

Министерство науки и высшего образования Российской Федерации  
 федеральное государственное автономное  
 образовательное учреждение высшего образования  
 «Национальный исследовательский Томский политехнический университет» (ТПУ)

Школа информационных технологий и робототехники  
 Направление подготовки 09.04.01 «Информатика и вычислительная техника»  
 Отделение школы (НОЦ) информационных технологий

### МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Тема работы
<b>Разработка рекомендательной системы для электронной коммерции</b>

УДК 004.65:004.451:339

Студент

Группа	ФИО	Подпись	Дата
8ВМ81	Аспомбитов Сабыржан Олегович		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Фадеев А.С.	К.Т.Н.		

### КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Конотопский В.Ю.	К.Э.Н.		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Горбенко М.В.	К.Т.Н.		

### ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Кочегурова Е.А.	К.Т.Н.		

Томск – 2020 г.

## ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

Код результата	Результат обучения (выпускник должен быть готов)
	<b>Общепрофессиональные компетенции</b>
P1	Воспринимать и самостоятельно приобретать, развивать и применять математические, естественнонаучные, социальноэкономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте.
P2	Владеть и применять методы и средства получения, хранения, переработки и трансляции информации посредством современных компьютерных технологий, в том числе в глобальных компьютерных сетях
P3	Демонстрировать культуру мышления, способность выстраивать логику рассуждений и высказываний, основанных на интерпретации данных, интегрированных из разных областей науки и техники, выносить суждения на основании неполных данных, анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями.
P4	Анализировать и оценивать уровни своих компетенций в сочетании со способностью и готовностью к саморегулированию дальнейшего образования и профессиональной мобильности. Владеть, по крайней мере, одним из иностранных языков на уровне социального и профессионального общения, применять специальную лексику и профессиональную терминологию языка.
	<b>Профессиональные компетенции</b>
P5	Выполнять инновационные инженерные проекты по разработке аппаратных и программных средств автоматизированных систем различного назначения с использованием современных методов проектирования, систем автоматизированного проектирования, передового опыта разработки конкурентно способных изделий.
P6	Планировать и проводить теоретические и экспериментальные исследования в области проектирования аппаратных и программных средств автоматизированных систем с использованием новейших достижений науки и техники, передового отечественного и зарубежного опыта. Критически оценивать полученные данные и делать выводы.
P7	Осуществлять авторское сопровождение процессов проектирования, внедрения и эксплуатации аппаратных и программных средств автоматизированных систем различного назначения.

	<b>Общекультурные компетенции</b>
P8	Использовать на практике умения и навыки в организации исследовательских, проектных работ и профессиональной эксплуатации современного оборудования и приборов, в управлении коллективом.
P9	Осуществлять коммуникации в профессиональной среде и в обществе в целом, активно владеть иностранным языком, разрабатывать документацию, презентовать и защищать результаты инновационной инженерной деятельности, в том числе на иностранном языке.
P10	Совершенствовать и развивать свой интеллектуальный и общекультурный уровень. Проявлять инициативу, в том числе в ситуациях риска, брать на себя всю полноту ответственности.
P11	Демонстрировать способность к самостоятельному обучению новым методам исследования, к изменению научного и научно-производственного профиля своей профессиональной деятельности, способность самостоятельно приобретать с помощью информационных технологий и использовать в практической деятельности новые знания и умения, в том числе в новых областях знаний, непосредственно не связанных со сферой деятельности, способность к педагогической деятельности.

Министерство науки и высшего образования Российской Федерации  
 федеральное государственное автономное  
 образовательное учреждение высшего образования  
 «Национальный исследовательский Томский политехнический университет» (ТПУ)

Школа Информационных технологий и робототехники  
 Направление подготовки (специальность) 09.04.01 Информатика и вычислительная техника  
 Отделение школы (НОЦ) Информационных технологий

УТВЕРЖДАЮ:  
 Руководитель ООП

\_\_\_\_\_  
 (Подпись)    (Дата)    (Ф.И.О.)

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

В форме:

Магистерской диссертации <small>(бакалаврской работы, дипломного проекта/работы, магистерской диссертации)</small>
---

Студенту:

Группа	ФИО
8ВМ81	Аспомбитову Сабыржану Олеговичу

Тема работы:

Разработка рекомендательной системы для электронной коммерции	
Утверждена приказом директора (дата, номер)	59-63/с от 28.02.2020

Срок сдачи студентом выполненной работы:	01.06.2020
--	------------

**ТЕХНИЧЕСКОЕ ЗАДАНИЕ:**

<b>Исходные данные к работе</b>	1. Требование реализовать рекомендательную систему для электронной коммерции.
<b>Перечень подлежащих исследованию, проектированию и разработке вопросов</b>	1. Исследование предметной области 2. Проектирование и реализация программных модулей 3. Финансовый менеджмент, ресурсоэффективность и ресурсосбережения 4. Социальная ответственность 5. Заключение по работе
<b>Перечень графического материала</b>	Презентация в формате *.pptx
<b>Консультанты по разделам выпускной квалификационной работы</b>	

Раздел	Консультант
Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	Конотопский В.Ю.
Социальная ответственность	Горбенко М.В.
<b>Названия разделов, которые должны быть написаны на русском и иностранном языках:</b>	
Рекомендательные системы	
Графовые базы данных	

Дата выдачи задания по линейному графику	10.03.2020
--	------------

**Задание выдал руководитель:**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Фадеев А.С.	К.Т.Н.		

**Задание принял к исполнению студент:**

Группа	ФИО	Подпись	Дата
8ВМ81	Аспомбитов Сабыржан Олегович		

Министерство науки и высшего образования Российской Федерации  
 федеральное государственное автономное  
 образовательное учреждение высшего образования  
 «Национальный исследовательский Томский политехнический университет» (ТПУ)

Школа Информационных технологий и робототехники  
 Направление подготовки (специальность) 09.04.01 Информатика и вычислительная техника  
 Уровень образования Магистратура  
 Отделение школы (НОЦ) Информационных технологий  
 Период выполнения \_\_\_\_\_ (осенний / весенний семестр 2019 /2020 учебного года) \_\_\_\_\_

Форма представления работы:

<b>Магистерская диссертация</b> (бакалаврская работа, дипломный проект/работа, магистерская диссертация)
---

### КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН выполнения выпускной квалификационной работы

Срок сдачи студентом выполненной работы:	01.06.2020
--	------------

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
29.05.2020	Основная часть	75
29.05.2020	Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	15
29.05.2020	Социальная ответственность	10

**СОСТАВИЛ:**

**Руководитель ВКР**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Фадеев А.С.	К.Т.Н.		

**СОГЛАСОВАНО:**

**Руководитель ООП**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Кочегурова Е.А.	К.Т.Н.		

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

Студенту:

Группа	ФИО
8ВМ81	Аспомбитову Сабыржану Олеговичу

Школа	ИШИТР	Отделение школы (НОЦ)	ОАР
Уровень образования	Магистратура	Направление/специальность	09.04.01 Информатика и вычислительная техника

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

1. Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих	Оклад руководителя – 33664 руб. Оклад инженера – 15470 руб. Общая сумма затрат – 452443 руб.
2. Нормы и нормативы расходования ресурсов	–
3. Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования	Ставка для расчета отчислений во внебюджетные фонды – 30 % Ставка для расчета накладных расходов – 10%

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. Оценка коммерческого и инновационного потенциала НТИ	Определение потенциальных потребителей, анализ конкурентных решений с позиции ресурсоэффективности и экономической эффективности, функционально-стоимостной анализ.
2. Разработка устава научно-технического проекта	–
3. Планирование процесса управления НТИ: структура и график проведения, бюджет, риски и организация закупок	Составление перечня этапов и работ, расчёт затрат на необходимые ресурсы, описание возможных рисков.
4. Определение ресурсной, финансовой, экономической эффективности	Оценка экономической эффективности использования результатов ВКР, характеристика других видов эффекта

Перечень графического материала (с точным указанием обязательных чертежей):

1. «Портрет» потребителя результатов НТИ
2. Сегментирование рынка
3. Оценка конкурентоспособности технических решений
4. Матрица SWOT
5. График проведения и бюджет НТИ
6. Оценка ресурсной, финансовой и экономической эффективности НТИ
7. Потенциальные риски

Дата выдачи задания для раздела по линейному графику	
--	--

**Задание выдал консультант:**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Конотопский В.Ю.	к.э.н.		

**Задание принял к исполнению студент:**

Группа	ФИО	Подпись	Дата
8ВМ81	Аспомбитов Сабыржан Олегович		

## ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

Группа	ФИО
8ВМ81	Аспомбитову Сабыржану Олеговичу

Школа	ИШИТР	Отделение (НОЦ)	ОИТ
Уровень образования	Магистратура	Направление/специальность	09.04.01 Информатика и вычислительная техника

Тема ВКР:

Разработка рекомендательной системы для электронной коммерции	
<b>Исходные данные к разделу «Социальная ответственность»:</b>	
1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения	Рабочая зона оборудована 6 местами, каждое из которых включает в себя персональный компьютер (ПК) с периферийными устройствами, расположенном на столе, стул. Технологический процесс представляет собой работу с ПК.
Перечень вопросов, подлежащих исследованию, проектированию и разработке:	
<b>1. Правовые и организационные вопросы обеспечения безопасности:</b> <ul style="list-style-type: none"> <li>- специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства;</li> <li>- организационные мероприятия при компоновке рабочей зоны.</li> </ul>	<ul style="list-style-type: none"> <li>- Трудовой кодекс Российской Федерации от 30.12.2001 N 197-ФЗ (Редакция от 24.04.2020)</li> <li>- ГОСТ 12.2.032-78 ССБТ</li> </ul>
<b>2. Производственная безопасность:</b> 2.1. Анализ выявленных вредных и опасных факторов 2.2. Обоснование мероприятий по снижению воздействия	<ul style="list-style-type: none"> <li>- Отклонение показателей микроклимата.</li> <li>- Недостаточная освещенность рабочей.</li> <li>- Превышение уровня шума.</li> <li>- Повышенный уровень электромагнитных излучений.</li> <li>- Психофизиологические факторы.</li> <li>- Электрический ток.</li> </ul>
<b>3. Экологическая безопасность:</b>	Анализ воздействия объекта на литосферу, гидросферу и атмосферу: утилизация отходов, связанные с выходом из строя ПК, люминесцентных ламп и др.
<b>4. Безопасность в чрезвычайных ситуациях:</b>	Возможной чрезвычайной ситуацией является возникновение пожара на рабочем месте, внезапное обрушение здания, землетрясение.

Дата выдачи задания для раздела по линейному графику	
--	--

**Задание выдал консультант:**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ООД	Горбенко Михаил Владимирович	к. т. н.		

**Задание принял к исполнению студент:**

Группа	ФИО	Подпись	Дата
8ВМ81	Аспомбитов Сабыржан Олегович		

## РЕФЕРАТ

Выпускная квалификационная работа содержит: 98 страниц, 24 рисунка, 29 таблиц, 36 источников, 1 приложения.

Ключевые слова: рекомендательная система, коллаборативная фильтрация, графовые базы данных, Neo4j, Cypher.

Объектом исследования является разрабатываемая рекомендательная система.

Целью работы является разработка рекомендательной системы на основе алгоритма коллаборативной фильтрации с использованием графовой базы данных.

В процессе разработки был проведено исследование рекомендательных систем, обзор вариантов построения рекомендательных систем, а так же изучена теория графовых баз данных и получен практический опыт работы с графовой базой данных Neo4j и с языком запросов Cypher.

В результате исследования был разработана рекомендательная система на основе алгоритма коллаборативной фильтрации, а также для дальнейшей интеграции, разработанной системы, с системой электронной коммерции был реализован прототип интернет-магазина.

Область применения: рекомендательная система ориентирована для интеграции с системами электронной коммерции.

Основные характеристики: рекомендательная система разработана на языке программирования Python и Cypher, для хранения и анализа данных для рекомендаций использовалась графовая база данных Neo4j. Прототип интернет-магазина разработана на языке PHP, для хранения данных использовалась СУБД MySQL. Для развертывания разработанных приложений использовался Docker и Docker Compose.

В будущем планируется интеграции рекомендательной системы с прототипом интернет-магазина и проведение комплексного тестирования.

## ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

**Рекомендательные системы** — программы, которые используя некоторую информацию о пользователе, пытаются предсказать какие объекты (новости, музыка, фильмы, книги, веб-сайты) будут ему интересны.

**Коллаборативная фильтрация** — это один из методов построения прогнозов в рекомендательных системах, использующий известные предпочтения группы пользователей для прогнозирования неизвестных предпочтений другого пользователя.

**Веб-приложение** — клиент-серверное приложение, в котором клиент взаимодействует с сервером при помощи браузера, а за сервер отвечает веб-сервер.

**Веб-сервер** — сервер, принимающий HTTP-запросы от клиентов, обычно веб-браузеров, и выдающий им HTTP-ответы, как правило, вместе с HTML-страницей, изображением, файлом, медиа-поток или другими данными.

**HTTP** (HyperText Transfer Protocol) — протокол прикладного уровня передачи данных (изначально — в виде гипертекстовых документов в формате «HTML», в настоящий момент используется для передачи произвольных данных).

**Yii2** — объектно-ориентированный компонентный фреймворк, написанный на PHP и реализующий парадигму MVC.

**Фреймворк** — программная платформа, определяющая структуру программной системы; программное обеспечение, облегчающее разработку и объединение разных компонентов большого программного проекта.

**MVC** (Model-View-Controller) — архитектурный паттерн, разделяющий данные приложения, пользовательский интерфейс и управляющую логику на три отдельных компонента: модель, представление и контроллер, позволяя осуществляться независимую модификация каждого компонента.

**Графовая база данных** — разновидность баз данных с реализацией сетевой модели в виде графа и его обобщений.

**Neo4j** — графовая база данных с открытым исходным кодом, реализованная на Java.

**Cypher** — это язык запросов декларативного графа, который обеспечивает выразительный и эффективный запрос данных в графе свойств.

## ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	14
1 РЕКОМЕНДАТЕЛЬНЫЕ СИСТЕМЫ.....	17
1.1 Контентная фильтрация (content-based filtering).....	19
1.2 Коллаборативная фильтрация.....	19
2 ГРАФОВЫЕ БАЗЫ ДАННЫХ.....	23
Neo4j.....	25
3 РЕАЛИЗАЦИЯ РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ.....	27
3.1 Технологии и инструментарий.....	27
3.1.1 Язык программирования.....	27
3.1.2 Фреймворки.....	28
3.1.3 Среда разработки.....	28
3.2 Поиск, фильтрация и подготовка данных для рекомендаций.....	30
3.3 Импорт данных «Клиент-Заказ-Товар».....	31
3.4 Реализация алгоритма коллаборативной фильтрации на основе схожести пользователей.....	33
4 РАЗРАБОТКА ИНТЕРНЕТ-МАГАЗИНА.....	37
4.1 Технологии и инструментарий.....	37
4.1.1 Язык программирования и фреймворки.....	37
4.1.2 Фреймворки.....	37
4.1.3 База данных.....	38
4.1.4 Среда разработки.....	39
4.2 Варианты использования.....	41
4.3 Описание разработанного веб-приложения.....	42
5 ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ.....	46
Введение.....	46
5.1 Оценка коммерческого потенциала и перспективности проведения научных исследований с позиции ресурсоэффективности и ресурсосбережения.....	46
5.1.1 Потенциальные потребители результатов исследования.....	46

5.1.2	Исследование целесообразности вложения денежных средств в научно-исследовательский проект.....	46
5.1.3	SWOT-анализ.....	48
5.1.4	Определение возможных альтернатив проведения исследований.....	49
5.2	Планирование научно-исследовательских работ.....	50
5.2.1	Структура работ в рамках научного исследования.....	50
5.2.2	Определение трудоемкости выполнения работ.....	51
5.2.3	Составление календарного план-графика работ.....	54
5.2.4	Бюджет научно-технического исследования.....	57
5.3	Определение ресурсной, финансовой и экономической эффективности исследования.....	61
	Вывод.....	62
6	СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ.....	63
	Введение.....	63
6.1	Правовые и организационные вопросы обеспечения безопасности.....	63
6.1.1	Правовые нормы охраны труда.....	63
6.1.2	Правовые и организационные вопросы обеспечения безопасности.....	65
6.2	Производственная безопасность.....	67
6.2.1	Отклонение показателей микроклимата.....	68
6.2.2	Недостаточная освещенность рабочей зоны.....	70
6.2.3	Повышенный уровень шума.....	77
6.2.4	Повышенный уровень электромагнитных излучений.....	78
6.2.5	Электрический ток.....	80
6.3	Экологическая безопасность.....	81
6.4	Безопасность в чрезвычайных ситуациях.....	83
	Вывод.....	84
	ЗАКЛЮЧЕНИЕ.....	85
	CONCLUSION.....	86
	СПИСОК ИСТОЧНИКОВ.....	87
	ПРИЛОЖЕНИЕ А.....	91

## ВВЕДЕНИЕ

Интернет-магазины являются одним из самых популярных онлайн-видов деятельности во всем мире. По данным отчета Statista.com в 2018 году больше полутора миллиарда человек во всем мире совершали покупки через Интернет. В течение того же года глобальные продажи через интернет-магазины составили 2,8 триллиона долларов США, а прогнозы аналитических служб показывают рост до 4,8 триллионов долларов США к 2021 году [1].

Одним из факторов такого роста является внедрение рекомендательных систем среди крупных ритейлеров. Исходя из того, что 70% интерфейса Amazon.com посвящены рекомендациям, можно сделать вывод, что рекомендации играют жизненно важную роль в процессах расширения покупательской корзины и стимулирования продаж. Инфографика от агентства Monetate демонстрирует, что рекомендательные системы могут обеспечить рост выручки на 300%, увеличить показатели конверсии на 150%, а размер среднего чека – на 50%. Эти цифры существенно варьируются в зависимости от выбора их первоначальных значений и от того, насколько широко используются на сайте рекомендательные блоки, но факт роста ключевых показателей эффективности магазинов с рекомендательными системами слишком очевиден, чтобы его игнорировать[2].

Большинство таких систем реализованы на алгоритмах с использованием матричной факторизации и машинного обучения, что требует больших вложений в инфраструктуру, на покупку дорогих серверов для обучения моделей, которое производится по мере обновления данных пользователей и товаров. Так же сложности может вызвать отсутствие достаточного объема обучающей выборки [3]. Модель для выдачи рекомендаций нельзя обучить на сторонних датасетах, как в случае, например с задачей распознавания определенных объектов на изображении.

Для решения вышеупомянутых проблем в данной работе предлагается разработка рекомендательной системы с использованием графовой базы

данных. Это позволит использовать графы как в качестве средства хранения данных, так и в качестве подхода к интеллектуальному анализу данных. Применение данной технологии даст возможность предоставлять рекомендации в режиме реального времени, без необходимости в предварительном вычислении в автономном режиме, тем самым устраняя задержку [4]. Использование графов позволяют реализовать как гибридный подход к рекомендациям так и отдельно подходы на основе контентной фильтрации и коллаборативного поиска. Графовые базы данных демонстрируют постоянную производительность при увеличении размера данных, что позволит внедрить систему рекомендаций как для интернет-магазина, который только недавно запущен, так и для уже крупного магазина.

**Цель работы.** Провести исследование рекомендательных систем, провести обзор вариантов построения рекомендательных систем, а так же реализовать алгоритм коллаборативной фильтрации с использованием графовой базы данных. Для выдачи рекомендаций написать REST API сервис. Разработать прототип интернет-магазина для проверки возможности взаимодействия рекомендательной системы с системой электронной коммерции.

Для достижения вышеуказанной цели были поставлены следующие задачи:

1. Провести анализ моделей и типов систем рекомендаций.
2. Провести обзор варианты построения рекомендательных систем.
3. Выбрать на основе проведенного анализа графовую БД и изучить соответствующий язык запросов к ней.
4. Осуществить поиск, фильтрацию и подготовку к импорту датасета, подходящего для целей данной работы.
5. Реализовать алгоритм коллаборативной фильтрации.

6. Оформить работу в виде REST API сервиса для выдачи рекомендаций.
7. Разработка прототипа интернет-магазина.

