

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

Институт кибернетики
Направление подготовки 01.04.02 Прикладная математика и информатика
Кафедра программной инженерии

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Тема работы
Применение математических методов для оценки качества рекомендательной системы библиотечного каталога ТПУ

УДК 025.343.2-021.465:519.876(571.16)

Студент

Группа	ФИО	Подпись	Дата
8БМ51	Ильмауров Вячеслав Олегович		

Руководитель

Должность	ФИО	Ученая степень, звание	Подпись	Дата
ведущий программист	Колобов Олег Сергеевич	К.Т.Н.		

КОНСУЛЬТАНТЫ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
доцент	Конотопский Владимир Юрьевич	К.Э.Н.		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
доцент	Извеков Владимир Николаевич	К.Т.Н.		

ДОПУСТИТЬ К ЗАЩИТЕ:

ИО зав. кафедрой	ФИО	Ученая степень, звание	Подпись	Дата
доцент	Иванов Максим Анатольевич	К.Т.Н.		

Томск – 2017 г.

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

Институт кибернетики
Направление подготовки Прикладная математика и информатика
Кафедра Прикладной математики

УТВЕРЖДАЮ:
Зав. кафедрой

(Подпись) (Дата) (Ф.И.О.)

ЗАДАНИЕ
на выполнение выпускной квалификационной работы

В форме:

магистерской диссертации

Студенту:

Группа	ФИО
8БМ51	Ильмаюров Вячеслав Олегович

Тема работы:

Применение математических методов для оценки качества рекомендательной системы библиотечного каталога ТПУ

Утверждена приказом директора Института кибернетики (дата, номер)

Срок сдачи студентом выполненной работы:
(дата)

ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

Исходные данные к работе

(наименование объекта исследования; документы конференции и отчеты НИИР; программное обеспечение).

Объектом исследования является рекомендательная система, данные представлены в файлах формата csv. Используются следующие программные платформы: Node.JS и PostgreSQL.

<p align="center">Перечень подлежащих исследованию, проектированию и разработке вопросов</p> <p align="center"><i>(аналитический обзор по литературным источникам с целью выяснения достижений мировой науки техники в рассматриваемой области; постановка задачи исследования; содержательная, концептуальная, математическая структурная модель; разработка алгоритма решения задачи; выбор программного обеспечения; индивидуальные выводы о практической значимости проведенных исследований; дополнительные разделы, подлежащие разработке; заключение по работе).</i></p>	<p>Целью исследования является оценка качества рекомендательной системы библиотечного каталога ТПУ. Исходя из этого были поставлены следующие задачи:</p> <ul style="list-style-type: none"> • выбор метода исследования для оценки качества рекомендательной системы • разработка и реализация информационной модели для оценки качества рекомендательной системы с использованием программных средств • анализ полученных результатов
--	--

<p align="center">Перечень графического материала</p> <p align="center"><i>(с точным указанием обязательных чертежей)</i></p>	
--	--

Консультанты по разделам выпускной квалификационной работы
(с указанием разделов)

Раздел	Консультант
Финансовый менеджмент	Конотопский Владимир Юрьевич
Ресурсоэффективность и ресурсосбережение	Конотопский Владимир Юрьевич
Социальная ответственность	Извеков Владимир Николаевич
Английский язык	Морозов Василий Сергеевич

<p align="center">Названия разделов, которые должны быть написаны на русском и иностранном языках:</p>
Введение
Обзор литературы
Объекты и методы исследования
Заключение
Список литературы

<p align="center">Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику</p>	
---	--

Задание выдал руководитель:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
ведущий программист	Колобов Олег Сергеевич	к. т. н.		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8БМ51	Ильмаюров Вячеслав Олегович		

РЕФЕРАТ

Выпускная квалификационная работа 77 с., 7 рис., 18 табл., 13 источников.

Ключевые слова: рекомендательная система, метрика подобия, коллаборативная фильтрация, K ближайших соседей, метрика Sensitivity (Recall)

Объектом исследования является рекомендательная система библиотечный каталог НТБ ТПУ

Цель работы – оценить качество рекомендательной системы библиотечного каталога НТБ ТПУ

В процессе исследования проводились: выбор метода для оценки качества рекомендательной системы библиотечного каталога НТБ ТПУ, построение информационной модели системы для оценки качества рекомендательной системы библиотечного каталога НТБ ТПУ, планирование и проведение вычислительного эксперимента по оценке качества рекомендательной системы

В результате исследования разработана программа для ЭВМ и произведена оценка рекомендательной системы библиотечного каталога НТБ ТПУ

Основные конструктивные, технологические и технико-эксплуатационные характеристики: используемые программные средства: Node.JS и PostgreSQL

Степень внедрения: разработано программа для ЭВМ, которая позволяет оценить качество рекомендательной системы библиотечного каталога НТБ ТПУ по заказам пользователей

Область применения: рекомендательные системы, используемые в разных информационных системах

Экономическая эффективность/значимость работы: данное исследование носит дидактический характер и не несет экономической выгоды

В будущем планируется улучшить программу для ЭВМ, которая применялась к различному роду рекомендательных систем.

ОПРЕДЕЛЕНИЯ

рекомендательная система – 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 рекомендаций (прогнозов) в рекомендательных системах, который используя известные предпочтения группы пользователей, прогнозирует неизвестные предпочтения другого пользователя;

релевантный – соответствующий интересам запросам.

СОКРАЩЕНИЯ

НТБ – научно-техническая библиотека;

ТПУ – Томский политехнический университет

Оглавление

Введение.....	8
1 Обзор литературы	11
2 Объект и методы исследования	16
2.1 Рекомендательные системы	16
3 Выбор меры оценки качества системы	19
3.1 Статистические меры оценивания	19
3.2 Мера классификации	19
3.3 Сравнительный анализ методов	23
3.4 Применение меры Sensitivity (Recall) и меры Precision.....	24
4 Реализация меры Sensitivity (Recall) для оценки качества системы.....	28
4.1 Информационная модель и используемые программные средства..	28
4.2 Вычислительный эксперимент	30
4.3 Анализ результатов.....	33
5 Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	35
5.1 Организация и планирование работ	35
5.1.1 Продолжительность этапов работ	36
5.1.2 Расчет накопления готовности проекта.....	41
5.2 Расчет сметы затрат на выполнение проекта	42
5.2.1 Расчет затрат на материалы	42
5.2.2 Расчет заработной платы.....	43
5.2.3 Расчет затрат на социальный налог	44
5.2.4 Расчет затрат на электроэнергию	44
5.2.5 Расчет амортизационных расходов.....	45
5.2.6 Расчет прочих расходов	46
5.2.7 Расчет общей себестоимости разработки.....	46
5.2.8 Расчет прибыли	47
5.2.9 Расчет НДС	47
5.2.10 Цена разработки НИР	47

5.3 Оценка экономической эффективности проекта.....	48
5.3.1 Оценка научно-технического уровня НИР	48
6 Социальная ответственность	50
6.1 Производственная безопасность	51
2. Напряженность зрения.....	51
6.1.1 Микроклимат.....	51
6.1.2 Шум на рабочем месте	53
6.1.3 Электромагнитные излучения	54
6.1.4 Освещенность рабочей зоны.....	55
6.1.5 Расчет системы искусственного освещения на рабочем месте оператора персональной ЭВМ.....	58
6.1.6 Электробезопасность.....	60
6.2 Экологическая безопасность	62
6.3 Безопасность в чрезвычайных ситуациях	63
6.4 Правовые и организационные вопросы обеспечения безопасности .	66
6.4.1 Психофизиологические факторы	66
6.4.2 Организационные мероприятия при компоновке рабочей зоны	67
Заключение	72
Список литературы	74

Введение

Одним из бурно развивающихся направлений совершенствования прикладных информационных технологий является развертывание рекомендательных систем – инструментов автоматической генерации рекомендаций по выбранному объекту на основе изучения персональных потребностей клиентов [1]. В обыденной жизни пользователя информационной системы рекомендации встречаются почти на каждом сайте: начиная от интернет-магазинов, заканчивая различными развлекательными порталами, посвященные музыке, фильмам и так далее. Одними из первых рекомендательных систем были система интернет-магазина Amazon и система компании Google. В системе интернет-магазина Amazon используется база данных предпочтений пользователей по отношению к различным объектам для создания рекомендаций пользователям. В 1992 г. в качестве основного алгоритма для рекомендательных систем был предложен метод коллаборативной фильтрации. Этот метод позволил решить задачу достаточно эффективно, оказался весьма совершенным и в настоящее время наблюдается улучшение показателя качества рекомендательной системы на 10% в конкурсе Netflix Prize [2].

Подмечено, что применение рекомендательных систем находят не только в развлекательной сфере жизни человека, но и также в образовательной сфере человека. Любому учащемуся приходилось искать информацию в библиотеке. Эта информация касалась как его учебной деятельности (научные статьи, учебники, задачки), так и его личных интересов (художественная литература, журналы). При поиске данной информации могут возникать определенные сложности. Построение рекомендаций тоже является поиском информации, что способно облегчить поиск объектов библиотечного каталога, и в отличие от поиска по запросу в библиотечном каталоге не требует человеческих затрат или сторонней помощи сотрудника библиотеки. Поэтому стоит отметить, что развертывание рекомендательной системы в библиотечном каталоге является

актуальной и немаловажной задачей, способной облегчить поиск информации большему числу пользователей.

Анализ публикаций показывает, что несмотря на различные исследования в области рекомендательных систем остается ряд принципиальных проблем построения систем, которые требуют разработки новых алгоритмов, специфических для этого класса систем обработки данных. Первая из этих проблем носит название проблемы масштабируемости, а вторая – «холодного старта». Наличие проблемы масштабируемости требует от алгоритмов рекомендательных систем возможность неограниченного наращивания числа как пользователей, так и числа вариантов возможных рекомендаций. Проблема «холодного старта» заключается в необходимости обеспечения работоспособности алгоритма для генерации рекомендаций пользователям, которые впервые вошли в систему, и информация о них отсутствует или чрезвычайно бедна. Разработка алгоритмов, разрешающих эти проблемы, позволяет строить рекомендательные системы, которые являются адаптивными по отношению к росту числа пользователей и товаров (в электронной коммерции) и к объему данных, имеющихся о пользователях.

Исследования рекомендательных систем для информационных ресурсов библиотек в отечественной практике не выполнялись и примеров реализаций рекомендательных систем для отечественных библиотек нам не известны. Поэтому, по причине отсутствия опыта разработки рекомендательной системы для библиотек на территории РФ была разработана рекомендательная система для библиотечного каталога научно-технической библиотеки Томского Политехнического Университета (ТПУ), которая способна выдать список рекомендаций для определенного пользователя библиотеки. Предположим, что в данной системе отсутствуют проблемы масштабируемости и холодного старта и данная рекомендательная система выдает рекомендации пользователям. В теории рекомендательная система должна выдать релевантные рекомендации из множества имеющихся данных. Тогда перед этапом внедрения данной рекомендательной системы в информационную систему НТБ ТПУ может

возникнуть следующая проблема: выдает ли данная рекомендательная система соответствующие рекомендации для пользователей библиотечного каталога, т. е. каково качество данной рекомендательной системы. Также пользователю важно получать не только какие-нибудь рекомендации из библиотечного каталога, но и получать их согласно его интересам, запросам. Может возникнуть ситуация, что рекомендательная система выдает рекомендации, а соответствие рекомендаций интересам, запросам пользователя оставляют желать лучшего. Целью данного исследования и будет являться оценка качества рекомендательной системы библиотечного каталога НТБ ТПУ. Существует множество разных подходов для оценки качества выдаваемых рекомендаций системой. У каждого подхода имеются свои плюсы и минусы. Тогда одной из задач будет являться выбор метода или подхода конкретно для рекомендательной системы библиотечного каталога НТБ ТПУ. Объектом нашего исследования является рекомендательная система библиотечного каталога НТБ ТПУ. Данная рекомендательная система работает с такими данными как пользователи и объекты, элементы заказов пользователей. Под объектами, элементами заказов пользователей понимаются различные материалы библиотечного каталога: книги, журналы, статьи, периодические издания и т. д.

Исходя из данной цели работы можно сформировать следующие задачи, решаемые в ходе данной работы:

- выбор метода исследования для оценки качества рекомендательной системы
- разработка и реализация информационной модели для оценки качества рекомендательной системы с использованием программных средств
- анализ полученных результатов

1 Обзор литературы

Про разработку рекомендательных систем и существующих алгоритмах построения рекомендательных систем посвящено много статей и книг. В частности, в статьях [3, 4] говорится о важности применения рекомендательных систем в информационных системах, которые в современном быстро развивающемся информационном обществе помогают среди большего объема информации найти пользователю нужную, соответствующую его интересам информацию. Также в статьях раскрыто понятие коллаборативной фильтрации, алгоритм работы данного подхода. Освещены такие подходы коллаборативной фильтрации как *item-based* (основанный на объектах) и *user-based* (основанный на пользователях).

Качество рекомендательных систем может оцениваться различными мерами. Существуют меры, основанные на классификации ответов системы и соответствию запросам пользователя [5]. Эта группа мер применяется и при анализе данных. Данными мерами являются Accuracy, Sensitivity (Recall), Specificity, Precision. Также существуют и меры, которые являются комбинациями исходных мер. Например, *G – Mean* (среднее геометрическое) *F-Measure*, которые рассчитывается по формуле:

$$G - Mean = \sqrt{Sensitivity * Specificity} \quad (1)$$

$$F - Measure = \frac{2 * Sensitivity * Precision}{Sensitivity + Precision} \quad (2)$$

В статье [5] приводится анализ данных мер классификации в зависимости от различных информационных моделей.

Статистические меры точности оценивают точность системы, сравнивая рекомендации и оценки пользователей для всех пар пользователь-объект в тестовой выборке. Например, мера средней абсолютной ошибки (Mean Absolute Error – MAE) между предсказаниями и оценками является широко используемой мерой. Мера средней абсолютной ошибки (Mean Absolute Error – MAE) есть мера

отклонения рекомендаций от их действительных пользовательских оценок. Существует также и схожая мера – это мера корень средней квадратичной ошибки (Root Mean Squared Error), которая также используется как статистическая мера точности [3, 4].

