



Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное
учреждение высшего образования
«Московский государственный технический университет имени
Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ «Информатика и системы управления»

КАФЕДРА «Программное обеспечение ЭВМ и информационные технологии»

Отчет по лабораторным работам по курсу "Системы искусственного интеллекта"

Студент Сёмина А.А.

Группа ИУ7-12М

Оценка (баллы) _____

Преподаватель Волкова Л.Л.

1 Блок заданий на рекомендательную систему

1.1 Выбор предметной области для рекомендательной системы

В качестве предметной области были выбраны книги. Книги в системе характеризуются следующими атрибутами:

- `title` - название (текстовое);
- `author` - автор (текстовое);
- `publisher` - издательство (текстовое);
- `year` - год издания (числовой);
- `language` - язык книги (текстовое);
- `age_restriction` - возрастное ограничение (категориальное: 0, 6, 12, 16, 18);
- `genre` - жанр (категориальное, древовидная структура);
- `pages` - количество страниц (числовой);
- `has_illustrations` - наличие иллюстраций (булевый: 0/1).

1.2 Вывод функции для нахождения расстояния между книгами

На основе анализа кода системы были выбраны следующие меры близости для различных типов характеристик:

1.2.1 Таксономическое расстояние для жанров

Для жанров, организованных в иерархическую древовидную структуру, применяется таксономическое расстояние. Дерево жанров включает три основных ветви:

- Художественная литература (с подкатегориями: проза, поэзия, драма)
- Нехудожественная литература (с подкатегориями: биография, научная, популярная наука, история, психология)
- Детская литература (с подкатегориями: сказка, развивающая, детские приключения)

Расстояние между двумя жанрами вычисляется по формуле:

$$d_{\text{genre}}(g_1, g_2) = \frac{L(g_1) + L(g_2) - 2 \times L_{\text{common}}(g_1, g_2)}{2 \times D_{\text{max}}}$$

где:

- $L(g)$ - глубина жанра в дереве (количество уровней от корня)
- $L_{\text{common}}(g_1, g_2)$ - глубина ближайшего общего предка
- D_{max} - максимальная глубина дерева (в системе = 4)

1.2.2 Манхэттенское расстояние для числовых признаков

В качестве меры близости для числовых характеристик книги (год издания, количество страниц) рассматривались следующие функции:

- евклидово расстояние;
- манхэттенское мера близости;
- косинусное расстояние;
- взвешенное евклидово расстояние.

