

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

Инженерная школа информационных технологий и робототехники
 Направление подготовки 09.04.01 «Информатика и вычислительная техника»
 Отделение информационных технологий

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Тема работы
Разработка информационной системы для оперативного обновления обучающей информации на портале

УДК 004.744:004.65:004.451:004.853

Студент

Группа	ФИО	Подпись	Дата
8BM81	Ахмеров Максим Шамильевич		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ	Саврасов Федор Витальевич	к.т.н.		

КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН	Конотопский Владимир Юрьевич	к.э.н.		

По разделу «Социальная ответственность»

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

ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ	Саврасов Федор Витальевич	к.т.н.		

ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

Код результата	Результат обучения (выпускник должен быть готов)
P1	Применять базовые и специальные естественнонаучные и математические знания в области информатики и вычислительной техники, достаточные для комплексной инженерной деятельности.
P2	Применять базовые и специальные знания в области современных информационных технологий для решения инженерных задач.
P3	Ставить и решать задачи комплексного анализа, связанные с созданием аппаратно-программных средств информационных и автоматизированных систем, с использованием базовых и специальных знаний, современных аналитических методов и моделей.
P4	Разрабатывать программные и аппаратные средства (системы, устройства, блоки, программы, базы данных и т. п.) в соответствии с техническим заданием и с использованием средств автоматизации проектирования.
P5	Проводить теоретические и экспериментальные исследования, включающие поиск и изучение необходимой научно-технической информации, математическое моделирование, проведение эксперимента, анализ и интерпретация полученных данных, в области создания аппаратных и программных средств информационных и автоматизированных систем.
P6	Внедрять, эксплуатировать и обслуживать современные программно-аппаратные комплексы, обеспечивать их высокую эффективность, соблюдать правила охраны здоровья, безопасность труда, выполнять требования по защите окружающей среды.
Универсальные компетенции	
P7	Использовать базовые и специальные знания в области проектного менеджмента для ведения комплексной инженерной деятельности.
P8	Владеть иностранным языком на уровне, позволяющем работать в иноязычной среде, разрабатывать документацию, презентовать и защищать результаты комплексной

	инженерной деятельности.
P9	Эффективно работать индивидуально и в качестве члена группы, состоящей из специалистов различных направлений и квалификаций, демонстрировать ответственность за результаты работы и готовность следовать корпоративной культуре организации.
P10	Демонстрировать знания правовых, социальных, экономических и культурных аспектов комплексной инженерной деятельности.
P11	Демонстрировать способность к самостоятельному обучению в течение всей жизни и непрерывному самосовершенствованию в инженерной профессии.

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

Инженерная школа информационных технологий и робототехники
 Направление подготовки 09.04.01 «Информатика и вычислительная техника»
 Отделение информационных технологий

УТВЕРЖДАЮ:
 Руководитель ООП

 (Подпись) (Дата) (Ф.И.О.)

ЗАДАНИЕ
на выполнение выпускной квалификационной работы

В форме:

Магистерской диссертации

Студенту:

Группа	ФИО
8BM81	Ахмерову Максиму Шамильевичу

Тема работы:

Разработка информационной системы для оперативного обновления обучающей информации на портале	
Утверждена приказом директора (дата, номер)	№ 59-63/с от 28.02.2020

Срок сдачи студентом выполненной работы:

01.06.2020

ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

Исходные данные к работе	Набор оценок пользователей по конкретным статьям, полученных с информационной системы.
---------------------------------	--

Перечень подлежащих исследованию, проектированию и разработке вопросов	<ul style="list-style-type: none"> — Обзор современных подходов к разработке информационных систем; — Изучение основных алгоритмов рекомендательных систем; — Разработка графического интерфейса системы; — Программная реализация информационной системы.
Перечень графического материала	<ul style="list-style-type: none"> — UML диаграмма компонентов; — UML диаграмма развертывания системы.
Консультанты по разделам выпускной квалификационной работы	
Раздел	Консультант
Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	Конотопский Владимир Юрьевич
Социальная ответственность	Горбенко Михаил Владимирович
Раздел на английском языке	Сидоренко Татьяна Валерьевна
Названия разделов, которые должны быть написаны на русском и иностранном языках:	
Анализ предметной области	

Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику	10.03.2020
---	------------

Задание выдал руководитель:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ	Саврасов Федор Витальевич	к.т.н.		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8BM81	Ахмеров Максим Шамильевич		

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

Инженерная школа информационных технологий и робототехники
 Направление подготовки 09.04.01 «Информатика и вычислительная техника»
 Уровень образования магистратура
 Отделение информационных технологий
 Период выполнения осенний / весенний семестр 2019 /2020 учебного года

Форма представления работы:

Магистерская диссертация

КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН выполнения выпускной квалификационной работы

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

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
15.11.19	Анализ существующих рекомендательных систем	20
25.02.20	Обзор современных подходов к разработке информационных систем	20
28.03.20	Разработка информационной системы для написаний и публикации статей	30
15.04.20	Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	10
10.05.20	Социальная ответственность	10
25.05.20	Раздел на английском языке	10

СОСТАВИЛ:

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ	Саврасов Федор Витальевич	к.т.н.		

СОГЛАСОВАНО:

Руководитель ООП

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ	Саврасов Федор Витальевич	к.т.н.		

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

Студенту:

Группа	ФИО
8ВМ81	Ахмеров Максим Шамильевич

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

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

1. Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих	Оклад руководителя – 33664руб. Оклад инженера – 15470руб Общая сумма затрат – 452443руб.
2. Нормы и нормативы расходования ресурсов	Тариф на электроэнергию – 6,59 руб/кВт Норма амортизации – 40%
3. Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования	Ставка для расчета отчислений во внебюджетные фонды –30 % Ставка для расчета накладных расходов –10%)

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

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

Перечень графического материала (с точным указанием обязательных чертежей):

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

Дата выдачи задания для раздела по линейному графику

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Конотопский В.Ю.	к.э.н.		25.02.2020 г.

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8ВМ81	Ахмеров Максим Шамильевич		

ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

Группа	ФИО
8ВМ81	Ахмерову Максиму Шамильевичу

Школа	Информационных технологий и робототехники	Отделение (НОЦ)	Отделение информационных технологий
Уровень образования	Магистратура	Направление/специальность	09.04.01 Информатика и вычислительная техника

Тема ВКР:

Разработка информационной системы для оперативного обновления обучающей информации на портале	
Исходные данные к разделу «Социальная ответственность»:	
1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения	Объектом исследования создание информационной системы для оперативного обновления обучающей информации на портале. Данная система предназначена для пользователей сети интернет, которые будут просматривать статьи портала для получения новой информации.
Перечень вопросов, подлежащих исследованию, проектированию и разработке:	
1. Правовые и организационные вопросы обеспечения безопасности: <ul style="list-style-type: none"> – специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства; – организационные мероприятия при компоновке рабочей зоны. 	1. Специальные правовые нормы трудового законодательства; 2. Организационные мероприятия при компоновке рабочей зоны.
2. Производственная безопасность: 2.1. Анализ выявленных вредных и опасных факторов 2.2. Обоснование мероприятий по снижению воздействия	Возможны следующие вредные факторы: – Нарушение параметров микроклимата. – Плохое освещение в помещении. Возможны следующие опасные факторы: – Высокое напряжение в электрической цепи, вероятность короткого замыкания.
3. Экологическая безопасность:	Воздействие объекта на атмосферу, гидросферу не происходит. В работе проведён анализ воздействия на литосферу (образование отходов при поломке оборудования).
4. Безопасность в чрезвычайных ситуациях:	Перечень возможных ЧС, которые могут возникнуть при работе в помещении офиса. Мероприятия по профилактике и недопущению пожаров.

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

Задание выдал консультант:

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

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8ВМ81	Ахмеров Максим Шамильевич		

РЕФЕРАТ

Выпускная квалификационная работа содержит 84 страницы, 20 рисунков, 19 таблиц, 22 источника и 1 приложение.

Ключевые слова: информационная система, веб-разработка, рекомендательная система, матричная факторизация.

Объектом исследования являются алгоритмы работы рекомендательных систем.

Цель работы: разработка информационной системы для написания и публикации статей и курсов. Так же в данной системе должен быть реализован рекомендательный поиск статей.

В процессе исследования были изучены основные виды рекомендательных систем, а так же основные алгоритмы для работы с ними.

В результате исследования была разработана готовая информационная система для написания и публикации статей. Так же была реализована рекомендательная система на основе коллаборативной фильтрации.

Основные конструктивные, технологические и технико-эксплуатационные характеристики заключаются в том, что разрабатываемая система представляет собой веб-приложение, позволяющее пользователям читать или добавлять новые статьи в систему.

Приложение будет располагаться в глобальной сети Интернет, и предназначено для публикации статей на тему программирования.

Значимость работы заключается в практической разработке информационной системы для пользователей сети Интернет. Благодаря данной системе увеличится качество подбора рекомендованных статей для пользователей.

Дальнейшим развитием работы является оптимизация рекомендательной системы, а именно увеличение эффективности работы алгоритма с целью повышения точности подбора статей.

ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

ИС – Информационная система

ВКР – Выпускная квалификационная работа

ALS – Alternating Least Square

MVC – Model View Controller

HTML – HyperText Markup Language

CSS – Cascading Style Sheets

PHP — Hypertext Preprocessor

Оглавление	
ВВЕДЕНИЕ.....	15
1. Анализ предметной области.....	16
1.1 Рекомендательная система.....	16
1.2 Виды оценок пользователя.....	17
1.2.1 Явная оценка.....	17
1.2.2 Неявная оценка.....	18
1.2.3 Гибридная оценка.....	18
1.3 Алгоритмы рекомендательных систем.....	19
1.3.1 Summary-based алгоритмы.....	19
1.3.2 Content-based алгоритмы.....	19
1.3.3 User-based алгоритмы.....	21
2. Разработка информационной системы.....	22
2.1 Методы и средства разработки.....	22
2.1.1 Языки разработки.....	22
2.1.2 Среда разработки.....	23
2.1.3 Используемые средства и фреймворки.....	23
2.2 Проектирование архитектуры приложения.....	26
2.3 База данных.....	28
2.4 Описание Alternating Least Squares алгоритма.....	29
3. Результаты работы.....	32
4. ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ.....	38
Введение.....	38

4.1	Оценка коммерческого потенциала и перспективности проведения научных исследований с позиции ресурсоэффективности и ресурсосбережения.....	38
4.1.1	Потенциальные потребители результатов исследования.....	38
4.1.2	Исследование целесообразности вложения денежных средств в научно-исследовательский проект	38
4.1.3	SWOT-анализ	40
4.2	Определение возможных альтернатив проведения исследований.....	41
4.3	Планирование научно-исследовательских работ	42
4.3.1	Структура работ в рамках научного исследования	42
4.3.2	Определение трудоемкости выполнения работ	43
4.3.3	Составление календарного план-графика работ	46
4.3.4	Бюджет научно-технического исследования.....	49
4.3.4.1	Расчет материальных затрат НТИ	49
4.3.4.2	Расчет амортизационных расходов	49
4.3.4.3	Расчет затрат на электроэнергию	50
4.3.4.4	Расчет основной и дополнительной заработной платы	51
4.3.4.5	Отчисления во внебюджетные фонды	52
4.3.4.6	Прочие расходы	52
4.3.4.7	Формирование бюджета НТИ	52
4.4	Определение ресурсной, финансовой и экономической эффективности исследования.....	53
5.	СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ.....	56
	Введение	56
5.1	Правовые и организационные вопросы обеспечения безопасности... ..	57
5.1.1	Специальные правовые нормы трудового законодательства	57

5.1.2	Организационные мероприятия при компоновке рабочей зоны	58
5.2	Производственная безопасность	59
5.2.1	Освещенность рабочей зоны	59
5.2.2	Микроклимат помещения	62
5.2.3	Уровень шума	64
5.2.4	Электромагнитное излучение	65
5.2.5	Электробезопасность	66
5.2.6	Обоснование мероприятий по защите окружающей среды	67
5.3	Экологическая безопасность	68
5.3.1	Загрязнение атмосферного воздуха	68
5.3.2	Загрязнение гидросферы	68
5.3.3	Отходы	69
5.4	Безопасность в чрезвычайных ситуациях	69
	ЗАКЛЮЧЕНИЕ	72
	CONCLUSION	73
	Список использованных источников	74
	Приложение А	76

ВВЕДЕНИЕ

В настоящее время в интернете существует проблема большого количества информации в интернете. Для ее решения необходимо фильтровать, расставлять приоритеты и эффективно доставлять информацию. Рекомендательные системы ищут в большом объеме динамически генерируемого контента нужную информацию, чтобы предоставить пользователям персонализированные товары и услуги. Информационно-поисковые системы, например, Google и Yandex частично решают данную проблему, предоставляя пользователям релевантный контент, но не так, как это делают рекомендательные системы. Для примера, большая часть выручки, около 75%, Netflix получает из продажи рекомендованных товаров. Для Amazon это процент составляет 35%. Благодаря рекомендательным системам пользу получают ее владельцы, так и сами клиенты.

Рекомендательные системы помогают пользователям быстрее найти необходимую информацию. С коммерческой точки зрения рекомендательные системы очень полезны, так как позволяют продавать большее количество товаров и услуг за счет рекомендаций.

1. Анализ предметной области

1.1 Рекомендательная система

Рекомендательная система — это программа, которая пытается предсказать, какие объекты (статьи, музыка, книги, новости, фильмы) будут интересны пользователю. Рекомендательные системы работают на основе информации о пользователе.

Система рекомендаций определяется как стратегия принятия решений для пользователей в сложных информационных средах [1]. Кроме того, система рекомендаций была определена с точки зрения электронной коммерции как инструмент, который помогает пользователям искать записи знаний, связанных с их интересами и предпочтениями. Также система рекомендаций была определена как средство помощи и расширения социального процесса использования рекомендаций других людей для принятия решений, когда нет достаточных личных знаний или опыта альтернатив.

В настоящее время системы рекомендаций основываются на таких подходах, как контентная, совместная или гибридная фильтрация. Гибридная фильтрация является самой популярной фильтрацией, так как включает в себя сильные стороны контентной и совместных фильтраций.

Совместная фильтрация предлагает элементы на основе других пользователей со схожими интересами. Контентная фильтрация же рекомендует элементы, основываясь только на предпочтениях пользователя. Например, если пользователь читает статьи на тему программирования, соответственно ему будет предлагаться больше статей по программированию.

Однако у данных подходов существуют ограничения. Для контентной фильтрации ограничением является отсутствие необходимых данных, что значительно уменьшает качество работы рекомендательной системы. Так же существуют проблемы масштабируемости и холодного запуска. Для решения данных проблем используется гибридная фильтрация, которая включает в

себя сильные стороны контентной и совместной фильтраций для повышения производительности и точности рекомендательных систем.