Для того, чтобы оценить качество рекомендательной системы нужно построить информационную модель для оценки качества системы. Информационная модель это по сути модель объекта. Она представлена в виде информации, что описывает существенные для конкретного случая параметры и переменные, связи между ними, а также входы и выходы для данных, при подаче на которые можно влиять на получаемый результат. Их нельзя увидеть или потрогать. В целом они не имеют материального воплощения, поскольку строятся на использовании одной информации. Сюда относятся данные, что характеризуют состояния объекта, существенные свойства, процессы и явления, а также связь с внешней средой. Это процесс называется описанием информационной модели. Это самый первый шаг проработки. Полноценной информационной моделью является обычно сложная разработка, которая может иметь много структур, что в рамках статьи сведены в три основных типа:

- **Описательная.** Сюда относятся модели, которые создаются на естественных языках. Они могут иметь любую произвольную структуру, которая удовлетворит составляющего их человека.
- **Формальная.** Сюда относят модели, которые создаются на формальных языках (научных, профессиональных или специализированных). В качестве примеров можно привести такое: все виды таблиц, формул, граф, карт, схемы и прочих подобных структурных формаций.
- **Хроматические.** Сюда относят модели, которые были созданы с применением естественного языка семантики цветовых концептов, а также их онтологических предикатов. Под последними понимают возможность распознавания значений цветовых канонов и смыслов. В качестве примера хроматических моделей можно привести те, что были

построены с использованием соответствующей теоретической базы и методологии.

Как видим, основной составляющей являются данные, их структура и процедура обработки. Развивая мысль, можно дополнить, что информационная модель является схемой, в которой описана суть определённого объекта, а также все необходимые для его исследования процедуры. Для более полного описания характеристик используют переменные. Они замещают атрибут цели, которая прорабатывается. И здесь имеет значительную важность структура информационной модели [6].

Для реализации информационной модели для оценки качества системы используется программная платформа Node.JS. Для работы платформы Node используется виртуальная машина V8, которая задействует Google Chrome для серверного программирования. Благодаря V8 производительность Node «взлетает до небес», поскольку устраняются промежуточные этапы создания исполняемого кода. Вместо генерирования байткода или использования интерпретатора выполняется непосредственная компиляция в собственный машинный код. В связи с тем, что Node применяет JavaScript на стороне сервера, появляются следующие преимущества:

- Разработчики могут создавать веб-приложения на одном языке, благодаря чему снижается потребность в переключении контекста при разработке серверов и клиентов. При этом обеспечивается совместное использование кода клиентом и сервером, например, кода проверки данных, вводимых в форму, или кода игровой логики.
- Популярнейший формат обмена данными JSON является собственным форматом JavaScript.
- Язык JavaScript применяется в различных базах данных NoSQL (например, в CouchDB и MongoDB), поэтому подключение к таким базам данных осуществляется в естественной форме. Например, оболочкой и языком запросов для базы данных MongoDB является язык

JavaScript; языком проецирования/сведения для базы данных CouchDB также является JavaScript.

- Целью компиляции в Node.js является JavaScript, к тому же в настоящее время существует ряд других языков программирования, компилируемых в JavaScript4.
- В Node используется единственная виртуальная машина (V8), совместимая со стандартом ECMAScript5. Другими словами, вам не придется ожидать, пока во всех браузерах станут доступны все новые средства языка JavaScript, связанные с платформой Node.

Еще несколько лет назад разработчики даже и предположить не могли, что серверные приложения будут создаваться на JavaScript. Притягательность Node для разработчиков объясняется высокой производительностью и некоторыми другими упомянутыми ранее преимуществами. Все эти преимущества обеспечивает не только JavaScript, но и то, как этот язык используется в Node [7].

PostgreSQL – свободная объектно-реляционная система управления базами данных (СУБД). PostgreSQL использует клиент-серверную модель.

Для каждого клиента на сервере создается новый процесс. Для работы с такими клиентскими процессами сервер использует семафоры.

Транзакционная модель построена на основе так называемого multi-version concurrency control (MVCC), что дает максимальную производительность. Благодаря максимальной производительности осуществляется быстрые запросы к базе данных, что очень важно при работе с высоконагруженными системами.

Для более детального изучения качества рекомендательной системы используется вычислительный эксперимент. Вычислительный эксперимент — метод изучения устройств или физических процессов с помощью математического моделирования. Он предполагает, что вслед за построением математической модели проводится ее численное исследование, позволяющее

«проиграть» поведение исследуемого объекта в различных условиях или в различных модификациях. Численное исследование модели дает возможность определять разнообразные характеристики процессов, оптимизировать конструкции или режимы функционирования проектируемых устройств. Более того, случается, что в ходе вычислительного эксперимента исследователь неожиданно открывает новые процессы и свойства, о которых ему ранее ничего не было известно [8, 9].

2 Объект и методы исследования

Как было сказано во введении объектом исследования оценки качества системы является рекомендательная система. Рассмотрим собой структуру данного объекта.

2.1 Рекомендательные системы

В настоящее время существуют множество подходов к разработке рекомендательных систем. Рекомендательные системы – это программы и сервисы, которые пытаются определить, что хотят видеть пользователи, и предоставить им это (или порекомендовать, откуда и название).

Рекомендательные сервисы собирают различную информацию о человеке, используя несколько методов, по которым и разделяют все системы.

Итак, первый тип — явный сбор данных. Пользователь сам предоставляет необходимые для работы материалы. Например, когда рекомендательные системы Яндекс или других поисковиков просят человека дать оценки разным элементам, составить список любимых объектов определенной сферы или же ответить на несколько вопросов. Если же человек отказывается дать информацию самостоятельно, актуальной будет следующая методика.

Второй тип – неявный сбор данных. Условно говоря, действия участника процесса фиксируются программой для дальнейшей обработки и применения. Что нужно для этого? Программа распознает покупки, оценки на сайтах, собирает информацию по просмотрам, комментариям.

Существуют также виды рекомендательных систем, определяемых по подходам, которые они применяют.

Первая базовая методика называется метод коллаборативной фильтрации (collaborative filtering) [3, 4]. Рекомендации с использованием данной методики выдаются, основываясь на поведенческих характеристиках одного человека или группы людей, последнее даже является более эффективным. В группы собираются люди, которые похожи между собой по поведению и

характеристикам. Например, каждый человек в некоторой группе оценивает определенное количество объектов – например, книги (каждой книге ставится оценка – «Очень понравилось», «Понравилось», «Так себе», «Скучновато», «Плохо»...). На основании данных о предпочтениях для каждого формируется круг единомышленников, вкус которых по отношению к оцененным им книгам совпадает с его личными пристрастиями. Далее выбираются произведения, прочитанные и высоко оцененные единомышленниками, но не прочитанные данным человеком – эти произведения и составляют выданные ему рекомендации. В самом методе коллаборативной фильтрации существуют два подхода: user-based и item-based подходы [3, 4, 5, 6]. При user-based подходе рекомендации строятся на основании схожести пользователей. То есть для определенного пользователя информационной системы находятся ближайшие по интересам соседи. И в качестве рекомендаций для данного пользователя рекомендуются те объекты у схожих пользователей, которые не использовались им ранее. А при item-based подходе рекомендации строятся на основании продуктов, документов, товаров, в общем смысле, каких-либо объектов, элементов системы. У данного пользователя имеется свой список объектов, которые заказывались им ранее. Для того, чтобы построить для данного пользователя рекомендации требуется найти схожие объекты для объектов из списка его заказов или предпочтений и выдать ему в качестве рекомендаций.

Второй подход называется метод контентной фильтрации (content-based filtering). Здесь рекомендации строятся на основе поведения пользователя в информационной системе. Этот подход использует различную информацию о пользователе и основано на содержании объектов и профиля пользователя. Содержание каждого объекта представлено в виде набор каких-либо характеристик и признаков, например, часто встречающиеся слова в содержании объекта. Например, рекомендательная система пытается найти фильмы на основе: характеристик фильмов (например, жанр, режиссер, год выхода), профиля каждого пользователя в терминах характеристик фильмов, характеристик пользователей (например, пол, профессия). Для каждой пары u, i

необходимо придумать признаки $f_{u,i}^n$, основанные на профиле пользователя, собранном на обучении, и характеристиках пользователей и фильмов, известных даже для новых пользователей и фильмов [10, 11].

Выделяют также смешанные подходы, в соответствии с которыми осуществляется разработка рекомендательной системы.

Смешанный подход – это сочетание коллаборативной и контентной фильтрации. Как известно, больше – лучше, поэтому смешение этих двух методик увеличивают эффективность систем рекомендации, а именно повышают точность прогнозов для конкретных людей.

Рекомендательная система, разработанная для библиотечного каталога НТБ ТПУ основана на методе коллаборативной фильтрации. У пользователей библиотечного каталога НТБ ТПУ нет возможности оценивать объекты (документы), то есть ставить какой-либо рейтинг объекту, следовательно, рекомендательная система является безрейтинговой. Данная рекомендательная система построена на item-based подходе. Схожесть объектов, в данном случае документов электронного каталога, находится при помощи условно-вероятностной метрики, где отбираются k ближайших соседей для объекта. Из этих выбранных объектов строятся рекомендации для пользователей длиной N .

Как было отмечено выше, одной из немаловажных задач при разработке рекомендательных систем является оценка качества получаемых рекомендаций системой. Как было сказано в обзоре литературы существуют множество методов оценки качества рекомендаций системы. Рассмотрим методы наиболее часто встречающиеся при оценке качества систем.

3 Выбор меры оценки качества системы

Как было упомянуто ранее существуют множество мер для оценки качества рекомендательной системы. Рассмотрим наиболее часто встречающиеся из них.

3.1 Статистические меры оценивания

Согласно обзору литературы, встречаются две наиболее популярные статистические меры оценивания: мера средней абсолютной ошибки (Mean Absolute Error – MAE) и мера корень средней квадратичной ошибки (Root Mean Squared Error), которые рассчитываются по формулам:

$$MAE = \frac{\sum_{\{u,i\}} |p_{u,i} - r_{u,i}|}{N}, \quad (3)$$

$$RMSE = \sqrt{\frac{\sum_{\{u,i\}} (p_{u,i} - r_{u,i})^2}{N}}, \quad (4)$$

где $p_{u,i}$ – оценка (рейтинг) предсказания для пользователя u с объектом i ;

$r_{u,i}$ – оценки (рейтинги), которые выставляли пользователи;

N – общее количество оценок (рейтингов).

Данные меры нашли широкое применение при статистическом анализе данных и позволяют оценить отклонения некоторых показателей от желаемых.

3.2 Мера классификации

Следующие методы оценки качества рекомендаций системы являются меры классификации. Меры классификации берут свое начало из информационного поиска. Информационный поиск – процесс поиска неструктурированной документальной информации, удовлетворяющей информационные потребности. Поиск информации представляет собой процесс выявления в некотором множестве документов (текстов) всех тех, которые

посвящены указанной теме (предмету), удовлетворяют заранее определенному условию поиска (запросу) или содержат необходимые (соответствующие информационной потребности) факты, сведения, данные [7]. Процесс поиска включает последовательность операций, направленных на сбор, обработку и предоставление информации.

Предположим, что знаем предпочтения пользователей по элементам системы. Тогда гипотетически можно представить все ответы системы по категориям:

- истинно-положительные (true positives) — категории элементов, которые ожидалось увидеть в рекомендациях и получили на выходе системы рекомендаций.
- ложно-положительные (false positives) — категории элементов, которых быть в рекомендациях не должно и система их ошибочно вернула на выходе рекомендаций.
- ложно-отрицательные (false negatives) — категории элементов, которые ожидали увидеть в рекомендациях, но система их не определила.
- истинно-отрицательные (true negatives) — категории элементов, которых быть в рекомендациях не должно и на выходе системы они совершенно правильно отсутствуют.

Назовем тестовой выборкой множество элементов, для которых знаем правильные ответы. Если подсчитать по каждой категории элементов число попаданий (считаем попадания по парам элемент — категория), получим каноническую табличку распределения ответов, представленную в таблице 1.

Таблица 1 – Классификация элементов

	Предпочтительно	Не предпочтительно
Рекомендовали	tp (true positive)	fp (false positive)
Не рекомендовали	fn (false negative)	tn (true negative)

Проанализируем данную классификацию, представленную в таблице 1. Левая колонка таблицы — это «правильные» сочетания элементов и категорий, то есть те элементы, которые мы ожидаем увидеть в рекомендациях (релевантные элементы), правая — неправильные элементы (нерелевантные элементы). Верхняя строка таблицы — положительные (positive) ответы системы, то есть те элементы, которые были выданы системой, нижняя строка — отрицательные, то есть те элементы, которые не были включены на выход системы рекомендаций. Если число всех пар элемент — категория равно N , то нетрудно увидеть, что $N = tp + fp + fn + tn$. Теперь после того, как была произведена классификация элементов выборки можно определить понятие мер, представленных в обзоре литературы, для оценки качества рекомендаций. Представим элементы из нашей выборки в виде областей (см. рисунок 1).

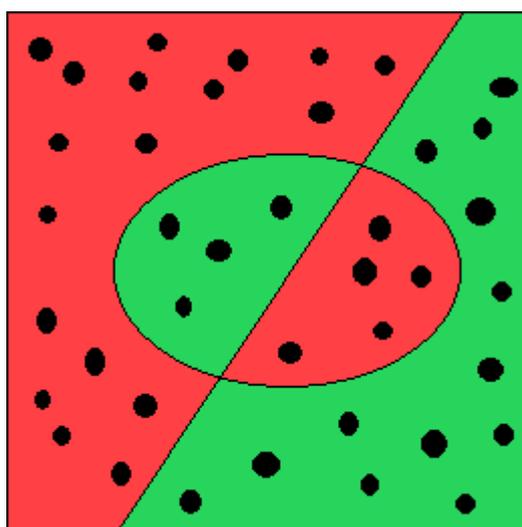


Рисунок 1 – Классификация элементов в виде областей

Здесь обозначено целых 4 непересекающихся класса элементов. Зеленые области обозначают правильные ответы, красные — неправильные. В овале находятся элементы, которые попали в рекомендацию системы. Весь прямоугольник целиком соответствует нашей выборке из N пар.

Согласно данной классификации можно определить меры для оценки качества рекомендательной системы электронного каталога библиотеки. Меры и формулы для их расчета представлены в таблице 2.