## 1 РЕКОМЕНДАТЕЛЬНЫЕ СИСТЕМЫ

Термин рекомендаций система относится ко всем программным инструментам и методикам, которые, используя знания, о пользователях и предметах, дают рекомендации по элементам, которые наиболее вероятно представляют интерес для конкретного пользователя. Рекомендации могут быть связаны с различными процессами принятия решений, например, какие продукты покупать, какую музыку слушать или какие фильмы смотреть. В этом контексте элемент — это общий термин, используемый для определения того, что система рекомендует пользователям. Система рекомендаций обычно ориентирована на определенный тип или класс предметов, таких как книги для покупки, новостные статьи для чтения или отели для бронирования [5].

Рекомендательные системы составляют одну из основных частей экосистемы электронной коммерции, позволяющий пользователям отсеивать большие объемы информации и продуктов. Исследования в этой области позволяют понять, как технология рекомендаций может применяться в конкретных областях. Наличие множества исследовательских работ, демонстрирующие различные алгоритмы рекомендаций и из сочетания, проведение конференций, таких как «The ACM Conference Series on Recommender System», указывают на то, что предоставление качественной рекомендации является актуальной проблемой [6].

Хотя главная цель рекомендательных систем для электронной коммерции — помочь компаниям продавать больше товаров, они также имеют много преимуществ с точки зрения пользователя. Пользователи постоянно перегружены выбором: какие новости читать, какие продукты покупать, какие смотреть и т.д. С увеличением объема каталога товаров пользователи больше не могут просматривать все из них. В этом смысле механизмы рекомендаций предоставляют «индивидуальный» опыт, помогая людям быстрее находить то, что они ищут или что может быть им интересно. Конечным результатом

является то, что удовлетворенность пользователей будет выше, потому что они получают релевантные результаты в более короткие сроки.

Существуют различные причины, по которым все больше и больше поставщиков услуг используют этот набор инструментов и методов, в том числе:

- Увеличение количества проданных товаров. Это одна из самых важных функций: помочь бизнесу продать больше товаров. Из-за огромных каталогов товаров доступных в большинстве магазинов, пользователи часто не могут найти то, что хотят.

- Повышение удовлетворенности пользователей. Правильно спроектированный интерфейс рекомендаций улучшает взаимодействие пользователя с приложением. Сочетание эффективных, точных рекомендаций и удобного интерфейса увеличивает субъективную оценку системы пользователем.

- Лучшее понимание того, что хочет пользователь. Рекомендательная система, создает модель предпочтений пользователя, которая может повторно использоваться бизнесом для ряда других целей, таких как улучшение рекламных кампаний или оптимизация управления запасами.

По некоторым из вышеперечисленных причин, графическое представление данных и анализ на основе графов может облегчить задачу, упрощая управление данными, анализ данных и обмен данными.

Системы рекомендаций различаются по способу, которым они анализируют данные, чтобы выявить отношения между пользователями и предметами. Большинство систем рекомендаций используют три основных подхода [7]:

- контентная фильтрация (content-based filtering);
- коллаборативная фильтрация (collaborative filtering);
- гибридные подход.

## **1.1 Контентная фильтрация (content-based filtering)**

Контентная фильтрация основана на предположении, что пользователи обычно не меняют свой профиль интересов из-за дня в день, а придерживаются в течении длительного периода. Системы рекомендаций по контентной фильтрации рассчитывают набор элементов, которые наиболее похожи на элементы, уже известные текущему пользователю [8].

Основной процесс формирования рекомендаций в данном подходе состоит в сопоставлении атрибутов профиля целевого пользователя, с атрибутами объектов, чтобы найти аналогичные тем, с которые пользователь взаимодействовал в прошлом. Обычно атрибуты для описания объекта - это элементы, извлеченные из метаданных или текстовые элементы, которые так или иначе связаны с этим элементом: описания, комментарии, ключевые слова и т.д. Эти элементы содержат много информации, которую можно использовать для сравнения или определения интересов пользователя на основе списка объектов, с которыми он взаимодействовал [9].

Одним из ключевых вопросов в содержательной рекомендации является качество описательных элементов товара или пользователя. Объекты должны быть описаны так, чтобы могло произойти осмысленное изучение пользовательских предпочтений. В идеале каждый объект должен быть описан с одинаковым уровнем детализации. Но в реальном мире это далеко не так. Описания могут быть частичными, или некоторые части объектов могут быть описаны более подробно, чем другие [10].

## **1.2 Коллаборативная фильтрация**

Данный подход является альтернативой контентной фильтрации и опирается только на поведение пользователей в прошлом, например, на предыдущие покупки или рейтинги элементов, или на мнение существующего сообщества пользователей, позволяющее предсказать, какие элементы

пользователям наиболее вероятно понравятся или будут им интересны, не требуя создания явного профиля для элементов и пользователей [9].

Основная идея заключается в предположении, что если пользователи разделяли одни и те же интересы в прошлом - например, если они просматривали или покупали одни и те же книги - у них в будущем также будут одинаковые вкусы. Таким образом, если, например, пользователь А и пользователь В имеют историю покупок, которая сильно пересекается, и пользователь А недавно купил товар 1, которую пользователь В еще не видел, то вполне возможно, что этот товар ему тоже понравится [8].

Методы совместной фильтрации, как правило, подразделяются на два основных подхода:

1. На основе соседства (neighborhood/memory - based).
2. На основе модели (model - based).

Методы, основанные на памяти, выдают рекомендации для пользователя основываясь на вычислении меры схожести по всем данным, полученных в ходе анализа. Этот метод можно разделить на два подтипа:

1. На основе схожести пользователей (user - based).
2. На основе схожести элементов (item — based).

Подход на основе схожести пользователей является наиболее часто используемым в коллаборативной фильтрации. Чтобы предсказать, какие элементы должны отображаться или рекомендоваться данному пользователю, система полагается на анализ соседства этого конкретного пользователя. Ближайшие соседи вычисляются на основе прошлых взаимодействий. К ним включаются пользователи, которые совершили аналогичный действия для других предметов. Затем для каждого предмета, который текущий пользователь А еще не видел, вычисляется прогноз на основе оценок  $r$ , сделанных ближайшими соседями [9].

Производительность этого подхода страдает при разреженных данных. При больших наборах элементов и небольшом количестве оценок или покупок возникают моменты, когда нельзя дать рекомендации для пользователя, у которого нет общих оценок с другими пользователями. Одним из решений проблемы с разреженностью данных является использование графа для отображения действий пользователей [12].

В отличие от выше рассмотренного алгоритма совместной фильтрации, основанного на пользователях, подход на основе элементов рассматривает набор элементов, оцениваемых целевым пользователем  $A$ , и вычисляет, насколько они похожи на целевой элемент  $1$ , а затем выбирает  $k$  наиболее похожих элементов.

Вторая группа алгоритмов это подход на основе модели. Создаются «модели» для пользователей и элементов, которые описывают их поведение с помощью набора «факторов» или «функций» и веса. Для пользователей каждый фактор выражает то, насколько пользователю нравятся элементы с высокой оценкой соответствующего фактора. В этих методах начальные данные (набор данных пользователь-элемент) сначала обрабатываются в автономном режиме, информация о рейтингах или предыдущих покупках используется для создания этой прогностической модели. Модели скрытого фактора представляют собой наиболее распространенные подходы в этом классе. Они пытаются объяснить рейтинги, характеризуя как предметы, так и пользователей [11].

Алгоритмы в этой категории используют вероятностный подход и рассматривают процесс совместной фильтрации как вычисление ожидаемой оценки пользователя, учитывая его оценки по другим элементам. Процесс построения модели выполняется с помощью различных алгоритмов машинного обучения, таких как байесовская сеть, кластеризация и подходы на основе правил. Модель байесовской сети формулирует вероятностную модель для проблемы совместной фильтрации. Модель кластеризации рассматривает совместную фильтрацию как проблему классификации [12] и работает путем

кластеризации похожих пользователей в одном классе и оценки вероятности того, что конкретный пользователь находится в определенном классе  $C$ , и оттуда вычисляет условную вероятность рейтингов. Подход, основанный на правилах, применяет алгоритмы обнаружения правил ассоциации, чтобы найти связь между купленными предметами, а затем генерирует рекомендации по предметам на основе силы связи между предметами [11].

Недавние исследования показывают, что современные подходы, основанные на моделях, превосходят алгоритмы на основе памяти при прогнозировании рейтингов [13, 14], однако также есть работы где сделаны выводы по поводу того, что одна только точность прогноза не гарантирует пользователям эффективный и удовлетворительный опыт [15].

Как было уже выше сказано, одни из основных причин, по которым компании внедряют рекомендательные системы, заключаются в повышении удовлетворенности и лояльности пользователей и продаже более разнообразных товаров. Рекомендация товара, представляет собой большую ценность, если пользователь не знал об этом товаре, но, скорее всего, обнаружил бы этот товар сам [5]. Этот пример показывает еще один важный фактор, который играет важную роль в оценке пользователями системы рекомендаций: это случайность [15]. С другой стороны, подходы на основе соседства фиксируют локальные ассоциации в данных. Следовательно, система рекомендации, основанная на таком подходе, может рекомендовать пользователю элементы, сильно отличающийся от его обычного вкуса, или элементы, которые не очень известны, если его ближайшие (пользователи) соседи дали ему большой рейтинг.

## 2 ГРАФОВЫЕ БАЗЫ ДАННЫХ

Графовые базы данных — это хранилище данных, построенная на основе теории графов, где граф представляет собой набор узлов и ребер. Узел в модели данных графа представляет собой сущность, а ребра, соединяющие два узла, являются отношениями. Семантические данные также могут быть прикреплены к узлам и ребрам и называются свойствами [16]. Примеры графа показаны на рисунке 1-2.

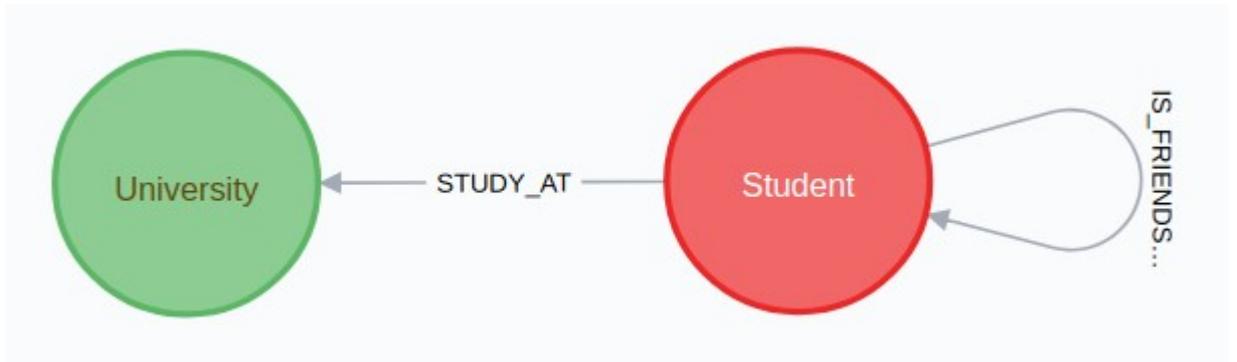


Рисунок 1 — Пример графа в БД Neo4j

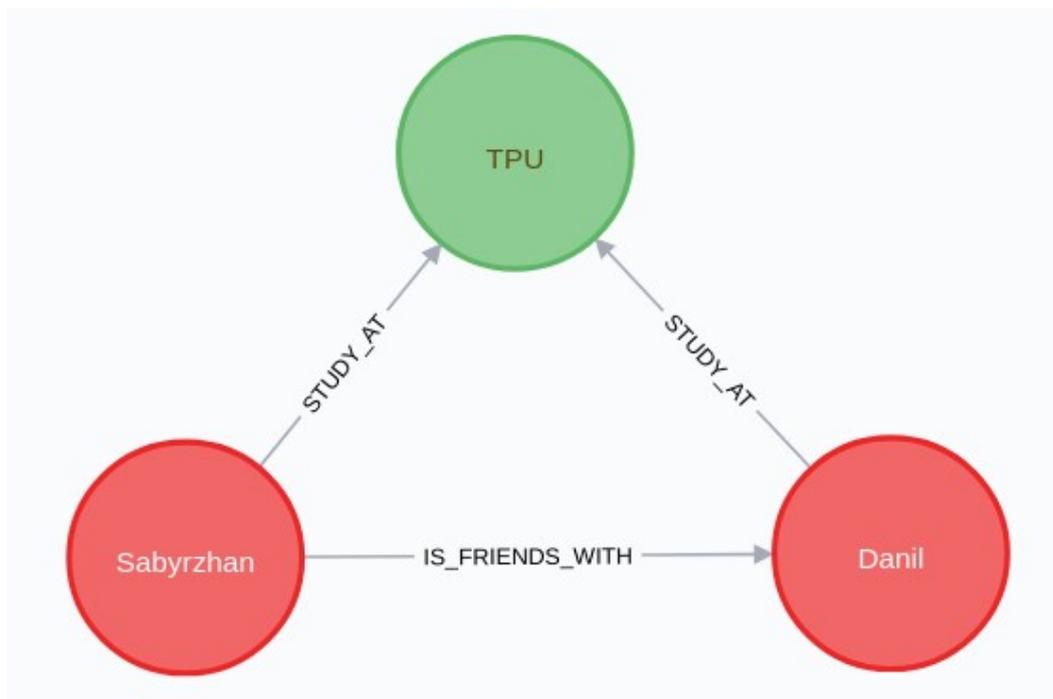


Рисунок 2 — Пример графа в БД Neo4j

Графовые базы данных являются базами данных NoSQL и предоставляют высокопроизводительный обход узлов и поиск данных. Для выполнения запросов в графовые базы данных не существует единого языка. Наиболее популярными языками запросов данных графа являются GraphQL, AQL, Gremlin, SPARQL и Cypher.

Графические базы данных используются многими крупнейшими компаниями в области компьютерных наук, такими как Facebook [17] и Google, благодаря их способности легко масштабироваться с точки зрения производительности независимо от размера набора данных, а также из-за легкости, с которой сложные запросы, такие как извлечение узлов, связанных через неизвестное число ребер, могут быть записаны и выполнены.

Графовые базы данных хороши для случаев использования, которые основаны на отношениях данных. Может использоваться для социальных сетей, рекомендательных систем, систем обнаружения мошенничества и как основное средство хранения данных. Хорошее практическое правило заключается в том, что, если отношения так же важны, как и сущности, следует рассмотреть базу данных графа [18].

Графовые базы данных могут быть классифицированы как нативные и не нативные, имеющие две основные особенности: обработка и хранение. Естественно, нативный код оптимизирован для обработки и хранения графиков, тогда как второй может быть, например, решение на основе реляционных баз данных, преобразованное в модель данных графов [19].

Обработка нативного графа обеспечивает высокую производительность обхода графа независимо от размера базы данных. Благодаря этому он подходит для сложных масштабируемых в реальном времени задач, в то время как реляционная база данных будет серьезно замедляться при объединении десятков таблиц [19].

## Neo4j

Neo4j - это графовая база данных с открытым исходным кодом, с собственным языком запросов Cypher, разработанная Neo4j Inc. Это графовая база данных как с точки зрения хранения, так и представления данных. Ebay, Adobe, Walmart - лишь некоторые из корпоративных клиентов этой технологии [20].

Neo4j использует модель графа, которая хранит данные в двух разных типах; узлы и отношения. Они оба могут содержать неограниченное количество свойств, которые являются парами ключ-значение. Не требуется, чтобы узел или отношение содержали свойства, но они часто находятся по крайней мере на узлах. Метки используются для группировки узлов, которые в основном эквивалентны именам таблиц в реляционных базах данных [21].

Одно из особенностей графовых баз данных состоит в том, что он не содержит схемы. Это означает, что узел может иметь произвольный набор свойств. Например, может быть два узла, user1 и user2, где user1 имеет такие свойства, как имя и номер телефона, а user2 имеет имя и адрес электронной почты. Концепция модели без схемы также применяется и к отношениям. В тоже время в Neo4j также есть понятие Constraints - Ограничения. Ограничения используются, чтобы убедиться, что данные соответствуют определенным правилам.

В Neo4j есть два типа ограничений:

- Ограничение уникальности: указывает, что свойство должно содержать уникальное значение.
- Ограничение существования свойства: гарантируется, что свойство будет существовать для всех узлов с определенной меткой или для всех отношений с определенным типом.

Neo4j использует свой язык запросов Cypher. Язык Cypher читабелен и понятен разработчикам, специалистам по базам данных и заказчикам. Простота

его использования основывается на сходстве способа описания графов с их интуитивным представлением в виде схем [21]. Пример запроса показан на рисунке 3.

```
1 MATCH (s:Student)-[rel:STUDY_AT]→(u:University)
2 RETURN s, rel, u
```

Рисунок 3 — Пример запроса на языке Cypher

Узлы заключены в круглые скобки, а отношения - в квадратные скобки. Свойства могут быть заданы внутри фигурных скобок. Направления отношений обозначены стрелками в стиле ASCII. Можно указать несколько узлов или шаблонов, разделяя их запятыми.

### **3 РЕАЛИЗАЦИЯ РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ**

После тщательного анализа моделей и типов систем рекомендаций был реализован алгоритм коллаборативной фильтрации на основе графовой базы данных, тем самым проверить гипотезу, которая была выдвинутую в начале работы.

Для построения рекомендательной системы, были выполнены следующие шаги:

1. Поиск, фильтрация и подготовка данных для рекомендаций.
2. Импорт данных «Клиент-Заказ-Товар».
3. Непосредственно реализация алгоритма коллаборативной фильтрации на основе пользователей и написание REST API для выдачи рекомендаций.

#### **3.1 Технологии и инструментарий**

##### **3.1.1 Язык программирования**

В качестве языка программирования был выбран Python, так как из языков программирования, с которыми есть наличие опыта работы, для данного языка есть официально поддерживаемый драйвер для работы с Neo4j последней версии по бинарному протоколу Bolt.

Python, как и любой другой язык программирования, имеет свои отличительные особенности. Итак, можно выделить следующие:

- Кроссплатформенность. Python – это интерпретируемый язык, его интерпретаторы существуют для многих платформ. Поэтому с запуском его на любой ОС не должно возникнуть проблем.
- С Python доступно огромное количество сервисов, сред разработки, и фреймворков. Легко можно найти подходящий продукт для работы.

- Возможность подключить библиотеки, написанные на С. Это позволяет повысить эффективность, улучшить быстродействие.

### **3.1.2 Фреймворки**

В качестве фреймворка для написания REST API был выбран FastAPI. FastAPI - это современный, высокопроизводительный фреймворк, которая отлично подходит для создания REST сервисов. Он может обрабатывать как синхронные, так и асинхронные запросы и имеет встроенную поддержку проверки данных, сериализации JSON, аутентификации и авторизации, а также документации OpenAPI.

### **3.1.3 Среда разработки**

Для удобства разработки и развертывания приложения использован Docker и Docker Compose.

Docker – это программная платформа для быстрой разработки, тестирования и развертывания приложений. Docker упаковывает приложение в стандартизованные блоки, называемыми контейнерами. Каждый такой контейнер позволяет включить все необходимое для обеспечения работы приложения: библиотеки, системные инструменты, код и среду исполнения. Docker дает возможность разворачивать и масштабировать приложения в любой среде.

Docker Compose предоставляет способ организации нескольких контейнеров, которые работают вместе.

Использование Compose в основном состоит из трех этапов:

- Определение среды приложения с помощью Dockerfile, чтобы его можно было запустить где угодно.
- Определение служб, из которых состоит приложение, в docker-compose.yml, чтобы их можно было запускать вместе в изолированной среде.
- Запуск docker-compose.

Конфигурации сервисов с помощью yaml-файла показана на листинге 1.

