Министерство науки и высшего образования Российской Федерации Федеральное государственное бюджетное образовательное учреждение высшего образования

ИРКУТСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

Институт информационных технологий и анализа данных

Допускаю к защите

Руководитель Л. С. Вахрушева

подпись, И.О. Фамилия

тема

Наименование темы

**Пояснительная записка к курсовому проекту**

по дисциплине

**«****Технологии разработки программных комплексов»**

1.008.00.00 ПЗ

обозначение документа

|  |  |  |
| --- | --- | --- |
| Разработал студент группы | подпись | М. О. Васильева  И. О. Фамилия |
| Нормоконтроль |  | Л. С. Вахрушева |
|  | подпись | И. О. Фамилия |

Курсовой проект защищен с оценкой

**Иркутск 2023**

Министерство науки и высшего образования Российской Федерации Федеральное государственное бюджетное образовательное учреждение высшего образования

ИРКУТСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

**ЗАДАНИЕ**

НА КУРСОВОЕ ПРОЕКТИРОВАНИЕ

По курсу Технологии разработки программных комплексов

Студенту Васильевой М. О.

(фамилия, инициалы)

Тема проекта Проектирование базы данных и разработка приложения для работы с ней

Рекомендуемая литература

1. Кузин А.В. Базы данных : учеб. пособие по направлению подгот. дипломир. специалистов "Информа- тика и вычисл. техника" / А. В. Кузин, С. В. Левонисова. - 4-е изд., стер.. - М. : Академия, 2010. - 311 с.
2. Базы данных : учеб. пособие / А. С. Дорофеев; Иркут. гос. техн. ун-т . – Иркутск: Изд-во ИрГТУ, 2008. – 99 с. : a-ил.
3. Базы данных : учебник для прикладного бакалавриата / Б. Я. Советов, В. В. Цехановский, В. Д. Чер- товской; С.-Петерб. гос. электротехн. ун-т. – 2-е изд. . – Москва: Юрайт, 2015. – 462 с.
4. Хомоненко А.Д. Базы данных: Учебник для ВУЗов / Хомоненко А.Д. Изд-во: "Корона-Век", 2010.
5. Дорофеев А.С. Базы данных (09.03.01) для набора с 2019 г. [Электронный ресурс]. [2021]. URL: https://el.istu.edu/course/view.php?id=5192 (дата обращения: 25.10.2021).
6. Разработка приложений баз данных на основе современных технологий : учебное пособие для вузов по направлению "Конструкторско-технологическое обеспечение машиностроительных производств" / А. С. Дорофеев [и др.]. - Старый Оскол : ТНТ, 2020. - 275 с.
7. Ставров, С. Г. Практикум по работе с базами данных в Microsoft Visio и СУБД Microsoft SQL Server : учебное пособие / С. Г. Ставров, А. Е. Кочетков. — Иваново : ИГЭУ, 2018. — 80 с. — Текст : электрон- ный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/154589> (дата обра- щения: 22.09.2021).
8. Волк, В. К. Базы данных : учебное пособие / В. К. Волк. — Курган : КГУ, 2018 — Часть 1 : Проекти- рование и программирование — 2018. — 178 с. — ISBN 978-5-4217-0472-0. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/177903 (дата обращения: 22.09.2021).

Графическая часть на листах.

Дата выдачи задания ―25‖ октября 2021 г.

Задание получил

подпись И.О. Фамилия

Дата представления проекта руководителю ―15‖ декабря 2021 г.