Система собирает информацию о пользователе для определения интересов пользователя и в дальнейшем для прогнозирования каких-либо элементов. Система учитывает поведение пользователя, содержимое ресурсов, которые пользователь просматривал. Система рекомендаций работает только при правильно построении моделей. Для хорошей точности подбора элементов системе необходимо предоставить как можно больше информации, чтобы дать разумную рекомендацию с самого начала. Для решение проблем малого количества информации используются различные типы ввода, такие как наиболее удобная высококачественная явная обратная связь, которая включает в себя явный ввод пользователей в отношении их интереса к предмету, или неявная обратная связь (путем косвенного определения предпочтений пользователя посредством наблюдения за его поведением).

1.2 Виды оценок пользователя

1.2.1 Явная оценка

Для создания рекомендаций системе необходимо достаточное количество информации о пользователе. Чем больше информации, тем точнее работает модель. Явный вид оценок является более надежным видом оценки, так как пользователь самостоятельно оценивает элемент. С другой стороны от пользователя требуются дополнительные усилия, чтобы получить эту оценку, поэтому обычно рекомендательной системе не хватает явных оценок для создания рекомендаций.

Ваш отзыв

Понравилась ли Вам еда? **10** Как Вам интерьер? **10** Как Вам обслуживание? **10** **Ваша оценка**
10

Потрясающее место, очень уютный интерьер, очень довольны.

+ Уютный интерьер, расположение ресторана очень удачное, отличная кухня, есть места для парковки.

- Нет

Добавить фото Опубликовать

Рисунок 1 – Пример явной оценки

1.2.2 Неявная оценка

Для получения большего количества информации об интересах пользователя так же используется неявная оценка. Для получения неявной оценки используются конкретные действия пользователя, такие как выбор конкретной статьи, прочтение статьи до конца или добавление комментария к статье. Основное преимущество неявной оценки заключается в том, что пользователю не надо совершать каких-либо дополнительных действий для добавления оценки. Однако такой способ оценки является менее точным.

1.2.3 Гибридная оценка

Сильные стороны как неявной, так и явной обратной связи, могут быть объединены в гибридной системе, чтобы минимизировать их слабые стороны и получить наиболее эффективную систему [2].

Это достигается благодаря использованию неявных данных в качестве проверки явной оценки, позволяя пользователю давать явную обратную связь только тогда, когда он решает обозначить явный интерес. Гибридная оценка является наилучшим решением для рекомендательных систем.

1.3 Алгоритмы рекомендательных систем

Существует множество алгоритмов рекомендательных систем. К наиболее популярным относятся: Summary-based (неперсонализированные), Content-based (модели, основанные на описании товара), Collaborative Filtering (коллаборативная фильтрация).

1.3.1 Summary-based алгоритмы

Неперсонализированные рекомендации самые простые в реализации. Потенциальный интерес пользователя высчитывается исходя из среднего рейтинга товара. Если товар нравится всем пользователям, то нравится и новому пользователю. По этому принципу работает большое количество сервисов, когда пользователь не аутентифицируется в системе.

Рейтинг товара может отображаться разными способами. Например, рейтинг товара может быть отображен количеством “лайков” или “звездочек”. На форумах обычно отображается разница положительных и отрицательных голосов.

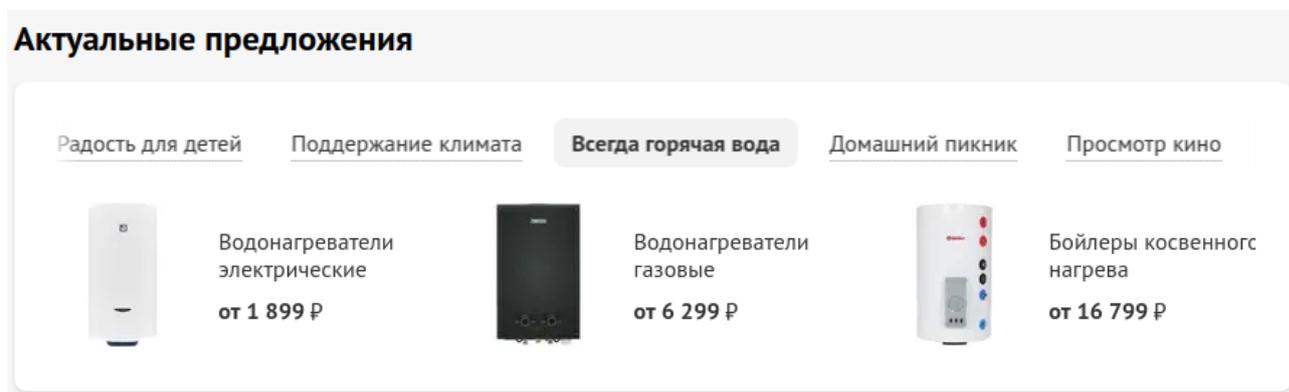


Рисунок 2 – Не персонализированные рекомендации

1.3.2 Content-based алгоритмы

Персональные рекомендации основываются на использовании информации о самом пользователе. В основном для расчетов используется информация о его предыдущих действиях, просмотрах и покупках. В данном подходе интересы пользователя сопоставляются с описанием предоставляемых товаров и услуг. Чем больше услуга или товар соответствует этим интересам, тем выше оценивается потенциальная

заинтересованность пользователя. Однако из этого следует, что у всех товаров и услуг должно быть описание.

В основном предметом Content-based-рекомендаций чаще бывают товары с неструктурированным описанием: статьи, книги, фильмы. Например, для фильмов это может быть краткое описание, состав актеров, для статей это могут быть теги, категории, или описание.

Неструктурированная информация может описываться векторами в пространстве (Vector-Space model). Каждый элемент такого вектора – признак, который представляет интересы пользователя. Сама сущность, например, статья или фильм, являются векторами.

По мере взаимодействия пользователя с системой (например, просмотр конкретной статьи), векторные признаки просмотренных статей объединяются (суммируются и нормализуются) в единый вектор и, в итоге, формируется вектор его интересов. Далее достаточно найти статью, вектор которой близок к вектору интересов пользователя, т.е. решить задачу поиска n -ближайших соседей.

Источником информации, с которой чаще всего используются контентные системы фильтрации, являются текстовые документы. Стандартный подход для разбора термина выбирает отдельные слова из документов. Модель векторного пространства и скрытая семантическая индексация — это два метода, которые используют эти термины для представления документов в качестве векторов в многомерном пространстве.

Обратная связь по релевантности, генетические алгоритмы, нейронные сети, и байесовский классификатор являются одними из методов обучения для изучения профиля пользователя. Модель векторного пространства и скрытая семантическая индексация могут использоваться этими методами обучения для представления документов. Некоторые из методов обучения также представляют профиль пользователя в виде одного или нескольких векторов в одном и том же многомерном пространстве, что позволяет легко сравнивать документы и профили. Другие методы обучения,

такие как байесовский классификатор и нейронные сети, не используют это пространство, а представляют профиль пользователя по-своему.

1.3.3 User-based алгоритмы

Данный вид систем начал активно развиваться в начале 90-ых годов 20-го века. В рамках этого подхода рекомендации подбираются на основе основании интересов других похожих пользователей. Такие рекомендации высчитываются благодаря «коллаборации» множества пользователей. Отсюда и произошло название метода. В основе подхода используется алгоритм k -ближайших соседей. Для каждого пользователя ищем k -наиболее похожих на него (на основании его предпочтений) и дополняем информацию о пользователе известными данными по его соседям.

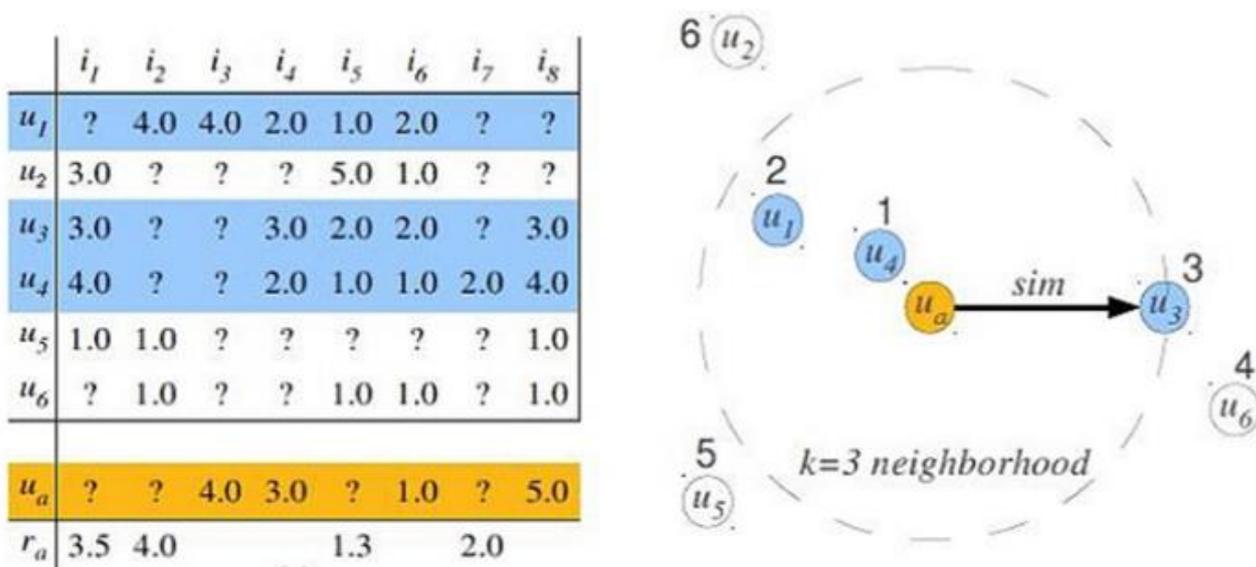


Рисунок 3 – Метод нахождения ближайших соседей

На рис. 3 желтым цветом выделен пользователь, для которого необходимо найти рекомендации. Синим цветом выделены три ближайших соседа, статьи которых будут предлагаться в качестве рекомендаций.

Схожесть интересов можно определять по формуле косинусного сходства.

$$similarity = \cos(\theta) = \frac{A * B}{\|A\| * \|B\|} = \frac{\sum_{i=1}^n A_i * B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} * \sqrt{\sum_{i=1}^n (B_i)^2}}$$

Так же существуют и другие алгоритмы для поиска ближайших соседей, например, корреляция Пирсона, расстояние Жаккара, расстояние Хэмминга. Однако данные методы плохо справляются с разреженностью данных, когда никто из соседей не оценил предмет, который система пытается предсказать для целевого пользователя. Кроме того, он неэффективен в вычислительном отношении, так как растет количество пользователей и продуктов.

2. Разработка информационной системы

Разработанная система предоставляет возможность пользователям писать свои собственные статьи и курсы и затем публиковать их. Создать свои статью или курс могут только аутентифицированные пользователи. Просматривать статьи может любой пользователь. Статьи должны быть разбиты на тематики. Основными тематиками будут современные направления frontend-разработки, такие как верстка, полезные инструменты, популярные библиотеки, дизайн, фреймворки, тестирование, архитектура приложения.

На данном сайте будет работать рекомендательная система подбора статей. Для не аутентифицированных пользователей будет использоваться неперсонализированная система, а для аутентифицированных пользователей будет использоваться коллаборативная фильтрация.

2.1 Методы и средства разработки

2.1.1 Языки разработки

Для серверной части был выбран язык программирования PHP. PHP — это популярный язык программирования общего назначения с открытым исходным кодом. Основным преимуществом данного языка является

встраиваемость в разметку HTML. Существует несколько основных областей применения PHP, одной из которых является создание скриптов для выполнения на стороне сервера. Данный вид применения является самым популярным. Для работы с PHP необходимо как минимум 3 вещи: интерпретатор PHP (в виде серверного модуля или программы CGI), веб-сервер и браузер. Для проверки работы PHP необходим браузер, который будет принимать ответы с сервера в виде готовых страниц.

Для написания клиентской части информационной системы использовался язык разметки HTML, а также таблицы стилей – CSS.

2.1.2 Среда разработки

Разработка проводилась в операционной системе Ubuntu 18.04.

В качестве редактора кода использовалась PHP Storm. PHP Storm — кросс-платформенная среда разработки кода, разрабатываемая компанией JetBrains. PHPStorm является очень популярным редактором кода, так как содержит в себе статический анализатор кода для PHP, HTML и JavaScript для уменьшения количества ошибок в коде, и универсальными средствами отладки для PHP и JavaScript.

2.1.3 Используемые средства и фреймворки

Фреймворк — инструмент, ускоряющий процесс написания и запуска веб-приложения. В основном, это набор библиотек или инструментов для разработки.

Для реализации информационной системы использовался Laravel. Это веб-фреймворк с открытым исходным кодом, предназначенный для разработки приложений с использованием архитектуры MVC. Использование Laravel помогает облегчить разработку, упрощая общие задачи, используемые в большинстве веб-проектов, такие как аутентификация, маршрутизация, сессии и кэширование. Laravel стремится сделать процесс разработки простым для разработчика, не жертвуя функциональностью

приложения. Laravel является доступным, но мощным инструментом, предоставляющим возможности, необходимые для больших и надежных приложений. Продвинутое инверсия контейнера управления, выразительная система миграции и тесно интегрированная поддержка модульного тестирования предоставляют инструменты, необходимые для создания любого приложения.

Для создания адаптивных страниц и выразительных графических элементов использовались следующие популярные библиотеки:

- jQuery;
- Bootstrap.

Для реализации рекомендательной системы был использован алгоритм Alternating Least Squares. Данный алгоритм реализован в библиотеке машинного обучения Spark Mllib, которая в свою очередь используется на сервере PredictionIO.

Apache PredictionIO — это сервер машинного обучения с открытым исходным кодом, созданный на основе современного стека, предназначенного для разработчиков и ученых, создающих прогностические механизмы для любой задачи машинного обучения. Это позволяет:

- быстро создать и развернуть движок в виде веб-службы на производстве с настраиваемыми шаблонами;
- отвечать на динамические запросы в режиме реального времени после развертывания в качестве веб-службы.

Apache PredictionIO может быть установлен как полный стек машинного обучения в комплекте с Apache Spark, MLib, HBase, Elasticsearch, что упрощает и ускоряет управление инфраструктурой масштабируемого машинного обучения.