Листинг 1.

```
version: '3.7'

services:
  #neo4j server
  neo4j:
    build:
      context: ../neo4j/
    restart: always
    ports:
      - 7474:7474
      - 7687:7687
    volumes:
      - ../neo4j/data:/data
      - ../neo4j/import:/import
      - ../neo4j/plugins:/plugins
    environment:
      - NE04J_dbms_memory_pagecache_size=1G
      - NE04J_dbms.memory.heap.initial_size=1G
      - NE04J_dbms_memory_heap_max__size=1G
      - NE04J_dbms_security_procedures_unrestricted=gds.*apoc.*
      - NE04J_dbms_security_procedures_whitelist=gds.*,apoc.*
      - NE04J_apoc_import_file_enabled=true
      - NE04JLABS_PLUGINS=["apoc"]
    container_name: neo4j
    networks:
      - app-network

  #recommendation engine api
  recom-engine:
    build:
      context: ../recom-engine/
    container_name: recom-engine
    restart: always
    depends_on:
      - neo4j
    working_dir: /var/www/
    command: bash -c "pip install --upgrade pip && pip install -r
/var/www/requirements.txt --user && uvicorn app.main:app --reload --workers 1 --
host 0.0.0.0 --port 8000"
    volumes:
      - ../recom-engine/:/var/www/
    ports:
      - 8002:8000
    networks:
      - app-network

# Docker networks
networks:
  app-network:
    driver: bridge
```

## 3.2 Поиск, фильтрация и подготовка данных для рекомендаций

На данном этапе был произведен поиск по открытым данным и выбран набор данных от Instacart, выложенный для соревнования на платформе Kaggle.

Набор данных для этого конкурса представляет csv-файлы с данными, описывающих заказы клиентов в течение определенного времени. Набор данных является анонимным и содержит выборку из более чем 3 миллионов продуктовых заказов от более чем 200 000 пользователей Instacart. Для каждого пользователя предоставлены от 4 до 100 их заказов с последовательностью продуктов, приобретенных в каждом заказе.

Для предварительного анализа и подготовки данных для импорта был использован Google Colab - бесплатный облачный сервис на основе Jupyter Notebook. Непосредственно работа с датасетами выполнялась с использованием библиотек языка Python: Pandas и Numpy. Пример обработки одного из файлов показан на листинге 2.

Листинг 2.

```
data_path = '/content/drive/My Drive/datasets/instacart/'
orders = pd.read_csv(data_path + 'orders/orders.csv')
orders = orders[orders['eval_set'] == 'prior']
user_order_count = orders.groupby('user_id')
                        .agg({'order_id': 'count'})
                        .rename(columns={'order_id': 'num_orders'})
                        .reset_index()
user_order_atleast_x = user_order_count
                        .loc[lambda df: (df['num_orders'] >= 5)
                             & (df['num_orders'] <= 10)]
```

### 3.3 Импорт данных «Клиент-Заказ-Товар»

С помощью встроенных инструментов Neo4j для импорта из различных источников данных, в том числе и из CSV-файлов были загружены данные в базу данных.

Пример загрузки одного из файлов показан на листинге 3.

Листинг 3.

```
LOAD CSV WITH HEADERS FROM 'file:///instacart/products.csv' AS row
MERGE (product:Product {productId: toInteger(row.product_id)})
  ON CREATE SET product.productName = row.product_name
MERGE (aisle:Aisle {aisleId: toInteger(row.aisle_id)})
MERGE (department:Department {departmentId: toInteger(row.department_id)})
MERGE (product)-[:IN_AISLE]->(aisle)
MERGE (product)-[:IN_DEPARTMENT]->(department)
```

Визуальное отображение модели загруженных данных показана на рисунке 4.

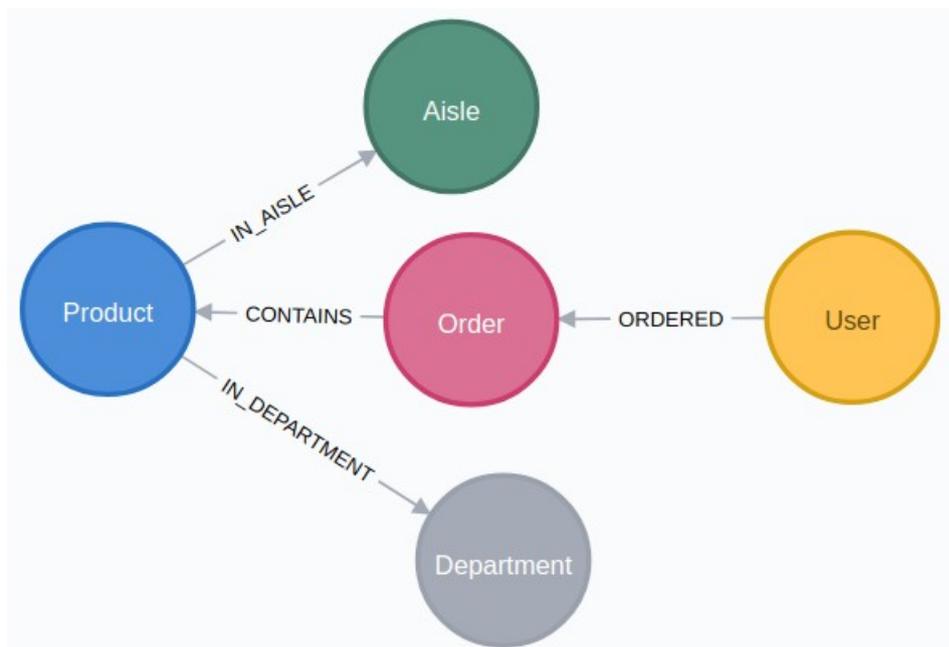


Рисунок 4 — Визуальное представление данных в базе Neo4j

Для реализации алгоритма коллаборативной фильтрации пользователей, важны отношения, которые существуют между пользователями на основе их предпочтений. В данном случае, имеются данные, состоящие из пользователей, которые совершали покупки, оформляя заказы. Эти заказы содержат продукты, которые находятся в отделе и в проходе. Эти отношения легко отображаются с помощью графа:

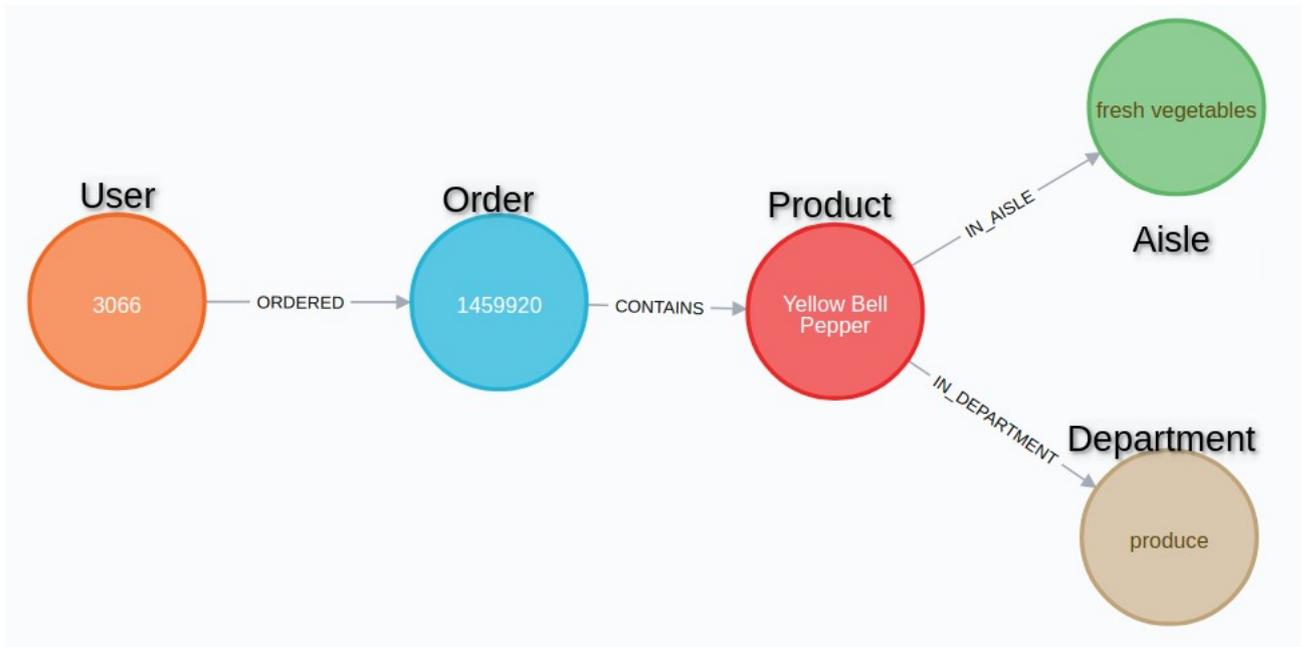


Рисунок 5 — Импортированные данные

Узлы представляют объекты данной системы: Пользователь, Заказ, Продукт, Отдел и Проход. Ребра представляют отношения: ORDERED, CONTAINS, IN\_DEPARTMENT и IN\_AISLE. Атрибуты в узлах представляют свойства, например, у пользователя есть свойство `userId`.

### 3.4 Реализация алгоритма коллаборативной фильтрации на основе схожести пользователей

Для реализация алгоритма коллаборативной фильтрации на основе схожести пользователей были выполнены следующие шаги:

- Расчет для целевого пользователя показателя сходства с остальными пользователями.
- Отбор  $k$  наиболее похожих пользователей.
- Определение продуктов, которые приобретены ближайшими соседями, но не приобретены целевым пользователем.
- Сортировка этих продуктов по количеству, купивших этот товар, соседей.
- Выдача рекомендации в виде  $n$  наиболее подходящих продуктов.

Расчет показателя сходства является одним из наиболее важные аспекты при построении рекомендаций на основе соседства, как это может оказать существенное влияние как на его точность, так и на производительность.

Метрикой подобия, которая использовать в данной реализации, является коэффициент подобия Жаккара или индекс Жаккара, также известный как пересечение над объединением. Индекс Жаккара измеряет сходство между двумя наборами со значениями в диапазоне от 0 до 1. Значение 0 указывает, что два набора не имеют общих элементов, в то время как значение 1 означает, что эти два набора идентичны. Для двух наборов  $A$  и  $B$  индекс Жаккара вычисляется по следующей формуле:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

Выбор данной метрики связан прежде всего с редким характером данных, когда существует большое количество продуктов, и каждый пользователь покупает лишь небольшую часть этих продуктов.

Если бы пользовательские предпочтения были бы смоделированы, используя двоичные атрибуты (то есть: 1, если пользователь купил продукт A, и 0, если пользователь не купил продукт A), было бы много нулей и очень мало единиц. Индекс Жаккара эффективен в этом случае, так как он устраняет совпадающие атрибуты, которые имеют значение 0 для обоих пользователей. Другими словами, при вычислении сходства между двумя пользователями учитываются только те продукты, которые были приобретены обоими пользователями или хотя бы одним из пользователей. Еще одна особенность индекса Жаккара заключается в том, что он учитывает случаи, когда один пользователь покупает значительно больше продуктов, чем тот, с которым его сравнивают.

Для выдачи рекомендаций был написан REST API сервис с двумя конечными точками:

- <http://localhost:8002/api/event/> - для добавления новых заказов пользователя.
- <http://localhost:8002/api/query/> - для выдачи рекомендаций.

Использование для написания сервиса фреймворка FastAPI позволило автоматически сгенерировать документацию согласно спецификации OpenAPI. Конечные точки сервиса показаны на рисунке 6.

# Recommendation Engine 0.0.1 OAS3

/openapi.json

## event

**POST** /api/event/ Put

## recommendation

**POST** /api/query/ Query

### Parameters

No parameters

**Request body** required

Example Value | Schema

```
{
  "item_id": 0,
  "number": 0
}
```

Рисунок 6 — Конечные точки сервиса

Для получения рекомендации нужно отправить POST запрос по адресу:

<http://localhost:8002/api/query/>

с телом запроса:

```
{
  "item_id": {идентификатор пользователя},
  "number": {кол-во рекомендации}
}
```

Пример ответа:

```
[  
  3066,  
  [  
    "Cucumber Kirby",  
    "Yellow Onions",  
    "Broccoli Crown",  
    "Feta Cheese Crumbles",  
    "Michigan Organic Kale"  
  ]  
]
```

## **4 РАЗРАБОТКА ИНТЕРНЕТ-МАГАЗИНА**

Для интеграции разработанной рекомендательной системы с системой электронной коммерции был написан прототип интернет-магазина.

### **4.1 Технологии и инструментарий**

#### **4.1.1 Язык программирования и фреймворки**

Разработка веб-приложения велась на языке программирования PHP с использованием фреймворка Yii2. PHP – это скриптовый язык выполняющийся на стороне сервера, особенно подходящий для создания динамических веб-сайтов. Этот язык программирования предлагает веб-разработчикам большой выбор инструментов. PHP совместим с большинством серверов и может работать как на Windows, так и на серверах UNIX. PHP также поддерживает все основные веб-серверы; будь то Apache или Microsoft IIS.

#### **4.1.2 Фреймворки**

В качестве фреймворка для написания интернет-магазина был выбран Yii2 - это высокопроизводительный, основанный на компонентах PHP фреймворк для быстро развивающихся современных веб-приложений. Данный фреймворк может использоваться для разработки всех видов веб-приложений. Благодаря своей компонентной архитектуре и сложной поддержке кэширования, он особенно подходит для разработки крупномасштабных приложений, таких как порталы, форумы, системы управления контентом (CMS), проекты электронной коммерции.

Для создания адаптивного дизайна использовались следующие популярные библиотеки:

- jQuery;
- Bootstrap.

### 4.1.3 База данных

Для хранения данных была использована СУБД MySQL. MySQL – это поддерживаемая Oracle система управления реляционными базами данных с открытым исходным кодом, основанная на языке структурированных запросов SQL. MySQL работает практически на всех платформах. Хотя он может использоваться в самых разных приложениях, MySQL чаще всего ассоциируется с веб-приложениями.

Схема базы данных представлена на рисунке 7.

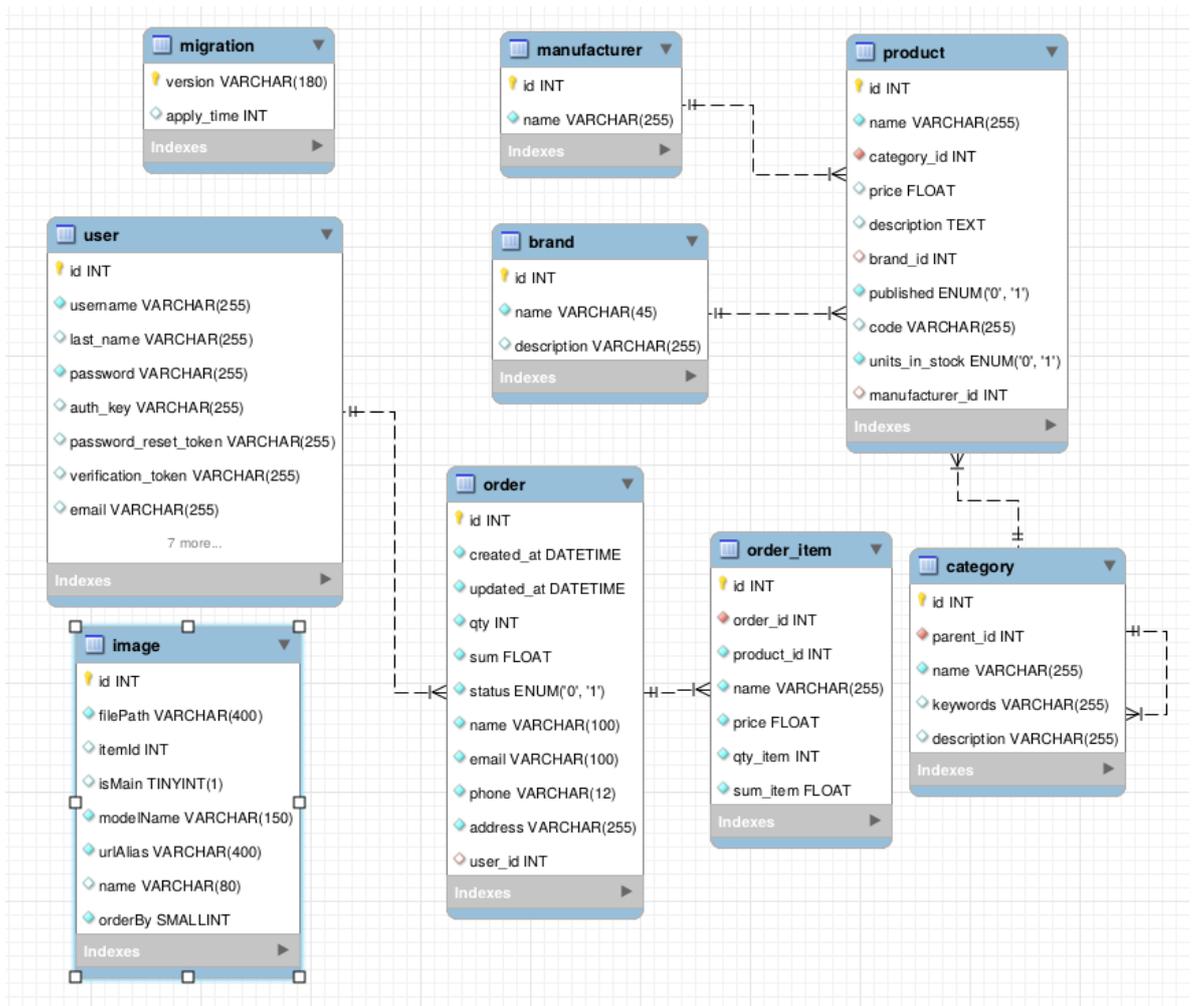


Рисунок 7 — Физическая схема базы данных

Основными сущностями являются «Product», «Order» и «OrderItem». В первой хранится вся необходимая информация для создания товара, во второй и третьей – информация для создания заказа и его дальнейшей обработки.

#### 4.1.4 Среда разработки

Для удобства разработки и развертывания приложения, как и при разработке рекомендательной системы, был использован Docker и Docker Compose. С помощью Docker Compose осуществлялось управление службами из которых состоит разработанное приложение. Сокращенная версия настройки запуска этих служб показана на листинге 4.

Листинг 4.

```
version: '3.7'
services:
  #PHP-Apache
  backend:
    container_name: backend
    ...
    networks:
      - app-network

  #Nginx Service
  frontend:
    container_name: frontend
    ports:
      - 80:8080
      - 443:1443
    ...
    depends_on:
      - backend
      - db
    ...

  # MySQL
  db:
    container_name: db
    ...
    environment:
      MYSQL_DATABASE: ${MYSQL_DATABASE}
    ...
    networks:
      - app-network

  # PhpMyAdmin
  pma:
    container_name: pma
    depends_on:
      - db
    ...
    networks:
      - app-network

# Docker networks
networks:
  app-network:
    driver: bridge

#Volumes
volumes:
  dbdata:
    driver: local
```

Схематично архитектура приложения показана на рисунке 8.

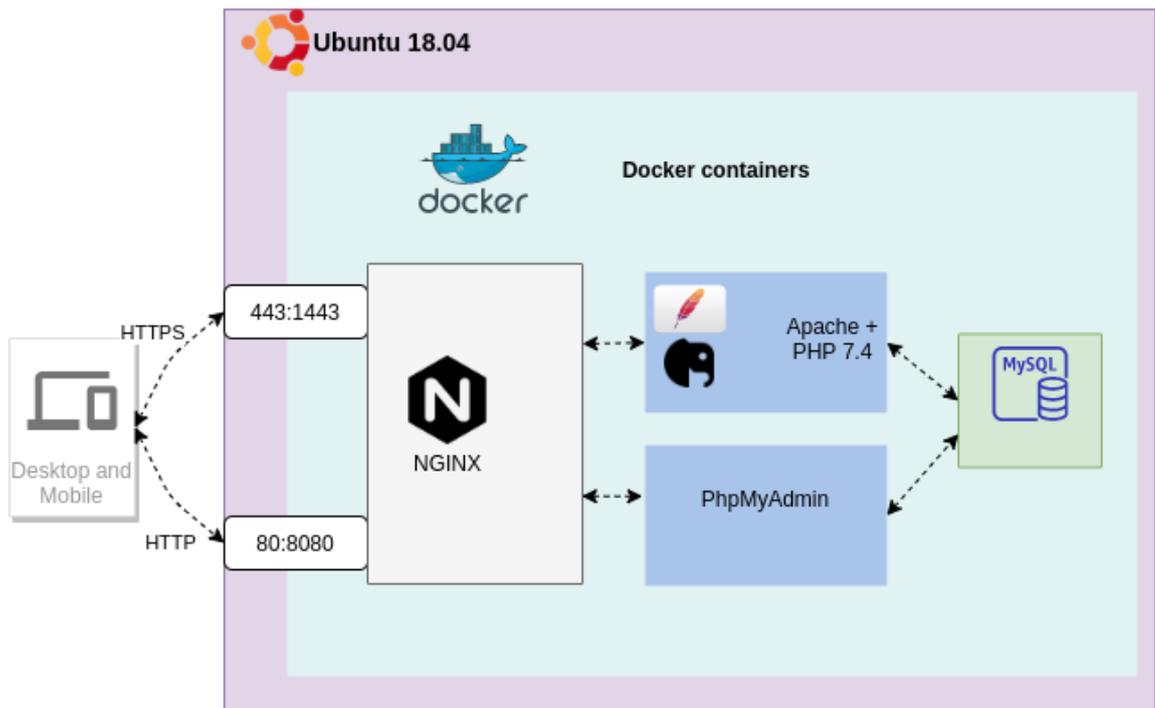


Рисунок 8 — Архитектура приложения

На рисунке видно, что используется распространенная схема размещения Nginx перед Apache в качестве реверс-прокси. В такой конфигурации Nginx называют фронтендом, а Apache — бэкендом.