Руководитель курсового проектирования Л. С. Вахрушева

подпись И.О. Фамилия

**Содержание**

[**«Технологии разработки программных комплексов»** 1](#_Toc153711286)

[**ЗАДАНИЕ** 2](#_Toc153711287)

[Введение 4](#_Toc153711288)

[1 Анализ 5](#_Toc153711289)

[1.1 Описание предметной области 5](#_Toc153711290)

[1.2 Обзор существующих программных средств 6](#_Toc153711291)

[1.3 Процесс TO BE 7](#_Toc153711292)

[1.4 Описание вариантов использования 8](#_Toc153711293)

[1.5 Выработка требований и постановка задачи 8](#_Toc153711294)

[2 Проектирование 9](#_Toc153711295)

[2.1 Выбор и обоснование средств проектирования и реализации 9](#_Toc153711296)

[2.2 Проектирование архитектуры приложения 9](#_Toc153711297)

[2.3 Проектирование хранилища данных 10](#_Toc153711298)

[3 Разработка 11](#_Toc153711299)

[4 Тестирование 12](#_Toc153711300)

[Заключение 13](#_Toc153711301)

[Список использованных источников 14](#_Toc153711302)

Введение

1. Анализ
   1. Описание предметной области

Рекомендательная система, также известная как система рекомендаций, - это информационная технология, разработанная для предоставления персонализированных рекомендаций пользователю. Основная цель рекомендательных систем - помочь пользователям открывать новые продукты, контент или услуги, которые могли бы заинтересовать их, на основе их предпочтений, истории взаимодействия и характеристик.

Рекомендательные системы широко применяются в различных сферах, включая электронную коммерцию, социальные сети, медиа, стриминговые платформы, онлайн-рекламу и многое другое. Они обеспечивают следующие основные функции:

* Фильтрация контента: Рекомендательные системы помогают пользователям избежать перебора огромного объема доступного контента, предоставляя им релевантные и интересные варианты.
* Персонализация: Системы анализируют предпочтения пользователя, учитывая их историю взаимодействия, оценки, покупки, просмотры и другие данные, чтобы предоставлять рекомендации, которые соответствуют его индивидуальным интересам.
* Прогнозирование: Рекомендательные системы используют алгоритмы и методы машинного обучения для прогнозирования, какие продукты или контент пользователь, скорее всего, предпочтет в будущем.
* Увеличение удержания клиентов: Путем предоставления релевантных рекомендаций системы способствуют удержанию пользователей и увеличению их вовлеченности.
* Кросс-продажи: Рекомендательные системы могут предлагать дополнительные продукты или услуги, которые могут заинтересовать пользователя, основываясь на его текущих предпочтениях.
* Улучшение пользовательского опыта: Путем предоставления более удовлетворительного и персонализированного опыта, рекомендательные системы способствуют увеличению удовлетворенности пользователей.

Рекомендательные системы могут использовать различные методы и алгоритмы, включая коллаборативную фильтрацию, контент-базированные методы, гибридные подходы и нейронные сети для анализа данных и создания рекомендаций. Важно, чтобы такие системы учитывали конфиденциальность данных пользователей и обеспечивали соответствие законодательству о защите данных.

**Целью** проекта является сокращение времени на подбор книг для дальнейшего чтения.

Для достижения данной цели необходимо выполнить следующие задачи:

1. выполнить анализ предметной области;
2. изучить работу рекомендательных систем и различные алгоритмы для их реализации;
3. изучить разные способы решения задач обработки естественного языка (NLP);
4. реализовать БД для хранения информации, необходимой для обучения модели;
5. реализовать рекомендательную систему:

* сформировать исходную выборку данных;
* обработать и нормализовать исходную выборку;
* разделить исходную выборку на тестовую и обучающую;
* определить тип рекомендательной системы;
* настроить параметры модели и алгоритм её обучения;
* получить модель;
* протестировать модель.

**Проблема**

Множество читателей сталкиваются с трудностью выбора книги из-за избытка информации и разнообразия литературных предложений. Это может привести к разочарованию в чтении. Наш проект целенаправленно решает эту проблему, предоставляя персонализированные рекомендации на основе интересов и предпочтений читателей, что облегчает поиск книг, способствует их лучшему использованию и создает более удовлетворительный литературный опыт.

**Актуальность**

В современном мире, на фоне огромного количества доступных книг и информационного шума, пользователи сталкиваются с серьезными трудностями в выборе литературы для чтения. Наш проект предоставляет рекомендательную систему, основанную на современных методах машинного обучения, которая персонализирует рекомендации для каждого пользователя. Мы решаем эту актуальную проблему, помогая людям находить книги, которые идеально соответствуют их интересам и предпочтениям.