Таблица 2 – Меры оценки качества

Мера	Формула
Accuracy	$\frac{tp + tn}{tp + tn + fp + fn}$
Sensitivity (Recall)	$\frac{tp}{tp + fn}$
Specificity	$\frac{tn}{tn + fp}$
Precision	$\frac{tp}{tp + fp}$
G – Mean	$\sqrt{Sensitivity * Specificity}$
F-Measure	$\frac{2 * Sensitivity * Precision}{Sensitivity + Precision}$

Мера Accuracy показывает отношение суммы соответствующих рекомендованных системой элементов и несоответствующих нерекондованных системой элементов к общему количеству элементов. Мера Sensitivity (Recall) показывает сколько соответствующих рекомендованных элементов среди всех соответствующих элементов. А мера Specificity в отличие от предыдущей меры наоборот показывает сколько несоответствующих нерекондованных элементов среди всех несоответствующих элементов. Как нетрудно заметить, мера Precision показывает сколько соответствующих рекомендованных элементов среди всех рекомендованных элементов.

G – Mean Measure представляет собой среднее геометрическое между Sensitivity и Specificity. F-Measure представляет собой гармоническое среднее между Sensitivity и Precision. Она стремится к нулю, если Sensitivity или Precision стремится к нулю.

Произведем сравнительный анализ описанных ранее мер оценки качества рекомендательной системы

3.3 Сравнительный анализ методов

Согласно параметрам рекомендательной системы электронного каталога ТПУ можно выбрать методы, которые возможно применить именно к этой системе. Под релевантностью объектов понимается соответствие объектов интересам запросам пользователя. Сравнительные характеристики методов приведены в таблице 3.

Таблица 3 – Сравнительные характеристики методов

Метод	Релевантность объектов	Наличие оценок пользователей
Мера средней абсолютной ошибки (Mean Absolute Error – MAE)	-	+
Мера корня средней квадратичной ошибки (Root Mean Squared Error – RMSE)	-	+
Мера Accuracy	+	-
Мера Sensitivity (Recall)	+	-
Мера Specificity	+	-
Мера Precision	+	-
G – Mean	+	-
F-Measure	+	-

Как видно из таблицы 3 для расчета мер требуется либо информация об оценках (рейтингах) пользователей элементам (объектам), либо наличие

релевантности элементов запросу пользователя. Так как данная рекомендательная система не имеет информации о релевантности и оценках пользователей объектам, то основываясь на таблицах 1 и 2 можно предположить, что те объекты, которые заказывал пользователь будем считать релевантными. Тогда согласно [12] и исходя из таблицы 1 нам известны только объекты, которые пользователь предпочитал, или в данном случае, заказывал (tp (true positive) и fn (false negative)). Согласно такой информации для оценки качества рекомендательной системы библиотечного каталога НТБ ТПУ можно использовать только меру Sensitivity (Recall). Таким образом, скрывая часть заказа пользователя, рекомендательная система пытается угадать данные скрытые элементы. Согласно определению меры Sensitivity (Recall) из таблицы 2 значения данной меры может принимать значения от 0 до 1. Насколько хороши выдаваемые рекомендации можно судить по значению данной меры. Классифицируем значения меры полноты следующим образом, чтобы в дальнейшем определить уровень качества рекомендательной системы (см. таблицу 4).

Таблица 4 – Уровень качества рекомендаций

Значения меры Sensitivity (Recall)	Уровень качества рекомендаций
0-0,3	низкий
0,3-0,7	средний
0,6-1	высокий

Рассмотрим вычисление меры Sensitivity (Recall) и меры Precision на простом примере.

3.4 Применение меры Sensitivity (Recall) и меры Precision

Рассмотрим пример оценки качества работы рекомендательной системы электронного каталога библиотеки. Допустим существует некий пользователь

библиотеки, который совершал заказы в электронном каталоге библиотеки.
Предположим, что данный пользователь ранее заказывал набор книг:

1. С. И. Баскаков Радиотехнические цепи и детали
2. Н. М. Изюмов Основы радиотехники
3. Г. Уэллс Морская дама

Пусть имеется выборка книг, которая доступна в электронном каталоге:

1. Д. В. Сивухин Механика, Том 1
2. А. С. Пушкин Капитанская дочка
3. Л. Д. Ландау, Е. М. Лифшиц Гидродинамика
4. Т. Карп Стальная воля. Как закалить свой характер
5. Учебник для вузов Биохимия
6. А. И. Струков, В. В. Серов Патологическая анатомия
7. Учебник – Теория систем и системный анализ
8. Д. В. Сивухин Атомная и ядерная физика
9. Учебник по философии
10. Учебник – Методы оптимизации
11. Магнус Эконометрика начальный курс
12. Учебник – Экономика России
13. Учебник – Квантовая физика
14. Учебник – Общая химия
15. С. Кинг Мизери
16. Дж. Оруэлл 1984
17. Александр Блок Собрание сочинений
18. Слотина учебное пособие Психология личности

Для выборки выше произведем классификацию по соответствию элементов выборки интересам пользователя, то есть разделим на релевантные и нерелевантные элементы. Релевантные элементы представляют собой документы, которые являются схожими или родственными по тематике

элементам из заказа пользователя. Поэтому в качестве релевантных документов могут выступать как документы по физике, так и художественная литература в жанре фантастика. Предположим, что его список релевантных документов может выглядеть следующим образом:

1. Д. В. Сивухин Механика, Том 1
2. Учебник – Квантовая физика
3. С. Кинг Мизери
4. Д. В. Сивухин Атомная и ядерная физика
5. Дж. Оруэлл 1984
6. Л. Д. Ландау, Е. М. Лифшиц Гидродинамика

Тогда остальные элементы выборки будут классифицированы как нерелевантные.

Предположим, что рекомендательная система на выход данному пользователю выдала рекомендацию книг:

1. Д. В. Сивухин Механика, Том 1
2. Слотина учебное пособие Психология личности
3. С. Кинг Мизери
4. Учебник для вузов Биохимия
5. Магнус Эконометрика начальный курс

Посчитаем меру полноты и точности для данного пользователя. Произведем классификацию ответов системы на следующие категории:

- истинно-положительные (true positives) — категории элементов, которые ожидалось увидеть в рекомендациях и получили на выходе системы рекомендаций. В данном случае это два элемента-документа: Стивен Кинг «Мизери», Л. В. Сивухин «Механика»
- ложно-положительные (false positives) — категории элементов, которых быть в рекомендациях не должно и система их ошибочно вернула на

выходе рекомендаций. В эту подгруппу вошли: Слотина «Психология личности», «Биохимия» - учебник для вузов, Магнус «Эконометрика».

- ложно-отрицательные (*false negatives*) — категории элементов, которые ожидали увидеть в рекомендациях, но система их не определила. В эту подгруппу попали следующие элементы: «Квантовая физика», «Атомная и ядерная физика», Дж. Оруэлл «1984», Ландау - «Гидродинамика»
- истинно-отрицательные (*true negatives*) — категории элементов, которых быть в рекомендациях не должно и на выходе системы они совершенно правильно отсутствуют.

Тогда определим меру точности, согласно указанной классификации:

$$P(\textit{precision}) = \frac{tp}{tp + fp} = \frac{2}{2 + 3} = \frac{2}{5} = 0,4 \quad (5)$$

Аналогично определим меру полноты:

$$Sensitivity(\textit{recall}) = \frac{tp}{tp + fn} = \frac{2}{2 + 4} = \frac{2}{6} = 0,33 \quad (6)$$

Согласно принятым уровням оценки качества из таблицы 4 мера Sensitivity (Recall) и мера Precision выдает рекомендации среднего уровня качества. Аналогичным образом можно посчитать меры для оценки качества системы рекомендательной системы электронного каталога библиотеки для других пользователей и, тем самым, получив среднюю оценку качества рекомендаций системы, можно оценить пригодность системы для использования.

4 Реализация меры Sensitivity (Recall) для оценки качества системы

Чтобы понять насколько хорошо или плохо рекомендательная система работает при различных условиях, следует разработать и произвести вычислительный эксперимент для рекомендательной системы, используя выбранную меру для оценки качества системы. Таким образом, вычислительный эксперимент должен ответить на вопрос: каким образом работает рекомендательная система. Для этого следует построить информационную модель системы, а на основании данной модели спланировать вычислительный эксперимент.

4.1 Информационная модель и используемые программные средства

Рекомендательную систему и систему оценки качества рекомендательной системы библиотечного каталога НТБ ТПУ можно выразить через информационную модель [9]. В данной информационной модели указываются основные потоки информации, при помощи которых достигается необходимая цель. Информационная модель для рекомендательной системы представлена на рисунке 2.

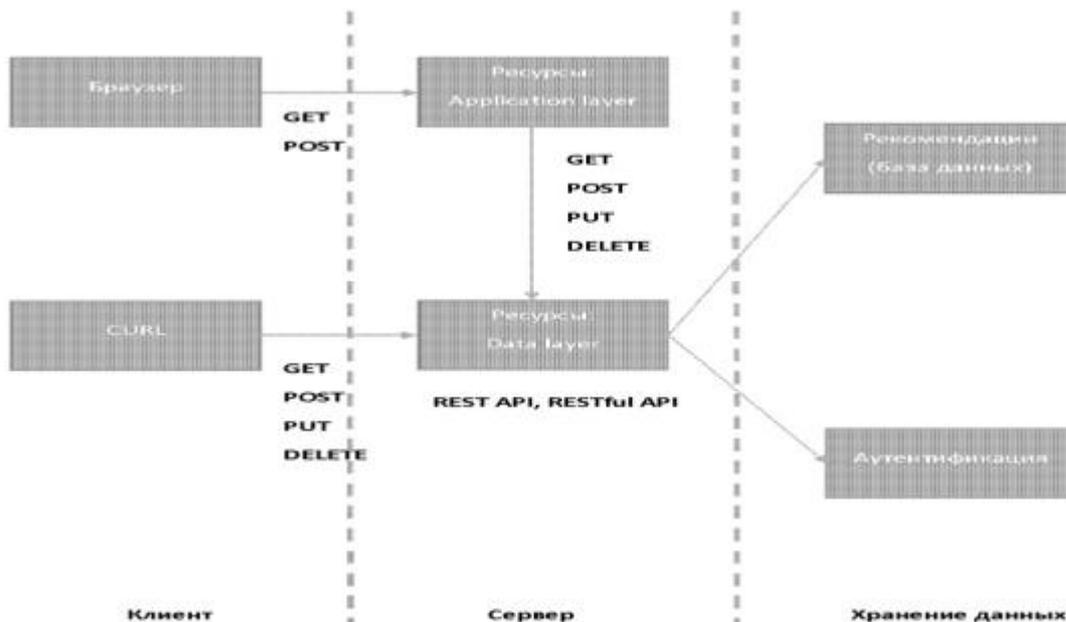


Рисунок 2 – Информационная модель рекомендательной системы библиотечного каталога НТБ ТПУ

Информационная модель для оценки качества рекомендательной системы библиотечного каталога НТБ ТПУ представлена на рисунке 3.

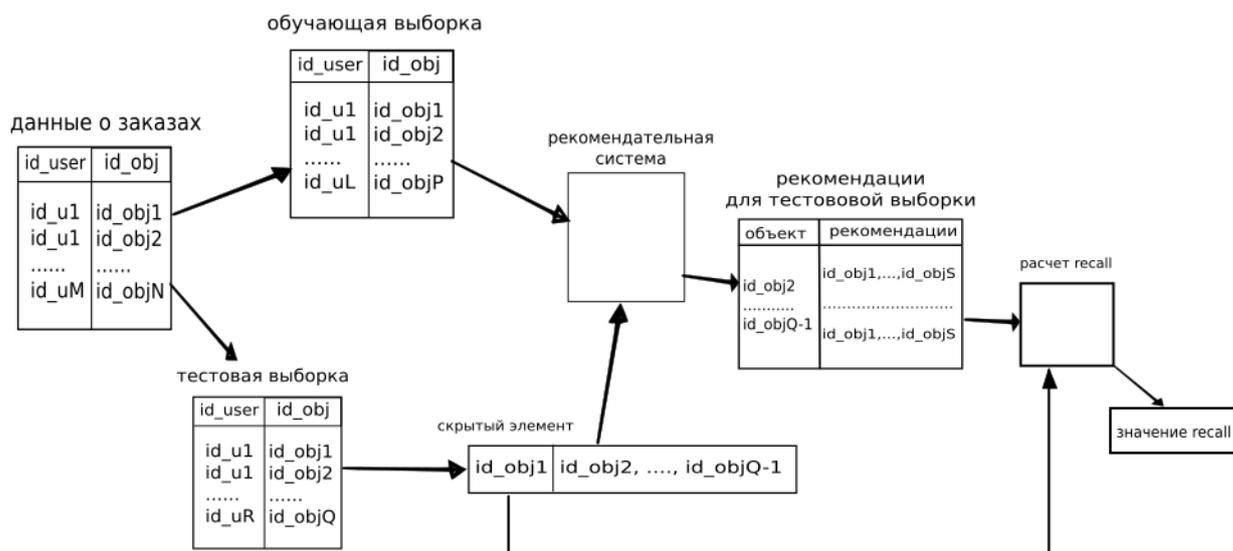


Рисунок 3 – Информационная модель для оценки качества рекомендательной системы библиотечного каталога НТБ ТПУ.

Для реализации данной информационной модели для оценки качества рекомендательной системы библиотечного каталога НТБ ТПУ используются PostgreSQL и NodeJS. При помощи данных программных средств построена рекомендательная система, поэтому оценивать качество системы можно также используя PostgreSQL и NodeJS. Не исключено также, что реализация оценки качества системы могла быть осуществлена и при помощи других программных средств, например, Python и MySQL, но пока остановимся на программных средствах, на которых реализована система.

NodeJS – программная платформа, основанная на движке V8 (транслирующем JavaScript в машинный код), превращающая JavaScript из узкоспециализированного языка в язык общего назначения. Node.js добавляет возможность JavaScript взаимодействовать с устройствами ввода-вывода через свой API (написанный на C++), подключать другие внешние библиотеки, написанные на разных языках, обеспечивая вызовы к ним из JavaScript-кода. Node.js применяется преимущественно на сервере, выполняя роль веб-сервера, но есть возможность разрабатывать на Node.js и десктопные оконные приложения (при помощи NW.js, AppJS или Electron для Linux, Windows и Mac OS) и даже программировать микроконтроллеры (например, tessel и espruino). В основе Node.js лежит событийно-ориентированное и асинхронное (или реактивное) программирование с неблокирующим вводом/выводом. Одним из плюсов выбора этой программной платформы является то, что NodeJS довольно быстро работает с высоконагруженными системами, что позволяет производить быстро расчеты с большими объемами данных, например, обработка больших объемов данных электронного каталога библиотеки.