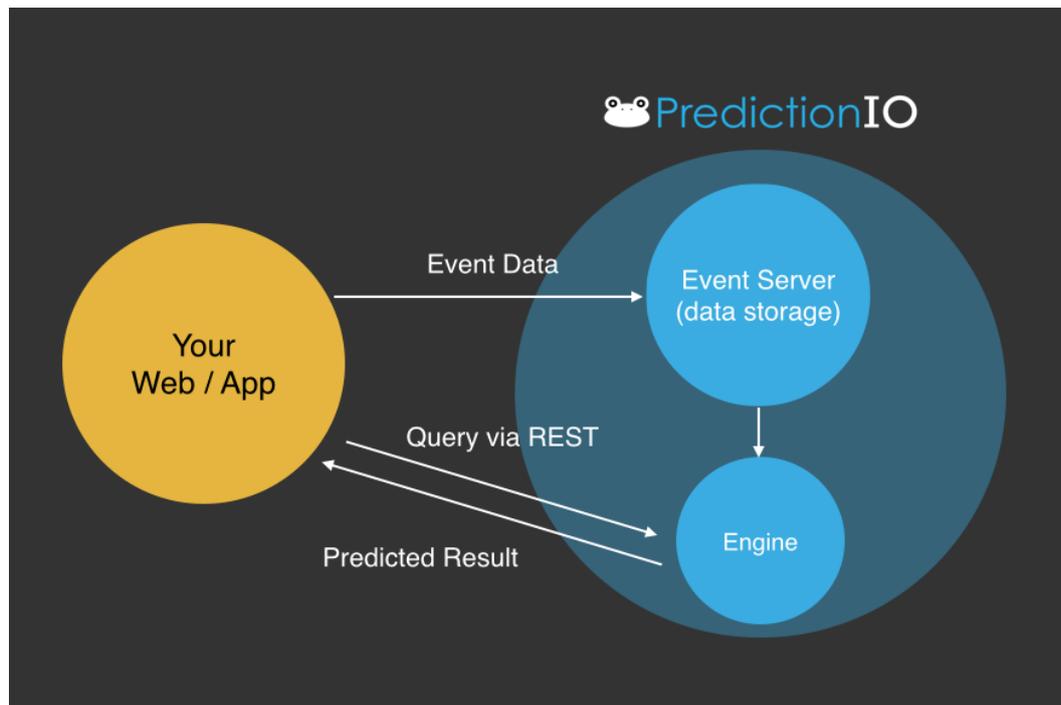


Рисунок 4 — Архитектура prediction io

В обычном сценарии сервер событий PredictionIO постоянно собирает данные из приложения [4]. Затем движок PredictionIO создает прогностические модели с одним или несколькими алгоритмами, используя данные. После развертывания в качестве веб-службы он прослушивает запросы из приложения и в режиме реального времени выводит данные в виде ожидаемых результатов.

Сервер событий собирает данные из приложения в режиме реального времени или в пакетном режиме. После того, как данные собраны, они в основном служат двум целям:

- предоставить данные Engine(s) для обучения и оценки модели;
- предоставить унифицированное представление данных для анализа.

Движок отвечает за прогнозирование. Он содержит несколько алгоритмов машинного обучения. Движок считывает данные обучения и строит прогностические модели. Затем он развертывается как веб-сервис. Развернутый механизм отвечает на запросы прогнозирования из приложения через REST API в режиме реального времени.

После того, как запущены EventServer и Engine, необходимо отправить данные на определенное имя приложения, идентифицируемое ключом доступа, через его API событий с HTTP-запросами или с одним из SDK. Для интеграции с разрабатываемым приложением используется Apache PredictionIO PHP SDK.

2.2 Проектирование архитектуры приложения

Взаимодействие компонентов информационной системы изображено на диаграмме компонентов.

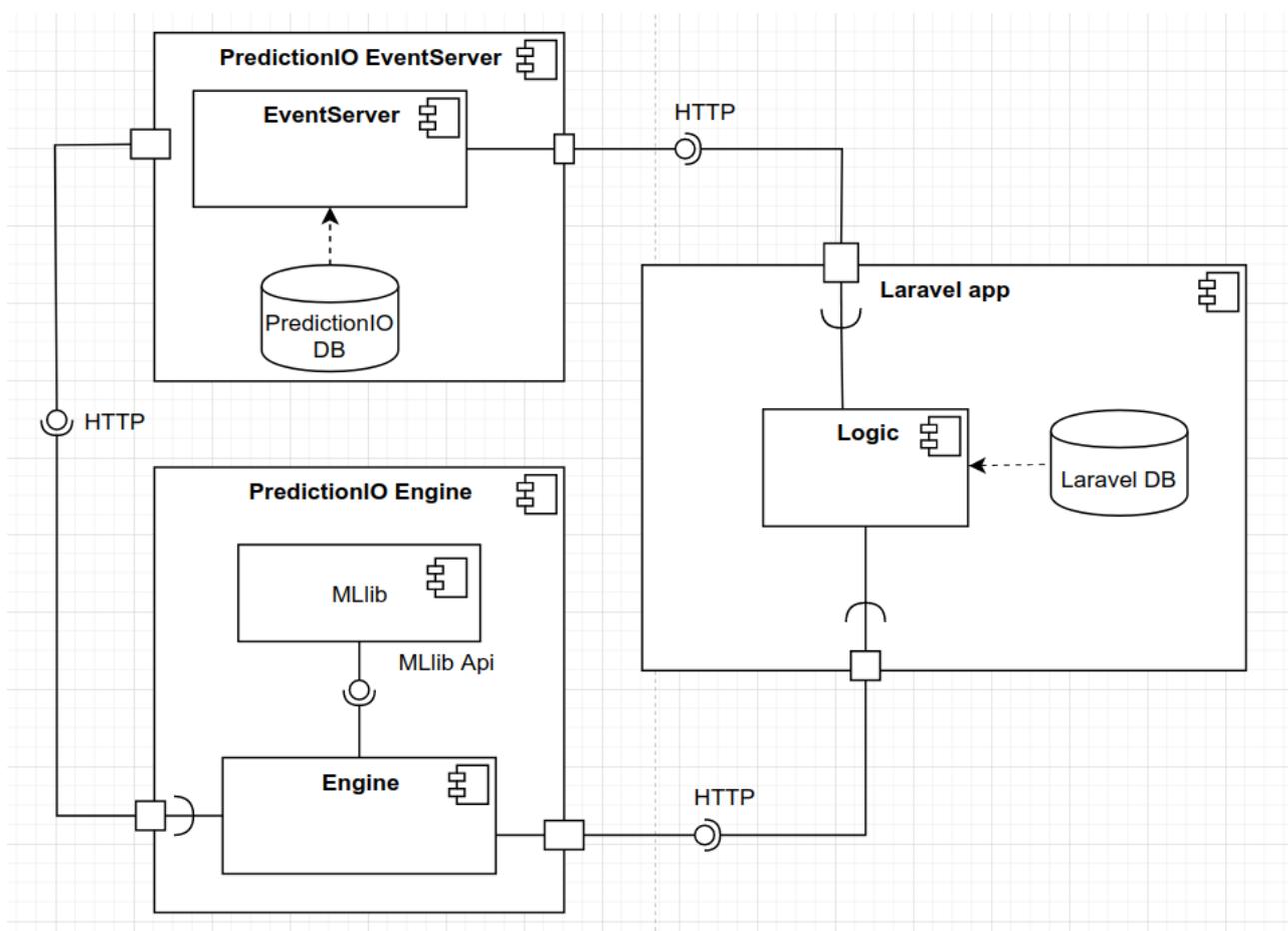


Рисунок 5 — Диаграмма компонентов приложения

Диаграмма компонентов системы включает в себя 3 основных элемента: PredictionIO EventServer, PredictionIO Engine, Laravel.

PredictioIO EventServer предоставляет интерфейс для обмена данными с помощью протокола HTTP и содержит в себе следующие компоненты:

EventServer – предназначен для обработки получаемой информации и ее хранения; PredictionIO DB — NoSQL база данных, которая хранит все данные, отправленные на EventServer. Используемая СУБД — Hbase.

PredictioIO Engine предоставляет интерфейс для обмена данными с помощью протокола HTTP и содержит в себе следующие компоненты: Engine – содержит в себе обучающую модель, начальные данные и функцию для запуска обучения; MLlib — Java-библиотека, в которой реализован алгоритм Alternating Least Squares.

Laravel app представляет собой веб-приложение, которое посредством протокола HTTP взаимодействует с PredictionIO для получения рекомендаций. Приложение так же содержит базу данных PostgreSQL для хранения статей и пользователей информационной системы.

Для представления общей конфигурации веб-приложения была использована диаграмма развертывания UML. Диаграмма развертывания предназначена для визуализации элементов приложения, которые существуют лишь на этапе ее исполнения. На диаграмме представлены четыре основных элемента: клиент, приложение, и два PredictionIO-сервера.

Клиент представлен в виде веб-браузера, который отправляет запрос на веб-сервер приложения для получения данных. Приложение представлено в виде веб-сервера. В качестве веб-сервера был выбран nginx. Nginx — это бесплатный высокопроизводительный HTTP-сервер с открытым исходным кодом и обратный прокси-сервер, а также прокси-сервер IMAP / POP3. Nginx известен своей высокой производительностью, стабильностью, богатым набором функций, простой настройкой и низким потреблением ресурсов. Веб-сервер представляет собой среду для исполнения, реализованную с помощью платформы PHP и веб-фреймворка Laravel. В основу фреймворка положены три модуля: Logic, Middlewares, Route Handler. Logic содержит в себе всю бизнес-логику приложения, описанную в контролерах приложения. Middlewares представляет собой промежуточный уровень, который осуществляет предобработку запросов. В Middlewares реализована отправка

данных для обучения моделей рекомендаций. Route Handler организует маршрутизацию клиентских запросов. Веб-сервера PredictionIO разворачиваются в Docker-контейнерах.

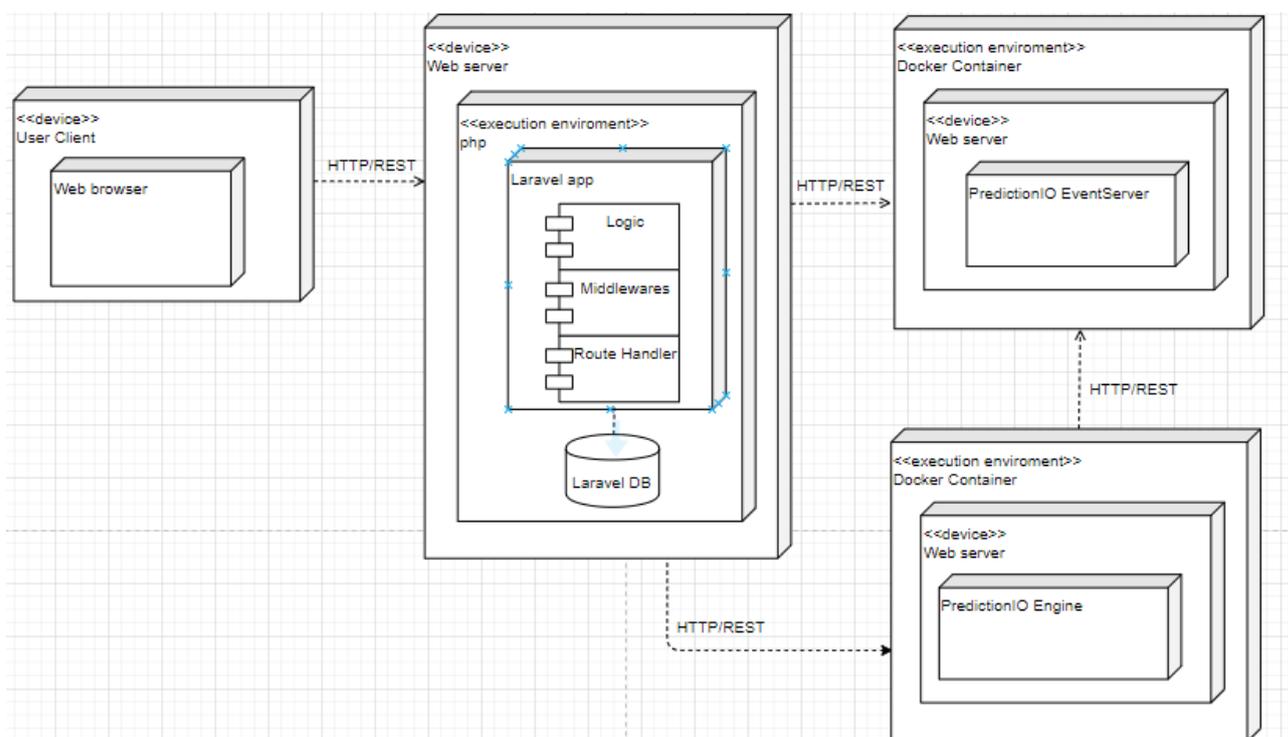


Рисунок 6 — Диаграмма развертывания

2.3 База данных

В качестве СУБД использовалась PostgreSQL. PostgreSQL — это мощная система объектно-реляционных баз данных с открытым исходным кодом с более чем 30-летним активным развитием, которая заслужила хорошую репутацию за стабильность, надежность функций и производительность. На рисунке представлена схема базы данных.

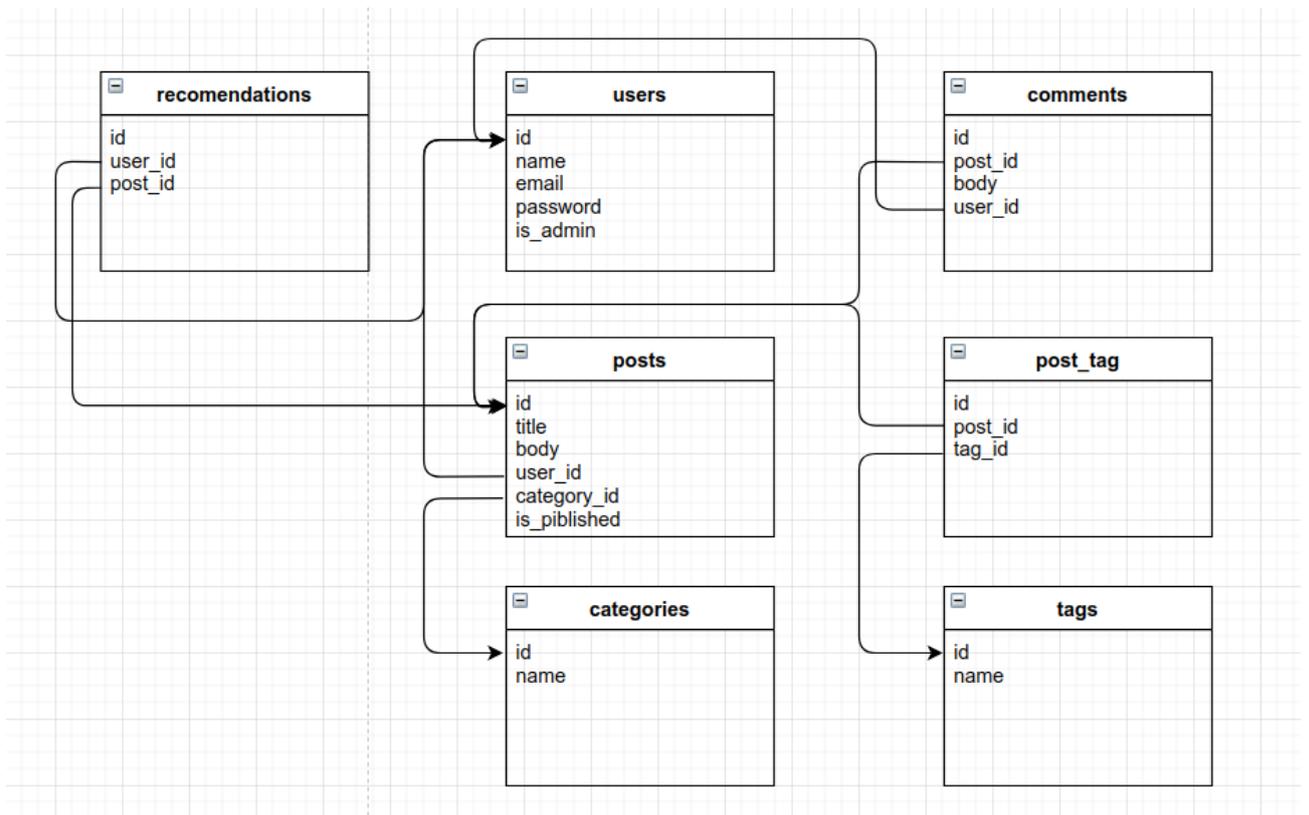


Рисунок 7 — Схема базы данных

2.4 Описание Alternating Least Squares алгоритма

Alternating Least Square (ALS) также является алгоритмом матричной факторизации и работает параллельно [3]. ALS реализован в Apache Spark ML и создан для масштабных задач совместной фильтрации.