* 1. Обзор существующих программных средств

Ниже перечислены некоторые существующие программные средства для подбора книг на основе прочтения пользователей:

* Goodreads: Это одна из самых популярных платформ для чтения и обсуждения книг. Goodreads предлагает пользователю возможность оценивать прочитанные им книги, смотреть рекомендации на основе предпочтений и интересов других пользователей, а также получать персональные рекомендации от системы;
* Amazon: Электронная коммерция Amazon предлагает функцию подбора книг на основе предыдущих покупок и интересов пользователя. Они используют алгоритмы машинного обучения, чтобы анализировать прочитанные книги и выдавать рекомендации, основанные на этом анализе;
* Bookmate: Это приложение для чтения электронных книг с функцией индивидуального подбора книг. Bookmate использует алгоритмы машинного обучения и анализирует прочитанные пользователем книги, чтобы предложить персонализированные рекомендации и новые книги, которые могут заинтересовать пользователя;
* LibraryThing: Платформа LibraryThing предлагает пользователям возможность вести списки прочитанных книг, оценивать их и получать рекомендации. Рекомендации основаны на анализе прочитанных книг и интересов других пользователей;
* BookBub: Это сервис, который предлагает персонализированные ежедневные предложения по книгам, основанные на предпочтениях и интересах пользователя. BookBub анализирует прочитанные книги, чтобы предложить новые заголовки, которые могут понравиться пользователю.

Это самые популярные из существующих программных средств для подбора книг на основе прочтения пользователей. Все из них являются зарубежными продуктами, ориентированные на англоязычных читателей. При условии существования большого массива данных оценок книг русскоязычныъ читателей, нашу систему можно будет легко переобучить и получать более релевантные рекомендации.

* 1. Процесс TO BE

На рисунке 1 изображена схема процесса TO BE проектируемой системы.