Nginx самостоятельно обслуживает статический контент, а для динамического контента, для запросов к PHP-скриптам, будет передавать запрос к Apache.

Эта конфигурация позволяет горизонтально масштабировать приложение: можно установить несколько бэкенд серверов за одним фронтендом и Nginx будет распределять нагрузку между ними, увеличивая тем самым отказоустойчивость приложения.

За реверс-прокси также размещен PhpMyAdmin. С помощью Nginx в данном случае обеспечивается безопасность, который ограничивает доступ по IP-адресу (Листинг 5).

## Листинг 5.

```
location ~* ^/phpmyadmin/(.*)$ {
    allow 188.162.15.122;
    deny all;
    set $query $1;
    set $phpmyadmin pma;
    proxy_set_header X-Real-IP $remote_addr;
    proxy_set_header X-Forwarded-For $remote_addr;
    proxy_set_header Host $host;
    proxy_pass http://$phpmyadmin/$query$is_args$args;
}
```

## 4.2 Варианты использования

Концепция разрабатываемой системы предполагает взаимодействие как с зарегистрированным пользователем, так и незарегистрированным.

Сценарий вариантов использования представлены на рисунке 9 - 10.



Рисунок 9 — Диаграмма вариантов для зарегистрированного пользователя



Рисунок 10 — Диаграмма вариантов использования для незарегистрированного пользователя

Также предполагается взаимодействие с администратором магазина (Рисунок 11).

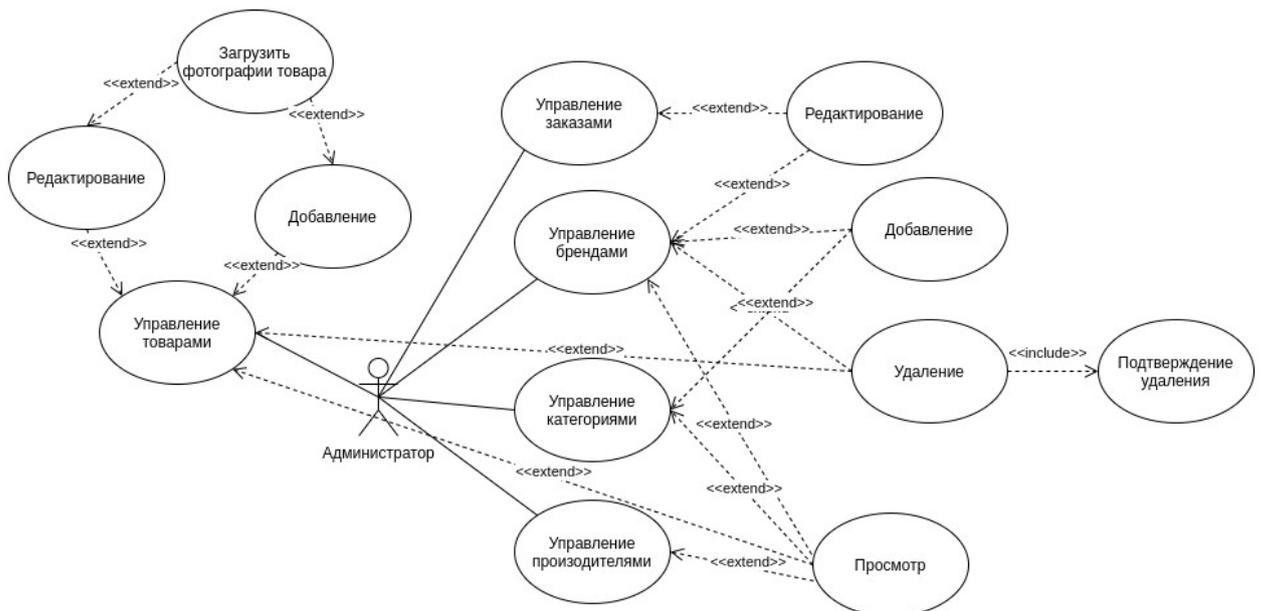


Рисунок 11 — Диаграмма вариантов использования для администратора

### 4.3 Описание разработанного веб-приложения

Ниже представлены текущие результаты разработки. На рисунке 12 изображена главная страница приложения. На ней в виде «карусели»

показываются основные категории, а также популярные товары. Пользователь может перейти к конкретному товару или перейти в каталог для расширенного поиска.

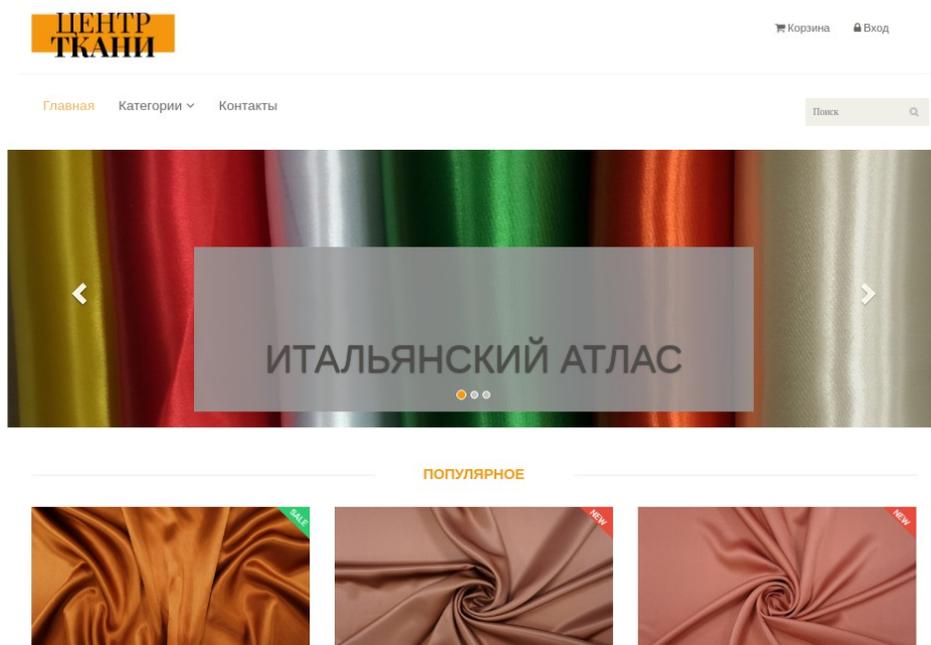


Рисунок 12 — Главная страница

Для авторизации пользователь может воспользоваться кнопкой «Вход» и ввести данные для входа, если пользователь не был зарегистрирован, то ему необходимо перейти по ссылке «Зарегистрироваться» (рисунок 13).

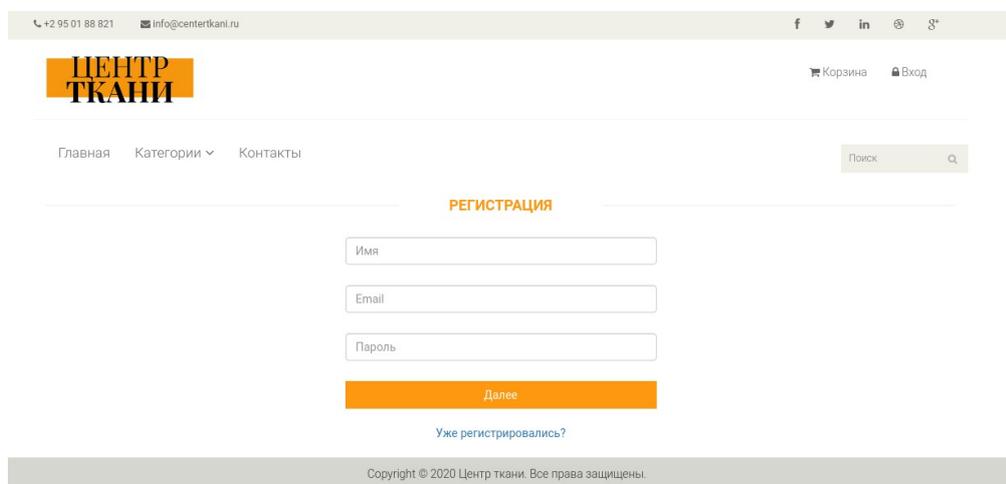


Рисунок 13 — Страница регистрации

После отправки формы, на указанный e-mail придет письмо со ссылкой для подтверждения (рисунки 14).

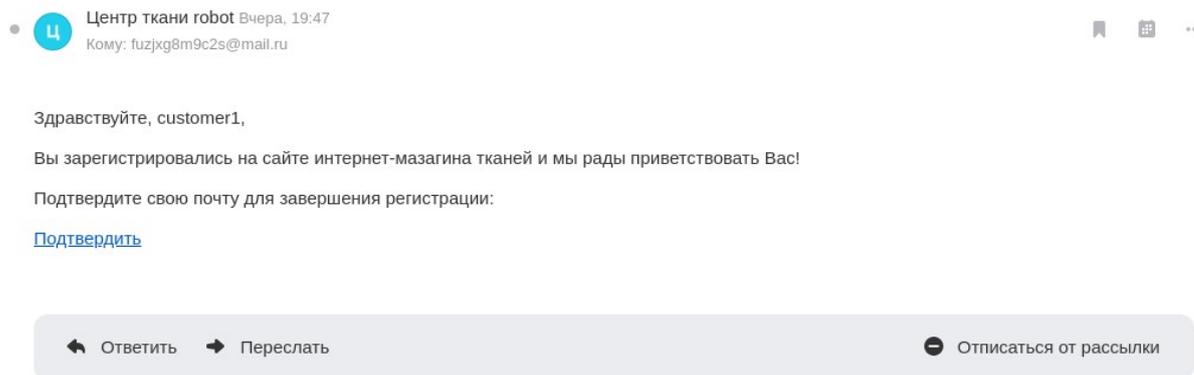


Рисунок 14 – Письмо для подтверждения электронной почты

На рисунке 15 представлен каталог товаров, доступный для любого пользователя. С помощью фильтров можно найти необходимые товары.

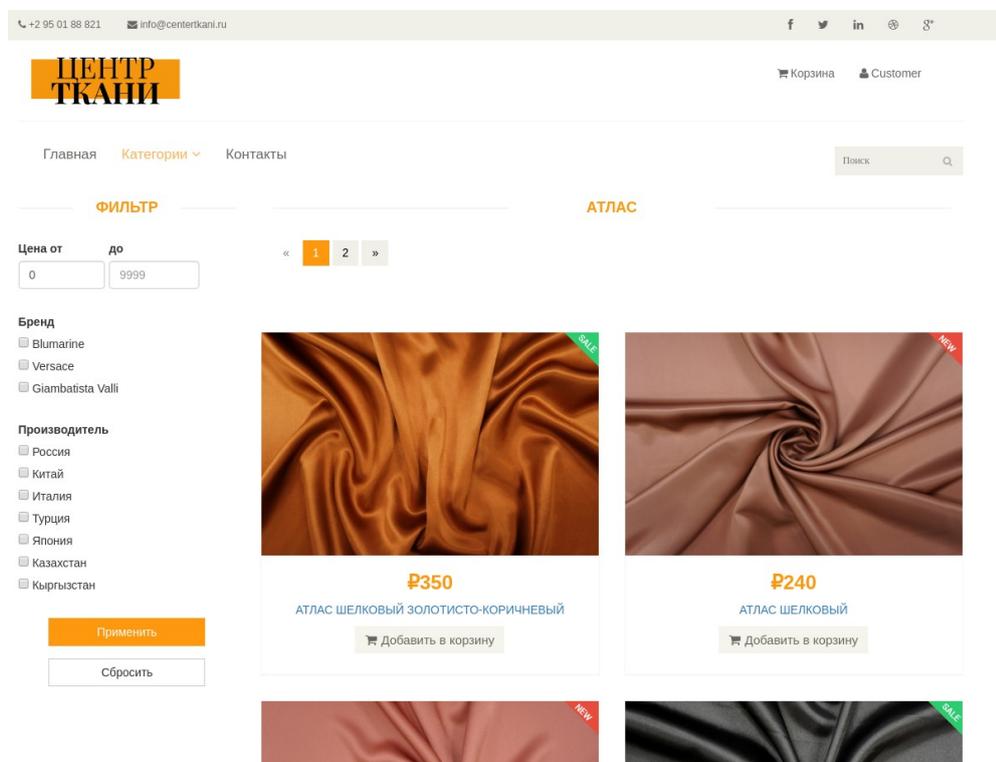


Рисунок 15 — Каталог товаров

Возможность добавления ткани в корзину доступна с любых страниц где присутствует данный товар. По нажатию кнопки «Добавить в корзину»

открывается модальное окно с корзиной, где пользователь может перейти на страницу оформления заказа либо продолжить покупку (рисунок 16).

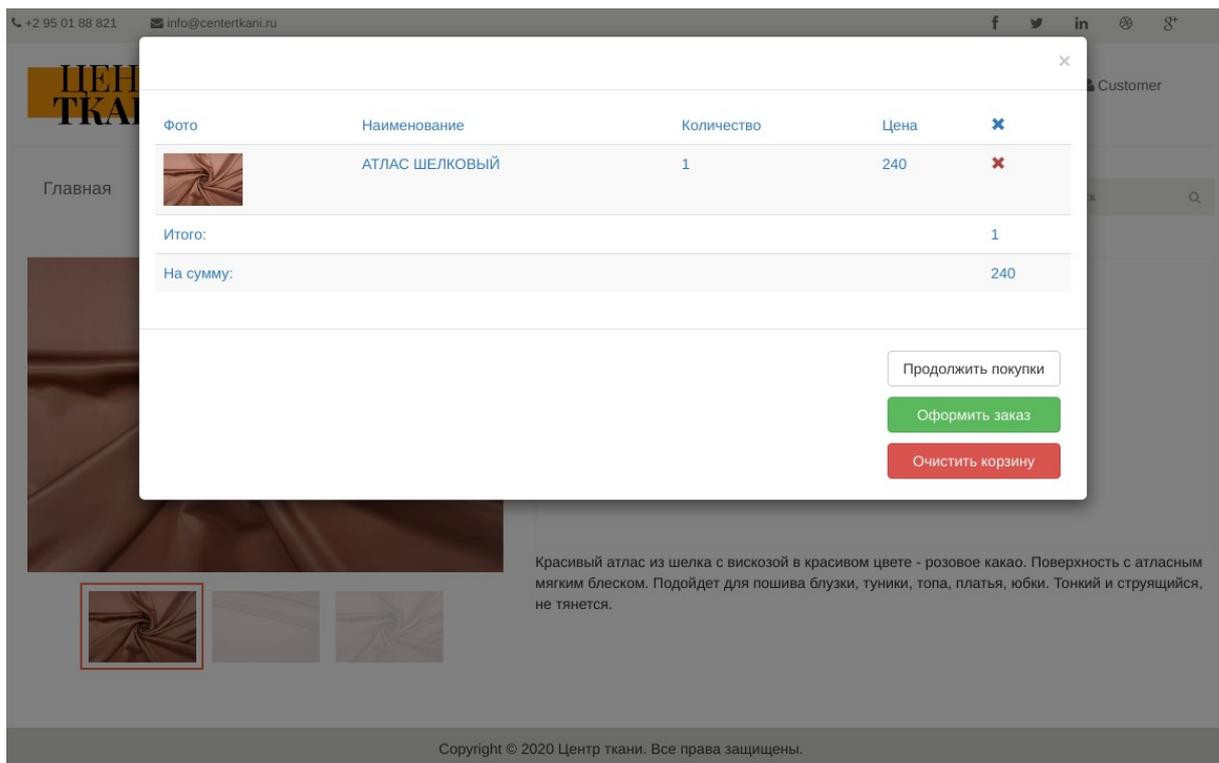


Рисунок 16 — Модальное окно с корзиной

Наполнение ассортимента магазина осуществляется через административную панель (рисунок 17).

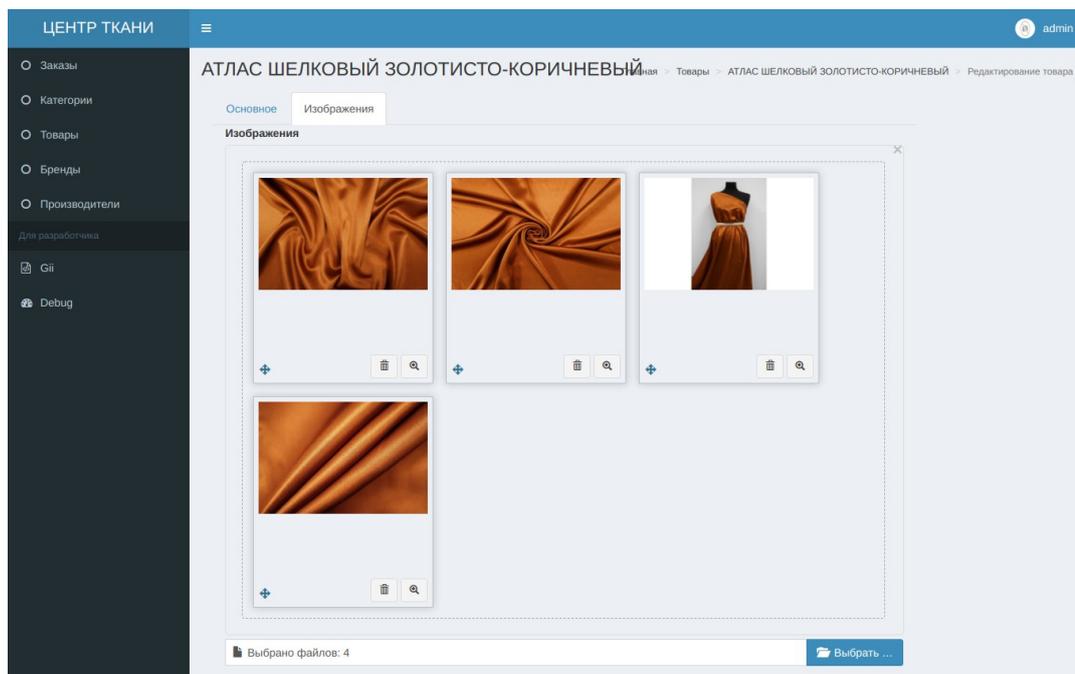


Рисунок 17 — Форма добавления/редактирования товара

## **5 ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ**

### **Введение**

При разработке научно-технического проекта одним из важных этапов является его технико-экономическое обоснование. При помощи него можно выделить преимущества и недостатки разработки данного программного продукта в разрезе экономической эффективности, социальной значимости и других аспектов.

Целью выполнения данного раздела является расчет затрат на создание рекомендательной системы для электронной коммерции.

### **5.1 Оценка коммерческого потенциала и перспективности проведения научных исследований с позиции ресурсоэффективности и ресурсосбережения**

#### **5.1.1 Потенциальные потребители результатов исследования**

Конечным потребителем результатов исследований являются клиенты интернет магазинов.