4.2 Вычислительный эксперимент

Расчет полноты для рекомендательной системы электронного каталога ТПУ будет производиться по данным заказов пользователей за 2015 год. Исходные данные заказов представлены в виде csv-файла объемом 101420 строк,

где в первом столбце указаны идентификаторы пользователей, а во втором столбце идентификаторы объектов, заказанные пользователями. Вычислительный эксперимент будет состоять из следующих этапов:

1. Фильтрация исходных данных. Из данной таблицы заказов убираются пользователи, которые заказывали 1 документ или более 1000 объектов. Отфильтрованные данные также записываются в таблицу.
2. Разбиение данных на две выборки. Отфильтрованные данные полученные в пункте 1 представляются в виде таблицы, где первый столбец – ID пользователя, второй столбец – список из ID объектов, которые заказывал пользователь. Полученную в таком виде таблицу делится случайным образом в процентном соотношении 70 на 30. Выборка, содержащая 70% случайных данных таблицы, называется обучающей, а оставшаяся часть выборки тестовой выборкой.
3. Получение полноты.
 - a. На основе обучающей выборки, полученной в пункте 2 строится матрица подобия объектов. Матрица подобия рассчитывает метрику схожести между всеми объектами, поданных на вход рекомендательной системы. Исходя из матрицы схожести можно выбрать, например, k наиболее схожих объекта для данного объекта. А на выходе системы вывести N рекомендованных объектов.
 - b. Из тестовой выборки получают массив уникальных объектов. В этом массиве скрывается один документ, а оставшиеся объекты по очереди подаются на вход рекомендательной системы. В результате для каждого поданного объекта получаем список k схожих объектов. Общая выборка объектов представляет собой объем меньше или равной количеству объектов умноженной на k . Среди данных объектов на рекомендации выдается N объектов. Среди этих рекомендованных объектов проверяем наличие скрытого объекта, если данный объект присутствует, то значит рекомендации были

построены верны и поэтому присваиваем значение 1, иначе, если отсутствует, присваиваем значение 0. Эта операция продлевается для каждого пользователя в тестовой выборке. Мера Sensitivity (Recall) будет рассчитываться как отношение количества угаданных рекомендациями скрытого объекта к общему числу пользователей в тестовой выборке, для которых получены рекомендации.

4. Получение сетки меры Sensitivity (Recall).
 - а. Зафиксируем в пункте 3а значение k – число ближайших соседей. И найдем значения средней меры Sensitivity (Recall), исходя из пункта 3б, для различного числа количества рекомендаций N . Занесем данные в таблицу.
 - б. Теперь возьмем изменим значения k и вновь посчитаем значение средней меры Sensitivity (Recall) согласно пункту 4а. Просчитав значения меры Sensitivity (Recall) для различных комбинаций k и N получим матрицу значений средней меры Sensitivity (Recall).
5. По полученным данным значений меры Sensitivity (Recall) построим трехмерный график зависимости средней меры Sensitivity (Recall) от значений k и N .
6. Произведем анализ полученных данных на трехмерном графике, построенными по данным представленным в пункте 5.

Согласно различным параметрам системы, в данном случае параметрам k и N , есть возможность оценить поведение рекомендательной системы, сделать выводы о том, возможно ли использовать рекомендательную систему в информационных системах. Также, если качество рекомендательной системы низкое, то ответить на вопрос: как изменить рекомендательную систему таким образом, чтобы качество рекомендованных документов было высоко.

4.3 Анализ результатов

Согласно спланированному вычислительному эксперименту получим график зависимости меры Sensitivity (Recall) от k и N . Параметр k – число ближайших соседей следует взять в интервале от 2 до 6, так как N – количество выдаваемых рекомендаций редко превышает 5-10 рекомендаций, то возьмем значения N от 2 до 10. График зависимости меры Sensitivity (Recall) от k и N для данной рекомендательной системы приведены на рисунке 3

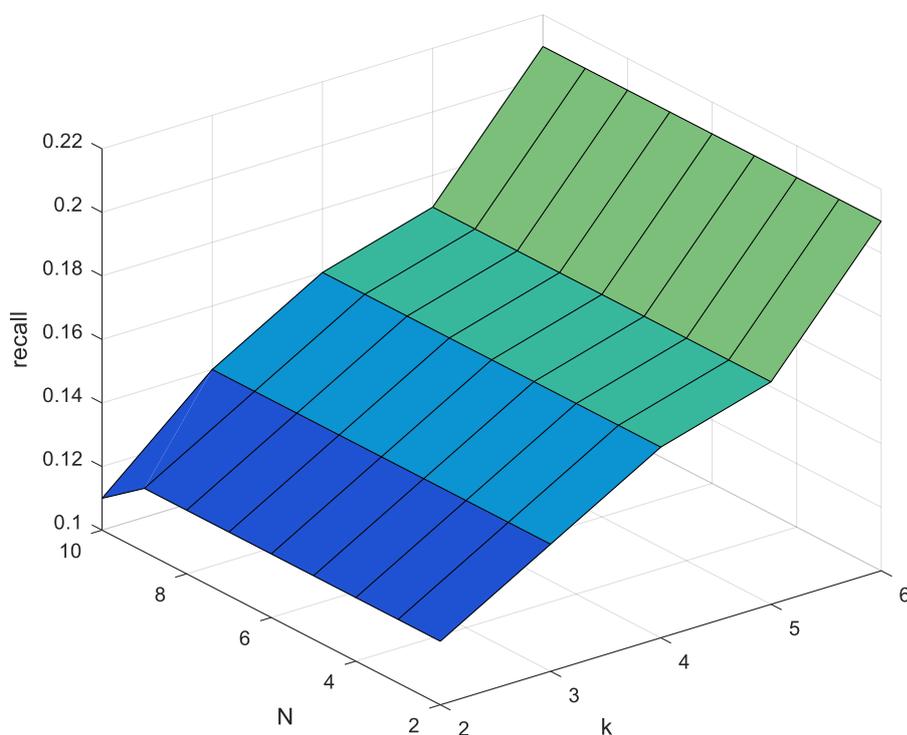


Рисунок 3 – Зависимость меры Sensitivity (Recall) от k и N для рекомендательной системы

Как видно из рисунка 3 с увеличением числа ближайших соседей k качество выдаваемых рекомендаций растет. Это вызвано тем, что чем больше число k , тем больше вероятность угадать, что данный объект будет угадан системой.

Рассмотрим значения меры Sensitivity (Recall) по всем пользователям в зависимости от числа k , при этом при фиксированном k усреднив значения N . Гистограмма распределения меры представлен на рисунке 4.

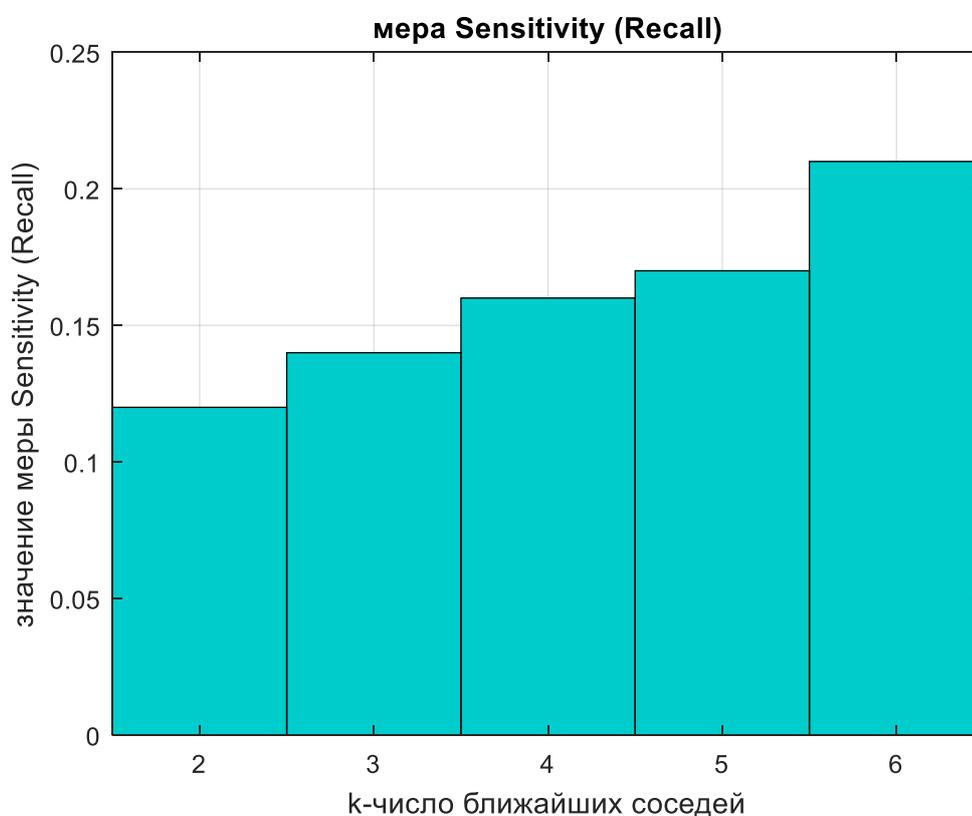


Рисунок 4 – Зависимость средней меры Sensitivity (Recall) от k .

Как видно из рисунка 4 значения меры Sensitivity (Recall) очень низкие. Исходя из таблицы 4 можно сделать вывод о том, что система выдает рекомендации низкого уровня качества. Данная ситуация может быть вызвана тем, что исходный объем данных недостаточен для получения более точной оценки качества рекомендательной системы.

5 Финансовый менеджмент, ресурсоэффективность и ресурсосбережение

Экономическая часть ВКР предназначена для комплексного описания и анализа финансово-экономических аспектов выполненной работы. Целью данного раздела является планирование и формирование бюджета научных исследований, определение ресурсной, финансовой, бюджетной, социальной и экономической эффективности исследования.

Целью данной ВКР является исследование оценки качества рекомендательной системы.

5.1 Организация и планирование работ

При организации процесса реализации конкретного проекта необходимо рационально планировать занятость каждого из его участников и сроки проведения отдельных работ. На данном этапе определяется полный перечень работ, определяются их исполнители и рациональная продолжительность. В качестве структуры, показывающей необходимые данные, используется линейный график работ, представленный в таблице 5.

Таблица 5 – Перечень этапов работ и распределение исполнителей

Этапы работы	Исполнители	Загрузка исполнителей
Постановка целей и задач	НР	НР – 100%
Составление и утверждение ТЗ	НР, И	НР – 100% И – 10%
Разработка календарного плана	НР, И	НР – 50% И – 100%
Изучение литературы и источников по данной теме	НР, И	НР – 50% И – 100%
Выбор подходящего метода для реализации поставленной задачи	НР, И	НР – 60% И – 100%

Разработка вычислительного эксперимента для оценки качества рекомендательной системы	НР, И	НР – 30% И – 80%
Изучение необходимых языков программирования и программных пакетов	И	И – 100%
Реализация консольного приложения.	НР, И	НР – 10% И – 100%
Тестирование разработанного приложения	И	И – 100%
Проведение вычислительного эксперимента	И	И – 100%
Анализ результатов эксперимента	НР, И	НР – 60% И – 100%
Оформление пояснительной записки	И	И – 100%
Подведение итогов	НР, И	НР – 40% И – 100%

НР – научный руководитель.

И – исполнитель Ильмауров Вячеслав Олегович.

5.1.1 Продолжительность этапов работ

Для определения трудоемкости выполнения работ необходимо на основе экспертной оценки ожидаемой трудоемкости выполнения каждой работы рассчитать длительность работ в рабочих и календарных днях для каждого из вариантов исполнения работ последующим формулам:

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

где $t_{ож}$ – ожидаемая продолжительность работы, дн.;

t_{min} – минимальная продолжительность работы, дн.;

t_{max} – максимальная продолжительность работы, дн.;

Исходя из ожидаемой продолжительности работ, определим продолжительность каждой работы в рабочих днях $T_{рд}$, учитывая параллельность выполнения работ несколькими исполнителями.

$$T_{PD} = \frac{t_{ож}}{K_{BH}} \cdot K_D,$$

где $t_{ож}$ – продолжительность работы, дн.;

K_{BH} – коэффициент выполнения работ, учитывающий влияние внешних факторов ($K_{BH} = 1$);

K_D – коэффициент, учитывающий дополнительное время на компенсацию непредвиденных задержек и согласование работ ($K_D = 1.2$).

Расчет продолжительности этапа в календарных днях ведется по формуле:

$$T_{KD} = T_{PD} \cdot T_K$$

где T_{PD} – продолжительность выполнения этапа в рабочих днях;

T_{KD} – продолжительность выполнения этапа в календарных днях;

T_K – коэффициент календарности.