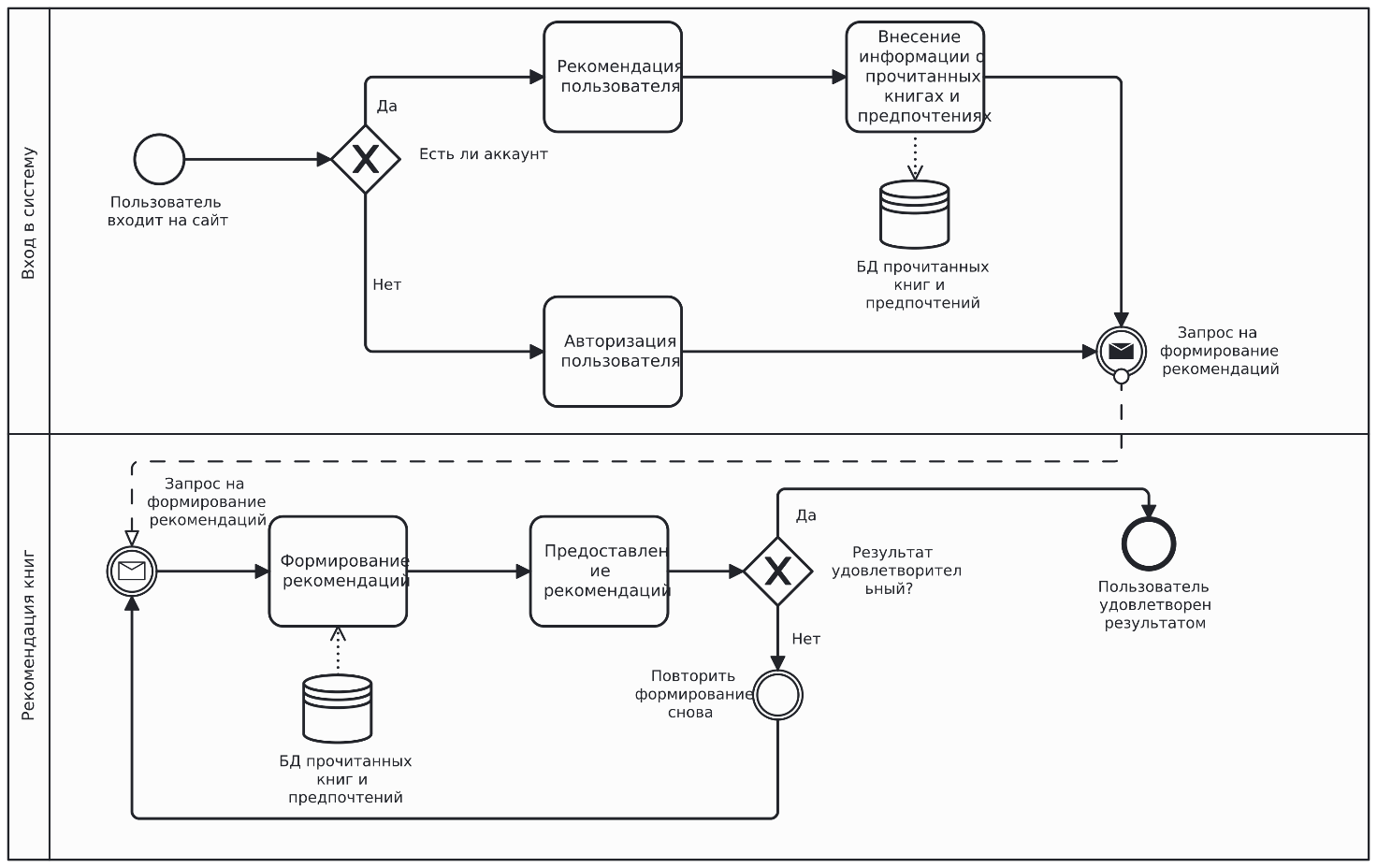


Рисунок 1 - Диаграмма TO BE

Когда пользователь заходит на сайт, ему надо либо авторизоваться в системе, либо зарегистрироваться. Если пользователь зарегистрирован в системе, то пользователь сразу переходит к формированию рекомендаций. Иначе новый пользователь вносит информацию о прочитанных книгах и предпочтениях, а после этого отправляется запрос на формирование рекомендаций.

Формирование рекомендаций происходит до тех пор, пока пользователь не будет удовлетворен результатом.

* 1. Описание вариантов использования

На рисунке 2 представлена диаграмма прецедентов для разрабатываемой системы.

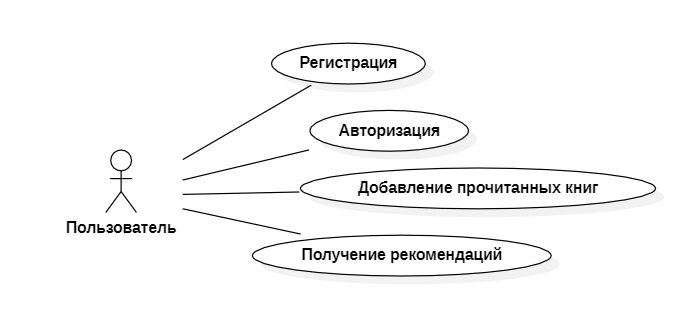


Рисунок 2 - Диаграмма прецедентов

* 1. Выработка требований и постановка задачи

Задачей является разработка системы для хранения информации о прочитанных пользователем книг и формирования рекомендаций на их основе с помощью модели, решающей задачу кластеризации.

К системе предъявляются следующие требования:

* регистрация новых пользователей;
* авторизация пользователей;
* ввод информации о прочитанных книгах;
* хранение информации о прочитанных книгах;
* формирование и выдача рекомендаций.

1. Проектирование
   1. Выбор и обоснование средств проектирования и реализации

Чтобы реализовать рекомендательную систему будет использоваться библиотека основанная на Python Surprise.

Surprise - это мощный инструмент, предоставляющий широкий набор алгоритмов и инструментов для построения и оценки рекомендательных систем.