#### **5.1.2 Исследование целесообразности вложения денежных средств в научно-исследовательский проект**

Для определения целесообразности инвестирования в реализацию проведенных исследований была использована технология Quad (Quality Advisor). Для анализа результатов выделим ряд наиболее важных характеристик проектов такого рода и их показателей. Результаты проведенного сравнения выбранных показателей представлены в таблице 1.

Оценка качества и перспективности по технологии Quad определяется по формуле:

$$P_{cp} = \sum B_i \times B_i,$$

где  $P_{ср}$  – средневзвешенное значение показателя качества и перспективности научной разработки;

$B_i$  – вес показателя (в долях единицы);

$B_i$  – средневзвешенное значение  $i$ -го показателя.

Значение  $P_{ср}$  позволяет говорить о перспективах разработки и качестве проведенного исследования. Если значение показателя  $P_{ср}$  получилось от 80 до 100, то такая разработка считается перспективной. Если от 60 до 79 – то перспективность выше среднего. Если от 40 до 59 – то перспективность средняя. Если от 20 до 39 – то перспективность ниже среднего.

Таблица 1 – Оценочная карта для сравнения конкурентных технических решений (разработок)

Критерии оценки	Вес критерия (В)	Баллы (Б)	Макс. значение	Относит. значение	Средневзвешенное значение ( $P_{ср}$ )
1	2	3	4	5 (3/4)	(5x2)
<b>Показатели оценки качества разработки</b>					
Надежность	0,09	70	100	0,7	6,3
Простота установки	0,4	60	100	0,6	24
Потребность в ресурсах памяти	0,05	95	100	0,95	4,75
Функциональные возможности	0,02	60	100	0,6	1,2
Скорость работы	0,03	45	100	0,45	1,35
<b>Показатели оценки коммерческого потенциала разработки</b>					
Конкурентоспособность продукта	0,1	55	100	0,55	5,5
Уровень проникновения на рынок	0,04	10	100	0,1	0,4
Перспективность рынка	0,1	50	100	0,5	5
Цена	0,15	75	100	0,75	11,25
Послепродажное обслуживание	0,01	60	100	0,6	0,6
Наличие сертификации разработки	0,01	5	100	0,05	0,05
<b>Итого</b>	<b>1</b>	<b>585</b>	<b>110</b>	<b>5,85</b>	<b>60,4</b>

Таким образом, по результатам применения технологии Quad, можно считать, что перспективность, созданной в рамках выполнения ВКР рекомендательной системы для электронной коммерции выше среднего, так как значение средневзвешенной оценки больше 60%.

### 5.1.3 SWOT-анализ

SWOT-анализ заключается в выявлении сильных и слабых сторон проекта, возможностей для дальнейшего развития и угроз существованию и развитию; направлен на исследование внутренней и внешней среды проекта.

Составим итоговую матрицу SWOT-анализа, представленную в таблице 2.

Таблица 2 - Матрица SWOT-анализа

	<b>Сильные стороны научно-исследовательского проекта:</b> С1. Проект уникальный и конкуренты отсутствуют. С2. Высококвалифицированный персонал. С3. Перспективный с точки зрения инвестирования проект. С4. Унифицированность разработки.	<b>Слабые стороны научно-исследовательского проекта:</b> Сл1. Сложность быстрого внедрения. Сл2. Отсутствие своевременного финансирования. Сл3. Возможность снижения спроса на технологию.
<b>Возможности:</b> В1. Появление дополнительного спроса на данный продукт. В2. Широкий спектр применения данной разработки в институтах. В3. Возможность использования контекстной информации для привлечения посетителей	Уникальность проекта и отсутствие конкурентов определяет перспективы проекта. Данные перспективы отражены в возможности внедрения данного модуля в программное обеспечение на сайты институтов других вузов.	Отсутствие своевременного финансирования, а также снижение спроса на данную технологию поставки могут помешать продвижению разработки и привлечению дополнительных средств с других источников для развития проекта. Подобное снижение может отодвинуть дату начала проекта или вовсе закрыть его.
<b>Угрозы:</b> У1. Сильно узкая специализация и область применения разработки. У2. Низкая скорость отклика сайта. У3. Снижение спроса в связи с появлением более простых в применении технологий.	В условиях нынешней конкуренции в разработке, возможны трудности с продвижением продукта. Так же усложняет ситуацию медленная скорость работы сайта.	При недостаточном финансировании разработки могут занять большее время, что приведёт к снижению спроса, в виду появления конкурентных разработок. Помимо этого, разработки конкурентов могут стать более привлекательными для покупателя, что приведёт к снижению спроса и потере доверия потребителя.

#### 5.1.4 Определение возможных альтернатив проведения исследований

Воспользуемся морфологическим подходом для определения возможных альтернатив проведения исследования. Упомянутый подход предполагает точную формулировку проблемы исследования, раскрытие важных морфологических характеристик объекта исследования и раскрытие вариантов по каждой характеристике. Составим морфологическую матрицу для определения морфологических характеристик объекта исследования.

Таблица 3 – Альтернативные варианты проведения исследований

	1	2	3
А. Интегрированная среда разработки	Microsoft Visual Studio 2015	VS Code	PHP Storm
Б. Язык программирования	C#	Python	PHP
В. Хранилище данных	MSSQL	Neo4j	MongoDB

Исходя из составленной морфологической матрицы, можно получить как минимум 3 варианта реализации и направления научных исследований при работе над проектом:

1. А1Б1В1
2. А2Б2В2
3. А3Б3В3

В качестве варианта реализации было выбрано «Исполнение 2». Данный вариант подразумевает использование языка Python и хранилища данных Neo4j. В качестве среды разработки используется VS Code.

## 5.2 Планирование научно-исследовательских работ

### 5.2.1 Структура работ в рамках научного исследования

Планирование комплекса предполагаемых работ осуществляется в следующем порядке:

- определение структуры работ в рамках научного исследования;
- определение участников каждой работы;
- установление продолжительности работ;
- построение графика научных исследований.

Рабочая группа, выполняющая научные исследования, состоит из двух человек: научного руководителя, непосредственного исполнителя – студента.

Таблица 4 – Перечень работ и распределение исполнителей

Основные этапы	№ раб	Содержание работ	Исполнители	Должность исполнителя
Подготовительный этап	1	Выбор темы ВКР	И	И – 100%
	2	Подбор и изучение материалов по теме	НР, И	НР – 100% И – 10%
	3	Формирование возможных решений поставленной задачи, выбор оптимального решения	НР, И	НР – 30% И – 100%
Основной этап	4	Проведение теоретических расчетов и обоснований	НР, И	НР – 100% И – 10%
	5	Описание мероприятий по социальной ответственности	НР, И	НР – 30% И – 100%
	6	Описание ресурсоэффективности ресурсосбережения разработки	НР, И	НР – 100% ИП – 70%
	7	Оценка полученных результатов исследований	НР, И	НР – 100% И – 80%
	8	Определение целесообразности проведения исследований	И	И – 100%
Заключительный этап	9	Составление отчета	И	НР – 60% И – 100%
	10	Защита ВКР	И	И – 100%

## 5.2.2 Определение трудоемкости выполнения работ

Трудовые затраты в большинстве случаев образуют основную часть стоимости разработки, поэтому важным моментом является определение трудоемкости работ каждого из участников научного исследования.

Трудоемкость выполнения научного исследования оценивается экспертным путем в человеко-днях и носит вероятностный характер, т.к. зависит от множества трудно учитываемых факторов. Для определения, ожидаемого (среднего) значения трудоемкости используется следующая формула:

$$t_{\text{ож}i} = \frac{3t_{\text{min}i} + 2t_{\text{max}i}}{5}$$

где  $t_{\text{ож}i}$  – ожидаемая трудоемкость выполнения  $i$ -ой работы чел.-дн.;

$t_{\text{min}i}$  – минимально возможная трудоемкость выполнения заданной  $i$ -ой работы (оптимистическая оценка: в предположении наиболее благоприятного стечения обстоятельств), чел.-дн.;

$t_{\text{max}i}$  – максимально возможная трудоемкость выполнения заданной  $i$ -ой работы (пессимистическая оценка: в предположении наиболее неблагоприятного стечения обстоятельств), чел.-дн.

Исходя из ожидаемой трудоемкости работ, определяется продолжительность каждой работы в рабочих днях  $T_p$ , учитывающая параллельность выполнения работ несколькими исполнителями:

$$T_{p_i} = \frac{t_{\text{ож}i}}{Ч_i}$$

где  $T_{p_i}$  – продолжительность одной работы, раб. дн.;

$t_{\text{ож}i}$  – ожидаемая трудоемкость выполнения одной работы, чел.-дн.

$\mathcal{C}_i$  – численность исполнителей, выполняющих одновременно одну и ту же работу на данном этапе, чел.

Длительность каждого из этапов работ из рабочих дней следует перевести в календарные дни. Для этого необходимо воспользоваться следующей формулой:

$$T_{ki} = T_{pi} \times k_{\text{кал}},$$

где  $T_{ki}$  – продолжительность выполнения  $i$ -й работы в календарных днях;

$T_{pi}$  – продолжительность выполнения  $i$ -й работы в рабочих днях;

$k_{\text{кал}}$  – коэффициент календарности.

Коэффициент календарности определяется по следующей формуле:

$$k_{\text{кал}} = \frac{T_{\text{кал}}}{T_{\text{кал}} - T_{\text{вых}} - T_{\text{пр}}},$$

где  $T_{\text{кал}}$  – количество календарных дней в году;

$T_{\text{вых}}$  – количество выходных дней в году;

$T_{\text{пр}}$  – количество праздничных дней в году.

В 2020 году 366 дней; из них 118 выходных и праздничных дней.

Коэффициент календарности рассчитаем следующим образом.

$$k_{\text{кал}} = \frac{366}{366 - 118} = 1,476 \approx 1,48.$$

Рассчитанные значения в календарных днях по каждой работе  $T_{ki}$  необходимо округлить до целого числа.

Результаты расчетов сведем в таблицу 5.

Таблица 5 – График проведения НТИ

№	Название	Продолжительность, дни	Дата начала работ	Дата окончания работ	Исполнители	Должность исполнителя
1	Составление и утверждение задач	14	03.09.19	18.09.19	НР, И	И – 500% НР – 100%
2	Составление технической документации	8	19.09.19	27.09.19	НР, И	НР – 100% И – 10%
3	Анализ литературы в исследуемой области	36	28.09.19	09.11.19	НР, И	НР – 30% И – 100%
4	Обзор существующих решений в исследуемой области	27	10.11.19	11.12.19	НР, И	НР – 100% И – 10%
5	Календарное планирование работ по теме	10	12.12.19	22.12.19	НР, И	НР – 30% И – 100%
6	Разработка алгоритма, решающего поставленную задачу	27	23.12.19	01.02.20	НР, И	НР – 100% ИП – 70%
7	Подготовка результатов для выполнения анализа	11	02.02. 20	14.02. 20	НР, И	НР – 100% И – 80%
8	Оценка результатов и дальнейшая работа над улучшением алгоритма	15/37	15.02. 20	01.04. 20	И	И – 100%
9	Сравнение эффективности алгоритма с аналогами	11	02.04. 20	14.04. 20	И	И – 100%
10	Анализ и оформление полученных результатов	15	15.04. 20	05.05. 20	НР, И	НР – 30% И – 100%
11	Оформление документации	21	06.05. 20	02.06. 20	И	И – 100%

На основе полученных данных, отраженных в таблице 5, строится календарный план график. Берется то исполнение, которое требует минимальное количество человеко-дней.

### **5.2.3 Составление календарного план-графика работ**

На основании расчетов трудоемкости проводимых исследований в различных исполнениях построим календарный план-график наиболее трудоемкого исполнения. Прямоугольниками со штриховкой будем обозначать трудозатраты студента, прямоугольниками без штриховки – трудозатраты научного руководителя.

Таблица 6 – Календарный план-график работ

№	Вид работ		Продолжительность выполнения работ																													
			Сент			Окт			Ноябрь			Дек			Янв.			Февр			Март			Апр			Май			Июнь		
			1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3			
1	Составление и утверждение технического задания	15	█	█																												
2	Составление технической документации	8			█																											
3	Анализ литературы в исследуемой области	42				█	█	█	█	█	█																					
4	Обзор существующих решений	31							█	█	█	█	█	█																		
5	Календарное планирование работ по теме	10													█	█																
6	Разработка алгоритма, решающего поставленную задачу	40																█	█	█	█	█	█									
7	Подготовка результатов для выполнения анализа	12																			█	█	█	█	█	█						



## 5.2.4 Бюджет научно-технического исследования

При проектировании бюджета научно-технического исследования должно быть обеспечено полное и достоверное отражение всех видов расходов, связанных с его выполнением. В процессе формирования бюджета НИИ используется следующая группировка затрат по статьям:

- материальные затраты НИИ;
- затраты на специальное оборудование для научных работ;
- основная заработная плата исполнителей темы;
- дополнительная заработная плата исполнителей темы;
- отчисления во внебюджетные фонды (страховые отчисления);
- накладные расходы.

### 5.2.4.1 Расчет материальных затрат НИИ

Данная статья включает в себя стоимость всех материалов, используемых при разработке проекта.

Таблица 7 – Материальные затраты

Наименование	Ед.изм.	Количество	Цена за ед., руб.	Затраты на материалы, (Зм), руб.
Флэш-карта	шт.	1	1000	1000
Итого				1000

### 5.2.4.2 Расчет амортизационных расходов

Данная статья включает в себя стоимость амортизация используемого оборудования за время выполнения проекта.

Используется формула:

$$C_{AM} = \frac{H_A * C_{OB} * t_{pf} * n}{F_d},$$

где  $H_A$  – годовая норма амортизации единицы оборудования;

$C_{OB}$  – балансовая стоимость единицы оборудования с учетом ТЗР;

$F_d$  – действительный годовой фонд времени работы соответствующего оборудования;

$t_{рф}$  – фактическое время работы оборудования в ходе выполнения проекта, учитывается исполнителем проекта;

$n$  – число задействованных однотипных единиц оборудования.

При использовании нескольких типов оборудования расчет по формуле делается соответствующее число раз, затем результаты суммируются.

Таблица 8 – Материальные затраты

Наименование	Балансовая стоимость единицы оборудования, руб.	Количество, шт	Годовой фонд времени работы, дней	Годовая норма амортизации	Фактическое время работы, в днях	Амортизация, руб.
Ноутбук	25000	1	248	0,4	220	8871
Итого						8871

#### 5.2.4.3 Расчет затрат на электроэнергию

Данный вид расходов включает в себя затраты на электроэнергию, потраченную в ходе выполнения проекта на работу используемого оборудования, рассчитываемые по формуле:

$$C_{эл.об.} = P_{об} \cdot t_{об} \cdot ЦЭ$$

$$C_{эл.об.} = P_{об} \cdot t_{об} \cdot ЦЭ$$

где  $P_{об}$  – мощность, потребляемая оборудованием, кВт;

$ЦЭ$  – тариф на 1 кВт·час;

$t_{об}$  – время работы оборудования, час.

Для ТПУ  $ЦЭ = 6,59$  руб./кВт·час (с НДС).

Таблица 9 – Затраты на электроэнергию

Наименование оборудования	Время работы оборудования $t_{об}$ , час	Потребляемая мощность $P_{об}$ , кВт	Затраты ЭОБ, руб.
Персональный компьютер	220*8*0,6	0,3	1056
Итого:			1056

#### 5.2.4.4 Расчет основной и дополнительной заработной платы

В рамках реализации проекта руководитель и исполнитель получают заработную плату.

Расчет заработной платы производится исходя из действительного годового фонда рабочего времени (таблица 10), оклада работника и длительности работ по двум исполнениям.

Таблица 10 - Баланс рабочего времени

Показатели рабочего времени	Руководитель	Разработчик
Календарное число дней	366	366
Количество нерабочих дней	118	118
Потери рабочего времени		
отпуск	28	28
невыходы по болезни	0	0
Действительный годовой фонд рабочего времени	220	220

Для учета в ее составе премий, дополнительной зарплаты и районной надбавки используется следующий ряд коэффициентов: К<sub>ПР</sub> = 1,1; К<sub>доп.ЗП</sub> = 1,188; К<sub>р</sub> = 1,3. Таким образом, для перехода от тарифной (базовой) суммы заработка исполнителя, связанной с участием в проекте, к соответствующему полному заработку (зарплатной части сметы) необходимо первую умножить на интегральный коэффициент  $K_{и(6)} = 1,1 * 1,188 * 1,3 = 1,699$ . Вышеуказанное значение К<sub>доп.ЗП</sub> применяется при 6-дневной рабочей неделе, при 5-дневной оно равно 1,113, соответственно в этом случае  $K_{и(5)} = 1,59$

Основная заработная плата работника – произведение среднедневной зарплаты на количество рабочих дней, затраченных на реализацию проекта, по трём исполнениям. Результаты расчетов приведены в таблице 11.

Таблица 11 - Расчёт основной заработной платы

Исполнитель	З <sub>тс</sub> , руб.	Средняя ставка./раб. день	Т <sub>р</sub> , раб. дни	Коэффициент	З <sub>осн</sub> , руб.
Научный руководитель	33664	1342,09	20,7	1,699	47200
Инженер	15470	745,54	220	1,59	260789

#### 5.2.4.5 Отчисления во внебюджетные фонды

Здесь отражаются обязательные отчисления по установленным законодательством Российской Федерации нормам органам государственного социального страхования (ФСС), пенсионного фонда (ПФ) и медицинского страхования (ФФОМС) от затрат на оплату труда.

Отчисления рассчитываются, как основная и дополнительная заработная плата сотрудника, умноженная на коэффициент отчислений. Для работников научно-образовательных учреждений данный коэффициент устанавливается равным 30%.

Результаты расчета отчислений во внебюджетные фонды сводится в таблицу 12.

Таблица 12 - Расчет отчислений во внебюджетные фонды

	З <sub>зп</sub> , руб.	З <sub>внеб.</sub> , руб.
Научный руководитель	47200	14160
Инженер	260789	78236
<b>Итого:</b>		<b>92396</b>

В статье «Прочие расходы» отражены расходы на выполнение проекта, которые не учтены в предыдущих статьях, их следует принять равными 10% от суммы всех предыдущих расходов, т.е.

$$C_{\text{проч.}} = (C_{\text{мат}} + C_{\text{зп}} + C_{\text{соц}} + C_{\text{эл.об.}} + C_{\text{ам}} + C_{\text{нп}}) \cdot 0,1$$
$$C_{\text{проч.}} = 411312 * 0,1 = 41131,2 \text{ руб.}$$

#### 5.2.4.6 Формирование бюджета НИИ

Рассчитанная величина затрат научно-исследовательской работы (темы) является основой для формирования бюджета затрат проекта, который при формировании договора с заказчиком защищается научной организацией в качестве нижнего предела затрат на разработку научно-технической продукции.

Определение бюджета затрат на научно-исследовательский проект по каждому варианту исполнения приведен в таблице 13.

Таблица 13 – Формирование бюджета НТИ

Наименование статьи	Сумма, руб.
1. Материальные затраты НТИ	1000
2. Затраты по заработной плате исполнителей темы	307989
3. Затраты на амортизацию	8871
4. Затраты на электроэнергию	1056
5. Отчисления во внебюджетные фонды	92396
6. Накладные расходы	41131
Итого	452443

### 5.3 Определение ресурсной, финансовой и экономической эффективности исследования

Для расчета интегрального показателя ресурсоэффективности используется сравнительная характеристика вариантов исполнения проекта (таблица 14).

Таблица 14 - Сравнительная оценка характеристик вариантов исполнения проекта

Критерии	Объект исследования	Весовой коэффициент параметра	Исп.1	Исп.2	Исп.3
Удобство в эксплуатации (соответствует требованиям потребителей)		0,20	4	3	5
Ресурсоемкость		0,10	5	3	2
Способствует росту производительности труда пользователя		0,20	4	2	5
Энергосбережение		0,10	4	4	3
Надежность		0,20	5	5	4
Скорость обработки информации		0,20	5	3	4

$$I_{p-исп1} = 4*0,2+5*0,1+4*0,2+4*0,1+5*0,2+5*0,2 = 4,5,$$

$$I_{p-исп2} = 3*0,2+3*0,1+2*0,2+4*0,1+5*0,2+3*0,2 = 3,3,$$

$$I_{p-исп3} = 5*0,2+2*0,1+5*0,2+3*0,1+4*0,2+4*0,2 = 4,1.$$

Интегральный показатель эффективности вариантов исполнения разработки ( $I_{исп}$ ) определяется на основании интегрального показателя ресурсоэффективности по формуле:

$$I_{исп.i} = \frac{I_{p-испi}}{I_{финр}}, \quad I_{исп1} = \frac{4,5}{0,8} = 5,625, \quad I_{исп2} = \frac{3,3}{0,91} = 3,63, \quad I_{исп3} = \frac{4,1}{1} = 4,1$$

Сравнительная эффективность для каждого исполнения рассчитывается по формуле:

$$\mathcal{E}_{cp} = \frac{I_{исп.1}}{I_{исп.i}}$$

Исходя из расчетов для реализации выбран первый вариант исполнения проекта, так как этот вариант является превосходит другие исполнения в ресурсоэффективности.