Коэффициент календарности рассчитывается по формуле:

$$T_K = \frac{T_{КАЛ}}{T_{КАЛ} - T_{ВД} - T_{ПД}}$$

где $T_{КАЛ}$ – календарные дни ($T_{КАЛ} = 365$);

$T_{ВД}$ – выходные дни ($T_{ВД} = 52$);

$T_{ПД}$ – праздничные дни ($T_{ПД} = 10$).

$$T_K = \frac{365}{365 - 52 - 10} = 1,205$$

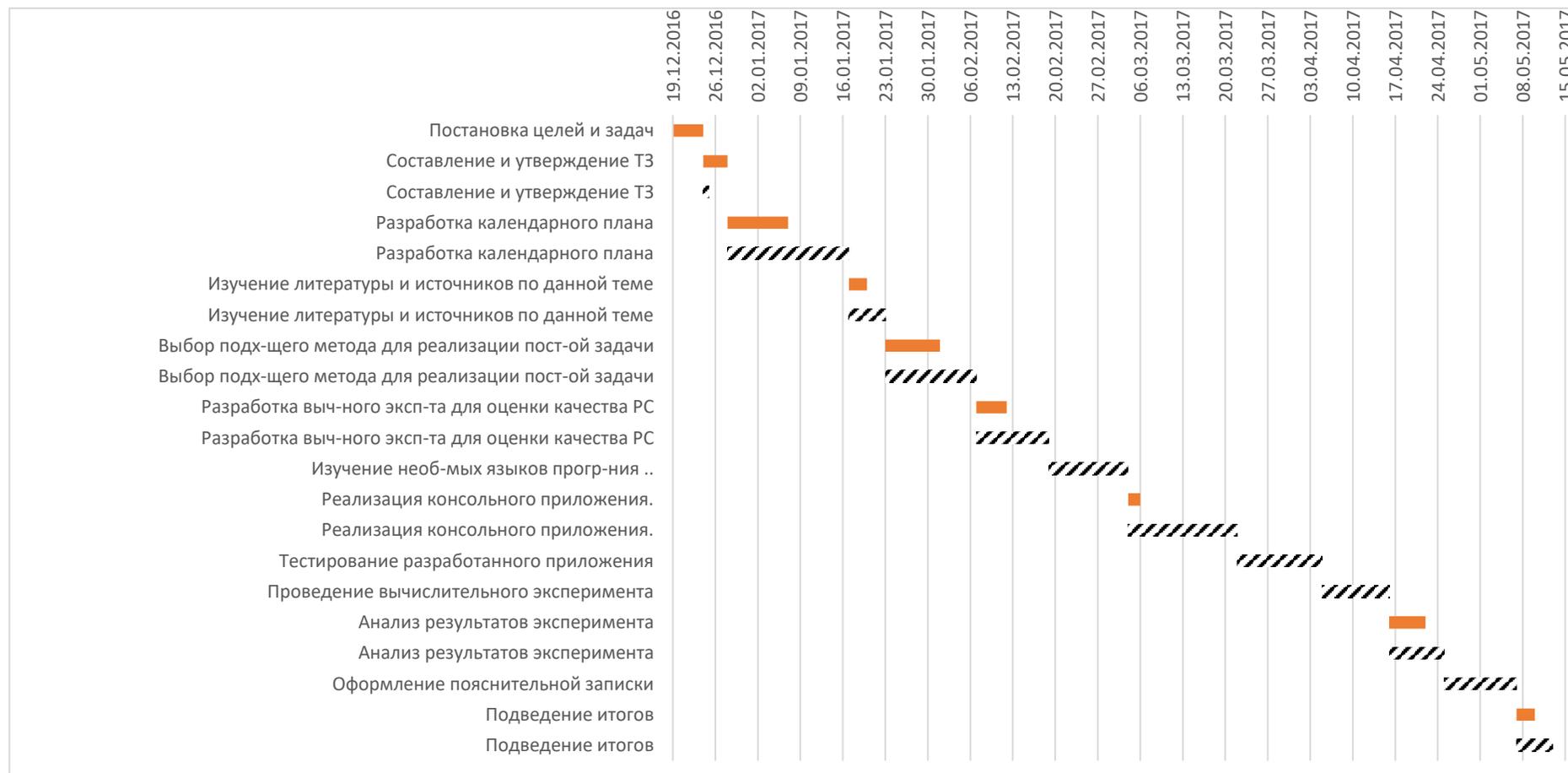
В таблице 6 приведены длительность этапов работ и число исполнителей, занятых на каждом этапе. Для наглядного отображения графика и распределения работ между участниками проекта использована диаграмма Ганта. Диаграмма Ганта представляет собой линейный график, на котором работы по теме представляются протяженными во времени отрезками, характеризующиеся датами начала и окончания выполнения того или иного этапа работ (таблица 7).

Таблица 6 – Трудозатраты на выполнение проекта

Этап	Исполнители	Продолжительность работ, дни			Трудоемкость работ по исполнителям чел.- дн.			
					$T_{РД}$		$T_{КД}$	
		t_{min}	t_{max}	$t_{ож}$	НР	И	НР	И
1	2	3	4	5	6	7	8	9
Постановка целей и задач	НР	2	4	2,8	3,36	–	4,05	–
Составление и утверждение ТЗ	НР, И	2	3	2,4	2,88	0,29	3,47	0,35
Разработка календарного плана	НР, И	12	15	13,2	7,92	15,84	9,54	19,09
Изучение литературы и источников по данной теме	НР, И	2	6	3,6	2,16	4,32	2,60	5,21
Выбор подходящего метода для реализации поставленной задачи	НР, И	9	12	10,2	7,34	12,24	8,85	14,75
Разработка вычислительного эксперимента для оценки	НР, И	8	13	10	3,60	9,60	4,34	11,57

качества рекомендательной системы								
Изучение необходимых языков программирования и программных пакетов	И	6	12	8,4	–	10,08	–	12,15
Реализация консольного приложения.	НР, И	10	16	12,4	1,49	14,88	1,79	17,93
Тестирование разработанного приложения	И	6	15	9,6	–	11,52	–	13,88
Проведение вычислительного эксперимента	И	5	10	7	–	8,40	–	10,12
Анализ результатов эксперимента	НР, И	5	8	6,2	4,46	7,44	5,38	8,97
Оформление пояснительной записки	И	7	10	8,2	–	9,84	–	11,86
Подведение итогов	НР, И	2	6	3,6	1,73	4,32	2,08	5,21
Итого:				97,6	34,94	108,77	42,11	131,07

Таблица 7 – Линейный график работ



■ – NP, ▨ – И.

5.1.2 Расчет накопления готовности проекта

Показатель технической готовности ВКР характеризует отношение продолжительности работ, выполненных на момент исчисления этого показателя, к общей запланированной продолжительности работ. Данный показатель позволяет точно знать, на каком уровне выполнения находится определенный этап или работа.

Степень готовности определяется формулой

$$СГ_{\square} = \frac{ТР_{\square}^H}{ТР_{\text{общ.}}} = \frac{\sum_{\square=1}^{\square} ТР_{\square}}{ТР_{\text{общ.}}} = \frac{\sum_{\square=1}^{\square} \sum_{\square=1}^{\square} ТР_{\square\square}}{\sum_{\square=1}^{\square} \sum_{\square=1}^{\square} ТР_{\square\square}},$$

где $ТР_{\text{общ.}}$ – общая трудоемкость проекта;

$ТР_i$ ($ТР_k$) – трудоемкость i -го (k -го) этапа проекта, $i = \overline{1, I}$;

$ТР_i^H$ – накопленная трудоемкость i -го этапа проекта по его завершении;

$ТР_{ij}$ ($ТР_{kj}$) – трудоемкость работ, выполняемых j -м участником на i -м этапе, здесь $j = \overline{1, m}$ – индекс исполнителя, в нашем примере $m = 2$.

Таблица 8 – Нарастание технической готовности работы и удельный вес каждого этапа

Этап	$ТР_i$, %	$СГ_i$, %
Постановка целей и задач	2,34	2,34
Составление и утверждение ТЗ	2,21	4,54
Разработка календарного плана	16,53	21,08
Изучение литературы и источников по данной теме	4,51	25,59
Выбор подходящего метода для реализации поставленной задачи	13,62	39,21
Разработка вычислительного эксперимента для оценки качества рекомендательной системы	9,19	48,40
Изучение необходимых языков программирования и программных пакетов	7,01	55,41

Реализация консольного приложения.	11,39	66,80
Тестирование разработанного приложения	8,02	74,82
Проведение вычислительного эксперимента	5,85	80,66
Анализ результатов эксперимента	8,28	88,94
Оформление пояснительной записки	6,85	95,79
Подведение итогов	4,21	100,00

5.2 Расчет сметы затрат на выполнение проекта

В состав затрат на создание проекта включается стоимость всех расходов, необходимых для реализации проекта. Расчет сметной стоимости на выполнение данной разработки производится по следующим статьям затрат:

- материалы и покупные изделия;
- заработная плата;
- социальный налог;
- расходы на электроэнергию (без освещения);
- амортизационные отчисления;
- командировочные расходы;
- оплата услуг связи;
- арендная плата за пользование имуществом;
- прочие услуги (сторонних организаций);
- прочие (накладные расходы) расходы.

5.2.1 Расчет затрат на материалы

Данная статья включает стоимость всех материалов, используемых при разработке проекта. Все расчеты сведены в таблицу 9.

Таблица 9 – Расчет затрат на материалы

Наименование материалов	Цена за ед., руб.	Кол -во	Сумма, руб.
Бумага для принтера формата А4	300	1 уп.	300
Заправка картриджа принтера	360	1 шт.	360
Итого:			660

ТЗР (транспортно-заготовительные расходы) составляют 5% от отпускной цены материалов, тогда расходы на материалы с учетом ТЗР равны $C_{\text{мат}} = 660 * 1,05 = 693$ руб.

5.2.2 Расчет заработной платы

Данная статья расходов включает в себя заработную плату научного руководителя и студента, а также премии и доплаты. Расчет выполняется на основе трудоемкости выполнения каждого этапа и величины месячного оклада (МО) исполнителя.

Среднедневная тарифная заработная плата ($ЗП_{\text{дн-т}}$) рассчитывается по формуле:

$$ЗП_{\text{дн-т}} = \text{МО} / 24,83$$

учитывающей, что в году 298 рабочих дня и, следовательно, в месяце в среднем 24,83 рабочих дней.

Расчеты затрат на полную заработную плату приведены в таблице 6. Затраты времени по каждому исполнителю в рабочих днях с округлением до целого взяты из таблицы 6. Для учета в ее составе премий, дополнительной зарплаты и районной надбавки используется следующий ряд коэффициентов: $K_{\text{ПР}} = 1,1$; $K_{\text{доп.ЗП}} = 1,188$; $K_{\text{р}} = 1,3$.

Таким образом, для перехода от тарифной суммы заработка исполнителя, связанной с участием в проекте, к соответствующему полному заработку (зарплатной части сметы) необходимо первую умножить на интегральный

коэффициент $K_{и}=1,1*1,188*1,3=1,699$. Вышеуказанное значение $K_{доп.ЗП}$ применяется при шестидневной рабочей неделе, при пятидневной оно равно 1,113, соответственно в этом случае $K_{и} = 1,592$.

Таблица 10 – Затраты на заработную плату

Исполнитель	Оклад, руб/мес	Среднедневная ставка, руб/раб.день	Затраты времени, раб.дни	Коэффициент	Фондз/платы, руб.
НР	23 264,86	924,31	27	1,699	42 400,83
С	14 584,32	587,37	109	1,62	103 717,80
Итого:					146 118,62

5.2.3 Расчет затрат на социальный налог

Затраты на единый социальный налог (ЕСН), включающий в себя отчисления в пенсионный фонд, на социальное и медицинское страхование, составляют 30 % от полной заработной платы по проекту, т.е. $C_{соц.} = C_{зп} * 0,3$. Итого, в нашем случае $C_{соц.} = 146 118,62 * 0,3 = 43 835,59$ руб.

5.2.4 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:

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

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

$Ц_{э}$ – тариф на 1 кВт·час;

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

$Ц_{э} = 5,782$ руб./квт·час (с НДС).

Время работы оборудования вычисляется на основе итоговых данных таблицы 2 из расчета, что продолжительность рабочего дня равна 8 часов.

$$t_{об} = T_{рд} * K_t,$$

где $K_t \leq 1$ – коэффициент использования оборудования по времени, равный отношению времени его работы в процессе выполнения проекта к $T_{рд}$, определяется исполнителем самостоятельно.

Мощность, потребляемая оборудованием, определяется по формуле:

$$P_{об} = P_{ном.} * K_C$$

где $P_{ном.}$ – номинальная мощность оборудования, кВт;

$K_C \leq 1$ – коэффициент загрузки, зависящий от средней степени использования номинальной мощности. Для технологического оборудования малой мощности $K_C = 1$.

Расчет затраты на электроэнергию для технологических целей приведен в таблице 11

Таблица 11 – Затраты на электроэнергию

Наименование оборудования	Время работы оборудования $t_{об}$, час	Потребляемая мощность $P_{об}$, кВт	Затраты $\text{Э}_{об}$, руб.
Персональный компьютер	769,6	0,3	1 334,95
Струйный принтер	2	0,1	1,16
Итого:			1336,11

5.2.5 Расчет амортизационных расходов

В статье «Амортизационные отчисления» от используемого оборудования рассчитывается амортизация за время выполнения работы для оборудования, которое имеется в наличии.

Амортизационные отчисления рассчитываются по формуле:

$$C_{ам} = \frac{N_A * C_{об} * t_{рф} * n}{F_D},$$

где H_A – годовая норма амортизации единицы оборудования;

$C_{ОБ}$ – балансовая стоимость единицы оборудования с учетом ТЗР.

F_D – действительный годовой фонд времени работы соответствующего оборудования, берется из специальных справочников или фактического режима его использования в текущем календарном году.

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

n – число задействованных однотипных единиц оборудования.

Для ПК: $H_A = 40\%$, $C_{ОБ} = 40000$ руб., $F_D = 2384$ часа, $t_{ВТ} = 769.6$ ч, $n = 1$. Для струйного принтера: $H_A = 50\%$, $C_{ОБ} = 12000$ руб., $F_D = 500$ часов, $t_{ВТ} = 2$ ч, $n = 1$.

Итак, затраты на амортизационные отчисления ПК составили:

$$C_{ам1} = \frac{0,4 \cdot 40000 \cdot 769,6 \cdot 1}{2384} = 5165,1.$$

Для струйного принтера:

$$C_{ам2} = \frac{0,5 \cdot 12000 \cdot 2 \cdot 1}{500} = 24.$$

В итоге амортизационные отчисления составили:

$$C_{ам} = 5165,1 + 24 = 5189,1.$$

5.2.6 Расчет прочих расходов

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. $C_{проч.} = (C_{мат} + C_{зп} + C_{соц} + C_{эл.об.} + C_{ам}) \cdot 0,1$

$$C_{проч.} = (693 + 146\,118,62 + 43\,835,59 + 1\,336,11 + 5\,189,1) \cdot 0,1 = 19\,717,24 \text{ руб.}$$

5.2.7 Расчет общей себестоимости разработки

Таким образом, на основании проведенных ранее подсчетов по отдельным статьям затрат вычислим общую плановую себестоимость исследования (таблица 12).