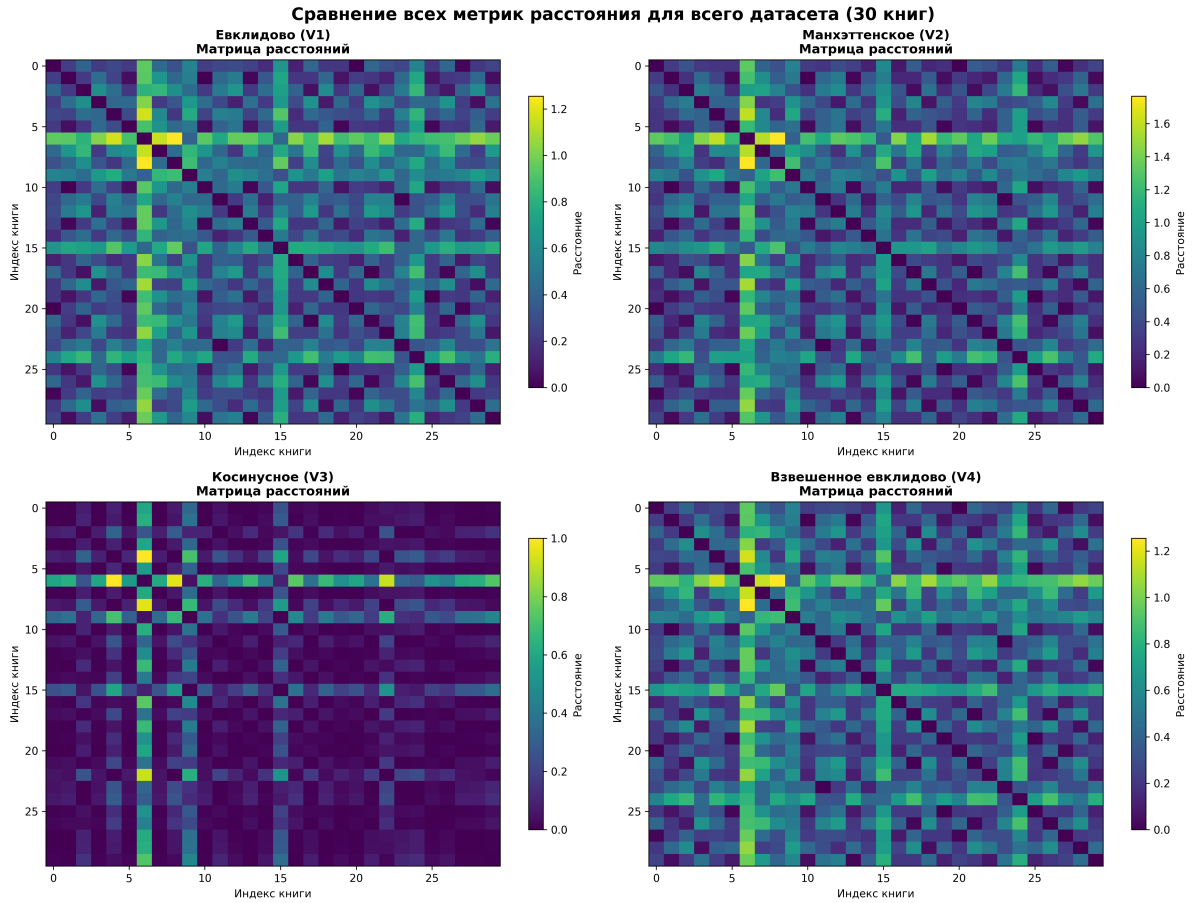


Рисунок 1.1 – Тепловые карты для различных мер близости.

Для числовых признаков (`year`, `pages`) в качестве меры близости было выбрано манхэттенское расстояние с предварительной нормализацией:

$$d_{\text{num}}(\mathbf{x}_1, \mathbf{x}_2) = \sum_{i=1}^k |\text{scaler}_i(x_{1i}) - \text{scaler}_i(x_{2i})|$$

где scaler_i - нормализатор для i -го числового признака.

1.2.3 Категориальное расстояние

Для категориальных признаков без иерархии (`author`, `publisher`, `language`, `age_restriction`) применяется простое расстояние:

$$d_{\text{cat}}(v_1, v_2) = \begin{cases} 0, & \text{если } v_1 = v_2 \\ 1, & \text{иначе} \end{cases}$$

1.2.4 Бинарное расстояние

Для булевого признака `has_illustrations` применяется аналогичная мера:

$$d_{\text{bin}}(v_1, v_2) = \begin{cases} 0, & \text{если } v_1 = v_2 \\ 1, & \text{иначе} \end{cases}$$

1.2.5 Композитное расстояние

Итоговое расстояние между двумя книгами вычисляется как взвешенная сумма расстояний по всем признакам:

$$D_{\text{total}}(b_1, b_2) = \sum_{i \in \text{признаки}} w_i \cdot d_i(b_1, b_2)$$

Где веса по умолчанию заданы в конфигурации системы:

- $\text{weight}_{\text{genre}} = 0.3$ - жанр
- $\text{weight}_{\text{numerical}} = 0.25$ - числовые признаки
- $\text{weight}_{\text{author}} = 0.2$ - автор
- $\text{weight}_{\text{publisher}} = 0.1$ - издательство
- $\text{weight}_{\text{language}} = 0.1$ - язык
- $\text{weight}_{\text{has_illustrations}} = 0.05$ - наличие иллюстраций

1.2.6 Оценка схожести

Для рекомендаций используется оценка схожести (similarity score), которая вычисляется как:

$$S(b_1, b_2) = 1 - D_{\text{total}}(b_1, b_2)$$

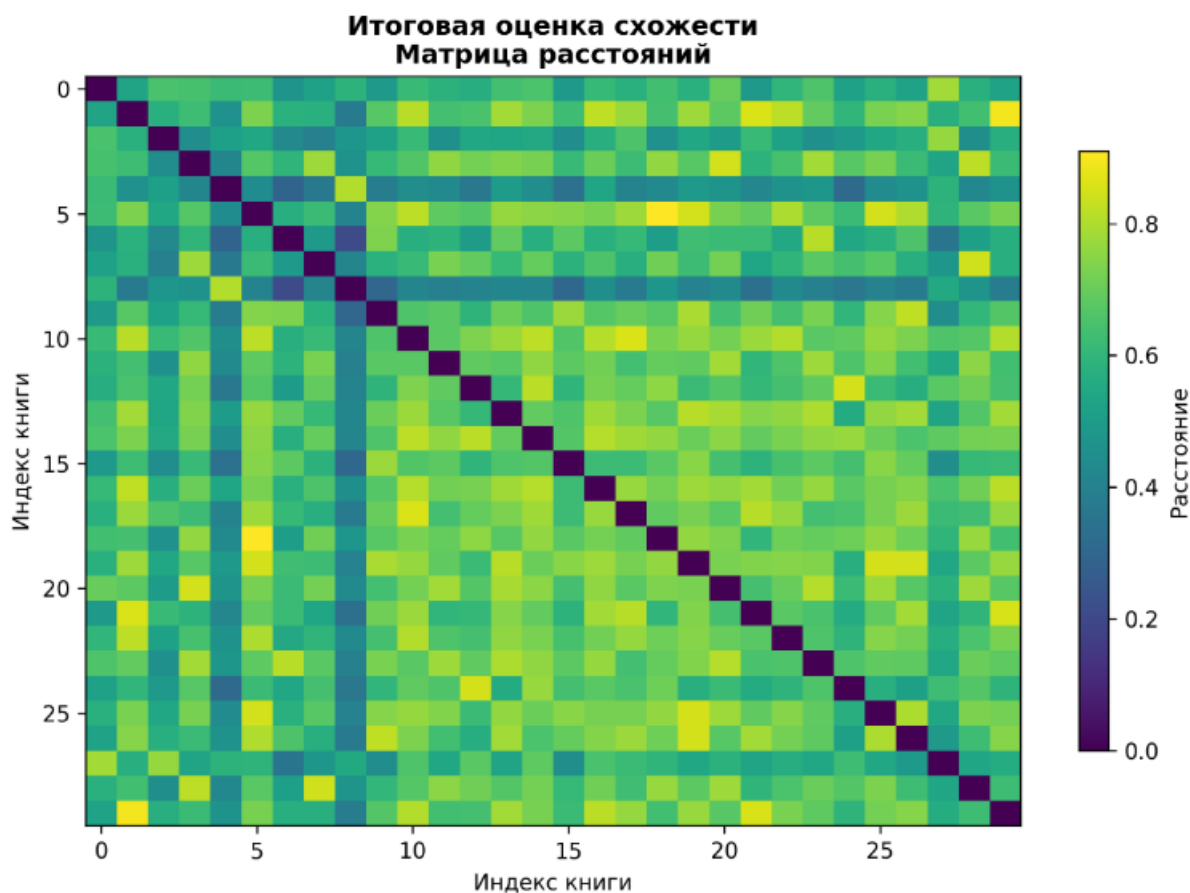


Рисунок 1.2 – Тепловая карта матрицы схожести между книгами

1.3 Реализация возможности ставить лайки и дизлайки

В системе сохраняются идентификаторы книг (id), которые пользователь оценил положительно (лайк) или отрицательно (дизлайк). Рекомендации формируются на основе максимизации схожести с книгами, получившими лайки, и минимизации схожести с книгами, получившими дизлайки. Реализована функция отката последнего действия (undo).

1.4 Реализация возможности поиска книг по характеристикам

Разработан интерфейс расширенного поиска с поддержкой:

- Поиска по точным совпадениям (автор, издательство, язык)
- Фильтрации по диапазонам (год издания, количество страниц)
- Фильтрации по возрастному ограничению
- Поиска по жанру (с учётом иерархии)
- Фильтрации по наличию иллюстраций

Также поиск в рекомендательной системе соединен с фильтрацией книг.

The screenshot shows a web application titled 'Система рекомендаций книг'. It features a sidebar with various filters and a main content area displaying a list of books.