### **Вывод**

В ходе написания части дипломной работы, затрагивающей финансовую и ресурсную эффективность, была проведена оценка потребителей. Также был проведен SWOT- и QuaD-анализ, анализ конкурентных решений, что позволило выявить слабые и сильные стороны разрабатываемого проекта и найти пути улучшения конкурентоспособности продукта. Также были рассмотрены статьи затрат на реализацию проекта.

## **6 СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ**

### **Введение**

Целью данной магистерской работы является разработка рекомендательной системы для электронной коммерции на основе алгоритма коллаборативной фильтрации. Для разработки системы проводились различные теоретические исследования, анализ и структуризация полученных данных, а также проектирование и программирование веб-приложения с использованием вычислительной техники. Следовательно, работу можно классифицировать как работу разработчика программного обеспечения.

Научно-исследовательская деятельность выполнялась в помещении компании «OFTE». Помещение оснащено персональными электронно-вычислительными машинами (ПЭВМ), компьютерными столами, стульями, столом для коллективной работы, огнетушителями, кондиционером, противопожарной сигнализацией и датчиками дыма.

В данном разделе рассмотрены вопросы, касающиеся соблюдения санитарных норм и правил в процессе разработки рекомендательной системы для электронной коммерции. Рассматриваются меры по защите сотрудника от негативного воздействия среды. Исследуются вредные и опасные факторы среды, а также вопросы охраны окружающей среды от негативного воздействия. Рассматриваются возможные чрезвычайные ситуации и действия, которые сотрудник должен выполнить в случае возникновения ЧС.

### **6.1 Правовые и организационные вопросы обеспечения безопасности**

#### **6.1.1 Правовые нормы охраны труда**

Порядок взаимоотношений работника и работодателя, в том числе и вопросах охраны труда, регулируется Трудовым кодексом РФ.

Нормальная продолжительность рабочего времени не может превышать 40 часов в неделю. Порядок исчисления нормы рабочего времени на определенные календарные периоды (месяц, квартал, год) в зависимости от установленной продолжительности рабочего времени в неделю определяется федеральным органом исполнительной власти, осуществляющим функции по выработке государственной политики и нормативно-правовому регулированию в сфере труда [23].

Сокращенная продолжительность рабочего времени устанавливается [23]:

- для работников в возрасте до шестнадцати лет - не более 24 часов в неделю;
- для работников в возрасте от шестнадцати до восемнадцати лет - не более 35 часов в неделю;
- для работников, являющихся инвалидами I или II группы, - не более 35 часов в неделю;
- для работников, условия труда на рабочих местах которых по результатам специальной оценки условий труда отнесены к вредным условиям труда 3 или 4 степени или опасным условиям труда, - не более 36 часов в неделю.

По соглашению сторон трудового договора работнику как при приеме на работу, так и впоследствии может устанавливаться неполное рабочее время (неполный рабочий день (смена) и (или) неполная рабочая неделя, в том числе с разделением рабочего дня на части). Неполное рабочее время может устанавливаться как без ограничения срока, так и на любой согласованный сторонами трудового договора срок. При работе на условиях неполного рабочего времени оплата труда работника производится пропорционально отработанному им времени или в зависимости от выполненного им объема работ. Работодатель обязан предоставить ежегодный отпуск длительностью в 28 календарных дней [23].

## 6.1.2 Правовые и организационные вопросы обеспечения безопасности

Рассмотрим основные требования по эргономике рабочего места при выполнении работ сидя.

ГОСТ 12.2.032-78 (ССБТ) «Рабочее место при выполнении работ сидя. Общие эргономические требования» регламентирует высоту сидения, ширину пространства для ног, а также высоту рабочей поверхности для человека, выполняющего работы сидя [24].

Все перечисленные параметры выбираются в зависимости от роста работник по номограмме, представленной на рисунке 6.1

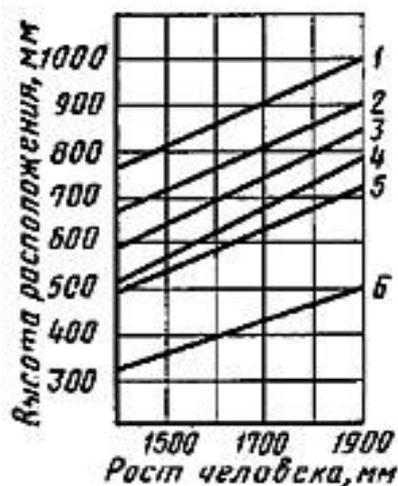


Рисунок 6.1 – Номограмма зависимости высоты рабочей поверхности для разных видов работ (1-4), пространства для ног (5) и высоты рабочего сиденья (6) от роста человека

При проектировании оборудования и организации рабочего места следует учитывать антропометрические показатели женщин (если работают только женщины) и мужчин (если работают только мужчины); если оборудование обслуживают женщины и мужчины - общие средние показатели женщин и мужчин.

Когда невозможно осуществить регулирование высоты рабочей поверхности и подставки для ног, допускается проектировать и изготавливать оборудование с нерегулируемыми параметрами рабочего места. В этом случае числовые значения этих параметров определяют по таблице 15, 16 и чертежу на рисунке 6.2.

Таблица 15 — Соответствие точности работ и высоты рабочей поверхности [24]

Наименование работ	Высота рабочей поверхности, мм			Номер номограммы
	для женщин	для мужчин	для женщин и мужчин	
Очень тонкие зрительные работы (сборка часов, гравировка, картография, сборка очень мелких деталей и др.)	930	1020	975	1
Тонкие работы (монтаж мелких деталей, станочные работы, требующие высокой точности, и др.)	835	905	870	2
Легкие работы (монтаж более крупных деталей, конторская работа, станочные работы, не требующие высокой точности, и др.)	700	750	725	3
Печатание на машинке, типографских станках, перфораторах, легкая сборочная работа более крупных деталей и др.	630	680	655	4

Таблица 16 — Соответствие пола работающего и высоты сидения [24]

Пол работающего	Высота сиденья, мм
Женщины	400
Мужчины и женщины	420
Мужчины	430

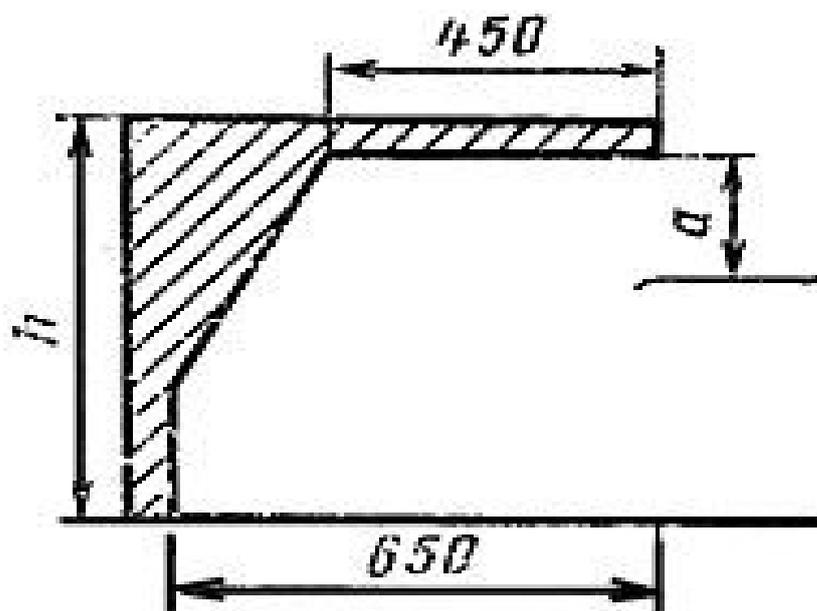


Рисунок 6.2 – Пространство для ног (ширина не менее 500 мм)

## 6.2 Производственная безопасность

Из физических факторов, указанных в ГОСТ 12.0.003-2015 [25], на своем рабочем месте разработчик ПО может подвергнуться влиянию следующих факторов:

- повышенная или пониженная влажность воздуха;
- недостаточная освещенность рабочей зоны;
- повышенный уровень шума на рабочем месте;
- повышенный уровень электромагнитных излучений;
- повышенное значение напряжения в электрической цепи, замыкание которой может произойти через тело человека.

Для каждого из выделенных факторов определены нормативные документы, которые регламентируют действие фактора, результат приведен в таблице 17.

Таблица 17 — Классификация вредных и опасных факторов

Факторы (ГОСТ 12.0.003- 2015)	Этапы работ			Нормативные документы
	Разработка	Изготовление	Эксплуатация	
1. Отклонение показателей микроклимата	+	+	+	1. СанПиН 2.2.4.548–96. Гигиенические требования к микроклимату производственных помещений [26]. 2. ГОСТ 12.1.005-88 ССБТ. Общие санитарно-гигиенические требования к воздуху рабочей зоны [27].
2. Превышение уровня шума	+	+	+	3. ГОСТ 12.1.003-2014 ССБТ. Шум. Общие требования безопасности [28].
3. Недостаточная освещенность рабочей зоны	+	+	+	4. СП 52.13330.2016 Естественное и искусственное освещение. Актуализированная редакция СНиП 23-05-95 [29].
4. Опасные и вредные производственные факторы, связанные с электромагнитными полями	+	+	+	5. СанПиН 2.2.4.1340-03. Гигиенические требования к персональным электронновычислительным машинам и организации работы [30]. 6. СанПиН 2.2.4.3359-16. Санитарно-эпидемиологические требования к физическим факторам на рабочих местах [31].
5.Повышенное значение напряжения в электрической цепи, замыкание которой может произойти через тело человека	+	+	+	7. ГОСТ 12.1.038-82 Система стандартов безопасности труда (ССБТ). Электробезопасность. Предельно допустимые значения напряжений прикосновения и токов [32]. 8. ГОСТ 12.4.011-89 ССБТ. Средства защиты работающих. Классификация [33].

### 6.2.1 Отклонение показателей микроклимата

Согласно ГОСТ 12.1.005-88, микроклимат характеризуется следующими показателями [5]:

- температура воздуха;
- относительная влажность воздуха;
- скорость движения воздуха;
- интенсивность теплового излучения.

При выполнении ВКР и при использовании ее результатов не применяется оборудование, создающее мощное тепловое излучение, следовательно, нет необходимости применять специальные нормы интенсивности теплового облучения от производственных источников.

Оптимальные значения температуры воздуха, поверхностей и скорости движения воздуха согласно СанПиН 2.2.4.548–96 приведены в таблице 18.

Таблица 18 — Оптимальные величины показателей микроклимата на рабочих местах производственных помещений [26]

Период года	Категория работ по уровню энергозатрат, Вт	Температура воздуха, °С	Температура поверхностей, °С	Относительная влажность воздуха, %	Скорость движения воздуха, м/с
Холодный	Ia (до 139)	22-24	21-25	60-40	0,1
	Iб (140-174)	21-23	20-24	60-40	0,1
	IIa (175-232)	19-21	18-22	60-40	0,2
	IIб (233-290)	17-19	16-20	60-40	0,2
	III (более 290)	16-18	15-19	60-40	0,3
Теплый	Ia (до 139)	23-25	22-26	60-40	0,1
	Iб (140-174)	22-24	21-25	60-40	0,1
	IIa (175-232)	20-22	19-23	60-40	0,2
	IIб (233-290)	19-21	18-22	60-40	0,2
	III (более 290)	18-20	17-21	60-40	0,3

Работа разработчика программного обеспечения является малоподвижной и соответствует категории Ia. Следовательно, в холодный период года температура воздуха на рабочем месте должна быть приблизительно равна 22–24°С; температура поверхностей – 21–25°С; относительная влажность воздуха – 60–40%; скорость движения воздуха – 0,1 м/с.

Оптимальная температура воздуха и поверхностей может быть обеспечена за счет централизованной системы отопления или индивидуальных теплогенераторов; качество воздуха – за счет естественной вентиляции или применения механических вентиляционных систем.

## 6.2.2 Недостаточная освещенность рабочей зоны

Освещение оказывает влияние на общее самочувствие и настроение, определяет эффективность труда. Нерационально организованное освещение может явиться причиной травматизма: недостаточно освещенные опасные зоны, слепящие источники света и блики от них, резкие тени и пульсации освещенности ухудшают видимость.

Работа разработчика программного обеспечения, как и любая другая работа с применением ПК, связана с потреблением большого количества зрительной информации. Следовательно, достаточная освещенность рабочего места – один из основных факторов, определяющих качество результатов и безопасность процесса работы специалиста. СП 52.13330.2016 «Естественное и искусственное освещение» устанавливают для работ высокой точности (размер различаемого объекта от 0,30 до 0,50 мм) допустимые значения освещенности, приведенные в таблице 19.

Таблица 19 — Показатели освещенности для работ III класса [29]

Подразряд зрительной работы	Контраст объекта с фоном	Характеристика фона	Искусственное освещение		
			Освещенность, лк		
			при системе комбинированного освещения		при системе общего освещения
			всего	в том числе от общего	
а	Малый	Темный	2000	200	500
			1500	200	400
б	Малый Средний	Средний Темный	1000	200	300
			750	200	200
в	Малый Средний Большой	Светлый Средний- Темный	750	200	300
			600	200	200
г	Малый Средний	Светлый- Средний	400	200	200

Таким образом, примем минимально допустимую достаточную для работы с ПК освещенность равной 200 лк.

Произведем расчет освещения помещения.

Расчёт общего равномерного искусственного освещения горизонтальной рабочей поверхности выполняется методом коэффициента светового потока, учитывающим световой поток, отражённый от потолка и стен.

Световой поток лампы определяется по формуле 2:

$$\Phi = \frac{E_n \cdot S \cdot K_z \cdot Z}{N \cdot \eta} \quad (2)$$

где  $E_n$  – нормируемая минимальная освещённость по СП 52.13330.2016, лк;

$S$  – площадь освещаемого помещения, м<sup>2</sup>;

$K_z$  – коэффициент запаса, учитывающий загрязнение светильника (источника света, светотехнической арматуры, стен и пр., т.е. отражающих поверхностей), наличие в атмосфере цеха дыма, пыли (таблица 6);

$Z$  – коэффициент неравномерности освещения, отношение  $E_{cp}/E_{min}$ . Для люминесцентных ламп при расчётах берётся равным 1,1 [34];

$N$  – число ламп в помещении;

$\eta$  – коэффициент использования светового потока.

Индекс помещения определяется по формуле 2:

$$i = S/h(A+B) \quad (3)$$

Коэффициенты отражения оцениваются субъективно (таблица 7).

Значения коэффициента использования светового потока  $\eta$  светильников для наиболее часто встречающихся сочетаний коэффициентов отражения и индексов помещения приведены в таблице 22.

Таблица 20 — Коэффициент запаса светильников с люминесцентными лампами [34]

Характеристика объекта	Коэффициент запаса
Помещения с большим выделением пыли	2,0
Помещения со средним выделением пыли	1,8
Помещения с малым выделением пыли	1,5

Таблица 21 - Значение коэффициентов отражения потолка и стен [34]

Состояние потолка	$\rho_n, \%$	Состояние стен	$\rho_{ст}, \%$
Свежепобеленный	70	Свежепобеленные с окнами, закрытыми шторами	70
Побеленный, в сырых помещениях	50	Свежепобеленные с окнами без штор	50
Чистый бетонный	50	Бетонные с окнами	30
Светлый деревянный (окрашенный)	50	Оклеенные светлыми обоями	30
Бетонный грязный	30	Грязные	10
Деревянный неокрашенный	30	Кирпичные неоштукатуренные	10

Таблица 22 — Коэффициенты использования светового потока светильников с люминесцентными лампами [34]

Тип светильника	ОД и ОДЛ			ОДР			ОДО			ОДОР			Л71БОЗ ОЛ1Б68		
	$\rho_n, \%$	$\rho_{ст}, \%$	$i$	$\rho_n, \%$	$\rho_{ст}, \%$	$i$	$\rho_n, \%$	$\rho_{ст}, \%$	$i$	$\rho_n, \%$	$\rho_{ст}, \%$	$i$	$\rho_n, \%$	$\rho_{ст}, \%$	$i$
	30	50	70	30	50	70	30	50	70	30	50	70	30	50	70
	10	30	50	10	30	50	10	30	50	10	30	50	10	30	50
	Коэффициенты использования, %														
0,5	23	26	31	21	24	28	21	25	30	18	21	26	14	16	19
0,6	30	33	37	27	30	34	27	31	36	23	27	32	18	20	22
0,7	35	38	42	32	35	38	32	36	41	27	31	35	21	23	25
0,8	39	41	45	35	37	41	36	39	44	30	33	38	23	25	27
0,9	42	44	48	38	40	43	39	42	46	32	36	40	25	27	29
1,0	44	46	49	40	42	45	41	44	48	34	38	42	26	28	30
1,1	46	48	51	41	43	46	42	46	50	36	39	43	27	29	31
1,25	48	50	53	43	45	48	44	48	52	38	41	45	29	30	32
1,5	50	52	56	45	48	51	46	50	55	40	43	47	30	31	34
1,75	52	55	58	47	50	53	49	52	58	42	45	50	31	33	35
2,0	55	57	60	50	52	54	51	55	60	43	47	52	33	34	36
2,25	57	59	62	52	54	56	53	57	62	45	49	54	34	35	37
2,5	59	61	64	53	55	58	55	58	64	47	50	56	35	36	39
3,0	60	62	66	54	56	60	56	60	66	48	52	58	36	37	40
3,5	61	64	67	56	57	61	58	62	67	49	53	59	37	38	40
4,0	63	65	68	57	58	62	59	63	68	50	54	60	38	39	41
5,0	64	66	70	58	60	63	60	64	70	51	56	62	38	40	42

Для общего освещения, как правило, применяются газоразрядные лампы как энергетически более экономичные и обладающие большим сроком службы. Наиболее распространёнными являются люминесцентные лампы. По спектральному составу видимого света различают лампы 41 дневной (ЛД), холодно-белой (ЛХБ), тепло-белой (ЛТБ) и белой цветности (ЛБ). Наиболее широко применяются лампы типа ЛБ. При повышенных требованиях к передаче цветов освещением применяются лампы типа ЛХБ, ЛД. Лампа типа ЛТБ применяется для правильной цветопередачи человеческого лица [34]. Характеристики люминесцентных ламп приведены в таблице 23.

Таблица 23 — Основные характеристики люминесцентных ламп [34]

Мощность, Вт	Напряжение сети, В	Световой поток, лм			
		ЛД	ЛХБ	ЛБ	ЛТБ
15	127	700	820	835	850
20	127	880	1020	1060	1060
30	220	1650	1940	2020	2020
40	220	2300	2700	2800	2850
65	220	3750	4400	4600	4600
80	220	4250	5000	5200	5200
125	220	-	8000	-	8150

При равномерном размещении люминесцентных светильников последние располагаются обычно рядами – параллельно рядам оборудования [34] (рисунок 6.3).