Таблица 12 – Смета затрат на разработку проекта

Статья затрат	Условное обозначение	Сумма, руб.
Материалы и покупные изделия	$C_{\text{мат}}$	693
Основная заработная плата	$C_{\text{зп}}$	146 118,62
Отчисления в социальные фонды	$C_{\text{соц}}$	43 835,69
Расходы на электроэнергию	$C_{\text{эл.}}$	1 336,11
Амортизационные отчисления	$C_{\text{ам}}$	5 189,1
Прочие расходы	$C_{\text{проч}}$	19 717,24
Итого:		216 959,76

Таким образом, затраты на разработку составили $C = 216\,959,76$ руб.

5.2.8 Расчет прибыли

Так как исполнитель работы не располагает данными для применения «сложных» методов, то прибыль составит 20% от полной себестоимости проекта, или **43 391,95** руб.

5.2.9 Расчет НДС

НДС составляет 18% от суммы затрат на разработку и прибыли. В нашем случае это $(216\,959,76 + 43\,391,95) * 0,18 = 46\,863,31$ руб.

5.2.10 Цена разработки НИР

Цена равна сумме полной себестоимости, прибыли и НДС:

$$C_{\text{НИР(КР)}} = 216\,959,76 + 43\,391,95 + 46\,863,31 = 307\,215,02 \text{ руб.}$$

5.3 Оценка экономической эффективности проекта

Данный проект не был ориентирован на экономический результат, главная цель проекта – оценить качество рекомендательной системы. Цель данного проекта носит сугубо дидактический характер.

Оценка качества рекомендательной системы способна дать ответ на вопрос: насколько выданные рекомендации подходят для пользователя. Отсюда следует, что пользователь будет меньше времени уделять поиску нужных ему вещей, элементов. Поэтому пользователь будет эффективно использовать свое время. В свою очередь работник библиотеки не будет лишним раз тратить время на пользователя, которому будет необходимо помочь с поиском литературы. Это позволит сэкономить время работников библиотеки и – по крайней мере условно, - затраты на их содержание.

5.3.1 Оценка научно-технического уровня НИР

Сущность метода заключается в том, что на основе оценок признаков работы определяется интегральный показатель (индекс) ее научно-технического уровня по формуле:

$$I_{НТУ} = \sum_{i=1}^3 R_i \cdot n_i,$$

где $I_{НТУ}$ – интегральный индекс научно-технического уровня;

R_i – весовой коэффициент i -го признака научно-технического эффекта;

n_i – количественная оценка i -го признака научно-технического эффекта, в баллах.

Таблица 13 – Весовые коэффициенты признаков НТУ

Признаки научно-технического эффекта НИР	Характеристика признака НИР	R _i
Уровень новизны	Систематизируются и обобщаются сведения, определяются пути дальнейших исследований	0,4
Теоретический уровень	Разработка способа (алгоритм, программа мероприятий, устройство, вещество и т.п.)	0,1
Возможность реализации	Время реализации в течение первых лет	0,5

Интегральный показатель научно-технического уровня для нашего проекта составляет:

$$I_{\text{нту}} = 0,4 \cdot 4 + 0,1 \cdot 6 + 0,5 \cdot 10 = 2 + 0,6 + 5 = 7,2$$

Таким образом, данный проект имеет средний уровень научно-технического эффекта.

6 Социальная ответственность

Представление понятия «Социальная ответственность» сформулировано в международном стандарте (МС) IS CSR-08260008000: 2011 «Социальная ответственность организации».

В соответствии с МС - Социальная ответственность - ответственность организации за воздействие ее решений и деятельности на общество и окружающую среду через прозрачное и этическое поведение, которое:

- содействует устойчивому развитию, включая здоровье и благосостояние общества;
- учитывает ожидания заинтересованных сторон;
- соответствует применяемому законодательству и согласуется с международными нормами поведения (включая промышленную безопасность и условия труда, экологическую безопасность);
- интегрировано в деятельность всей организации и применяется во всех ее взаимоотношениях (включая промышленную безопасность и условия труда, экологическую безопасность).

Объект исследования – рекомендательная система электронного каталога библиотеки.

Согласно техническому заданию (ТЗ) планируется оценить качество рекомендательной системы электронного каталога библиотеки. Оценка качества рекомендательной системы выявит эффективность рекомендательной системы, что в свою очередь, позволит в будущем улучшить рекомендательную систему. Таким образом можно добиться повышения эффективности рекомендательной системы, оценив качество выдаваемых ее рекомендаций.

В разделе будут рассмотрены опасные и вредные факторы, оказывающие влияние на производственную деятельность, рассмотрены воздействия разрабатываемой системы на окружающую среду, правовые и организационные вопросы, а также мероприятия в чрезвычайных ситуациях.

6.1 Производственная безопасность

Сначала необходимо выявить источники опасности, связанные с техническими системами и окружающей средой, обуславливающие возникновение чрезвычайных ситуаций техногенного характера.

Все опасные и вредные производственные факторы в соответствии с ГОСТ 12.0.003-74 подразделяются на физические, химические, биологические и психофизиологические. Рассмотрим наиболее опасные факторы при работе за ПК (таблица 14).

Таблица 14 – Опасные и вредные производственные факторы при использовании методики

Источник фактора, наименование видов работ	Факторы (по ГОСТ 12.003-74)		Нормативные документы
	Вредные	Опасные	
Работа за ПК	<ol style="list-style-type: none">1. Микроклимат2. Напряженность зрения3. Шум на рабочем месте4. Электромагнитные излучения5. Освещенность рабочей зоны	<ol style="list-style-type: none">1. Электрический ток.	<ul style="list-style-type: none">• Параметры микроклимата устанавливаются СанПиН 2.2.4-548-96• Параметры уровня шума устанавливаются СН 2.2.4/2.1.8.562-96.• Параметры уровня электромагнитных излучений устанавливаются ГОСТ 12.1.006-84 ССБТ• Нормы естественного и искусственного освещения предприятий, СНиП 23-05-95• Параметры освещенности устанавливаются СП 52.13330.2011

6.1.1 Микроклимат

Микроклимат производственных помещений определяется сочетанием температуры, влажности, подвижности воздуха, температуры окружающих

поверхностей и их тепловым излучением. Параметры микроклимата определяют теплообмен организма человека и оказывают существенное влияние на функциональное состояние различных систем организма, самочувствие, работоспособность и здоровье.

Метеорологические условия для рабочей зоны производственных помещений регламентируются ГОСТ 12.1.005-88 «Общие санитарно-гигиенические требования к воздуху рабочей зоны» и СанПиН 2.2.4.548—96 «Гигиенические требования к микроклимату производственных помещений».

Принципиальное значение в нормах имеет отдельное нормирование каждого компонента микроклимата: температуры, влажности, скорости движения воздуха. В рабочей зоне должны обеспечиваться параметры микроклимата, соответствующие оптимальным и допустимым значениям. На основе интенсивности общих энергозатрат организма в ккал/ч (Вт) вид данной работы относится к категории Ia, который характеризуется интенсивностью энергозатрат до 120 ккал/ч, сидячим состоянием и незначительным физическим напряжением. В таблице 15 приведены оптимальные параметры микроклимата, в таблице 16 – допустимые. В нашем случае работа проводилась в теплое время года. Температура колебалась в пределах 21-25 градусов, относительная влажность в пределах 45-55. Следовательно, рабочая зона имела допустимые нормы.

Таблица 15 – Оптимальные значения характеристик микроклимата

Период года	Температура воздуха, °С	Температура поверхностей, °С	Относительная влажность воздуха, %	Скорость движения воздуха, м/с
Холодный	22-24	21-25	40-60	0,1
Тёплый	23-25	22-26	40-60	0,1

Таблица 16 – Допустимые значения характеристик микроклимата

Период года	Температура воздуха, °С	Температура поверхностей, °С	Относительная влажность воздуха, %	Скорость движения воздуха, м/с
Холодный	20,0–25,0	19,0–26,0	15–75	0,1
Теплый	21,0-28,0	20,0–29,0	15–75	0,1-0,2

К мероприятиям по оздоровлению воздушной среды в производственном помещении относятся правильная организация вентиляции и кондиционирования воздуха, отопление помещений. Вентиляция может осуществляться естественным и механическим путём. В зимнее время в помещении необходимо предусмотреть систему отопления. Также возможно использование ионизации воздуха, что позволит увеличить в воздухе количество аэроионов.

6.1.2 Шум на рабочем месте

Интенсивное шумовое воздействие на организм человека неблагоприятно влияет на протекание нервных процессов, способствует развитию утомления, изменениям в сердечно-сосудистой системе и появлению шумовой патологии, среди многообразных проявлений которой ведущим клиническим признаком является медленно прогрессирующее снижение слуха по типу кохлеарного неврита. Источниками шума на предприятии являются сами вычислительные машины (встроенные в стойки ЭВМ вентиляторы, принтеры и т.д.), центральная система вентиляции и кондиционирования воздуха и другое оборудование. Также присутствуют шумы источников извне предприятия, например, шум автомобилей, соседних объектов.

Допустимые шумовые характеристики рабочих мест регламентируются ГОСТ 12.1.003-83 «Шум. Общие требования безопасности» и СанПиН

2.2.4/2.1.8.562–96 «Шум на рабочих местах, в помещениях жилых, общественных зданий и на территории застройки» и СанПиН 2.2.2/2.4.1340-03 «Гигиенические требования к персональным электронно-вычислительным машинам и организации работы».

При применении методики на ПК уровень шума на рабочем месте не должен превышать 50 дБА. Допустимые уровни звукового давления в помещениях для медицинского персонала приведены в таблице 17

Таблица 17 – Допустимые уровни звука, создаваемого ПЭВМ

Уровни звукового давления в октавных полосах со среднегеометрическими частотами									Уровни звука в дБА
31,5 Гц	63 Гц	125 Гц	250 Гц	500 Гц	1000 Гц	2000 Гц	4000 Гц	8000 Гц	
86 дБ	71 дБ	61 дБ	54 дБ	49 дБ	45 дБ	42 дБ	40 дБ	38 дБ	50

Для снижения уровня шума стены и потолок помещений, где установлены компьютеры, могут быть облицованы звукопоглощающими материалами.

6.1.3 Электромагнитные излучения

Воздействие электромагнитных полей на человека зависит от напряженностей электрического и магнитного полей, потока энергии, частоты колебаний, размера облучаемого тела. Источниками электромагнитных излучений на предприятии являются ПЭВМ, мониторы и любая другая техника, подключенная к электросети.

Опасное воздействие на работников могут оказывать электромагнитные поля радиочастот (60 кГц-300 ГГц).

Допустимые значения норм напряженности электромагнитного поля указаны в таблице 18

Таблица 18 – Допустимые уровни электромагнитных полей

Наименование параметров	Допустимые значения
Напряженность электромагнитного поля на расстоянии 50 см вокруг ВДТ по электрической составляющей должна быть не более: в диапазоне частот 5 Гц – 2 кГц в диапазоне частот 2 – 400 кГц	25 В/м 2.5 В/м
Плотность магнитного потока должна быть не более: в диапазоне частот 5 Гц – 2 кГц в диапазоне частот 2 – 400 кГц	250 нТл 25 нТл
Напряженность электростатического поля:	15 кВ/м

Для понижения уровня напряженности электромагнитного поля следует использовать мониторы с пониженным уровнем излучения (MPR-II, TCO-92, TCO-99), устанавливать защитные экраны, устранять неисправности. А для профилактики компьютерного зрительного синдрома, улучшения визуальных показателей видеомониторов, повышения работоспособности, снижения зрительного утомления возможно применение защитных очков со спектральными фильтрами.

6.1.4 Освещенность рабочей зоны

Правильно спроектированное и выполненное освещение обеспечивает высокую работоспособность, оказывает положительное психологическое воздействие на работающих, способствует повышению производительности труда. О важности вопроса производственного освещения говорит тот факт, что основной объем информации (около 90%) человек получает с помощью зрения.

К системам освещения предъявляют следующие требования:

- соответствие уровня освещенности рабочих мест характеру выполняемой зрительной работы;
- достаточно равномерное распределение яркости на рабочих поверхностях и в окружающем пространстве;
- отсутствие резких теней, прямой и отраженной блескости;
- постоянство освещенности во времени;
- оптимальная направленность излучаемого осветительными приборами светового потока;
- долговечность, экономичность, электробезопасность и пожаробезопасность, эстетичность, удобство и простота эксплуатации.

Для обеспечения освещенности рабочего места применяют естественное уличное освещение и только при его нехватке подключаются внутренние системы освещения.

Искусственное освещение в помещениях при использовании методики должно осуществляться системой общего равномерного освещения. В производственных и административно-общественных помещениях при работе с ПК следует применять системы комбинированного освещения (к общему освещению дополнительно устанавливаются светильники местного освещения, предназначенные для освещения зоны расположения документов).

Требования к освещению на рабочих местах, оборудованных ПЭВМ регламентируются СанПиН 2.2.2/2.4.1340—03 "Гигиенические требования к ПЭВМ и организации работы". Освещенность на поверхности стола в зоне размещения рабочего документа должна быть 300 — 500 лк. Освещение не должно создавать бликов на поверхности экрана. Освещенность поверхности экрана не должна быть более 300 лк.

Следует ограничивать неравномерность распределения яркости в поле зрения пользователя ПЭВМ, при этом соотношение яркости между рабочими поверхностями не должно превышать 3:1 — 5:1, а между рабочими поверхностями и поверхностями стен и оборудования 10:1.

В качестве источников света при искусственном освещении следует применять преимущественно люминесцентные лампы типа ЛБ и компактные люминесцентные лампы (КЛЛ). При устройстве отраженного освещения в производственных и административно-общественных помещениях допускается применение металлогалогенных ламп. В светильниках местного освещения допускается применение ламп накаливания, в том числе галогенные.

В помещениях медицинских учреждений обычно применяют одностороннее боковое естественное освещение. С целью уменьшения солнечной инсоляции светопроемы устраивают с северной, северо-восточной или северо-западной ориентацией. Мониторы располагают подальше от окон и таким образом, чтобы окна находились сбоку.

Если экран дисплея расположен к окну, необходимы специальные экранящие устройства (светорассеивающие шторы, регулируемые жалюзи, солнцезащитная пленка с металлизированным покрытием).

Разряд зрительных работ врача относится к разряду «г» (высокой точности), параметры искусственного освещения указаны в таблице 19.