ALS демонстрирует довольно хорошую производительность по решению масштабируемости и разреженности данных Ratings, и это просто и оптимально масштабируется для очень больших наборов данных.

userid	4	5	10	14	15	18	19	26	31	34	...	283199	283204	283206	283208	283210	283215	283219	283222	283224		
movieid																						
1	4.0	NaN	5.0	4.5	4.0	NaN	NaN	NaN	5.0	NaN	...	5.0	NaN	NaN	4.5	NaN	4.0	4.0	NaN	NaN		
2	4.0	NaN	NaN	4.0	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN									
3	NaN	NaN	NaN	NaN	NaN	NaN	4.0	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	5.0	NaN	NaN	4.0		
4	NaN	...	NaN	NaN																		
5	2.0	NaN	...	NaN	NaN	NaN	NaN	NaN	5.0	NaN	NaN	NaN										
6	4.5	NaN	NaN	NaN	NaN	3.0	4.0	NaN	NaN	NaN	...	NaN	NaN									
7	NaN	...	NaN	NaN	NaN	NaN	NaN	5.0	NaN	NaN	NaN											
8	NaN	...	NaN	NaN																		
9	NaN	NaN	NaN	NaN	NaN	NaN	4.0	NaN	NaN	NaN	...	NaN	NaN									
10	4.0	NaN	NaN	NaN	NaN	3.0	4.0	NaN	NaN	NaN	...	NaN	3.0									
11	3.5	NaN	NaN	NaN	NaN	NaN	4.0	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	3.0	NaN	NaN	NaN	NaN	
12	NaN	NaN	NaN	NaN	NaN	NaN	3.0	NaN	NaN	NaN	...	NaN	NaN									
13	NaN	...	NaN	NaN	NaN																	
14	NaN	...	NaN	NaN	3.0																	
15	NaN	NaN	NaN	NaN	NaN	NaN	3.0	NaN	NaN	NaN	...	NaN	NaN	NaN								

Рисунок 8 – Разреженная матрица

Для этого применяется факторизация матрицы. В итоге получается матрица пользователя и матрица элементов. Чтобы высчитать эти элементы, изначально коэффициенты заполняются случайным образом, и затем одна часть матрицы фиксируется, а вторая используется для вычисления коэффициентов. Затем подбираются коэффициенты, где ошибка будет минимальной, после этого вторая часть матрицы фиксируется, а другая используется для высчитывания коэффициентов. В итоге получается 2 матрицы, которые перемножаются и заполняются пропущенными значениями из итоговой матрицы.

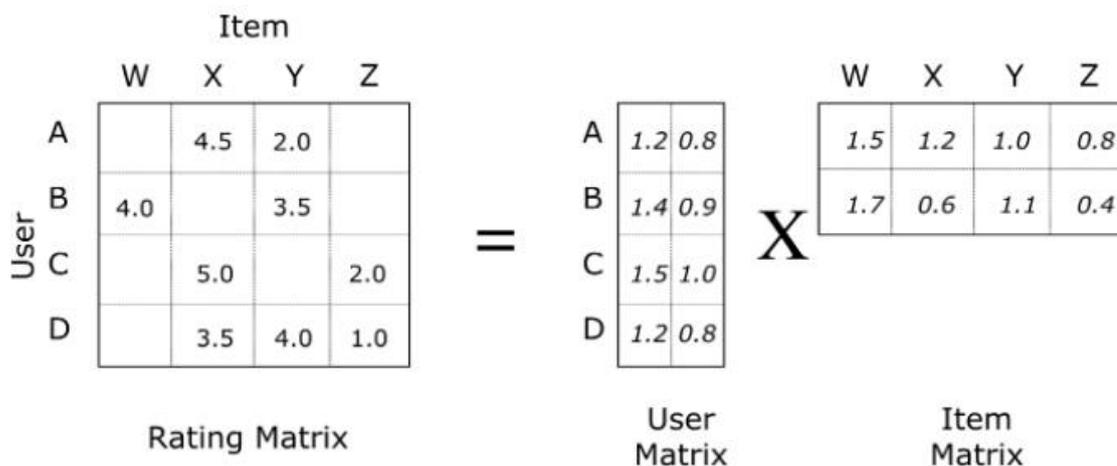


Рисунок 9 – Матричная факторизация

Для интегрирования с сервером PredictionIO используется Apache PredictionIO PHP SDK. Для добавления информации для обучения моделей используется объект класса EventClient. Отправка осуществляется в методе createEvent, куда передается массив данных.

```
$response = $client->createEvent([
    'event' => 'rate',
    'entityType' => 'user',
    'entityId' => $id,
    'targetEntityType' => 'item',
    'targetEntityId' => $postId,
    'properties' => ['rating' => $value]
]);
```

Заинтересованность пользователя в конкретной статье оценивается по 10-балльной шкале. Для добавления неявных оценок используются определенные действия пользователя. Все виды неявных оценок показаны в таблице 1.

Таблица 1 — оценка действий пользователя.

Действие пользователя	Неявная оценка
Пользователь просмотрел статью	4
Пользователь прочитал статью	6
Пользователь добавил комментарий к статье	8

3. Результаты работы

Для работы информационной системы необходимо запустить сервера PredictionIO. Результат запуска показан на рисунках 10-11.

```
dev@m-akhmerov:~/files/git$ cd ../../predictionio/predictionio/docker/
dev@m-akhmerov:~/predictionio/predictionio/docker$ docker-compose -f docker-compose.yml \
> -f pgsql/docker-compose.base.yml \
> -f pgsql/docker-compose.meta.yml \
> -f pgsql/docker-compose.event.yml \
> -f pgsql/docker-compose.model.yml \
>
up
Starting docker_postgres_1 ...
Starting docker_postgres_1 ... done
Starting docker_pio_1 ...
Starting docker_pio_1 ... done
Attaching to docker_postgres_1, docker_pio_1
postgres_1 |
postgres_1 | PostgreSQL Database directory appears to contain a database; Skipping initialization
pio_1 | Waiting for PredictionIO...
postgres_1 |
postgres_1 | LOG: database system was interrupted; last known up at 2020-05-26 16:35:19 UTC
postgres_1 | LOG: database system was not properly shut down; automatic recovery in progress
postgres_1 | LOG: invalid record length at 0/23FBE78: wanted 24, got 0
postgres_1 | LOG: redo is not required
postgres_1 | LOG: MultiXact member wraparound protections are now enabled
postgres_1 | LOG: database system is ready to accept connections
postgres_1 | LOG: autovacuum launcher started
pio_1 | [INFO] [Management$] Inspecting PredictionIO...
pio_1 | [INFO] [Management$] PredictionIO 0.13.0 is installed at /usr/share/predictionio
pio_1 | [INFO] [Management$] Inspecting Apache Spark...
pio_1 | [INFO] [Management$] Apache Spark is installed at /usr/share/spark-2.2.2-bin-hadoop2.7
pio_1 | [INFO] [Management$] Apache Spark 2.2.2 detected (meets minimum requirement of 1.6.3)
pio_1 | [INFO] [Management$] Inspecting storage backend connections...
pio_1 | [INFO] [Storage$] Verifying Meta Data Backend (Source: PGSQL)...
pio_1 | [INFO] [Storage$] Verifying Model Data Backend (Source: PGSQL)...
pio_1 | [INFO] [Storage$] Verifying Event Data Backend (Source: PGSQL)...
pio_1 | [INFO] [Storage$] Test writing to Event Store (App Id 0)...
pio_1 | [INFO] [Management$] Your system is all ready to go.
pio_1 | [INFO] [Management$] Creating Event Server at 0.0.0.0:7070
pio_1 | [INFO] [HttpListener] Bound to /0.0.0.0:7070
pio_1 | [INFO] [EventServerActor] Bound received. EventServer is ready.
```

Рисунок 10 – Запуск EventServer

PredictionIO Engine Server at 0.0.0.0:8000
org.example.recommendation.RecommendationEngine (default)

Engine Information

Training Start Time	Monday, May 25, 2020 7:15:13 AM UTC
Training End Time	Monday, May 25, 2020 7:15:17 AM UTC
Variant ID	default
Instance ID	11d5d30f-4b67-4131-8f69-e162aa0c4698

Server Information

Start Time	Saturday, May 30, 2020 3:40:39 PM UTC
Request Count	12
Average Serving Time	0.0356 seconds
Last Serving Time	0.0350 seconds
Engine Factory Class (Scala/Java)	org.example.recommendation.RecommendationEngine

Data Source

Parameters	DataSourceParams (myApp1, None)
------------	---------------------------------

Рисунок 11 – Запуск Engine

Также была разработана информационная система для написания и публикации статей. В данной системе пользователи могут регистрироваться и аутентифицироваться. Для не аутентифицированных пользователей работает Summary-based-рекомендательная система (подбираются 3 статьи, по которым больше всего откликов).

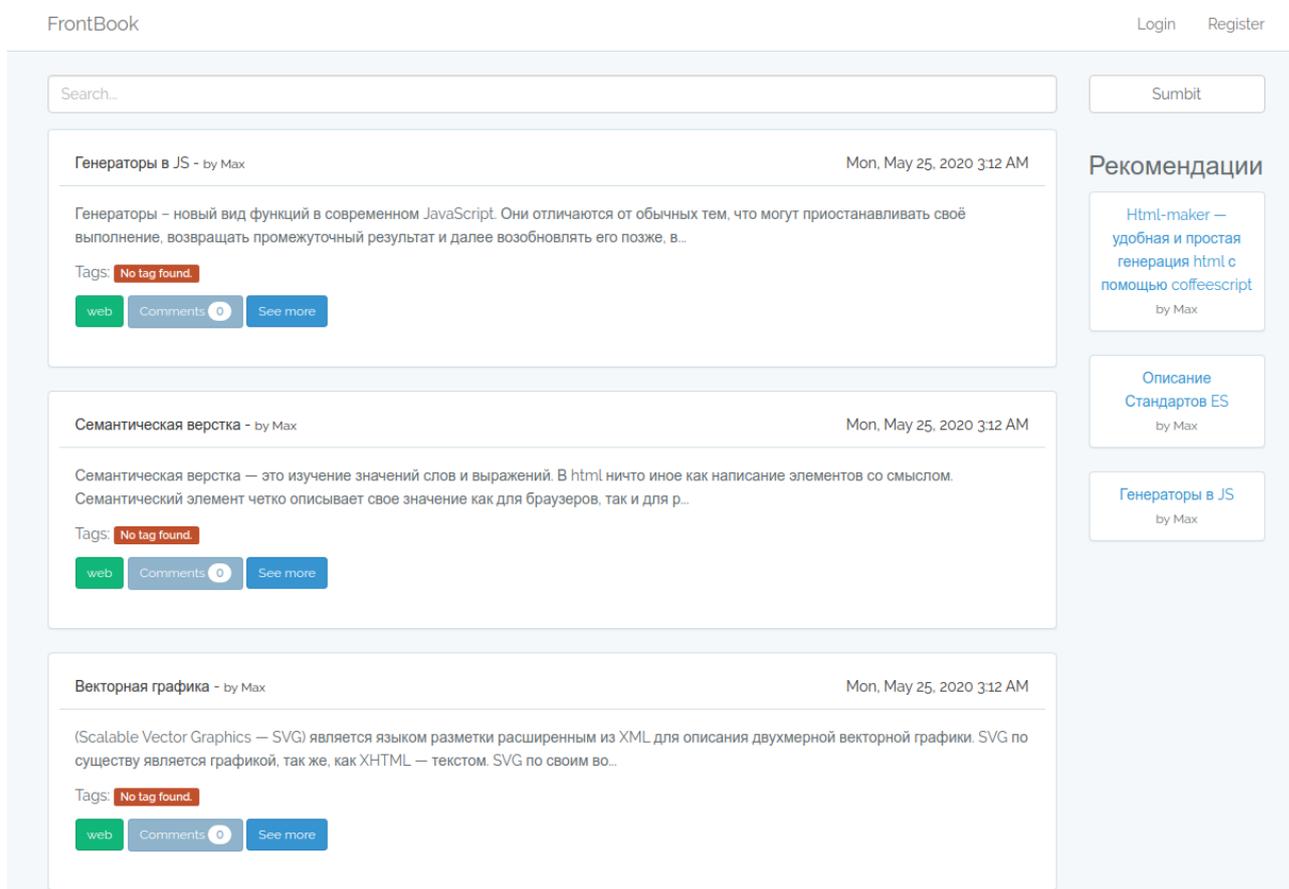


Рисунок 12 – Статьи для не аутентифицированных пользователей

Для аутентифицированных пользователей предлагаются статьи других пользователей на основе схожих интересов.