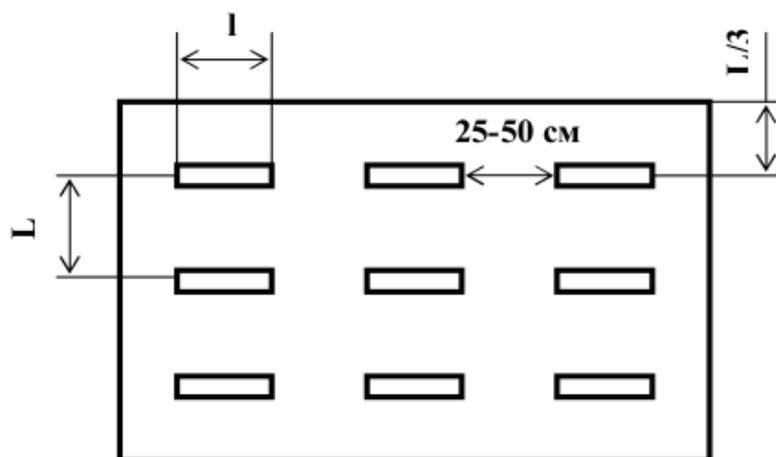


Рисунок 6.3 – Схема размещения светильников в помещении для люминесцентных ламп

Размещение светильников в помещении определяется следующими параметрами, м (рисунок 6.4):

$H$  – высота помещения;

$h_c$  – расстояние светильников от перекрытия (свес);

$h_n = H - h_c$  – высота светильника над полом, высота подвеса;

$h_{rp}$  – высота рабочей поверхности над полом;

$h = h_n - h_{rp}$  – расчётная высота, высота светильника над рабочей поверхностью.

$L$  – расстояние между соседними светильниками или рядами (если по длине ( $A$ ) и ширине ( $B$ ) помещения расстояния различны, то они обозначаются  $L_A$  и  $L_B$ ),

Для создания благоприятных зрительных условий на рабочем месте, для борьбы со слепящим действием источников света введены требования ограничения наименьшей высоты светильников над полом (таблица 24);

$l$  – расстояние от крайних светильников или рядов до стены.

Оптимальное расстояние  $l$  от крайнего ряда светильников до стены рекомендуется принимать равным  $L/3$ .

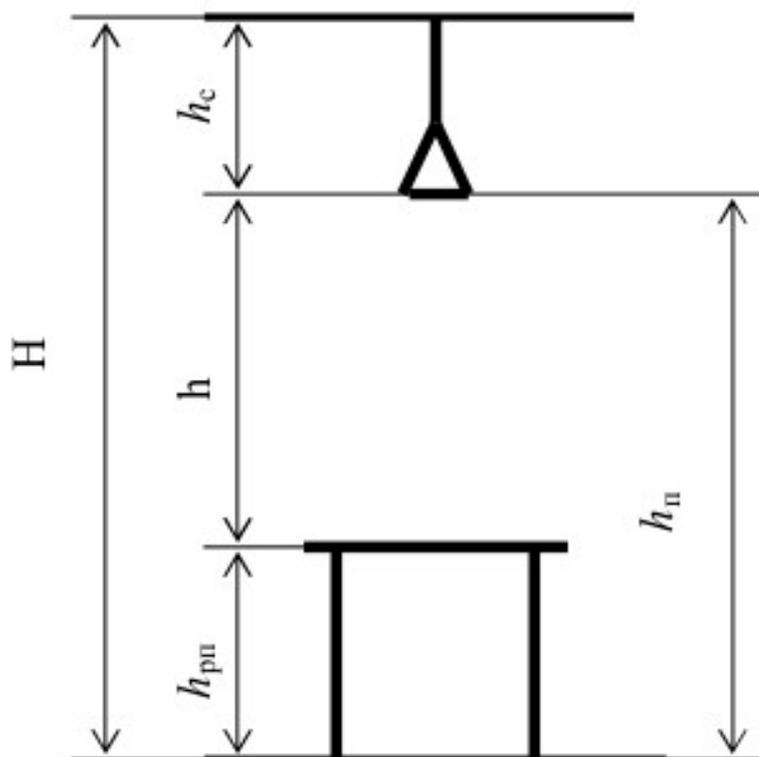


Рисунок 6.4 – Основные расчетные параметры

Таблица 24 — Наименьшая допустимая высота подвеса светильников с люминесцентными лампами

Наименование светильников	Наименьшая допустимая высота подвеса над полом, м
Двухламповые светильники ОД, ОДР, ОДО, ОДОР при одиночной установке или при непрерывных рядах из одиночных светильников	3,5
Двухламповые светильники ОД, ОДР, ОДО, ОДОР при непрерывных рядах из сдвоенных светильников	4,0
Двухламповые светильники ШЛД, ШОД	2,5
Двухламповые уплотнённые светильники ПВХ	3,0

Интегральным критерием оптимальности расположения светильников является величина  $\lambda = L/h$ , уменьшение которой удорожает устройство и обслуживание освещения, а чрезмерное увеличение ведёт к резкой

неравномерности освещённости [34]. В таблице 25 приведены значения  $\lambda$  для разных светильников.

Таблица 25 — Наивыгоднейшее расположение светильников

Наименование светильников	$\lambda$
Люминесцентные с защитной решёткой ОДР, ОДОР, ШЛД, ШОД	1,1 – 1,3
Люминесцентные без защитной решётки типов ОД, ОДО	1,4
ПВЛ	1,5
ГС, ЛЦ	1,6
У	1,8
ШМ	2,3

Кабинет, где выполнялась научно-исследовательская работа имеет следующие размеры: длина  $A = 15$  м, ширина  $B = 12$  м, высота  $H = 4$  м. Высота рабочей поверхности  $h_{pn} = 0,8$  м. Из таблицы 5 видно, что требуется создать освещённость как минимум  $E = 200$  лк. Из таблице 21 имеем следующие субъективно оцененные коэффициенты: коэффициент отражения стен  $R_c = 30\%$  (оклеенные светлыми обоями), коэффициент отражения потолка  $R_n = 50\%$  (чистый бетонный). По таблице 6 коэффициент запаса  $k = 1,5$  (помещения с малым выделением пыли), коэффициент неравномерности  $Z = 1,1$  [34].

Рассчитываем систему общего люминесцентного освещения.

В качестве источника света используются светильники типа ОД, из таблицы 25 видно, что  $\lambda = 1,4$ .

Приняв  $h_c = 0,5$  м, получаем

$$h = 4 - 0,5 - 0,8 = 2,7 \text{ м;}$$

$$L = 1,4 \cdot 2,7 = 3,78 \text{ м; } L/3 = 1,26 \text{ м.}$$

Размещаем светильники в три ряда. В каждом ряду можно установить 7 светильников типа ОД мощностью 40 Вт (с длиной 1,23 м [34]), при этом разрывы между светильниками в ряду составят 50 см. Учитывая, что в каждом светильнике установлено две лампы, общее число ламп в помещении  $N = 42$ .

Находим индекс помещения:

$$i = 180 / (2,7 (15 + 12)) = 2,47.$$

По таблице 22 определяем коэффициент использования светового потока:

$$\eta = 0,61.$$

$$\Phi = \frac{200 \cdot 180 \cdot 1,5 \cdot 1,1}{42 \cdot 0,61} = 2319$$

Определяем потребный световой поток ламп в каждом из рядов:

По таблице 23 выбираем ближайшую стандартную лампу – ЛТБ 40 Вт с потоком 2850 лм. Делаем проверку выполнения условия:

$$-10\% \leq \frac{\Phi_{л.станд} - \Phi_{л.расч}}{\Phi_{л.станд}} \cdot 100\% \leq +20\%$$

Получаем:

$$-10\% \leq 19\% \leq +20\%$$

Рассчитанный поток лампы входит в рамки диапазона.

### 6.2.3 Повышенный уровень шума

Шум –это совокупность различных звуков, возникающих в процессе производства и неблагоприятно воздействующих на организм. Шум может привести к нарушениям слуха (в случае постоянного нахождения при шуме более 85 децибел(dB)), может являться фактором стресса и повысить систолическое кровяное давление. Дополнительно, он может способствовать несчастным случаям, маскируя предупреждающие сигналы и мешая сконцентрироваться.

Для рассматриваемого помещения основными источниками шума являются персональные компьютеры. Нормативным документом, регламентирующим уровни шума для различных категорий рабочих мест служебных помещений, является ГОСТ 12.1.003- 2014. ССБТ. Шум. Общие требования безопасности [28]. Помещения, в которых для работы используются ПК не должны граничить с помещениями, в которых уровни шума превышают нормируемые значения. В помещениях, оборудованных ПК, которые являются основным источником шума при выполнении данных видов работ, уровень

шума на рабочем месте не должен превышать 50 дБА. Для снижения уровня шума могут быть использованы следующие средства:

- низкошумовые устройства вентиляции и кондиционирования;
- звукопоглощающий корпус ПК;
- звукоизолирующая прокладка для корпуса ПК;
- вентиляторы охлаждения корпуса ПК.

#### **6.2.4 Повышенный уровень электромагнитных излучений**

Электромагнитное излучения – возмущение электромагнитного поля, создающееся любой техникой, подключенной к сети питания с переменным напряжением, например, к бытовой электрической сети ~220 В, 50 Гц.

Основные возможные источники питания при работе с ПК – монитор и бесперебойный источник питания. При выполнении работы применялся компьютер, не оборудованный бесперебойным источником питания и оснащенный жидкокристаллическим монитором.

СанПиН 2.2.4.3359-16 определяет нормы допустимых уровней напряженности электрических полей. Они зависят от времени пребывания человека в контролируемой зоне. Время допустимого пребывания в рабочей зоне в часах рассчитывается по формуле:

$$T = (50/E)^{-2}, \quad (4)$$

Где  $E$  — напряженность ЭП в контролируемой зоне, кВ/м;

$T$  — допустимое время пребывания в ЭП при соответствующем уровне напряженности, ч.

Если напряженность электрического поля лежит в диапазоне 20–25 кВ/м, то работа не может продолжаться более 10 минут. При напряженности, не превышающей 5 кВ/м деятельность людей в рабочей зоне может осуществляться в течение 8 часов [31].

Для защиты от электромагнитного излучения монитора рекомендуется использовать защитные экраны; обеспечить заземление массивных

металлических частей компьютера и монитора; делать перерывы во время работы за компьютером.

СанПиН 2.2.4.1340-03 «Гигиенические требования к персональным электронно-вычислительным машинам и организации работы», регламентирует безопасные уровни излучений. В таблицах 19-20 представлены предельно-допустимые уровни напряженности на рабочих местах и допустимые уровни электромагнитных полей.

Таблица 26 — Временные допустимые уровни ЭМП, создаваемых ПЭВМ на рабочих местах [30]

Наименование параметров		ВДУ
Напряженность электрического поля	в диапазоне частот 5 Гц - 2 кГц	25 В/м
	в диапазоне частот 2 кГц - 400 кГц	2,5 В/м
Плотность магнитного потока	в диапазоне частот 5 Гц - 2 кГц	250 нТл
	в диапазоне частот 2 кГц - 400 кГц	25 нТл
Напряженность электростатического поля		15 кВ/м

Таблица 27 — ПДУ постоянного магнитного поля на рабочих местах [31]

Время воздействия	Условия воздействия			
	общее		локальное	
за рабочий день, мин	ПДУ напряженности, кА/м	ПДУ магнитной индукции, мТл	ПДУ напряженности, кА/м	ПДУ магнитной индукции, мТл
≤10	24	30	40	50
11-60	16	20	24	30
61-480	8	10	12	15

### 6.2.5 Электрический ток

Работа разработчика ПО предполагает использование ПК, который является электрической установкой. Следовательно, основным опасным фактором является электрический ток. Поражение электрическим током может спровоцировать электролитическое разложение физиологических жидкостей (крови), непроизвольное сокращение мышц, разрывы и обугливание тканей, поражение глаз. Степень опасности влияния электрического тока зависит от его рода и величины, частоты переменного тока, пути тока через тело человека, продолжительности воздействия тока на тело человека и дополнительных условий внешней среды. Согласно ГОСТ 12.1.038-82 значения токов и напряжений, протекающих через тело человека при нормальном режиме работы электроустановки не должны превышать значений, указанных в таблице 28 [32].

Таблица 28 — Максимально допустимые значения токов и напряжений, протекающих через тело человека в нормальном режиме работы электроустановок [32]

Род тока	Напряжение, В	Ток, А
	(не более)	
Переменный, 50 Гц	2,0	0,3
Переменный, 400 Гц	3,0	0,4
Постоянный	8,0	1,0

Отдельно регламентируются предельно допустимые токи и напряжения в аварийных режимах для производственных и бытовых электроустановок. Персональный компьютер можно отнести к бытовым электроустановкам до 1000 В, питаемых током 50 Гц. Предельно допустимые значения токов и напряжений, проходящих через тело человека в аварийном режиме работы таких электроустановок приведены в таблице 29.

Таблица 29 — Предельно допустимые значения токов и напряжений, протекающих через тело человека в аварийных режимах работы бытовых электроустановок, питаемых переменным током 50 Гц [32]

Нормируемая величина	Предельно допустимые значения при разных продолжительностях воздействия t, с											
	0,01-0,08	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0	Св. 1,0
Напряжение, В	550	340	160	135	120	105	95	85	75	70	60	20
Ток, мА	650	400	190	160	140	125	105	90	75	65	50	6

Согласно ГОСТу 12.4.011-89 «Средства защиты работающих. Общие требования и классификация» [33] к средствам защиты от повышенного уровня статического электричества относятся:

- заземляющие устройства;
- нейтрализаторы;
- увлажняющие устройства;
- антиэлектростатические вещества;
- экранирующие устройства.

Для защиты пользователя ПК от электрического тока применено несколько из перечисленных ранее методов. Так, например, все части компьютера, находящиеся под высоким напряжением находятся в недоступных для пользователя защитных корпусах и снабжены предупреждающими знаками; пользователь физически взаимодействует только с устройствами ввода информации, находящимися под малыми напряжениями (5 В). Для дополнительной защиты от поражения электрическим током предлагается проводить регуляторные инструктажи по технике безопасности.

### 6.3 Экологическая безопасность

Разработка рекомендательной системы выполнялась в офисном помещении. Офис является источником следующих видов отходов:

- Твердые отходы бумага, канцелярские принадлежности, комплектующие;

- Жидкие отходы: сточные воды;
- Люминесцентные лампы.

При использовании компьютера и другой вычислительной техники неизбежно происходит устаревание оборудования. В результате этого возникает необходимость замены и обновления. Чтобы уменьшить негативное воздействие на экологическую обстановку, утилизировать электронику нужно наиболее безопасным для окружающей среды способом, то есть обратиться в специализированную компанию по утилизации. Такие компании действуют на всей территории Российской Федерации, в том числе и в Томской области.

Бумага может быть переработана и использована в качестве вторсырья. Для сбора макулатуры в России существуют специальные пункты приема. Некоторые из них предоставляют услугу вывоза макулатуры.

Сточные воды – один из источников загрязнения природных водоемов, так как они содержат различные загрязнения, в том числе мусор и примеси. В результате деятельности офисного предприятия происходит образование сточных вод. Для последующей очистки в системах канализации применяются отстаивание и фильтрация. Возможна дополнительная очистка с использованием озонаторов и ультрафиолета.

Отдельного внимания заслуживает вопрос утилизации люминесцентных ламп. Они покрыты люминесцентным веществом, имеют стеклянную оболочку и электроды. Внутри таких ламп находится инертный газ с парами ртути. В случае повреждения корпуса лампы, пары ртути попадают в атмосферу. Поэтому после окончания срока службы люминесцентные лампы необходимо сдавать на специальные предприятия по утилизации, имеющие специальную лицензию на данный вид деятельности. Необходимо отметить, что в целом при работе с компьютером существенного загрязнения окружающей среды не происходит и вредные выбросы не сравнимы с производственными.

#### 6.4 Безопасность в чрезвычайных ситуациях

На официальном портале города Томск указаны наиболее возможные чрезвычайные ситуации на территории города [35]:

- обусловленные бытовыми пожарами и ДТП, источник – человеческий фактор, неблагоприятные погодные условия.
- обусловленных обрушением строительных конструкций зданий, источник – ветхость жилого фонда.
- обусловленных авариями на объектах ЖКХ, источник – техническое состояние оборудования, неблагоприятные погодные условия.

Наиболее характерной чрезвычайной ситуацией для объекта, где выполнялась разработка рекомендательной системы, является пожар. Рассмотрим подробнее меры противопожарной защиты.

Причинами возникновения данного вида ЧС могут являться:

- возникновением короткого замыкания в электропроводке;
- возгоранием устройств ПЭВМ из-за неисправности аппаратуры;
- возгоранием устройств искусственного освещения;
- возгоранием мебели по причине нарушения правил пожарной безопасности, а также неправильного использования дополнительных бытовых электроприборов и электроустановок.

Помещение для работы операторов ПЭВМ по системе классификации категорий помещений по взрывопожарной и пожарной опасности относится к категории Д (из 5-ти категорий А, Б, В1-В4, Г, Д), т.к. относится к помещениям с негорючими веществами и материалами в холодном состоянии [36].

Пожарная безопасность подразумевает надлежащее состояние объекта с исключением возможности возникновения очага возгорания (пожара) и его распространения в пространстве. Обеспечение пожарной безопасности — приоритетная задача для любого предприятия. Создание системы защиты регламентировано законом и нормативными документами различных ведомств.

Каждый сотрудник организации должен быть ознакомлен с инструкцией по пожарной безопасности, пройти инструктаж по технике безопасности и строго соблюдать его.

### **Вывод**

В данном разделе проанализированы вредные и опасные производственные факторы. Было установлено, что офисное помещение, в котором производилась разработка рекомендательной , соответствует нормам СНиП и СанПиН. Дополнительных средств защиты сотрудников не требуется. При работе в офисе производятся отходы: бумага, канцелярские принадлежности, люминесцентные лампы и т.д. При надлежащей утилизации этих отходов (с помощью специальных фирм, имеющих лицензию на осуществление утилизации) загрязнение окружающей среды мало. Рассмотрена наиболее распространенная чрезвычайная ситуация – пожар. Пожар может быть следствием короткого замыкания или неверной эксплуатации электроприборов, а также несоблюдения техники безопасности. Поэтому сотрудники офиса проходят инструктаж по правилам пожарной безопасности, проходят учебные эвакуации. Правовые и организационные вопросы обеспечения безопасности регулируются государственными органами. Правила и нормы для обеспечения нормальных условий труда всех сотрудников устанавливаются на государственном уровне.

## ЗАКЛЮЧЕНИЕ

В ходе данной работы был проведен анализ моделей и типов систем рекомендаций, изучены варианты построения рекомендательных систем. Для реализации алгоритма коллаборативной фильтрации были выбраны необходимые программные инструменты. Для выдачи непосредственно самих рекомендации был написан REST API сервис на языке Python. Реализации рекомендательного движка в виде отдельно сервиса даст возможность внедрить в уже готовый интернет магазин.

Для проверки возможности взаимодействия рекомендательной системы с системой электронной коммерции был написан прототип интернет-магазина на языке PHP 7.4 с использованием фреймворков Yii2 и Bootstrap 3. Однако на текущий момент эти системы не интегрированы. В дальнейшем планируется реализовать взаимодействие между ними. Синхронизацию данных между MySQL и Neo4j планируется обеспечить с использованием компонента Kafka Connect брокера сообщений Apache Kafka.

Результаты этой работы показали, что механизм качественных рекомендаций не обязательно нуждается в машинном обучении, как это часто считается. Однако для оценки качества рекомендаций, разработанной системы, необходимо провести комплексное тестирование с реальными пользователями. Одним из вариантов такого тестирования может стать A/B тестирование.

## CONCLUSION

In the course of this work, an analysis of models and types of systems was carried out. To implement the algorithm, the needed software tools were selected. REST API service has been implemented on Python to receive the recommendations. Implementation of the recommendation engine in the form of a separate service will make it possible to integrate it in a ready-made online store.

The prototype of an online store was written on PHP 7.4 using the Yii2 and Bootstrap 3 frameworks, to test the possibility of interaction between the recommendation system and the e-commerce system. However, these systems are currently not integrated. In the future, it is planned to realize the interaction between them. It is planned to provide data synchronization between MySQL and Neo4j using the Kafka Connect component of the Apache Kafka message broker.

The results of this work showed that the system of high-quality recommendations does not necessarily need machine learning. However, to assess the quality of recommendations of the developed system, it is necessary to conduct comprehensive testing with real users. One of the options for such testing may be A/B testing.

