристики, как эффективность, конкурентоспособность и т.д.

В ходе QuaD анализа для получения оценки были выделены следующие критерии:

- Восприимчивость к масштабированию аппаратной части системы:

Конечный продукт должен иметь способность анализировать весь доступный объем информации и использовать его для получения рекомендаций. Вычислительная сложность используемого алгоритма должна позволять осуществлять это за приемлемое время. Также система должна иметь возможность к горизонтальному масштабированию (при увеличении количества вычислительных центров, возможный объем информации для обработки должен увеличиваться).

- Способность к масштабированию программной части системы:

Исходный код проекта должен быть организован таким образом, чтобы при появлении новых разработчиков проекта, они смогли использовать наработки при возникновении новых функциональных требований.



- Адаптируемость к ресурсам:

Итоговая система должна быть способна обеспечивать приемлемую скорость работы на заданном вычислительном центре.

- Отказоустойчивость:

Рекомендательная система предназначена для запуска на более чем одном вычислительном центре. Таким образом в случае падения любого из них нужно обеспечить сохранность данных.

- Качество рекомендаций:

Качество рекомендаций должно удовлетворять потребности клиентов

Оценочная карта для сравнения конкурентных технических решений приведена в таблице 1.:

Таблица 1. – Оценочная карта

Критерии оценки	Вес	Баллы	Макс. балл	Отн. знач.	Ср.-взвеш. знач.
<b>Показатели оценки качества разработки</b>					
1. Восприимчивость к масштабированию аппаратной части системы	0,15	80	100	0,8	0,12
2. Способность к масштабированию программной части системы	0,1	60	100	0,6	0,06
3. Адаптируемость к ресурсам	0,1	70	100	0,7	0,07
4. Отказоустойчивость	0,05	60	100	0,6	0,03
5. Качество рекомендаций	0,2	90	100	0,9	0,18

Показатели оценки коммерческого потенциала разработки					
6. Перспективность	0,1	80	100	0,8	0,08
7. Коммерческая ценность	0,1	80	100	0,8	0,08
8. Правовая защищенность	0,05	70	100	0,7	0,035
Итого:					0,655

Оценка качества и перспективности по технологии QuaD определяется по формуле:

$$P_{\text{ср}} = \sum B_i \cdot B_i,$$

где  $P_{\text{ср}}$  – средневзвешенное значение показателя качества и перспективности научной разработки;  $B_i$  – вес показателя (в долях единицы);  $B_i$  – средневзвешенное значение  $i$ -го показателя.

Итоговый результат для интегрального показателя конкурентоспособности данной разработки составляет 0,655, что является критерием средней перспективности. При дальнейшей разработке данный показатель может быть существенно увеличен.

## 4. SWOT-анализ

SWOT-анализ нужен для выявления сильных и слабых сторон проекта, определения угроз и возможностей. После выявления данных особенностей нужно провести для них сравнительный анализ.

Сводная таблица для SWOT анализа представлена в таблице 2:

Таблица 2. – Результаты SWOT-анализа

	<b>Сильные стороны</b>	<b>Слабые стороны</b>
	1. Использование современных фреймворков распределенных вычислений (SPARK) и языков разработки (SCALA)	1. Высокая конкуренция (причем аналоги зачастую разрабатываются большой командой разработчиков, гораздо более длительное время)
	2. Возможность распараллелить вычисления и как следствие потенциал к масштабированию	2. Нет онлайнности (алгоритм, применяющийся сейчас хоть и может выполняться параллельно, но не имеет возможности динамически обновлять модель делающую рекомендации)
	3. Спрос на рынке на подобные системы	
	4. Гибкость системы (можно использовать как	

	<p>самостоятельный проект, так и внедрять в некоторый веб-сервис)</p> <p>5. Большие возможности по конфигурированию системы (настройка параметров отвечающих за качество рекомендаций)</p>	
<p><b>Возможности</b></p> <p>1. Потенциал к развитию системы</p> <p>2. Продажа продукта</p>	<p>Использование современных фреймворков для обработки и анализа больших данных позволит быть в курсе последних исследований в этой области и иметь возможность улучшить проект</p>	<p>При развитии системы возможно будет пересмотрена/дополнена алгоритмическая база и как следствие появится поддержка онлайнности</p>

Анализируя сводную таблицу можно прийти к выводу, что основными рисками при дальнейшей разработке является наличие конкурентов с гораздо большими финансовыми возможностями и как следствие со схожими или же превосходящими функциональными возможностями.

## 5. Определение возможных альтернатив проведения научных исследований

Также был произведен морфологический анализ. Его идея заключается в определении параметров рассматриваемого объекта и рассмотрении различных комбинаций этих параметров, а также рассмотрении возможных аналогов на замену некоторых параметров. Результаты анализа представлены в таблице ниже.

Таблица 3. – Морфологическая матрица

	1	2	3
А. Фреймворк для параллельных вычислений	SPARK	HADOOP	...
Б. Фреймворки для предобработки данных	SPARK	Pandas	Ninjas
В. Сайты предоставляющие данные для обучения моделей	Last FM	Vkontakte	Sky FM
Г. Фреймворки для построения моделей	Scikit learn	ML Lib	Spark-sklearn
Д. Способ размещения системы	Самостоятельная система	Часть веб-сервиса	...
Е. Способ сбора обучающей выборки	Last fm API/ Vk API	Реализация собственного приложения для сбора информации	...
Ж. Возможные рекомендательные сервисы	Рекомендация фильмов	Рекомендации книг	Рекомендация музыки

### Список публикаций студента

1. Бенц, А. С. Исследование зависимости качества классификации от различных параметров ИНС [Электронный ресурс] / А. С. Бенц, А. Я. Браневский, П. А. Хаустов // Молодежь и современные информационные технологии : сборник трудов XI Международной научно-практической конференции студентов, аспирантов и молодых ученых, г. Томск, 13-16 ноября 2013 г. / Национальный исследовательский Томский политехнический университет (ТПУ) ; под ред. Е. А. Сикоры и др. . — Томск; : Изд-во ТПУ , 2013 . — [С. 139-141] . Схема доступа: <http://www.lib.tpu.ru/fulltext/c/2013/C04/060.pdf>
2. Браневский А.Я. Разработка рекомендательной систем по подбору музыки / Браневский.А.Я, науч. рук. Наламвар Хитеш // Технологии Microsoft в теории и практике программирования: сборник трудов XIII Всероссийской научно-практической конференции студентов, аспирантов и молодых ученых, г.Томск, 22-23 марта 2016 г. / В печати.
3. Браневский А.Я. Разработка рекомендательной систем по подбору музыки / Браневский.А.Я, науч. рук. Наламвар Хитеш // Информационные технологии в науке, управлении, социальной сфере и медицине: сборник трудов III Международной научной конференции, г. Томск, 23-26 мая 2016 г. / В печати.