Filters:

- Жанры:** roman, сатира, сказка, фантастика (selected)
- Год издания от:** 2017 **до:** 2022
- Страницы от:** 96 **до:** 1504
- Авторы:** Филип Пулман, Фрэнк Герберт, Харпер Ли, Эрих Мария Ремарк (selected)
- Издательство:** АСТ
- Язык:** Все
- Возрастное ограничение:** Все
- Иллюстрации:** Все
- Название (поиск):** (empty)

Buttons: Применить фильтры, Сбросить фильтры, Показать статистику

Book List:

Название	Автор	Издательство	Год	Язык	Возраст	Жанр	Страницы	Иллюстрации
Мастер и Маргарита	Михаил Булгаков	АСТ	2020	русский	16	классика	480	Есть
Автостопом по галактике	Дуглас Адамс	АСТ	2019	русский	12	фантастика	320	Нет
Игра престолов	Джордж Мартин	АСТ	2021	русский	18	фэнтези	768	Нет
Дюна	Фрэнк Герберт	АСТ	2020	русский	16	фантастика	704	Нет
Понедельник начинается в субботу	Аркадий и Борис Стругац	АСТ	2018	русский	12	фантастика	320	Нет
Властелин колец	Джон Рональд Руэл Толки	АСТ	2021	русский	12	фэнтези	1216	Нет

Рисунок 1.3 – Интерфейс поисковой системы

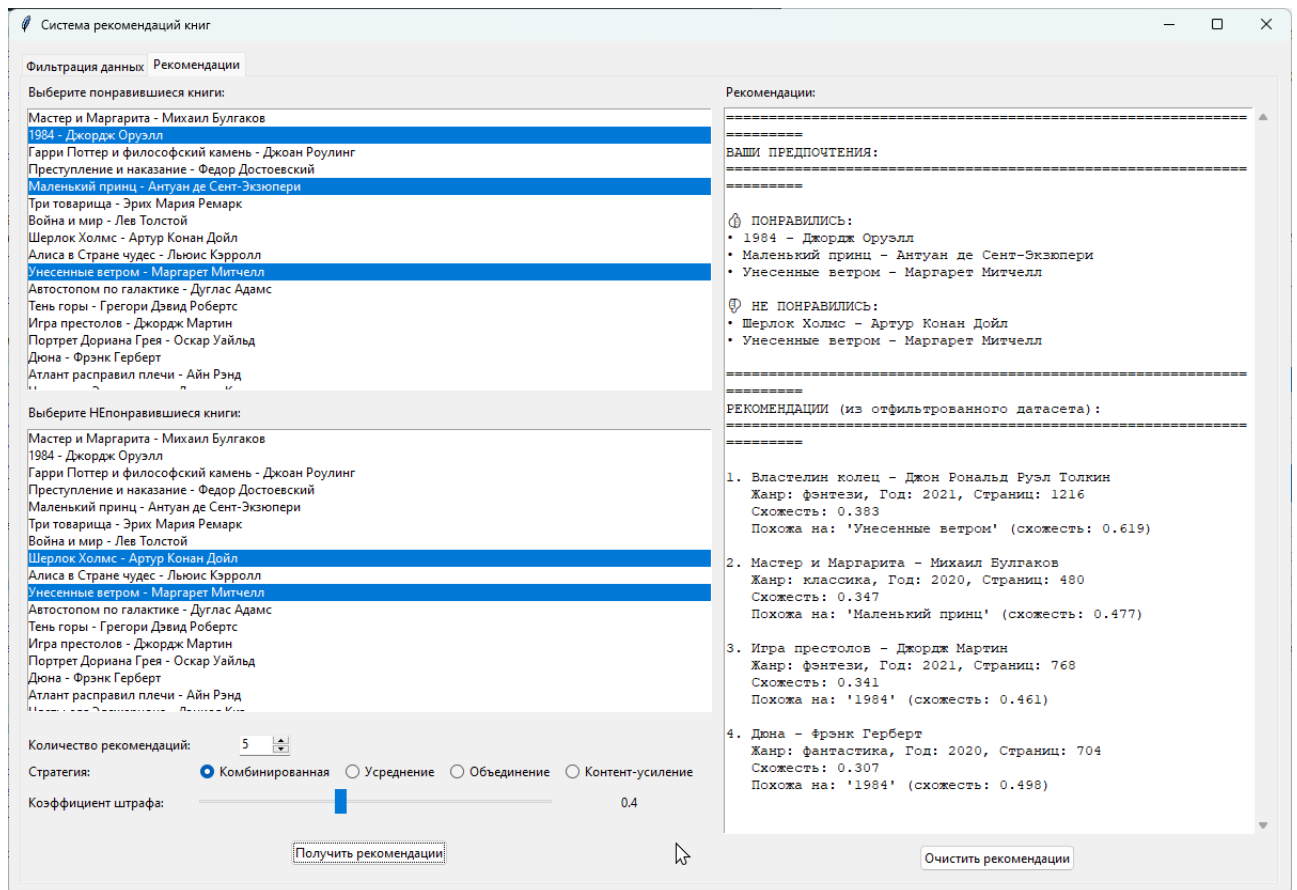


Рисунок 1.4 – Интерфейс рекомендательной системы

2 Блок заданий на диалоговую систему

2.1 Составление списка вопросов, на которые диалоговая система должна уметь отвечать