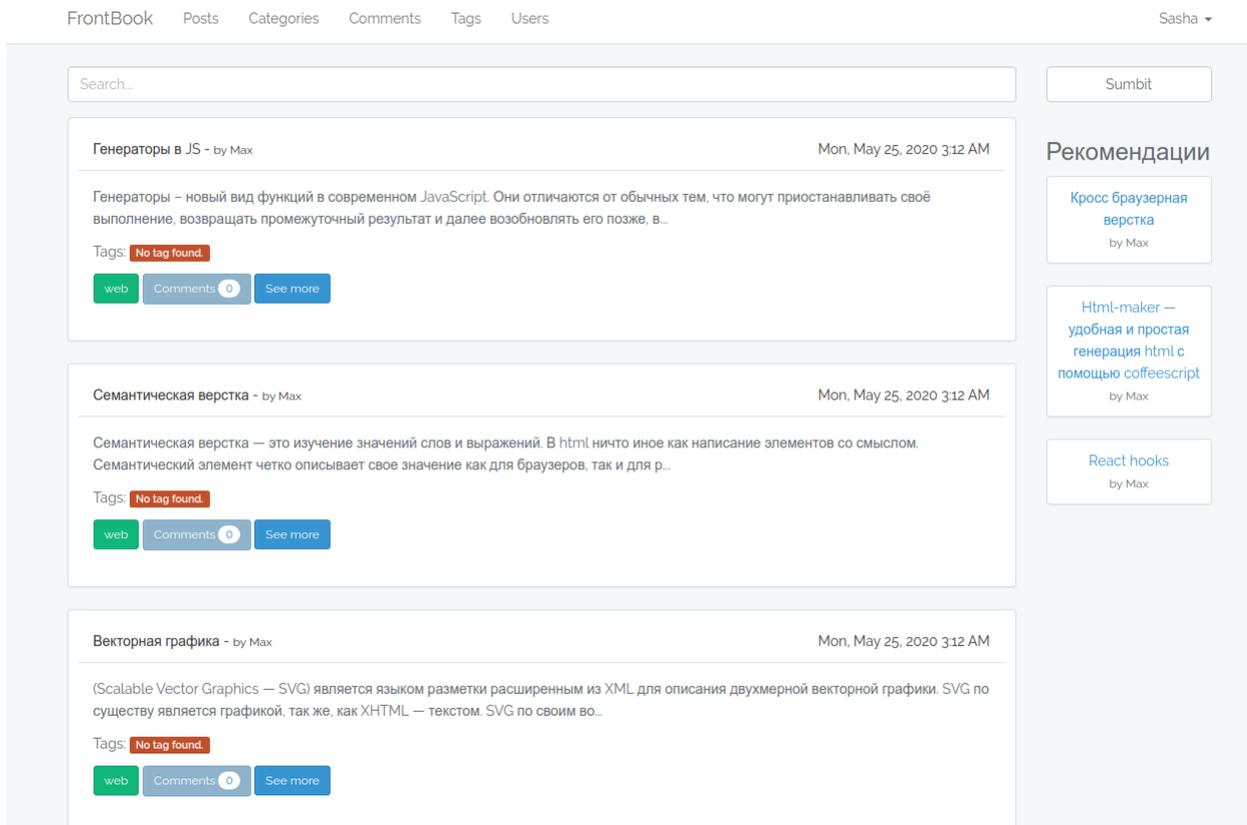


Рисунок 13 – Статьи для аутентифицированных пользователей

Публикация статей происходит вручную при помощи администратора.

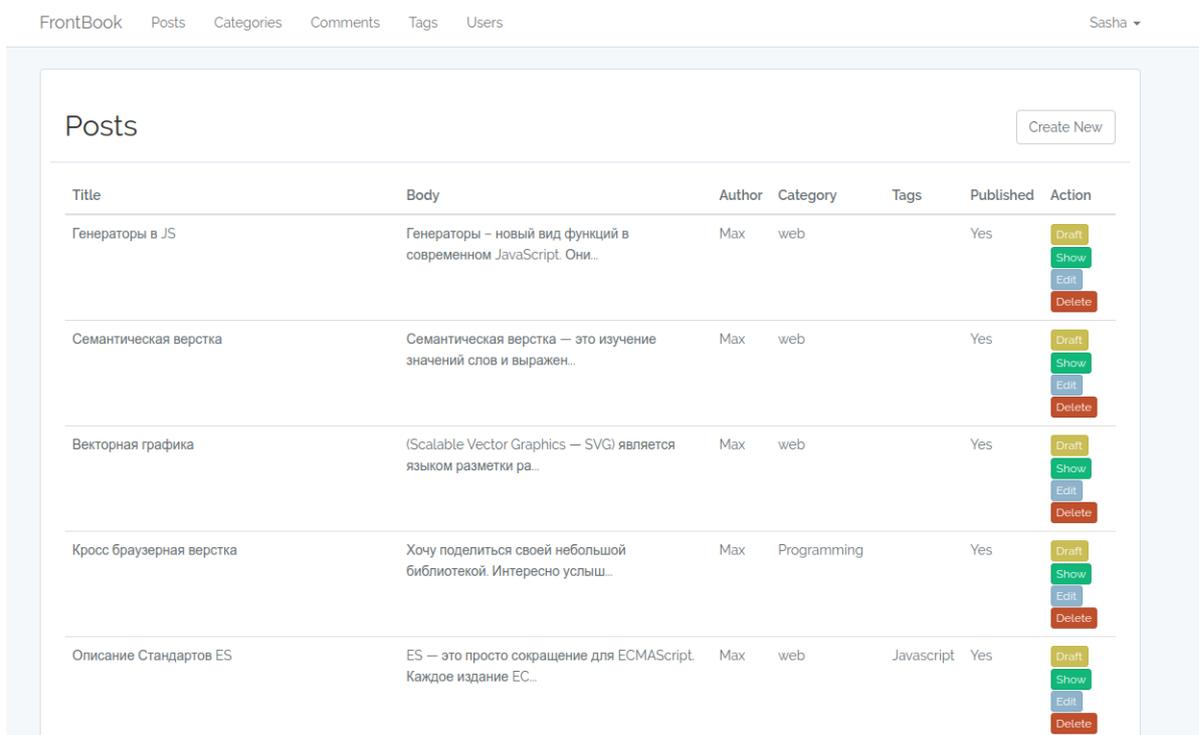


Рисунок 14 – Страница просмотра статей администратором

Администратор может добавлять, редактировать, и удалять любые статьи.

FrontBook Posts Categories Comments Tags Users Sasha ▾

Edit Post Go Back

Title

Body

Генераторы – новый вид функций в современном JavaScript. Они отличаются от обычных тем, что могут приостанавливать своё выполнение, возвращать промежуточный результат и далее возобновлять его позже, в произвольный момент времени.

Для объявления генератора используется новая синтаксическая конструкция: `function*` (функция со звёздочкой).

Её называют «функция-генератор» (generator function).

При создании генератора код находится в начале своего выполнения.

Основным методом генератора является `next()`. При вызове он возобновляет выполнение кода до

Category

Tag

Рисунок 15 – Страница редактирования статьи администратором

После прочтения статьи пользователь может оценить её по 5-балльной шкале. После оценки пользователя все данные отправляются на EventServer для дальнейшего обучения модели и подбора рекомендательных статей.

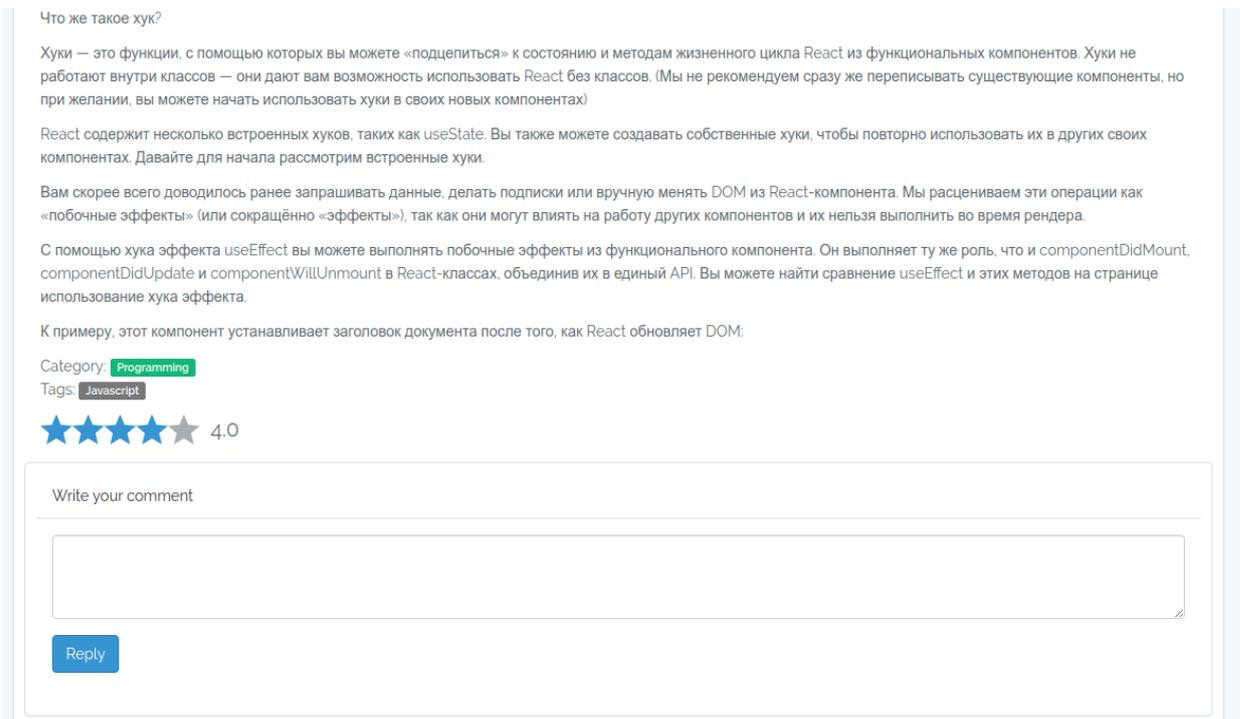


Рисунок 16 – Страница оценки статьи пользователем

Администратор может создавать новые теги для статей.

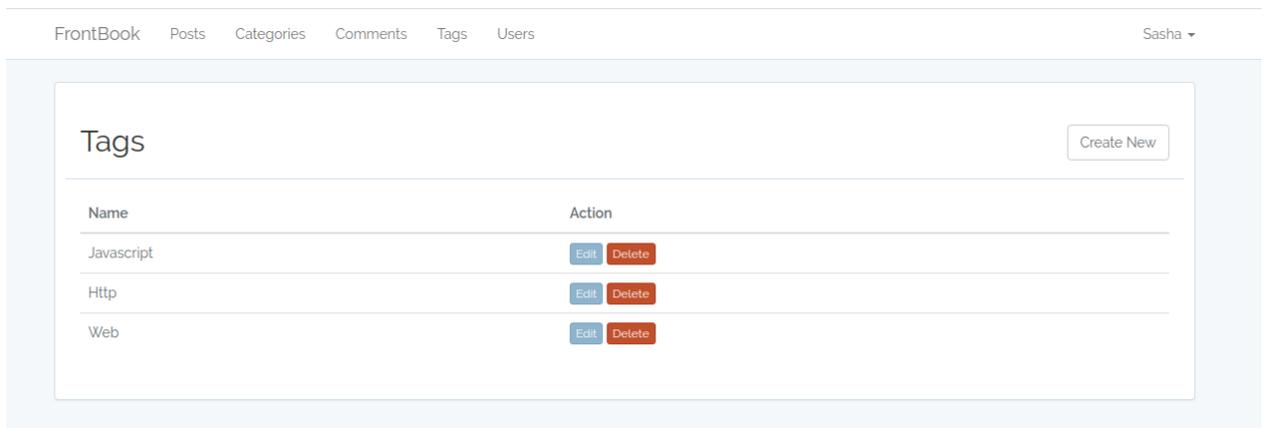


Рисунок 17 – Страница оценки статьи пользователем

В панели администратора отображается статистика по всем пользователям. Также администратор имеет возможность удалить пользователей.

Users

Name	Email	Admin?	No of Posts	Action
Nuruzzaman Milon	contact@milon.im	Yes	0	Delete
Max	contact@milon.im	Yes	11	Delete
Misha	misha@mail.ru	No	0	Delete
Sasha	sasha@mail.ru	Yes	0	Delete
Pasha	pasha@mail.ru	No	0	Delete

Рисунок 18 – Страница оценки статьи пользователем

4. ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ

Введение

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

Целью выполнения данного раздела является расчет затрат на создание информационной системы для оперативного обновления обучающей информации на портале.

4.1 Оценка коммерческого потенциала и перспективности проведения научных исследований с позиции ресурсоэффективности и ресурсосбережения

4.1.1 Потенциальные потребители результатов исследования

Конечным потребителем результатов исследований являются обычные пользователи интернета.

4.1.2 Исследование целесообразности вложения денежных средств в научно-исследовательский проект

Для определения целесообразности инвестирования в реализацию проведенных исследований была использована технология Quad (Quality Advisor). Для анализа результатов выделим ряд наиболее важных характеристик проектов такого рода и их показателей. Результаты проведенного сравнения выбранных показателей представлены в таблице 2.

Оценка качества и перспективности по технологии Quad определяется по формуле:

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

где P_{cp} – средневзвешенное значение показателя качества и перспективности научной разработки; V_i – вес показателя (в долях единицы); B_i – средневзвешенное значение i -го показателя. Значение P_{cp} позволяет

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

Таблица 2 – Оценочная карта для сравнения конкурентных технических решений (разработок)

Критерии оценки	Вес критерия (В)	Баллы (Б)	Макс. значение	Относит. значение	Средневзвешенное значение (P_{cp})
1	2	3	4	5 (3/4)	(5x2)
Показатели оценки качества разработки					
Надежность	0,09	70	100	0,7	6,3
Дизайн сайта	0,4	60	100	0,6	24
Потребность в ресурсах памяти	0,05	95	100	0,95	4,75
Функциональные возможности	0,02	60	100	0,6	1,2
Скорость работы	0,03	45	100	0,45	1,35
Показатели оценки коммерческого потенциала разработки					
Конкурентоспособность продукта	0,1	55	100	0,55	5,5
Уровень проникновения на рынок	0,04	10	100	0,1	0,4
Перспективность рынка	0,1	50	100	0,5	5
Цена	0,15	75	100	0,75	11,25
Послепродажное обслуживание	0,01	60	100	0,6	0,6
Наличие сертификации разработки	0,01	5	100	0,05	0,05
Итого	1	585	1100	5,85	60,4

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

4.1.3 SWOT-анализ

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

Составим итоговую матрицу SWOT-анализа, представленную в таблице 3.

Таблица 3 — Матрица SWOT-анализа

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

4.2 Определение возможных альтернатив проведения исследований

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

Таблица 4 – Альтернативные варианты проведения исследований

	1	2	3
А. Название систем	Drupal	Joomla	WordPress
Б. Возможность дополнения функционала	отлично	хорошо	нет
В. Удобство повседневного управления сайтом	хорошо	хорошо	хорошо

Можно составить несколько вариантов исполнения задачи:

- 1) АЗБ1В1
- 2) А1Б1В3
- 3) А2Б2В2

4.3 Планирование научно-исследовательских работ

4.3.1 Структура работ в рамках научного исследования

Планирование комплекса предполагаемых работ осуществляется в следующем порядке:

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

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

Таблица 5 – Перечень работ и распределение исполнителей

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

	8	Определение целесообразности проведения исследований	И	И – 100%
Заключительный этап	9	Составление отчета	И	НР – 60% И – 100%
	10	Защита ВКР	И	И – 100%

4.3.2 Определение трудоемкости выполнения работ

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

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

$$t_{ожі} = \frac{3t_{mini} + 2t_{maxі}}{5},$$

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

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

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

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

$$T_{pі} = \frac{t_{ожі}}{Ч_i},$$

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

$t_{ожі}$ – ожидаемая трудоемкость выполнения одной работы, чел.-дн.