Таблица 19 – Нормативные значения освещённости

Характеристика зрительной работы	Наименьший или эквивалентный размер объекта различения, мм	Разряд зрительной работы	Подразряд зрительной работы	Контраст объекта с фоном	Характеристика фона	Искусственное освещение		
						Освещённость, лк		
						При системе комбинированного освещения		При системе общего освещения
						всего	В том числе от общего	
Высокой точности	От 0,3 до 0,5	III	г	Средний,	Светлый <<	400	200	200

				боль шой <<	средний			
--	--	--	--	-------------------	---------	--	--	--

Если нормы освещенности рабочей зоны не совпадают с нормативными, то следует добавить дополнительные источники света в рабочую зону.

6.1.5 Расчет системы искусственного освещения на рабочем месте оператора персональной ЭВМ

Длина помещения составляет 6 метров, ширина – 5 м, высота – 3 м. Высота рабочей поверхности $h_{рп} = 0,8$ м.

В помещении установлены светильники типа ARS/R 418 ($\Phi_{ст} = 1200$ лм) (Санитарные правила и нормы: СанПиН 2.1.7.1322-03. Гигиенические требования к размещению и обезвреживанию отходов производства и потребления), укомплектованные четырьмя люминесцентными лампами мощностью 20 Вт. Геометрические размеры светильников 595x595x36 мм, $\lambda = 1,4$. Светильники размещены, как представлено на рисунке 5.

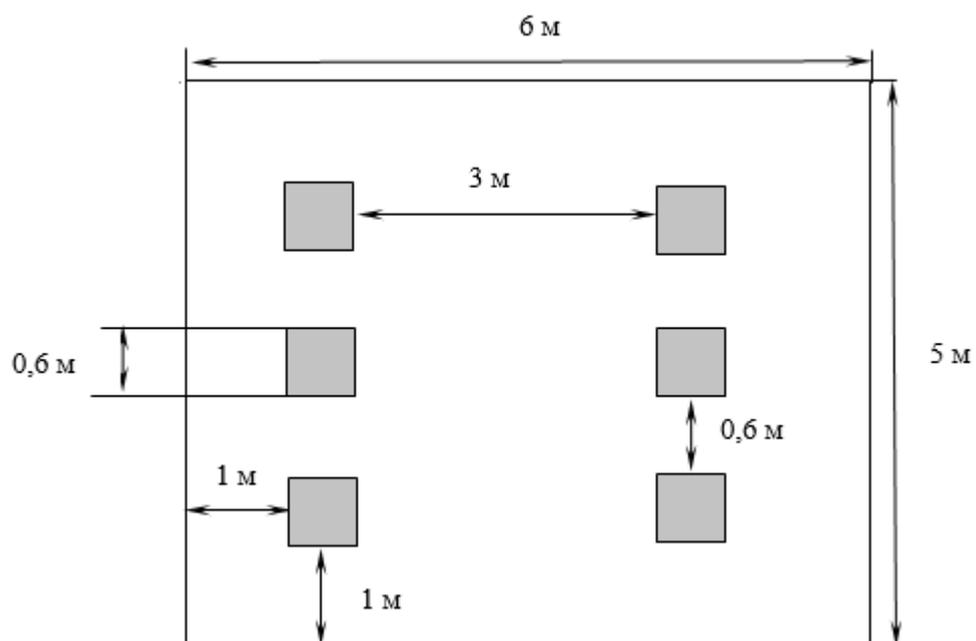


Рисунок 5 – Схема освещения

Учитывая, что в каждом светильнике установлено по 4 лампы, количество ламп составит $N = 24$. Определим электрическую мощность установки:

$$P = 24 \times 20 \text{ Вт} = 480 \text{ Вт}$$

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

Коэффициент Z (отношение средней освещенности к минимальной) примем равным $Z = 1.1$. Коэффициент запаса определяется по таблице из СанПин 2.2.2/2.4.1340-03 «Гигиенические требования к персональным электронно-вычислительным машинам и организации работы» в зависимости от запылённости помещения, в нашем случае $K = 1.5$ (коэффициент запаса). Коэффициент использования, выражается отношением светового потока, падающего на расчетную поверхность, к суммарному потоку всех ламп и исчисляется в долях единицы. Он зависит от характеристик светильника, размеров помещения, окраски стен и потолка, характеризуемых коэффициентами отражения от стен (P_C) и потолка (P_{II}), значение коэффициентов P_C и P_{II} определим по таблице из СНиП 23-05-95 [23], $P_C = 70\%$, $P_{II} = 50\%$.

Значение η определим по таблице коэффициентов использования различных светильников из СНиП 23-05-95 Естественное и искусственное освещение. Для этого вычислим индекс помещения по формуле:

$$I = \frac{S}{h \cdot (a+b)}$$

Площадь помещения составляет 30 м^2 .

$$I = \frac{30}{2,25 \cdot (5+6)} = 1,21$$

Зная индекс помещения I , Pc и Pn , определим коэффициент использования светового потока из таблицы взятой из СНиП 23-05-95, $\eta = 0,57$. Стандартный световой поток возьмём из таблицы, зная мощность и тип люминесцентной лампы, и примем его равным $\Phi_{ст} = 1060$. Теперь можно рассчитать фактическую минимальную освещённость E_{ϕ} .

$$E_{\phi} = \frac{N \times n \times \Phi_{cm} \times \eta}{S \times k \times z}$$

$$E_{\phi} = \frac{6 \times 4 \times 1060 \times 0,57}{30 \times 1,5 \times 1,1} = 293 \text{ лк}$$

$$\Delta E = \frac{E_{\phi} - E_n}{E_n} \times 100\%$$

$$\Delta E = \frac{390 - 400}{400} \times 100\% = -26,57\%$$

По нормам, установленным СНиП 23-05-95, минимальная освещённость рабочих поверхностей в офисных помещениях для работы с дисплеями и видеотерминалами при общем освещении должна быть равна 300-500 лк. Рассчитанное значение оказалось меньше необходимого. Для соблюдения норм установленных СНиП 23-05-95, необходимо установить дополнительное освещение на рабочих столах, либо установить лампы большей мощности и светоотдачи.

6.1.6 Электробезопасность

В зависимости от условий в помещении опасность поражения человека электрическим током увеличивается или уменьшается. Не следует работать с ПЭВМ в условиях повышенной влажности (относительная влажность воздуха длительно превышает 75%), высокой температуры (более 35°C), наличии токопроводящей пыли, токопроводящих полов и возможности одновременного

соприкосновения к имеющим соединению с землёй металлическим элементам и металлическим корпусом электрооборудования.

Согласно ПУЭ рабочее помещение по степени опасности поражения электрическим током можно отнести к классу помещений без повышенной опасности.

Специфическая опасность электроустановок: токоведущие проводники, корпуса стоек ЭВМ и прочего оборудования, оказавшегося под напряжением в результате повреждения (пробоя) изоляции, не подают каких-либо сигналов, которые предупреждали бы человека об опасности. К работе в электроустановках должны допускаться лица, прошедшие инструктаж и обучение безопасным методам труда, проверку знаний правил безопасности и инструкций в соответствии с занимаемой должностью применительно к выполняемой работе с присвоением соответствующей квалификационной группы по технике безопасности и не имеющие медицинских противопоказаний.

Согласно ПУЭ и ГОСТ 12.1.019-79 «ССБТ. Электробезопасность. Общие требования и номенклатура видов защиты» для обеспечения безопасности работ в действующих электроустановках должны выполняться следующие организационные мероприятия:

- назначение лиц, ответственных за организацию и безопасность производства работ;
- оформление наряда или распоряжения на производство работ;
- осуществление допуска к проведению работ;
- организация надзора за проведением работ;
- оформление окончания работы, перерывов в работе, переводов на другие рабочие места;
- установление рациональных режимов труда и отдыха.

Конкретные перечни работ, которые должны выполняться по наряду или распоряжению, следует устанавливать в отраслевой нормативной документации.

Для обеспечения безопасности работ в электроустановках следует выполнять:

- отключение установки (части установки) от источника питания;
- проверку отсутствия напряжения;
- механическое запираание приводов коммутационных аппаратов, снятие предохранителей, отсоединение концов питающих линий и другие меры, исключающие возможность ошибочной подачи напряжения к месту работы;
- заземление отключенных токоведущих частей (наложение переносных заземлителей, включение заземляющих ножей);
- ограждение рабочего места или остающихся под напряжением токоведущих частей, к которым в процессе работы можно прикоснуться или приблизиться на недопустимое расстояние.

6.2 Экологическая безопасность

При применении методики используется ПК, который потребляет сравнительно небольшое количество электроэнергии (мощностью 220 Вт).

Под охраной окружающей среды характеризуется различного рода мероприятиями, влияющие на следующие природные зоны:

- атмосфера;
- гидросфера;
- литосфера.

При работе с ПК возникает необходимость утилизировать производственные отходы, в качестве которых в данном случае выступают бумажные отходы (макулатура) и неисправные детали персональных компьютеров, плат, контроллеров. Экологическая безопасность при работе с отходами регламентируется ГОСТ Р 55090-2012 «Ресурсосбережение.

Обращение с отходами. Рекомендации по утилизации отходов бумаги» и ГОСТ Р 53692-2009 «Ресурсосбережение. Обращение с отходами. Этапы технологического цикла отходов». Бумажные отходы должны передаваться в соответствующие организации для дальнейшей переработки во вторичные бумажные изделия. Неисправные комплектующие персональных компьютеров должны передаваться либо государственным организациям, осуществляющим вывоз и уничтожение бытовых и производственных отходов, либо организациям, занимающимся переработкой отходов. Важнейшим этапам обращения с отходами является их сбор, а в дальнейшем переработка, утилизация и захоронение.

6.3 Безопасность в чрезвычайных ситуациях

Вероятные чрезвычайные ситуации классифицируются:

- преднамеренные/непреднамеренные;
- техногенные: взрывы, пожары, обрушение помещений, аварии на системах жизнеобеспечения;
- природные – связанные с проявлением стихийных сил природы.
- экологические – это аномальные изменения состояния природной среды, такие как загрязнения биосферы, разрушение озонового слоя, кислотные дожди;
- антропогенные – являются следствием ошибочных действий людей.
- биологические – различные эпидемии, эпизоотии, эпифитотии;
- комбинированные.

При применении методики основным источником чрезвычайных опасностей является возникновение пожара.

Обеспечение пожарной безопасности в помещениях научных учреждений достигается, прежде всего, установлением жесткого противопожарного режима и обучением обслуживающего персонала и учащихся мерам пожарной безопасности и действиям во время пожара. Согласно НПБ 105-03 «Определение

категорий помещений, зданий и наружных установок по взрывопожарной и пожарной опасности» рабочее помещение относится к классу Д, где применяются негорючие вещества и материалы в холодном состоянии.

Территория научного учреждения, а также участки, прилегающие к нему, должны своевременно очищаться от горючих отходов, мусора, которые следует собирать на специально выделенных площадках в контейнеры или ящики, а затем вывозить на свалку.

Важно контролировать состояние дорог, проездов, подъездов и проходов к зданиям, следить за тем, чтобы они ничем не загромождались, а в зимнее время регулярно очищались от снега и льда.

В зданиях, относящихся к объектам с массовым пребыванием людей, особое внимание должно уделяться содержанию путей эвакуации. Каждое здание должно иметь не менее двух эвакуационных выходов: если один из них отрезан огнем, для спасения используется другой. Запасные выходы должны быть свободны и иметь надпись «Запасный выход». Все двери эвакуационных выходов свободно открываются в сторону выхода из помещений.

На случай отключения электроэнергии у обслуживающего персонала должны быть электрические фонари – не менее одного на каждого работника дежурного персонала.

На каждом этаже здания, на видном месте должен быть вывешен план эвакуации с этажа (здания). На плане эвакуации кроме путей выхода (стрелками) указываются места размещения средств пожаротушения, телефонов.

Необходимо проводить следующие пожарно-профилактические мероприятия:

1. Организационные мероприятия:
 - a. противопожарный инструктаж обслуживающего персонала;
 - b. обучение персонала правилам техники безопасности;
2. Эксплуатационные мероприятия:
 - a. соблюдение эксплуатационных норм оборудования;

- b. обеспечение свободного подхода к оборудованию
- c. содержание в исправности изоляции токоведущих проводников.

3. Технические мероприятия:

- a. соблюдение противопожарных мероприятий при устройстве электропроводок, оборудования, систем отопления, вентиляции и освещения. В здании должны присутствовать порошковые огнетушители, установлен рубильник, обесточивающий все помещение. Если возгорание произошло в электроустановке, для его устранения должны использоваться углекислотные огнетушители или порошковые;
- b. профилактический осмотр, ремонт и испытание оборудования.

В случае возникновения пожара сотрудники должны предпринять следующие меры

- сообщить о пожаре в пожарную охрану, задействовать систему оповещения;
- задействовать план эвакуации (открыть запасные двери и включить светоуказатели эвакуационных путей);
- вывести людей в безопасное место в соответствии с планом эвакуации;
- проверить поименно, все ли эвакуированы;
- приступить к тушению пожара первичными средствами;
- встретить пожарные подразделения и сообщить, где могли остаться люди, как туда можно подойти;
- принять меры к эвакуации имущества.

План эвакуации людей при пожаре из помещения представлен на рисунке

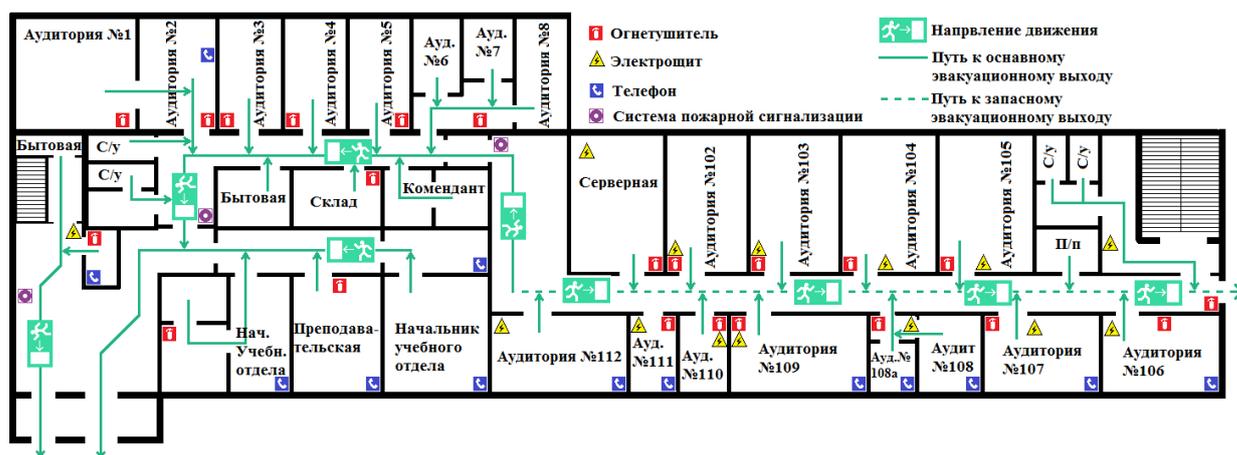


Рисунок 6 - План эвакуации при пожаре

Ответственность за нарушение Правил пожарной безопасности, согласно действующему федеральному законодательству, несет руководитель объекта.