Так же использование Python упрощает взаимодействие с другими частями проекта.

* 1. Проектирование архитектуры приложения

Рекомендательные системы относятся к классу алгоритмов, где обучение происходит "без учителя", и решают задачу кластеризации.

Для решения поставленных задач отлично подходят коллаборативные фильтрации, основанные на пользователе. Для создания модели необходимо определить несколько параметров, таких как модель алгоритма и метрика оценки.

Будет использован метод ближайшего соседа с учетом средних оценок пользователей. Метод основан на идее использования информации о предпочтениях ближайших соседей для формирования рекомендаций. Он оценивает оценки (рейтинги) элементов для пользователя на основе оценок его "ближайших соседей" - пользователей, у которых похожие предпочтения.

Для измерения точности прогноза будут использоваться метрики Mean Absolute Error (MAE) и Root Mean Square Error (RMSE), которые определяют насколько близко система предсказывает рейтинг, который пользователь дал элементу.

На рисунке 3 изображен пример реализации рекомендательной системы.

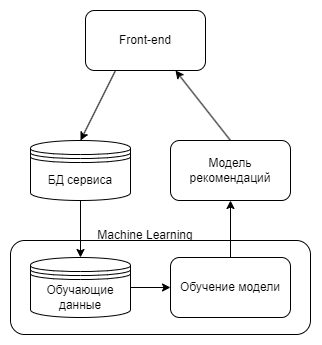


Рисунок 3 - Архитектура сервиса

Через формы интерфейса, с которыми взаимодействует пользователь, в БД сервиса поступают данные. На их основе формируется матрица товарных предпочтений. Далее с помощью матрицы проходит обучение модели. В обученную модель подается идентификатор пользователя и формируется заданное количество рекомендованных книг. Полученные рекомендации отсылаются на формы.

* 1. Проектирование хранилища данных

На рисунке 4 представлена логическая модель базы данных для всего сервиса.

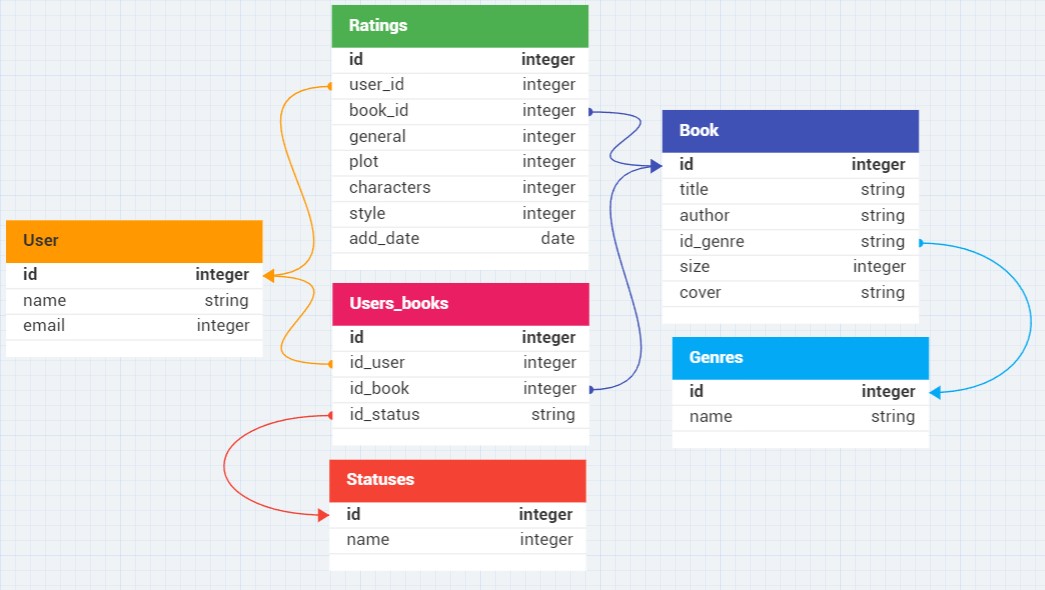


Рисунок 4 - Логическая модель БД сервиса

Для обучения модели необходимы данные из таблицы "Ratings", которые и будут преобразованы в матрицу товарных предпочтений.