$Ч_i$ – численность исполнителей, выполняющих одновременно одну и ту же работу на данном этапе, чел.

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

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

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

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

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

Коэффициент календарности определяется по следующей формуле:

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

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

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

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

В 2020 году 366 дней; из них 118 выходных и праздничных дней. Коэффициент календарности рассчитаем следующим образом.

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

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

Результаты расчетов сведем в таблицу 6.

На основе полученных данных, отраженных в таблице 6, строится календарный план график. Берется то исполнение, которое требует минимальное количество человеко-дней.

Таблица 6 – График проведения НИИ.

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

4.3.3 Составление календарного план-графика работ

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

Таблица 7 – Календарный план-график работ

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

4.3.4 Бюджет научно-технического исследования

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

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

4.3.4.1 Расчет материальных затрат НТИ

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

Таблица 8 – Материальные затраты

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

4.3.4.2 Расчет амортизационных расходов

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

Используется формула

$$C_{AM} = \frac{N_A * C_{OB} * t_{pф} * n}{F_D},$$

где N_A – годовая норма амортизации единицы оборудования;

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

F_D – действительный годовой фонд времени работы соответствующего оборудования.

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

n – число задействованных однотипных единиц оборудования.

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

Таблица 9 – Материальные затраты

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

4.3.4.3 Расчет затрат на электроэнергию

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

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

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

$Ц_{э}$ – тариф на 1 кВт·час;

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

Для ТПУ $Ц_{э} = 6,59$ руб./квт·час (с НДС).

Таблица 10 – Материальные затраты

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

4.3.4.4 Расчет основной и дополнительной заработной платы

В рамках реализации проекта руководитель и исполнитель получают заработную плату.

Расчет заработной платы производится исходя из действительного годового фонда рабочего времени (таблица 11), оклада работника и длительности работ по двум исполнениям.

Таблица 11 — Баланс рабочего времени

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

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

Основная заработная плата работника – произведение среднедневной зарплаты на количество рабочих дней, затраченных на реализацию проекта, по трём исполнениям. Результаты расчетов приведены в таблице 12.

Таблица 12 — Расчёт основной заработной платы

Исполнитель	$Z_{тс}$, руб.	Средняя ставка /раб. день	T_p , раб. дни	Коэффициент	$Z_{осн}$, руб.
Научный руководитель	33664	1342,09	20,7	1,699	47200

Инженер	15470	745,54	220	1,59	260789
---------	-------	--------	-----	------	--------

4.3.4.5 Отчисления во внебюджетные фонды

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

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

Результаты расчета отчислений во внебюджетные фонды сводится в таблицу 13.

Таблица 13 — Расчет отчислений во внебюджетные фонды

	З _{зп} , руб.	З _{внеб} , руб.
Научный руководитель	47200	14160
Инженер	260789	78236
Итого:		92396

4.3.4.6 Прочие расходы

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

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

$$C_{\text{проч.}} = 411312 \cdot 0,1 = 41131,2 \text{ руб.}$$

4.3.4.7 Формирование бюджета НИИ

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

Определение бюджета затрат на научно-исследовательский проект по каждому варианту исполнения приведен в таблице 14.

Таблица 14 – Формирование бюджета НИИ

Наименование статьи	Сумма, руб.
Материальные затраты НИИ	1000
Затраты по заработной плате исполнителей темы	307989
Затраты на амортизацию	8871
Затраты на электроэнергию	1056
Отчисления во внебюджетные фонды	92396
Накладные расходы	41131
Итого	452443

4.4 Определение ресурсной, финансовой и экономической эффективности исследования

Для расчета интегрального показателя ресурсоэффективности используется сравнительная характеристика вариантов исполнения проекта (таблица 15).

Таблица 15 — Сравнительная оценка характеристик вариантов исполнения проекта

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

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

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

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

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

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

Сравнительная эффективность для каждого исполнения рассчитывается по формуле:

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

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

Вывод

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

5. СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ

Введение

В данном разделе ВКР рассматриваются вопросы, касающиеся соблюдения санитарных норм и правил в процессе создания и использования информационной системы для оперативного обновления обучающей информации на портале. Рассматриваются меры по защите сотрудника от негативного воздействия среды. Исследуются вредные и опасные факторы среды, а также вопросы охраны окружающей среды от негативного воздействия. Рассматриваются возможные чрезвычайные ситуации и действия, которые сотрудник должен выполнить в случае возникновения ЧС. Созданная информационная система будет использоваться как обычными пользователями интернета, так и студентами, преподавателями в офисных помещениях.

Работы проводятся в учебно-научной лаборатории (далее – офис), находящейся в 108-й аудитории десятого корпуса Томского Политехнического университета. Предоставлено рабочее место за столом с креслом, ЖК монитором и системным блоком компьютера. Работа пользователя при использовании портала классифицируется как работа высокой точности.

Выделены и рассмотрены такие вредные факторы как: неоптимальный микроклимат, повышенный уровень шума, неправильное освещение, высокий уровень электромагнитного излучения. К опасным факторам при работе с компьютером относится высокое напряжение в электрической цепи и возможность короткого замыкания, влекущего за собой опасность поражения сотрудника электрическим током.

Рассматриваются вопросы правового регулирования трудовых отношений, связанных с использованием разработанной системы. Потенциальные причины и риски, а также конкретный набор мероприятий по их устранению, определяются спецификой выполняемых работ и априорными условиями труда (в частности, видом и состоянием рабочих мест исполнителей) [6].

5.1 Правовые и организационные вопросы обеспечения безопасности

5.1.1 Специальные правовые нормы трудового законодательства

Функции государственного надзора и контроля в организациях независимо от организационно-правовых форм и форм собственности осуществляются специально уполномоченными на то государственными органами и инспекциями согласно федеральным законам. В ТК РФ и СанПиН 2.2.2/2.4.1340-03:

- Продолжительность рабочего дня не должна превышать 40 часов в неделю.
- Продолжительность непрерывной работы за компьютером без регламентированного перерыва не должна превышать 1 час;
- Рекомендуется делать перерывы в работе за ПК продолжительностью 10-15 минут через каждые 45-60 минут работы;
- Во время регламентированных перерывов целесообразно выполнять комплексы упражнений и осуществлять проветривание помещения;

Не рекомендуется работать за компьютером более 6 часов за смену. Для того чтобы ПЭВМ соответствовали нормам, осуществляется производственный контроль и надзор внутри предприятия-производителя. Эксплуатирующие предприятия также следят за характеристиками используемой аппаратуры.

5.1.2 Организационные мероприятия при компоновке рабочей зоны

Большое значение для профилактики статических физических перегрузок имеет правильная организация рабочего места человека, работающего с ПЭВМ. Рабочее место должно быть организовано в соответствии с требованиями стандартов, технических условий и (или) методических указаний по безопасности труда. Оно должно удовлетворять следующим требованиям:

- обеспечивать возможность удобного выполнения работ;
- учитывать физическую тяжесть работ;
- учитывать размеры рабочей зоны и необходимость передвижения;
- учитывать технологические особенности процесса выполнения работ.

Невыполнение требований к расположению и компоновке рабочего места может привести к получению пользователем производственной травмы или развития у него профессионального заболевания. Рабочее место программиста должно соответствовать требованиям СанПин 2.2.2/2.4.1340-03.

Конструкция оборудования и рабочего места при выполнении работ в положении сидя должна обеспечивать оптимальное положение работающего, которое достигается регулированием высоты рабочей поверхности, высоты сидения, оборудованием пространства для размещения ног и высотой подставки для ног. Схемы размещения рабочих мест с персональными компьютерами должны учитывать расстояния между рабочими столами с мониторами: расстояние между боковыми поверхностями мониторов не менее 1,2 м, а расстояние между экраном монитора и тыльной частью другого монитора не менее 2,0 м. Клавиатура должна располагаться на поверхности стола на расстоянии 100-300 мм от края, обращенного к пользователю. Быстрое и точное считывание информации обеспечивается при расположении плоскости экрана ниже уровня глаз пользователя, предпочтительно перпендикулярно к нормальной линии взгляда (нормальная линия взгляда 15 градусов вниз от горизонтали). Рабочие места с ПЭВМ при выполнении творческой работы,

требующей значительного умственного напряжения или высокой концентрации внимания, рекомендуется изолировать друг от друга перегородками высотой 1,5 — 2,0 м.

5.2 Производственная безопасность

5.2.1 Освещенность рабочей зоны

Под освещенностью понимается отношение светового потока, падающего на элемент поверхности, к площади этого элемента. Обозначается освещенность буквой E , измеряется в люксах.

Недостаточная освещенность рабочего места увеличивает напряжение глаз сотрудников, что может привести к ухудшению зрения [8]. При плохой освещенности труднее различать цвета, возможно снижение способности к концентрации. Также недостаток света в помещении приводит к снижению уровня работоспособности, бодрости и ухудшает настроение сотрудников. Кроме того, низкое или чересчур пульсирующее освещение может способствовать появлению головных болей или мигреней [7]. К пульсации приводит, как правило, использование газоразрядных ламп, работающих на частоте 50 Гц.

Для комфортной работы сотрудника необходимо отсутствие пульсации света, обеспечение достаточной контрастности в цветопередаче монитора, отсутствие бликов на поверхностях офисного оборудования, а также соответствующее направление светового потока и его спектр [5].

Поскольку работа сотрудника офиса относится к работе высокой точности, необходимо, чтобы параметры освещенности рабочего места соответствовали требованиям СНиП 23-05-95, представленным в таблице 1.

Для производственных помещений всех назначений применяются системы общего равномерного освещения.

Источником света для общего освещения являются люминесцентные лампы белой цветности (ЛБ). Мощность (ЛБ) 40 Вт. Световой поток 2800 лм.

В светильники используется 2 люминесцентные лампы. Мощность лампы 40 Вт. Длина светильника 1228 мм. Ширина светильника 284 мм.

Параметры рабочего места инженера: $A = 10\text{ м}$ – длина помещения; $B = 5\text{ м}$ – ширина помещения; $H = 3\text{ м}$ – высота помещения; $S = 50\text{ м}^2$ – площадь помещения.

Размещение светильников в помещении определяется следующими параметрами:

$h_c = 0,2\text{ м}$ – расстояние светильников от перекрытия (свес);

$h_n = H - h_c = 3 - 0,2 = 2,8\text{ м}$ – высота светильника над полом, высота подвеса;

$h_{рп} = 0,8\text{ м}$ – высота рабочей поверхности над полом;

$h = h_n - h_{рп} = 2,8 - 0,8 = 2\text{ м}$ – расчётная высота, высота светильника над рабочей поверхностью.

L – расстояние между соседними светильниками или рядами.

План размещения светильников в рабочем помещении показан на рисунке 19.

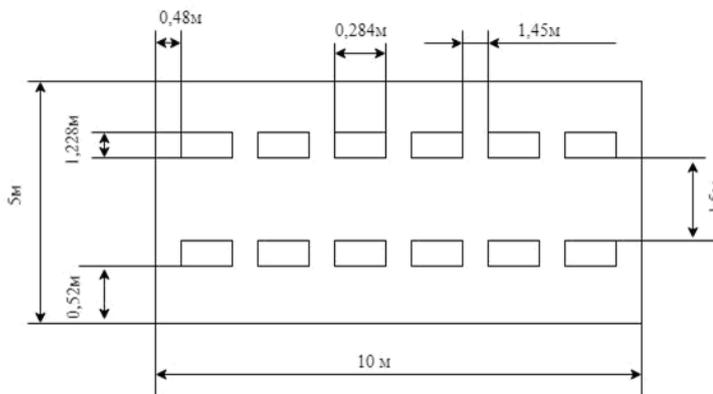


Рисунок 19 – План размещения светильников в помещении

Уровень необходимого освещения определяется степенью точности зрительных работ. В таблице 16 приведены требования к освещению помещений промышленных зданий согласно СП 52.13330.20 СНиП 23-05-95.

Таблица 16 — Требования к освещению помещений жилых и общественных зданий при зрительной работе высокой точности.

Характеристика зрительной работы	Наименьший или эквивалентный размер объекта различения, мм	Разряд зрительной работы	Подразряд зрительной работы	Относительная продолжительность зрительной работы при направлении зрения на рабочую поверхность, %	Искусственное освещение				Естественное освещение	
					Освещённость на рабочей поверхности от системы общего освещения, лк	Цилиндрическая освещённость, лк	Объединённый показатель UGR, не более	Коэффициент пульсации освещённости КП, %, не более	КЕО еН, %, при	
									Верхнем или комбинированном	Боковом
Высокой точности	От 0,3 до 0,5	Б	1	Не менее 70	400	100	18	15	3,0	1

Индекс помещения находится по формуле:

$$i = \frac{S}{h * (A + B)} = \frac{50}{2 * (10 + 5)} = 1,667$$

Индекс помещения $i = 1,667$. Согласно таблице 1.9.3 из пособия МГСН 2.06-99 коэффициент отражения побеленного потолка $R_n = 70\%$, коэффициент отражения для побеленных стен при незанавешенных окнах $R_c = 50\%$ [18].

В каждом ряду установлено 6 светильников ШОД-2-40. Учитывая, что в каждом светильнике установлено две лампы белой цветности (ЛБ). Мощность (ЛБ) 40 Вт. Световой поток 2800 лм., общее число ламп в помещении $N = 24$. Согласно таблице 11 из учебного пособия для соответствующих значений i , R_n , R_c для данного типа светильников ШОД η – коэффициент использования светового потока = 0,48[17].

По формуле определяем световой поток лампы:

$$\Phi = \frac{E_n S K_3 Z}{N \eta}$$

где Φ – световой поток лампы;

E_n – норма освещенности, 400 Лк;

S – площадь помещения, 50 м²;

K_3 – коэффициент запаса, 1,5 – для люминесцентных ламп;

Z – коэффициент неравномерности освещения, 1,1 – для люминесцентных ламп;

N – число ламп в помещении, 24;

$$\Phi = \frac{400 * 50 * 1,5 * 1,1}{24 * 0,48} = 2865 \text{ лм}$$

Делаем проверку выполнения условия:

$$-10\% \leq \frac{\Phi_{\text{расч}} - \Phi_{\text{лампы}}}{\Phi_{\text{расч}}} * 100\% \leq +20\%$$

$$\frac{2865 - 2800}{2865} * 100\% = 2,27\%$$

$$-10\% \leq 2,27\% \leq +20\%$$

Показатели в рабочем помещении соответствуют требованиям.