6.4 Правовые и организационные вопросы обеспечения безопасности

6.4.1 Психофизиологические факторы

К психофизиологическим вредным факторам относятся статические физические перегрузки, умственное перенапряжение, монотонность труда, эмоциональные перегрузки.

Продолжительность непрерывной работы с компьютером без регламентированного перерыва не должна превышать 2-х часов. А во время регламентированных 15-минутных перерывов с целью снижения нервно-эмоционального напряжения, утомления зрительного анализатора, устранения влияния гиподинамии и гипокинезии, предотвращения развития познотонического утомления (при статических нагрузках) необходимо выполнять комплексы упражнений

6.4.2 Организационные мероприятия при компоновке рабочей зоны

Для того чтобы люди плодотворно трудились нужно создать подходящие условия труда. Для организации рабочих мест следует руководствоваться ГОСТ 12.2.032-78, ГОСТ 12.2.033-78, СанПиН 2.2.2/2.4.1340-03, ГОСТ 12.2.049-80, ГОСТ 12.2.061-81, ГОСТ 12.4.026-2001 межотраслевыми и отраслевыми рекомендациями по их обустройству.

Эргономическими аспектами проектирования рабочих мест программистов, в частности, являются: высота рабочей поверхности, размеры пространства для ног, требования к расположению документов на рабочем месте (наличие и размеры подставки для документов, возможность различного размещения документов, расстояние от глаз пользователя до экрана, документа, клавиатуры и т.д.), характеристики рабочего кресла, требования к поверхности рабочего стола, регулируемость элементов рабочего места.

Главными элементами рабочего места программиста или оператора являются стол и кресло. Основным рабочим положением является положение сидя.

Рациональная планировка рабочего места предусматривает четкий порядок и постоянство размещения предметов, средств труда и документации. То, что требуется для выполнения работ чаще, расположено в зоне легкой досягаемости рабочего пространства.

Моторное поле - пространство рабочего места, в котором могут осуществляться двигательные действия человека.

Максимальная зона досягаемости рук - это часть моторного поля рабочего места, ограниченного дугами, описываемыми максимально вытянутыми руками при движении их в плечевом суставе.

Оптимальная зона - часть моторного поля рабочего места, ограниченного дугами, описываемыми предплечьями при движении в локтевых суставах с опорой в точке локтя и с относительно неподвижным плечом.

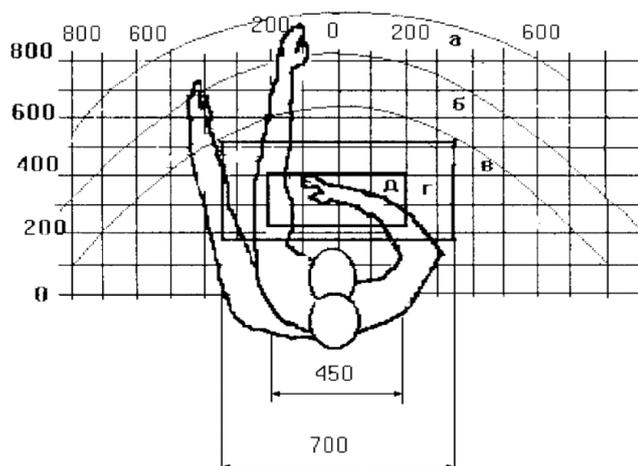


Рисунок 7 - Зоны досягаемости рук в горизонтальной плоскости

- а - зона максимальной досягаемости;
- б - зона досягаемости пальцев при вытянутой руке;
- в - зона легкой досягаемости ладони;
- г - оптимальное пространство для грубой ручной работы;
- д - оптимальное пространство для тонкой ручной работы.

Оптимальное размещение предметов труда и документации в зонах досягаемости:

- дисплей размещается в зоне "а" (в центре);
- системный блок размещается в предусмотренной нише стола;
- клавиатура - в зоне "г"/"д";
- манипулятор "мышь" - в зоне "в" справа;
- документация: необходимая при работе - в зоне легкой досягаемости ладони — "в", а в выдвижных ящиках стола - литература, неиспользуемая постоянно.

Для комфортной работы стол должен удовлетворять следующим условиям:

- высота стола должна быть выбрана с учетом возможности сидеть свободно, в удобной позе, при необходимости опираясь на подлокотники;
- нижняя часть стола должна быть сконструирована так, чтобы программист мог удобно сидеть, не был вынужден поджимать ноги;
- поверхность стола должна обладать свойствами, исключающими появление бликов в поле зрения программиста;
- конструкция стола должна предусматривать наличие выдвижных ящиков (не менее 3 для хранения документации, листингов, канцелярских принадлежностей).
- высота рабочей поверхности рекомендуется в пределах 680-760 мм. Высота поверхности, на которую устанавливается клавиатура, должна быть около 650 мм.

Большое значение придается характеристикам рабочего стула (кресла).

Рабочий стул (кресло) должен быть подъемно-поворотным и регулируемым по высоте и углам наклона сиденья и спинки, а также регулируемым по расстоянию спинки от переднего края сиденья. Конструкция стула должна обеспечивать:

- ширину и глубину поверхности сиденья не менее 400 мм;
- поверхность сиденья с закругленным передним краем;
- регулировку высоты поверхности сиденья в пределах 400 - 550 мм и углов наклона вперед до 15° и назад до 5° ;
- высоту опорной поверхности спинки 300 ± 20 мм, ширину - не менее 380 мм и радиус кривизны горизонтальной плоскости - 400 мм;
- угол наклона спинки в вертикальной плоскости в пределах $0 \pm 30^\circ$;

- регулировку расстояния спинки от переднего края сиденья в пределах 260-400 мм;
- стационарные или съемные подлокотники длиной не менее 250 мм и шириной - 50-70 мм;
- регулировку подлокотников по высоте над сиденьем в пределах 230 ± 30 мм и внутреннего расстояния между подлокотниками в пределах 350 - 500 мм.

Поверхность сиденья, спинки и других элементов стула (кресла) должна быть полумягкой с нескользящим, неэлектризующимся и воздухопроницаемым покрытием, обеспечивающим легкую очистку от загрязнения.

Кресло следует устанавливать на такой высоте, чтобы не чувствовалось давления на копчик (это может быть при низком расположении кресла) или на бедра (при слишком высоком).

Работающий за ПЭВМ должен сидеть прямо, опираясь в области нижнего края лопаток на спинку кресла, не сутулясь, с небольшим наклоном головы вперед (до 5-7°). Предплечья должны опираться на поверхность стола, снимая тем самым статическое напряжение плечевого пояса и рук.

Рабочее место должно быть оборудовано подставкой для ног, имеющей ширину не менее 300 мм, глубину не менее 400 мм, регулировку по высоте в пределах до 150 мм и по углу наклона опорной поверхности подставки до 20°. Поверхность подставки должна быть рифленой и иметь по переднему краю бортик высотой 10 мм.

Необходимо предусматривать при проектировании возможность различного размещения документов: сбоку от видеотерминала, между монитором и клавиатурой и т.п. Кроме того, в случаях, когда видеотерминал имеет низкое качество изображения, например, заметны мелькания, расстояние от глаз до экрана делают больше (около 700 мм), чем расстояние от глаза до документа (300 - 450 мм). Вообще при высоком качестве изображения на

видеотерминале расстояние от глаз пользователя до экрана, документа и клавиатуры может быть равным.

Положение экрана определяется:

- расстоянием считывания (0,6...0,7 м);
- углом считывания, направлением взгляда на 20° ниже горизонтали к центру экрана, причем экран перпендикулярен этому направлению.

Должна также предусматриваться возможность регулирования экрана:

- по высоте +3 см;
- по наклону от -10° до $+20^\circ$ относительно вертикали;
- в левом и правом направлениях.

Большое значение также придается правильной рабочей позе пользователя. При неудобной рабочей позе могут появиться боли в мышцах, суставах и сухожилиях.

Требования к рабочей позе пользователя видеотерминала следующие:

- голова не должна быть наклонена более чем на 20° ;
- плечи должны быть расслаблены;
- локти - под углом $80^\circ \dots 100^\circ$;
- предплечья и кисти рук - в горизонтальном положении.

Причина неправильной позы пользователей обусловлена следующими факторами:

- нет хорошей подставки для документов;
- клавиатура находится слишком высоко, а документы – низко;
- некуда положить руки и кисти;
- недостаточно пространство для ног.

Создание благоприятных условий труда и правильное эстетическое оформление рабочих мест на производстве имеет большое значение как для облегчения труда, так и для повышения его привлекательности, положительно влияющей на производительность труда.

Заключение

В ходе данного исследования была проведена оценка качества существующей рекомендательной системы на примере библиотечного каталога НТБ ТПУ. Были выполнены следующие задачи:

- выбран наиболее подходящий метод исследования для оценки качества рекомендательной системы библиотечного каталога НТБ ТПУ – это мера Sensitivity (Recall). Для этого был проведен сравнительный анализ существующих методов оценки качества
- разработана и реализована информационная модель для оценки качества рекомендательной системы с использованием программных средств Node.JS и PostgreSQL. Данная модель позволяет получить значения меры Sensitivity (Recall), которые дают оценку качества системы

Анализ качества рекомендательной системы на основе меры Sensitivity (Recall), который был проведен согласно вычислительному эксперименту, показал, что данная рекомендательная система для библиотечного каталога НТБ ТПУ имеет рекомендации низкого уровня качества. Также выявлено, что для порядка 90% пользователей из тестовой выборки система выдает низкий уровень качества рекомендаций. Такая ситуация может быть вызвана тем, что для оценки качества рекомендательной системы для библиотечного каталога НТБ ТПУ использовался достаточно малый объем данных. В данном исследовании использовались данные о 101420 заказах. Для получения более точной оценки рекомендательной системы желательно использовать данные о 500 тысячах и более заказов. В ходе данной работе качество выдаваемых рекомендаций оказалось недостаточным для того, чтобы использовать данную рекомендательную систему в информационной системе НТБ ТПУ.

В дальнейшем, для улучшения качества рекомендательной системы библиотечного каталога НТБ ТПУ могут быть предприняты следующие меры:

1. Введение системы оценивания объектов библиотеки по шкале. То есть для пользователя библиотечного каталога НТБ ТПУ будет предложено оценить тот или иной объект по шкале («Очень понравилось», «Понравилось», «Нейтрально», «Не понравилось», «Плохо») или по балльной шкале (от 1 до 5 баллов)
2. Возможность добавления в избранное. Для пользователя информационной системы НТБ ТПУ при просмотре тех или иных объектов библиотечного фонда будет возможность добавить тот или иной документ в избранное.
3. Просмотр веб-страницы объекта. Для пользователя библиотечного каталога НТБ ТПУ будут фиксироваться системой объекты, которые пользователь просматривал, а также подсчитываться количество посещений веб-страницы для данного объекта.

Данные меры позволят получить информацию об интересах и предпочтениях по каждому пользователю. В дальнейшем это позволит дать более точную оценку качества рекомендательной системы. Также стоит отметить, что разработанная система для оценки качества возможно использовать и на других рекомендательных системах.

Список литературы

1. Melville, P. Content-Boosted Collaborative Filtering for Improved Recommendations / P. Melville, R. Mooney, R. Nagarajan. – Схема доступа: <http://www.aaai.org/Papers/AAAI/2002/AAAI02-029.pdf>.
2. Федоровский А. Н. Архитектура рекомендательной системы, работающей на основе неявных пользовательских оценок / А. Н. Федоровский, В. К. Логачева // Труды 13й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции». RCDL'2011. -- Воронеж, 2011.
3. Melville, Prem, Recommender Systems / Prem Melville, Vikas Sindhwani. Схема доступа: .
4. Sarwar, Badrul. Item-Based Collaborative Filtering Recommendation Algorithms / Badrul Sarwar, George Karypis, Joseph Konstan, John Riedl. Схема доступа: <http://www.ra.ethz.ch/cdstore/www10/papers/pdf/p519.pdf>.
5. Akosa, Josephine S. Predictive Accuracy: A Misleading Performance Measure for Highly Imbalanced Data / Josephine S Akosa. Схема доступа: <http://support.sas.com/resources/papers/proceedings17/0942-2017.pdf>.
6. Информационная модель: описание, структура, виды, типы информационных моделей, разработка, создание, использование информационной модели. – Схема доступа: <https://businessman.ru/new-informacionnaya-model-opisanie-struktura-vidy-tipy-informacionnyh-modelej-razrabotka-sozdanie-ispolzovanie-informacionnoj-modeli.html> (дата обращения 19.04.2017)
7. Кантелон, Майк. Node.JS в действии / Майк Кантелон.
8. Самарский А.А. Математическое моделирование и вычислительный эксперимент // Вестник АН СССР. — 1979. — № 5. — С. 38–49.
9. Горбунов-Посадов М.М. Системное обеспечение пакетов прикладных программ / М.М. Горбунов-Посадов, Д.А. Корягин, В.В. Мартынюк. — М.: Наука, 1990. — 208 с.

10. Recommender System. – Схема доступа: <http://recommender-systems.org/content-based-filtering/> (дата обращения 19.04.2017)
11. Pazzani, Michael J. Content-based Recommendation Systems / Michael J. Pazzani, Daniel Billsus. – Схема доступа: <https://pdfs.semanticscholar.org/3444/6adc7d701a2c3a89c2fc5f6d3479eef407b0.pdf>.
12. Asela Gunawardana, Guy Shani A Survey Evaluation Metrics of Recommendation Tasks / Asela Gunawardana, Guy Shani.