## СПИСОК ИСТОЧНИКОВ

1. E-commerce worldwide - Statistics & Facts [Электронный ресурс]. – Режим доступа: <http://www.statista.com/topics/871/online-shopping/> (дата обращения 20.03.2020).
2. Шесть подходов к рекомендациям товаров в интернет-магазине [Электронный ресурс]. — Режим доступа: <https://retailrocket.ru/blog/shest-podhodov-k-rekomendatsiyam-tovarov-v-internet-magazine/> (дата обращения 20.03.2020).
3. Is It Worth It? ROI of Recommender Systems [Электронный ресурс]. — Режим доступа: <https://dzone.com/articles/is-it-worth-it-roi-of-recommender-systems> (дата обращения 05.04.2020).
4. What Are the Major Advantages of Using a Graph Database? [Электронный ресурс]. — Режим доступа: <https://dzone.com/articles/what-are-the-pros-and-cons-of-using-a-graph-databa> (дата обращения 10.04.2020).
5. Francesco Ricci, Lior Rokach, BrachaShapira. Recommender Systems Handbook. Springer, 2015.
6. Ekstrand M., Riedl J., Konstan J. Collaborative Filtering Recommender Systems / Foundations and Trends in Human-Computer Interaction, 2011.
7. G. Adomavicius, A. Tuzhilin. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. IEEE Transactions on Knowledge and Data Engineering, June 2005 <https://doi.org/10.1109/TKDE.2005.99>
8. M. Robillard, W. Maalej, R. Walker, T. Zimmermann. Recommendation Systems in Software Engineering - Basic Approaches in Recommendation Systems. Springer, 2014.
9. Recommender systems: an introduction / D. Jannach, M. Zanker, A. Felfernig, G. Friedrich. – Cambridge University Press, 2010.

10. Robin Burke, Alexander Felfernig, Mehmet H. Göker. Recommender Systems: An Overview. September 2011 *Ai Magazine* 32:13-18.
11. Item-Based Collaborative Filtering Recommendation Algorithms. WWW '01: Proceedings of the 10th international conference on World Wide Web April 2001. <https://doi.org/10.1145/371920.372071> .
12. Graph-based real-time recommendation systems / [Medium.com](https://medium.com) [Электронный ресурс]. URL: <https://medium.com/quantyca/graph-based-real-time-recommendation-systems-8a6b3909b603> (дата обращения 23.04.2020)
13. Koren, Yehuda. Factorization Meets the Neighborhood: A Multifaceted Collaborative Filtering Model. Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2008: 426–434. DOI: <https://doi.org/10.1145/1401890.1401944>
14. Takacs, Gabor, Istvan Pitaszy, Bottyan Nemeth, and Domonkos Tikk. Major Components of the Gravity Recommendation System. SIGKDD Explorations Newsletter 9:2. 2007. 80–83. DOI: <https://doi.org/10.1145/1345448.1345466>
15. Herlocker, Joseph A. Konstan, Loren G. Terveen, and John T. Riedl. Evaluating Collaborative Filtering Recommender Systems. ACM Transactions on Information Systems 2004. <https://dl.acm.org/doi/10.1145/963770.963772>
16. Most Popular Graph Databases [Электронный ресурс]. URL: <https://www.c-sharpcorner.com/article/most-popular-graph-databases/> (дата обращения 21.04.2020)
17. TAO: Facebook’s Distributed Data Store for the Social Graph. Nathan Bronson, Zach Amsden, George Cabrera, Prasad Chakka, Peter Dimov, Hui Ding, Jack Ferris, Anthony Giardullo, Sachin Kulkarni, Harry Li, Mark Marchukov, Dmitri Petrov, Lovro Puzar, Yee Jiun Song, and Venkat Venkataramani, Facebook, Inc. Distributed Data Store for the Social Graph. USENIX [Электронный ресурс]. URL: <https://www.usenix.org/conference/atc13/technical-sessions/presentation/bronson> (дата обращения: 25.04.2020)

18. Why you should use a graph database. [Электронный ресурс]. URL: <https://www.infoworld.com/article/3251829/nosql/why-you-should-use-a-graph-database.html>. (Дата обращения 05.04.2020).
19. Graph Databases for Beginners: Native vs. Non-Native Graph Technology. [Электронный ресурс]. URL: <https://dzone.com/articles/graph-databases-for-beginners-native-vs-non-native>. (Дата обращения 21.04.2020).
20. Neo4j Customers/ [Электронный ресурс]. URL: <https://neo4j.com/customers/>. (Дата обращения 04.04.2020)
21. Ian Robinson, Jim Webber, Emil Eifrem. Graph databases, 2 nd edition, 2015. O'Reilly Media.
22. The Instacart Online Grocery Shopping Dataset 2017. URL <https://www.instacart.com/datasets/grocery-shopping-2017> on 20.04.2020. <https://www.instacart.com/datasets/grocery-shopping-2017>
23. «Трудовой кодекс Российской Федерации» от 30.12.2001 N 197-ФЗ (ред. от 24.04.2020).
24. ГОСТ 12.2.032-78 ССБТ. Рабочее место при выполнении работ сидя. Общие эргономические требования.
25. ГОСТ 12.0.003-2015 ССБТ. Опасные и вредные производственные факторы. Классификация.
26. СанПиН 2.2.4.548-96 Гигиенические требования к микроклимату производственных помещений.
27. ГОСТ 12.1.005-88 ССБТ. Общие санитарно-гигиенические требования к воздуху рабочей зоны.
28. ГОСТ 12.1.003- 2014 ССБТ. Шум. Общие требования безопасности».
29. СП 52.13330.2016 Естественное и искусственное освещение. Актуализированная редакция СНиП 23-05-95
30. СанПиН 2.2.4.1340-03. Гигиенические требования к персональным электронно-вычислительным машинам и организации работы.

31. СанПиН 2.2.4.3359-16. Санитарно-эпидемиологические требования к физическим факторам на рабочих местах.
32. ГОСТ 12.1.038-82 Система стандартов безопасности труда (ССБТ). Электробезопасность. Предельно допустимые значения напряжений прикосновения и токов.
33. ГОСТ 12.4.011-89 ССБТ. Средства защиты работающих. Классификация.
34. Безопасность жизнедеятельности: практикум / Ю.В. Бородин, М.В. Василевский, А.Г. Дашковский, О.Б. Назаренко, Ю.Ф. Свиридов, Н.А. Чулков, Ю.М. Федорчук. — Томск: Изд-во Томского политехнического университета, 2009. — 101 с.
35. ЕДДС города Томска. Краткосрочный прогноз возможных чрезвычайных ситуаций на территории г. Томска на период с 14.02.2020 по 20.02.2020 [Электронный ресурс]. URL: <http://admin.tomsk.ru/pgs/bzk>
36. НПБ 105-03 Определение категорий помещений, зданий и наружных установок по взрывопожарной и пожарной опасности.

## ПРИЛОЖЕНИЕ А

(справочное)

### Recommendation systems

Студент

Группа	ФИО	Подпись	Дата
8ВМ81	Аспомбитов Сабыржан Олегович		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ	Фадеев А.С.	к.т.н.		

Консультант – лингвист отделения иностранных языков ШБИП

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ШБИП	Сидоренко Т.В.	к.т.н.		

## 1. RECOMMENDATION SYSTEM

Recommendation Systems are software tools and techniques that provide suggestions for items that are most likely of interest to a particular user. The suggestions can be related to various decision-making processes, such as what products to buy, which music to listen to, or which films to watch. In this context, item is the general term used to identify what the system recommends to users. A recommendation system normally focuses on a specific type or class of items, such as books to buy, news articles to read, or hotels to book. The overall design and the techniques used to generate the recommendations are customized to provide useful and relevant suggestions for that specific type of item. Other times, it is possible to use information gathered from a class of items to provide recommendations for other types of items. For example, someone who buys suits might be interested in business books or expensive phones.

Recommendation systems are an important part of the e-commerce ecosystem. They represent a powerful method for enabling users to filter through large information and product spaces. Research in this area makes it possible to understand how technology recommendations can be applied in specific areas. The presence of many research papers demonstrating various recommendation algorithms and from a combination of conferences, holding conferences such as “The ACM Conference Series on Recommender System”, indicates that providing high-quality recommendations is the actual problem.

The benefits from a recommendation system for a business is an increase of key performance indicators. Such indicators may include the number of goods sold, sales, sales revenue, duration of the user session on the site, the degree of user involvement, user loyalty to the service.

Although the main purpose of recommendation systems is to help companies to sell more items, they also have a lot of advantages from the user’s perspective. Users are continually overwhelmed by choice: what news to read, products to buy, shows to watch, etc. The set of “items” offered by different providers it is growing

quickly, and users can no longer sift through all of them. In this sense, recommendation engines provide a “customized” experience, helping people to find what they are looking for or what could be of interest to them more quickly. The end result is that user satisfaction will be higher, because they will get relevant results in a shorter amount of time.

Recommendation systems vary in the way they analyze these data sources to develop notions of affinity between users and items, which can be used to identify well-matched pairs.

Most recommendation systems use four basic approaches:

- content-based filtering;
- collaborative filtering;
- hybrid techniques.

### **1.1 Content-based filtering**

Content-based filtering is based on the assumption of monotonic personal interests. For example, users interested in the topic A are typically not changing their interest profile from one day to another but will also be interested in the topic in the (near) future. In online scenarios, content-based recommendation approaches are applied, for example, when it comes to the recommendation of websites (news items with a similar content compared to the set of already consumed news).

Content-based filtering (see Figure 2) relies on two different types of background data: (a) a set of users and (b) a set of categories (or keywords) that have been assigned to (or extracted from) the available items (item descriptions). Content-based filtering recommendation systems calculate a set of items that are most similar to items already known to the current user.

The basic approach of content-based filtering is to compare the content of already consumed items (e.g., a list of news articles) with new items that can potentially be recommended to the user, i.e., to find items that are similar to those already consumed (positively rated) by the user. The basis for determining such a similarity are keywords extracted from the item content descriptions (e.g., keywords

extracted from news articles) or categories in the case that items have been annotated with the relevant meta-information.

One of the key issues in content-based recommendation is feature quality. The objects to be recommended need to be described so that meaningful learning of user preferences can occur. Ideally, every object would be described at the same level of detail and the feature set would contain descriptors that correlate with the discriminations made by users. Unfortunately, this is often not the case. Descriptions may be partial or some parts of the object space may be described in greater detail than others.

## **1.2 Collaborative filtering**

An alternative to content filtering relies only on past user behavior, for example, previous transactions or item ratings, or the opinions of an existing user community to predict which items the users will most probably like or be interested in, without requiring the creation of explicit profiles for both items and users based on item features.

The fundamental assumption behind this method is that other users' opinions can be selected and aggregated in such a way as to provide a reasonable prediction of the active user's preference. Intuitively, they assume that, if users agree about the quality or relevance of some items, then they will likely agree about other items — if a group of users likes the same things as user A, then user A is likely to like the things they like which she hasn't yet seen.

Collaborative filtering methods are usually divided into two main approaches:

- neighborhood/memory – based;
- model – based.

Memory – based methods gives recommendations to the user based on the evaluation of a measure of similarity for all data obtained during the analysis. This approach can be divided into two subtypes:

- user – based.

- item — based.

Memory-based collaborative filtering is the most commonly used approach. This approach based on an analysis of the neighborhood of that particular user to predict which items should be displayed or recommended for him.

Nearest neighbors are compute based on the user past interactions. Nearest neighbors is users who have performed similar actions for other items. Then, for each item that the current user A has not yet seen, a forecast is calculated based on estimates of  $p$  made by the nearest neighbors.

Neighborhood approaches capture local associations in the data. Consequently, it is possible for a recommendation system based on this type of approach to recommend the user a item very different from their usual taste or a item that is not well known, if one of their nearest (user) neighbors has given it a high rating.

The performance of this approach degrades with sparse data. With large sets of products and a small number of ratings or purchases, there are cases when it is impossible to give recommendations for a user who does not have common ratings with other users. One solution to the problem of sparseness of data is to use a graph to display user actions.

Unlike the user-based collaborative filtering algorithm discussed above, the item-based approach considers the set of items evaluated by the target user A and computes how they are similar to target item 1, and then selects the  $k$  most similar items.

The second group of algorithms is a model-based approach. Model-based collaborative filtering algorithms provide item recommendation by first developing a model of user ratings. The models are created for users and elements that describe their behavior using a set of «factors» or «functions» and weight. For users, each factor expresses how much the user likes elements with a high rating of the corresponding factor. In these methods, the initial data (user-element dataset) is first processed offline, ratings or previous purchases are used to create this predictive

model. Hidden factor models are the most common approaches in this class. They are trying to explain the ratings, characterizing both objects and users.

Algorithms in this category take a probabilistic approach and envision the collaborative filtering process as computing the expected value of a user prediction, given his/her ratings on other item. The model building process is performed by different machine learning algorithms such as Bayesian network, clustering, and rule-based approaches. The Bayesian network model formulates a probabilistic model for collaborative filtering problem. Clustering model treats collaborative filtering as a classification problem and works by clustering similar users in same class and estimating the probability that a particular user is in a particular class  $C$ , and from there computes the conditional probability of ratings. The rule-based approach applies association rule discovery algorithms to find association between co-purchased items and then generates item recommendation based on the strength of the association between items.

While recent investigations show state-of-the-art model-based approaches as being superior to neighborhood ones at the task of predicting ratings, there is also an emerging understanding that good prediction accuracy alone does not guarantee users an effective and satisfying experience.

As said above, some of the main reasons why companies implement recommendation engines are to increase user satisfaction and loyalty and to sell more diverse items. Moreover, recommending to a user a item constitutes a novel recommendation if the user was not aware of that item, but most probably, the user would have discovered it on his own. This example shows another relevant factor that has been identified as playing an important role in users appreciation of the recommendation system: the serendipity.

## 2. GRAPH DATABASE

Graph databases are developed based on graph theory where a graph is a set of nodes and edges. A node in a graph data model represents an entity. The edges connecting two nodes is a relationship. Semantic data can be attached to nodes and edges called properties. Some examples of graph are shown in the figure 1-2.

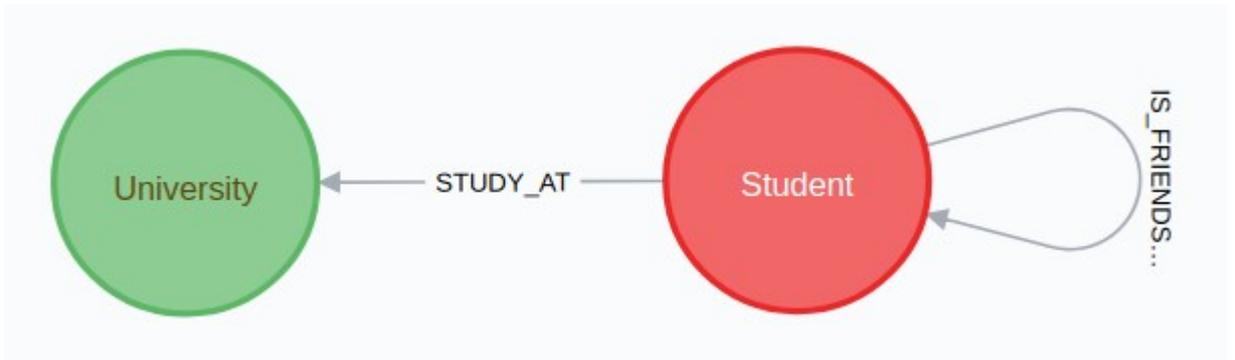


Figure 1 - Graph in the Neo4j database

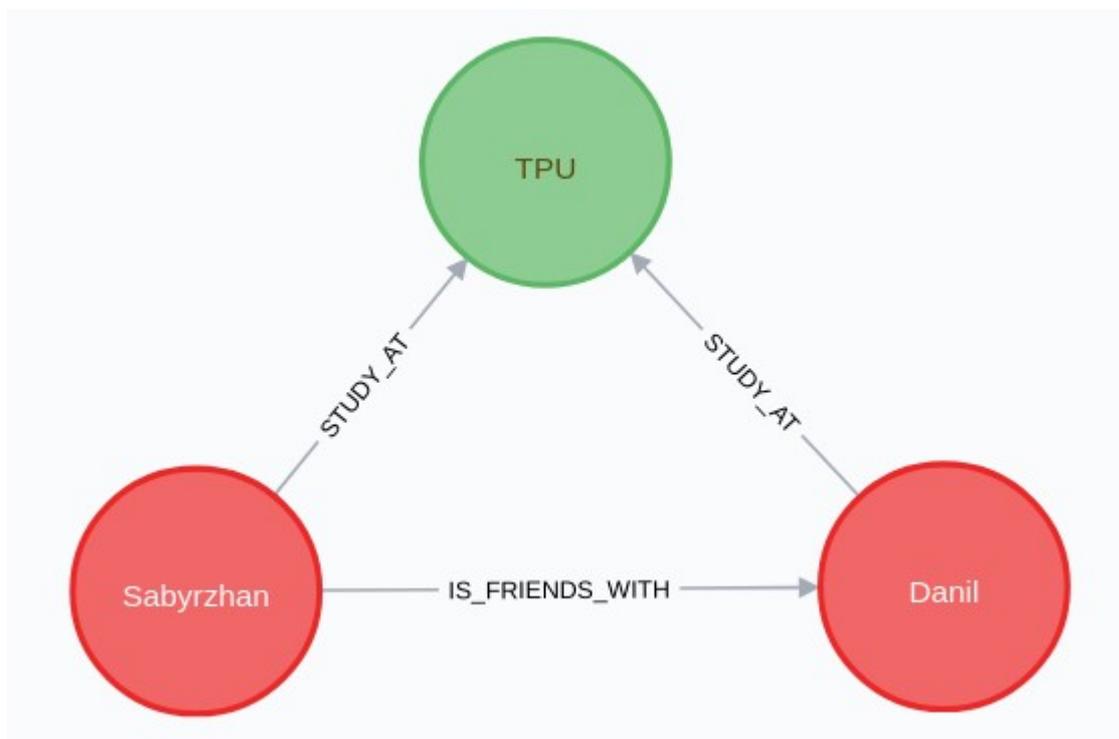


Figure 2 - Graph in the Neo4j database

Graph databases are NoSQL databases and provide high-performance node traversal and data retrieval. To query graph databases, there is no one single language.

Most popular graph data query languages are GraphQL, AQL, Gremlin, SPARQL, and Cypher.

Graph databases have been employed by many giants in the field of computer science such as Facebook or Google, because graph databases can easily scale in terms of performance regardless of the size of a dataset as well as because graph databases can do easily a complex queries, such as retrieving nodes linked through an unknown number of edge.

Graph databases are good for use cases that rely on data relationships. They are commonly used for social networks, recommendations, fraud detection and master data management among other things. A good rule of thumb is that if the relationships are as important as the nodes, a graph database should be considered.

Graph databases can be categorized as native and non-native, having two main features to differentiate these: processing and storage. Naturally the native is optimized for processing and storing graphs, whereas non-native might be for example a relational database-based solution converted to a graph data model.

Native graph storages are specifically designed for storing and managing graphs, whereas non-native means that the storage is provided elsewhere. It could be provided from a relational or NoSQL database, and therefore the algorithms being not optimized for storing graphs, resulting in a lack of efficiency.

### **Neo4j**

Neo4j is an open source graph database developed by Neo4j Inc. It is a graph database both in terms of storage as well as the representation of the data. Walmart, eBay and Medium are just some of the corporate customers of this technology.

Neo4j uses property graph model, which stores data into two different types; nodes and relationships. They both may contain unlimited amount of properties, which are basic key-value pairs. It is not a requirement for a node or a relationship to contain properties, but they are often found at least on the nodes. Labels are used for grouping the nodes, which are basically the equivalent of table names in relational databases.

Neo4j uses its Cypher query language. Cypher is designed to be easily read and understood by developers, database professionals, and business stakeholders. Its ease of use derives from the fact that it is in accord with the way we intuitively describe graphs using diagrams. An example request is shown in Figure 3.

```
1 MATCH (s:Student)-[rel:STUDY_AT]→(u:University)
2 RETURN s, rel, u
```

Figure 3 – An example query in Cypher

The nodes are wrapped in parentheses and relationships in square brackets. Properties may be given inside curly braces. Relationship directions are indicated with ASCII style arrows. Multiple nodes or patterns may be given for clauses by separating them with commas.