5.2.2 Микроклимат помещения

Микроклимат – это комплекс физических факторов внутренней среды помещений, оказывающий влияние на тепловой обмен организма и здоровье человека[7].

Воздействие комплекса микроклиматических факторов отражается на теплоощущении человека и обуславливает особенности физиологических реакций организма. Температурные воздействия, выходящие за пределы нейтральных колебаний, вызывают изменения тонуса мышц, периферических сосудов, деятельности потовых желез, теплопродукции. При этом постоянство теплового баланса достигается за счет значительного напряжения терморегуляции, что отрицательно сказывается на самочувствии, работоспособности человека, его состоянии здоровья.

К рассматриваемым параметрам микроклимата относятся: температура воздуха в помещении, температура поверхностей, относительная влажность воздуха, скорость движения воздуха.

В случае если пользователь испытывает дискомфорт, связанный со слишком высокой или низкой температурой, или влажностью, а также при высокой скорости движения воздуха, вероятно ослабление внимательности и концентрации, ухудшение настроения, снижение работоспособности.

Увеличивается риск возникновения простудных заболеваний, а также их последующее распространение в коллективе.

Работа офисного сотрудника относится к категории Ia, поскольку производится сидя и сопровождается незначительным физическим напряжением. Оптимальные значения параметров микроклимата производственных помещений для категории Ia содержатся в и показаны в таблице 17.

Таблица 17 — Оптимальные параметры микроклимата на рабочих местах производственных помещений

Период года	Категория работ по уровню энергозатрат, Вт	Температура воздуха, 0С	Относительная Влажность воздуха, %	Скорость движения воздуха, м/с
Холодный	Категория Ia(до 139)	23-25	40-60	0,1
Теплый	Категория Ia(до 139)	20-22	40-60	0,1

5.2.3 Уровень шума

Шум — это совокупность звуков, неблагоприятно воздействующих на организм человека и мешающих его работе и отдыху. При высоком уровне шума наблюдается снижение слуха у пользователей, снижение концентрации и работоспособности[9]. К источникам шума в офисе можно отнести разговоры других людей, звук работающего кондиционера, принтера, факса, системы охлаждения ПЭВМ, звук шагов, отодвигаемых стульев, шелест бумаг.

Шум характеризуется уровнем звукового давления для различных частот. Уровень шума на рабочем месте сотрудника не должен превышать описанные в нормы и правила. Предельный уровень звукового давления для работ, требующих высокого умственного напряжения, представлен в таблице 18.

Таблица 18 — Предельно допустимые уровни звукового давления, уровни звука и эквивалентные уровни звука для работ, требующих высокой степени внимания и концентрации.

Вид трудовой деятельности, рабочее место	Уровни звукового давления, дБ, в октавных полосах со среднегеометрическими частотами, Гц									Скорость движения воздуха, м/с
	31,5	63	125	250	500	1000	2000	4000	8000	
Творческая деятельность, руководящая работа с повышенными требованиями, научная деятельность, конструирование и проектирование, программирование, преподавание и обучение, врачебная деятельность. Рабочие места в помещениях дирекции, проектноконструкторских бюро, расчетчиков, программистов вычислительных машин, в лабораториях для теоретических работ и обработки данных.	86	71	61	54	49	45	42	40	38	50

Для защиты от шума используют звукоизоляцию помещений. К индивидуальным средствам защиты от шума относятся беруши, однако не представляется необходимым использовать их в офисе, поскольку уровень шума там, как правило, не превышает норму.

5.2.4 Электромагнитное излучение

Электромагнитным излучением называется распространяющееся в пространстве возмущение (изменение состояния) электромагнитного поля. Источником излучения в офисе могут стать системные блоки персональных компьютеров. Негативное воздействие электромагнитного излучения на организм человека проявляется в повышении температуры тела, а также может спровоцировать микропроцессы в организме, изменяющие свойства молекул [11]. Длительное воздействие повышенного уровня электромагнитного излучения повышает утомляемость, может вызвать гипертонию, катаракту,

изменения в крови, повышает сонливость. Соответствующие значения показаны в таблице 19.

Таблица 19 — ПДУ энергетических экспозиций ЭМП диапазона частот ≥ 30 кГц — 300 ГГц.

Диапазоны частот	Предельно допустимые уровни энергетической экспозиции		
	По электрической составляющей, (В/м) ² × ч	По магнитной составляющей, (А/м) ² × ч	По плотности потока энергии (мкВт/см ²) × ч
30 кГц — 3 МГц	20000,0	200	-
3 — 30 МГц	7000,0	-	-
30 — 50 МГц	800,0	0,72	-
50 — 300 МГц	800,0	-	-
300 МГц — 300 ГГц	-	-	200

Для снижения вредного воздействия ЭМП пользователю необходимо соблюдать расстояние от глаз до монитора от 0.5 до 1 метра. Также в целях уменьшения ЭМП рекомендуется пользоваться жидкокристаллическими мониторами вместо устаревших устройств с электронно-лучевой трубкой.

5.2.5 Электробезопасность

Электробезопасность – система организационных и технических мероприятий и средств, обеспечивающих защиту людей от вредного и опасного воздействия электрического тока, электрической дуги, электромагнитного поля и статического электричества [10].

При работе с компьютером возможно поражение электрическим током, что ведет к появлению ожогов, нагреву сосудов, механическим повреждениям тканей и сосудов, раздражающим воздействиям на ткани. Причиной поражения человека током может быть:

- Непосредственное прикосновение к токоведущим частям, оказавшимся под напряжением;
- Соприкосновение с конструктивными частями, оказавшимися под напряжением.
- Офисное помещение относится к категории помещений без повышенной опасности, однако необходимо соблюдать меры

предосторожности при работе с компьютером. Так, не рекомендуются следующие действия:

- Закладывать провода и шнуры за газовые и водопроводные трубы, за батареи отопительной системы;
- Выдергивать штепсельную вилку из розетки за шнур, усилие должно быть приложено к корпусу вилки;
- Работать на средствах вычислительной техники и периферийном оборудовании, имеющих нарушения целостности корпуса, нарушения изоляции проводов, неисправную индикацию включения питания, с признаками электрического напряжения на корпусе;
- Класть на средства вычислительной техники и периферийное оборудование посторонние предметы.

5.2.6 Обоснование мероприятий по защите окружающей среды

Отработанная техника утилизируется с уничтожением информации согласно ГОСТ Р 50739-95 «Средства вычислительной техники. Защита от несанкционированного доступа к информации» [12]. Отработанные люминесцентные лампы утилизируются согласно ГОСТ 30772-2001 «Ресурсосбережение. Обращение с отходами. Термины и определения» [13].

В случае выхода из строя они списываются и отправляются на специальный склад, который при необходимости принимает меры по утилизации списанной техники и комплектующих.

В целом, утилизация ламп предполагает передачу использованных ламп предприятиям – переработчикам, которые с помощью специального оборудования перерабатывают вредные лампы в безвредное сырье – сорбент, которое в последующем используют в качестве материала для производства, например тротуарной плитки.

Утилизация макулатуры должна производиться в соответствии с ГОСТ Р55090-2012 «Ресурсосбережение. Обращение с отходами. Рекомендации по утилизации отходов бумаги» [15] и ГОСТ 10700-97 «Макулатура бумажная и картонная. Технические условия» [16], согласно которым:

- макулатура должна разделяться на три группы: А – высокого качества, Б – среднего качества, В – низкого качества;
- макулатура каждой группы в зависимости от состава, источников поступления, цвета и способности к роспуску должна соответствовать маркам;
- макулатура должна содержать не более определённого количества примесей, определённого для каждой группы.

Все мероприятия по утилизации выполняются посредством сторонних компаний, имеющих соответствующие лицензии на такую деятельность (согласно ст. 17 Федерального закона от 8 августа 2001 г. №128-ФЗ «О лицензировании отдельных видов деятельности» [14]).

5.3 Экологическая безопасность.

5.3.1 Загрязнение атмосферного воздуха

Атмосфера всегда содержит определенное количество примесей, поступающих от естественных и других источников. К числу примесей, производимых естественными источниками, относят: пыль, туман, дым, газы от лесных и степных пожаров и др.

Основное загрязнение атмосферного воздуха производит отрасль промышленности.

Выполнение ВКР не осуществляет выбросов вредных веществ в атмосферу. Загрязнение атмосферного воздуха может возникнуть в случае возникновения пожара в помещении, в этом случае дым и газы от пожара будут являться антропогенным загрязнением атмосферного воздуха.

5.3.2 Загрязнение гидросферы

Сточная вода – это вода, бывшая в бытовом или производственном употреблении, а также прошедшая через какую-либо загрязненную территорию.

В ходе выполнения ВКР образовывались хозяйственно – бытовые воды. Бытовые сточные воды помещения образуются при эксплуатации туалетов, столовой, а также при мытье рук и проведении влажной уборки и т.п. Данные воды отправляются на городскую станцию очистки.

5.3.3 Отходы

Основные виды загрязнения литосферы – твердые бытовые и промышленные отходы, а также отходы возникающие в случае поломки ноутбука.

В ходе выполнения ВКР, образовывались различные твердые отходы. К ним можно отнести: бумагу, лампочки, отходы от продуктов питания и личной гигиены, отходы от канцелярских принадлежностей и т.д.

Защита почвенного покрова и недр от твердых отходов реализуется за счет сбора, сортирования и утилизации отходов и их организованного захоронения.

Главными нормативными актами, регулирующими вопрос утилизации ноутбуков, являются федеральные законы РФ «Об охране окружающей среды» и «Об отходах производства и потребления». А по ним вся оргтехника подлежит утилизации с соблюдением определенных правил: демонтаж запчастей, сортировка отходов и утилизация.

5.4 Безопасность в чрезвычайных ситуациях

В ходе работы за ПЭВМ может возникнуть чрезвычайная ситуация – пожар. Причинами пожаров могут быть:

- Игнорирование основных правил пожарной безопасности;
- Неисправность электрической проводки;
- Возгорание электроприборов — неисправных, самодельных или оставленных без присмотра;
- Курение в неположенных местах.

Были рассмотрены требования к системам противопожарной защиты, мероприятия по обеспечению пожарной безопасности.

Для защиты от пожаров необходимо иметь в наличии такое пожарное оборудование как пожарные шкафы, пожарные щиты и огнетушители. Сотрудники должны уметь пользоваться таким оборудованием. Углекислотные огнетушители применяются для ликвидации пожаров, вызванных возгоранием

электрооборудования. На рисунке 20 представлена принципиальная схема углекислотного огнетушителя.

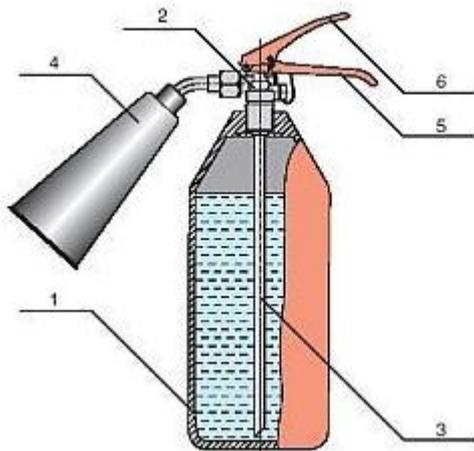


Рисунок 20 – Принципиальная схема ОУ.

На рисунке выше введены следующие обозначения:

- 1) Стальной баллон;
- 2) Запорно-пусковое устройство (ЗПУ);
- 3) Сифонная трубка;
- 4) Раструб;
- 5) Ручка для переноски огнетушителя;
- 6) Рычаг ЗПУ.

Для использования ОУ необходимо направить раструб на очаг возгорания и открыть ЗПУ.

Сотрудники должны знать план эвакуации из помещения, расположение выходов из здания. Также необходимо проводить плановые эвакуации из здания, для того чтобы подготовить сотрудников к действиям в чрезвычайной ситуации.

Заключение

В данной части работы изложены основные требования к рабочему месту инженера-программиста. Они должны создавать условия для комфортной работы. Выполнение условий, необходимых для оптимальной организации рабочего места инженера-программиста, позволяет в течение всего рабочего дня сохранять высокую работоспособность, повышать производительность труда работника.

В части экологии производителям и потребителям компьютерной техники необходимо серьезно задуматься о снижении вредного воздействия на окружающую среду посредством разработки регламентирующих документов по утилизации данного оборудования, отправки его в переработку, для возможного дальнейшего вторичного использования.

ЗАКЛЮЧЕНИЕ

В ходе выполнения ВКР была разработана информационная система для написания и публикации статей. Так же был развернут PredictionIO-сервис для реализации рекомендательной системы подбора статей на основе алгоритма ALS.

Благодаря данной ИС увеличивается качество подбора рекомендованных статей для пользователей. Такие системы уменьшают издержки поиска необходимых товаров и услуг, помогая пользователям быстрее найти нужную информацию. В сфере электронной коммерции рекомендательные системы увеличивают прибыль, так как они являются эффективным средством продажи большего количества товаров и услуг.

В ходе выполнения ВКР были продемонстрированы результаты работы информационной системы. Так же были рассмотрены основные алгоритмы рекомендательных систем, такие как: Summary-based, Content-based и User-based алгоритмы. Были использованы два вида алгоритмов: Summary-based для неаутентифицированных и User-based для аутентифицированных пользователей.

CONCLUSION

Thus, during the master's thesis an information system was developed for writing and publishing articles. PredictionIO was also deployed to implement a recommendation system for selecting articles based on the ALS algorithm.

This information system improves the quality of the selection of recommended articles for users. Such systems reduce the cost of searching for the necessary goods and services, helping users find the necessary information faster. In e-Commerce, recommendation systems increase profits because they are an effective means of selling more products and services.

During the master's thesis, the results of the information system were demonstrated. The main algorithms of recommendation systems, such as Summary-based, Content-based, and User-based algorithms, were also considered. Two types of algorithms were used: Summary-based for non-authenticated users and User-based for authenticated users.

Список использованных источников

1. Deep learning concepts — PART 1 — Towards Data Science [Электронный ресурс]: информационный сайт – URL: <https://towardsdatascience.com/prototyping-a-recommender-system-step-by-step-part-2-alternating-least-square-als-matrix-4a76c58714a1> (дата обращения 12.05.2020).
2. Deep learning concepts — PART 2 — Towards Data Science [Электронный ресурс]: информационный сайт – URL: <https://towardsdatascience.com/prototyping-a-recommender-system-step-by-step-part-1-knn-item-based-collaborative-filtering-637969614ea> (дата обращения 12.05.2020).
3. ALS Implicit Collaborative Filtering — Rn Engineering — Medium [Электронный ресурс]: информационный сайт – URL: <https://medium.com/radon-dev/als-implicit-collaborative-filtering-5ed653ba39fe> (дата обращения 12.05.2020).
4. Welcome to Apache PredictionIO – [Электронный ресурс]: официальный сайт PredictionIO – URL: <http://predictionio.apache.org/> (дата обращения 12.05.2020).
5. ГОСТ 12.0.003-74 (с измен. №1, октябрь 1978г., переиздание 1999 г.) «Классификация вредных и опасных производственных факторов».
6. Безопасность жизнедеятельности: Учебник для вузов /Под ред. К.З. Ушакова. – М.: Изд-во Московского гос. горного университета, 2000.– 430 с.
7. СанПиН 2.2.4.548 – 96. «Гигиенические требования к микроклимату производственных помещений».
8. СНиП 23-05-95. «Естественное и искусственное освещение».
9. СН 2.2.4/2.1.8.562-96 «Шум на рабочих местах, в помещениях жилых, общественных зданий и на территории жилой, застройки».
10. ГОСТ 12.1.009-76 «Электробезопасность. Термины и определения».
11. СанПиН 2.2.4.1191-03. «Электромагнитные поля в производственных условиях».