Так как на данный момент в системе отсутствует необходимый объем данных, будут использоваться данные, собранные во время бук-кроссинга, в которые входит 62 657 оценок от 1 295 пользователей на 14 684 книги.

1. Разработка

В начале загрузим данные в датафрейм, как показано на рисунке 5.

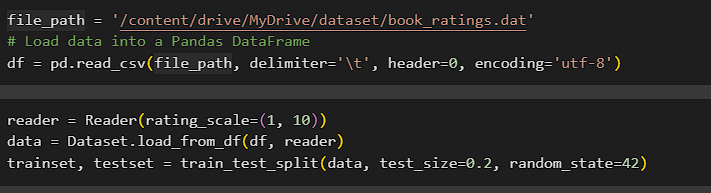


Рисунок 5 - Загрузка данных

Далее укажем в параметрах модели, что сходство считается между пользователями, а его мерой выбран косинусный коэффициент. На рисунке 6 указан массив параметров.

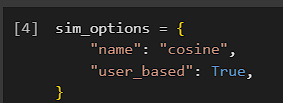


Рисунок 6 - Массив параметров для обучения

На рисунке 7 показаны команды для подготовки обучающей и тестовой выборки и обучения модели. Получим по 5 рекомендаций для каждого пользователя.

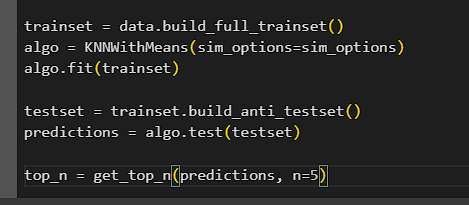


Рисунок 7 - Обучение модели

Посмотрим результаты на примере одного пользователя. Для удобства добавим в вывод названия рекомендуемых книг. На рисунке 8 показаны рекомендации пользователя с идентификатором 2.

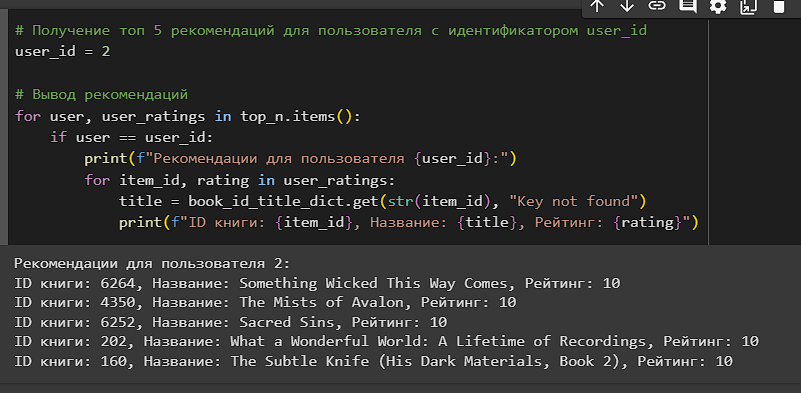


Рисунок 8 - Получение рекомендаций

1. Тестирование

Для расчёта эффективности моделей часто используют следующие метрики:

* средняя асолютная ошибка (MAE);
* среднеквадратическая ошибка (RMSE).

Чем ниже их значения, тем лучше лучше модель соответствует набору данных.

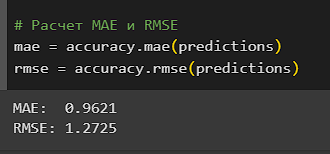


Рисунок 9 - Оценка модели

Средняя абсолютная разница между прогнозируемыми значениями модели и фактическими значениями составляет 0,9621.

Квадратный корень из среднеквадратичной разницы между предсказанными набранными баллами оценки и фактическими набранными очками равен 1.2725.

Эти значения не слишком велики, значит построенная модель эффективна.

Заключение

Список использованных источников