Вопросы сгруппированы по типам с учётом реализованной системы:

2.1.1 Вопросы к рекомендательной системе

- **По одной книге:** "Мне понравилась книга [название]. Что похожее посоветуешь?"
- **По нескольким книгам:** "Мне нравятся книги [список]. Что ещё почитать?"
- **С отрицательными примерами:** "Мне нравится [книга X], но не нравится [книга Y]. Что посоветуешь?"
- **Общие рекомендации:** "Подбери мне что-нибудь почитать"

2.1.2 Вопросы на поиск с фильтрацией

- **По атрибутам:** "Найди книги автора [имя] издательства [название]"
- **По жанру:** "Покажи книги в жанре фантастика"
- **По году:** "Книги, изданные после 2020 года"
- **По объёму:** "Покажи короткие книги" (использует лингвистическую переменную)

2.1.3 Вопросы на сравнение

- "Чем похожи книги [X] и [Y]?"
- "Сравни [X] и [Y]"
- "Что общего у [X] и [Y]?"

2.1.4 Вопросы на шаги по истории

- "Вернись назад"
- "Отмени последнее"
- "Заново"
- "Начать сначала"

2.1.5 Иные вопросы

- Все остальные вопросы не относящиеся к рекомендательной системе книг

2.2 Реализация диалоговой системы

Диалоговая система реализована с использованием:

- Нейросетевого распознавания намерения выражения
- Нейросетевого извлечения сущностей из текста запроса
- Нечёткого поиска по названиям книг и авторов (нейросеть настроена на распознавание опечаток)

2.2.1 Механизм обработки запросов

1. На основе промпта нейросеть "llama3.1:8b-instruct-q4_0" извлекает параметры (названия книг, имёна авторов, числовые значения), преобразовывает лингвистические термины в числовые диапазоны, исправляет опечатки, убирает склонения
2. Формирование и выполнение запроса к системе рекомендаций/поиска
3. Форматирование ответа
4. Вывод результата

2.3 Реализация поддержки лингвистической переменной

Для характеристики "длины книги" была реализована лингвистическая переменная. Числовые диапазоны для каждого её терма определялись на основании опроса 4-х респондентов ИУ7-12М. Результаты опроса представлены в Таблице 2.1.

Таблица 2.1 – Результаты опроса респондентов ИУ7-12М о числовых диапазонах для термов лингвистической переменной "Длина книги"(в страницах).

Терм	Дрёмин	Сёмина	Виногр.	Горынкин	Среднее
очень короткая книга	< 20	< 20	до 100	< 50	0 – 48
короткая книга	до 250	< 50	100-150	< 100	0 – 138
не очень короткая книга	> 200	> 40	151-250	> 100	123 – 1504
средняя книга по длине	300-800	50-350	251-400	100-300	175 – 463
не очень длинная книга	< 500	< 550	401-550	< 200	0 – 450
длинная книга	> 800	> 350	551-700	> 300	500 – 1504
очень длинная книга	> 2000	> 550	> 700	> 650	975 – 1504

На основании рассчитанных средних значений были установлены итоговые числовые диапазоны для каждого термина лингвистической переменной:

- **"очень короткая книга"**: от 0 до 48 страниц
- **"короткая книга"**: от 0 до 138 страниц
- **"не очень короткая книга"**: от 123 до 1504 страниц
- **"средняя книга по длине"**: от 175 до 463 страниц
- **"не очень длинная книга"**: от 0 до 450 страниц
- **"длинная книга"**: от 500 до 1504 страниц
- **"очень длинная книга"**: от 975 до 1504 страниц

Полученные диапазоны позволяют системе корректно интерпретировать пользовательские запросы, содержащие качественные описания дли-

ны книги, и преобразовывать их в точные числовые фильтры для поиска в базе данных.

2.3.1 Особенности реализации

- Перекрывающиеся диапазоны позволяют плавные переходы между термами
- Интеграция с системой рекомендаций: можно комбинировать с лайками