12. ГОСТ Р 50739-95 «Средства вычислительной техники. Защита от несанкционированного доступа к информации».
13. ГОСТ 30772-2001 «Ресурсосбережение. Обращение с отходами. Термины и определения».
14. Федеральный закон от 8 августа 2001 г. №128-ФЗ «О лицензировании отдельных видов деятельности».
15. ГОСТ Р 55090-2012 «Ресурсосбережение. Обращение с отходами. Рекомендации по утилизации отходов бумаги».
16. ГОСТ 10700-97 «Макулатура бумажная и картонная. Технические условия».
17. «Безопасность жизнедеятельности. Практикум: учебное пособие по выполнению индивидуальных заданий для студентов всех специальностей» [Электронный ресурс] – Режим доступа http://window.edu.ru/catalog/pdf2txt/965/73965/53126?p_page=6 (дата обращения 24.05.2020).
18. «Пособие к МГСН 2.06-99. Расчет и проектирование искусственного освещения помещений общественных зданий» [Электронный ресурс] – Режим доступа <https://meganorm.ru/Data2/1/4294846/4294846582.htm> (дата обращения 25.05.2020).
19. Tryolabs – интернет портал для статей [Электронный ресурс] – Режим доступа: <https://tryolabs.com/blog/introduction-to-recommender-systems/> (дата обращения: 24.05.2020).
20. KDnuggets – информационный портал [Электронный ресурс] – Режим доступа: <https://www.kdnuggets.com/2019/09/machine-learning-recommender-systems.html> (дата обращения: 24.05.2020).
21. G. Adomavicius and A. Tuzhilin, “Towards the next generation of recommender systems: a survey of the state-of-the-art and possible extensions”, 2005.
22. G. Linden, B. Smith, and J. York, “Amazon.com recommendations: itemto-item collaborative filtering”, 2003.

Приложение А
(справочное)

Information system development for learning portal operational updating

Студент

Группа	ФИО	Подпись	Дата
8ВМ81	Ахмеров М.Ш.		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ	Саврасов Ф.В.	к.т.н.		

Консультант-лингвист отделения иностранных языков ШБИП

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИЯ	Сидоренко Т.В.	к.п.н		

The graduation work contains x pages, x figure, x tables, x source and x applications.

Keywords: an information system, a web-development, a recommendation system, a matrix factorization.

The object of the research is the algorithms that are used in recommendation systems for selecting articles.

A purpose of study: to develop the information system for writing and publishing articles and courses. Also in this system it is necessary to introduce the recommendation system for articles.

During the research it is necessary to study the main types of recommendation systems, as well as the basic algorithms for working with them.

As a result of the study, this ready-made information system was developed for writing and publishing articles. The recommendation system was also implemented based on collaborative filtering.

The main design, technological and technical and operational characteristics: a web application that allows users to read or add new articles to the system.

Extent of implementation: the application will be located on the global Internet.

Scope: web application is intended for the publication of articles on programming.

The significance of the work lies in the development of the information system for Internet users. The quality of the selection of recommended articles for users will increase with this information system.

As a further development of the work, it is possible to optimize the algorithm, increasing the efficiency of the algorithm in order to increase the accuracy of the selection of articles.

Introduction

Currently, there is a problem with a large amount of information on the Internet. To solve this problem, you need to filter, prioritize, and deliver information efficiently. Recommendation systems search for the necessary information in a large volume of dynamically generated content to provide users with personalized products and services. Information search engines such as Google and Yandex partially solve this problem by providing users with relevant content, but not in the way that recommendation systems do. For example, most of the revenue, about 75%, Netflix receives from the sale of recommended products. For Amazon, this percentage is 35%. Both customers and its owners benefit with recommendation systems.

Recommendation systems are useful for both clients and service providers. They reduce the cost of searching for the necessary products and services, helping users find the information they need faster. In e-Commerce, recommendation systems increase profits because they are an effective means of selling more products and services.

1. Analysis of subject area

1.1 Recommendation system

Recommendation system — a program that tries to predict which objects (articles, music, books, news, movies) will be interesting to the user. Recommendation systems work on the basis of information about the user.

A recommendation system is defined as a decision-making strategy for users in complex information environments. In addition, the recommendation system has been defined from an e-Commerce perspective as a tool that helps users search for knowledge records that are related to user interests and preferences. The recommendation system has been defined as a means of helping and expanding the social process of using other people's recommendations to make decisions when there is not enough personal knowledge or experience of alternatives. Recently, various approaches have been developed to build recommendation systems that can use either joint filtering, content filtering, or hybrid filtering. The joint filtration technique is the most Mature and most frequently used. Collaborative filtering recommends elements by identifying other users with similar tastes. Content filtering methods usually make predictions based on the user's interests, and they do not take into account other users, as in the case of collaboration methods.

However, there are limitations, despite the success of these two filtering methods. The main problems are related to content filtering features, such as the lack of necessary data and limited content analysis. In addition, collaborative approaches reveal problems of cold start, rarity, and scalability. These problems usually reduce the quality of recommendations. To alleviate some of the identified problems, hybrid filtering has been proposed, which combines two or more filtering methods in different ways to improve the accuracy and performance of recommendation systems. These methods combine two or more filtering approaches to exploit their strengths by aligning their respective weaknesses. Currently, collaborative filtering and content filtering approaches are widely used, which implement content-based and collaboration-based methods in different ways, and then combine the results of their prediction or add content filtering characteristics to collaborative filtering and Vice versa.

The system collects relevant user information to generate a user profile or model for forecasting tasks, including user attributes, behavior, or the content of resources that the user accesses.

Recommendation system cannot work correctly until the user profile or model is built correctly. The system needs to know full information of a user to give a reasonable recommendation from the start. Recommendation systems rely on various types of input, such as explicit feedback, which includes explicit assessment from users regarding their interest in a subject, or implicit feedback by indirectly determining user preferences by observing user behavior.

1.2 Types of user ratings

1.2.1 Explicit estimate

The system usually offers the user an evaluation option for elements to build and improve their model. The accuracy of the recommendation depends on the number and accuracy of ratings provided by the user. The only drawback is that there are not enough evaluations to make recommendations. Explicit user assessment is more reliable data, since it does not involve to extract user interests from actions, and it also makes recommendation process more clear, resulting in a slightly higher perceived quality of recommendations, but it requires more actions from users. However, explicit feedback requires more effort from the user, since the user must independently set a rating.

Ваш отзыв

Понравилась ли Вам еда? **10** Как Вам интерьер? **10** Как Вам обслуживание? **10** Ваша оценка **10**

Потрясающее место, очень уютный интерьер, очень довольны.

+ Уютный интерьер, расположение ресторана очень удачное, отличная кухня, есть места для парковки.

- Нет

Добавить фото Опубликовать

Picture 1 — Explicit evaluation.

1.2.2 Implicit evaluation

The system independently calculates the user's interests by tracking various user actions in the system, such as selecting a specific article or when the user has read the article to the end. Implicit feedback decreases the load on users by extracting their user interests from their behavior in the system. The method, however, does not require any effort on the part of the user, but it is less accurate. There is also an opinion that implicit evaluation can be more objective, since there is no bias arising from the fact that users can react in a socially desirable way.

1.2.3 Hybrid evaluation

The advantages of both implicit and explicit feedback can be combined in a hybrid system to minimize their weaknesses and get the most efficient system [22]. This can be achieved by using implicit data as a test of explicit evaluation, or by allowing the user to give explicit feedback only when they decide to express explicit interest.

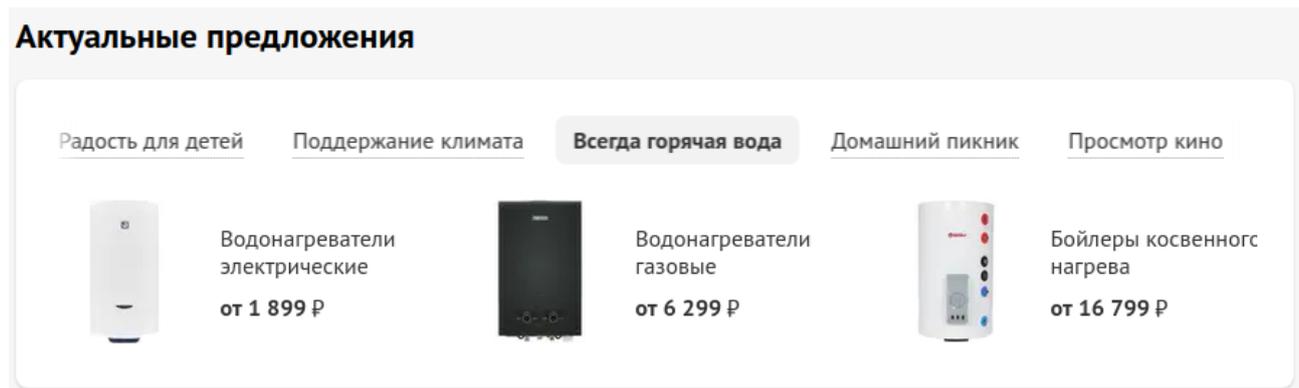
1.3 Algorithms for recommendation systems

There are many algorithms for recommendation systems. The most popular are: Summary-based (not personal), Content-based (models based on the product description), User-based (models based on the interests of other users).

1.3.1 Summary-based algorithms

Nonpersonalized recommendations are the easiest to implement. The user's potential interest is calculated based on the average product rating. If all users like it, then the new user also likes it. This principle is used by a large number of services when the user is not authenticated in the system.

The product rating can be displayed in different ways. For example, it can be asterisks next to the product, the number of likes, or the difference between positive and negative votes (as is usually done on forums).



Picture 2— summary-based recommendation system.

1.3.2 Content-based algorithms

Personal recommendations are based on the use of information about the user, primarily about their previous actions, purchases, and views. In this approach, the user's interests are compared with the description of the products and services provided [2]. The more a service or product matches these interests, the higher the user's potential interest is evaluated. However, it follows that all products and services must have a description.

The main subject of Content-based recommendations was more often products with unstructured descriptions: articles, books, and movies. For example, for movies, it can be a short description, the cast, and for articles, it can be tags, categories, or a description.

Unstructured information can be described by vectors in space (Vector-Space model). Each element of this vector is a feature that represents the user's interests. The entity itself, such as an article or movie, are vectors.

As the user interacts with the system (for example, he is viewing a specific article), the vector features of the viewed articles are combined (summarized and

normalized) into a single vector and, as a result, the vector of his interests is formed. Next, it is enough to find an article whose vector is close to the user's interests vector, i.e. solve the problem of finding the nearest neighbors.

The source of information that content filtering systems are most often used with is text documents. The standard approach to parsing a term selects individual words from documents [19]. The vector space model and hidden semantic indexing are two methods that use these terms to represent documents as vectors in a multidimensional space.

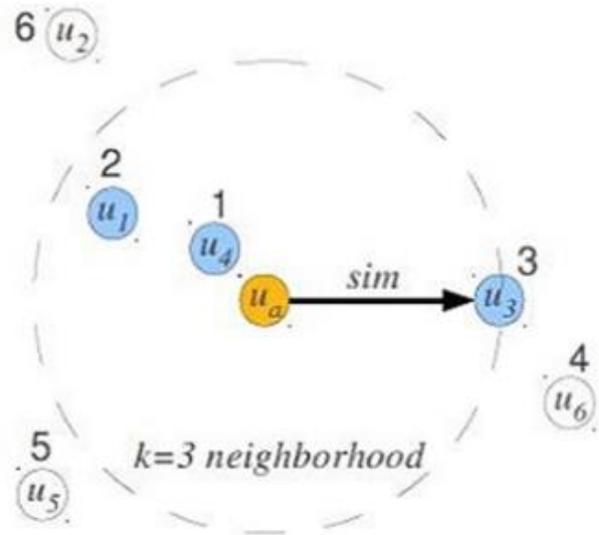
Relevant feedback, genetic algorithms, neural networks, and Bayesian classifier are some of the training methods for learning the user profile. The vector space model and hidden semantic indexing can be used by these training methods to represent documents. Some of the training methods also represent the user profile as one or more vectors in the same multidimensional space, which makes it easy to compare documents and profiles. Other learning methods, such as the Bayesian classifier and neural networks, do not use this space, but represent the user profile in their own way[22].

1.3.3 User-based algorithms

This type of system began to develop actively in the early 90's. In this approach, recommendations are selected based on the interests of other similar users. Such recommendations are calculated thanks to the "collaboration" of many users. Hence the name of the method.

The approach is based on the K nearest neighbor algorithm. On the fingers – for each user, we search for the k most similar to them (based on their preferences) and Supplement the information about the user with known data on their neighbors.

	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8
u_1	?	4.0	4.0	2.0	1.0	2.0	?	?
u_2	3.0	?	?	?	5.0	1.0	?	?
u_3	3.0	?	?	3.0	2.0	2.0	?	3.0
u_4	4.0	?	?	2.0	1.0	1.0	2.0	4.0
u_5	1.0	1.0	?	?	?	?	?	1.0
u_6	?	1.0	?	?	1.0	1.0	?	1.0
u_a	?	?	4.0	3.0	?	1.0	?	5.0
r_a	3.5	4.0			1.3		2.0	



Picture 3 — knn algorithm.

In the image above, the yellow color indicates the user to find recommendations for. Blue color indicates 3 nearest neighbors, whose articles will be offered as recommendations.

The similarity of interests can be determined by the cosine similarity formula.

$$similarity = \cos(\theta) = \frac{A * B}{\|A\| * \|B\|} = \frac{\sum_{i=1}^n A_i * B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} * \sqrt{\sum_{i=1}^n (B_i)^2}}$$

There are also other algorithms for searching for nearest neighbors, such as the Pearson correlation, the jacquard distance, and the Hamming distance. However, these methods do not handle data sparseness well when none of the neighbors have evaluated the item you are trying to predict for the target user. In addition, it is computationally inefficient, as the number of users and products increases. Collaborative filtering systems recommend items based on similarity measures between users and/or items[21].