



Казанский федеральный  
УНИВЕРСИТЕТ  
ВЫСШАЯ ШКОЛА  
информационных технологий  
и интеллектуальных систем

# Разработка рекомендательного сервиса для научных цифровых библиотек

---

Мустаев Артур Рустемович 11-502

Научный руководитель:

Елизаров Александр Михайлович

# Проблема

Проблема информационного поиска релевантных научных документов в коллекциях цифровых научных библиотек.

# Цели и задачи

Цель: разработка рекомендательного сервиса для научных цифровых библиотек.

Задачи:

- Исследовать виды рекомендательных систем и принципы их построения.
- Проанализировать основные алгоритмы, применяемые в рекомендательных системах.
- Спроектировать и реализовать рекомендательный сервис.
- Разработать веб-приложение для научной цифровой библиотеки с реализованным рекомендательным сервисом.

# Подходы

- Коллаборативная фильтрация.  
Формирует рекомендации на основе пользовательских оценок похожих пользователей.
- Контентная фильтрация.  
Для формирования рекомендаций пользовательские оценки не требуются, вместо этого используются связи между документами.

# Коллаборативная фильтрация

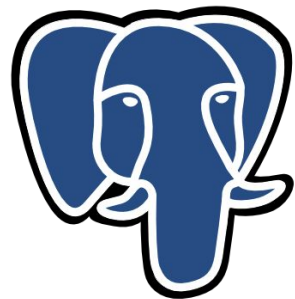
- Формирование матрицы user-user, где в ячейках матрицы расположены значения мер схожести между пользователями.
- Поиск 20 ближайших соседей для целевого пользователя.
- Вычисление прогнозируемых оценок с использованием оценок соседей.
- Выбор топ 10 рекомендаций.

# Контентная фильтрация

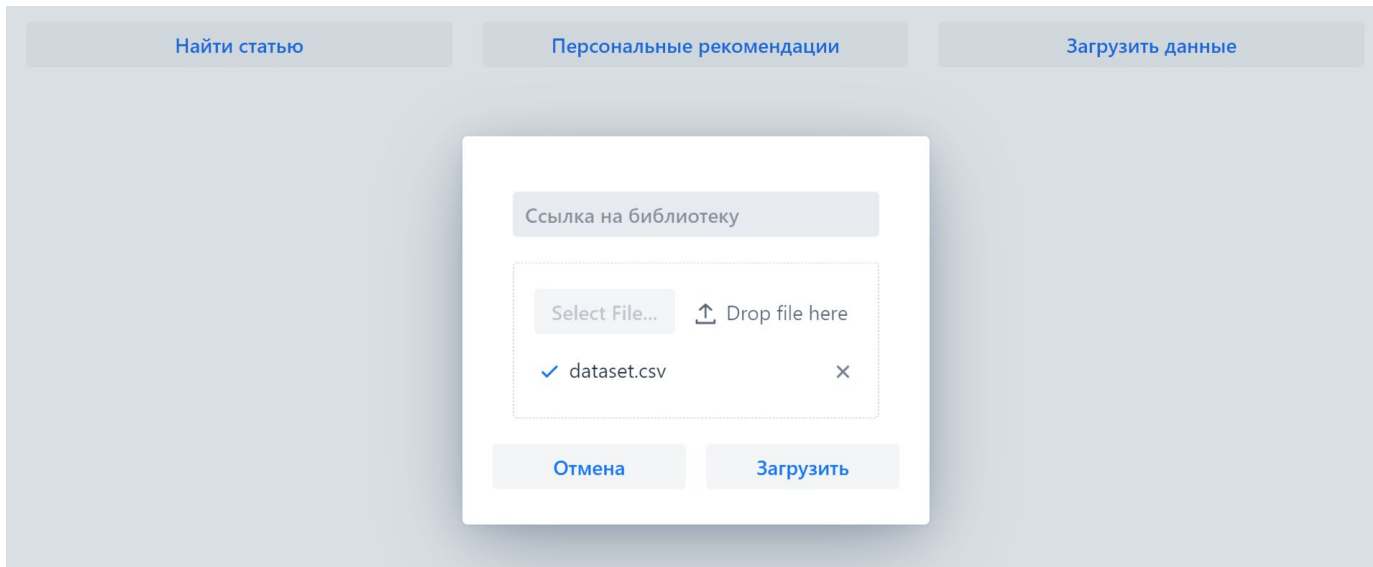
- Выделяются 15 самых популярных слов документа с использованием меры TF, которые представляют собой вектор документа.
- Сравнение векторов документов с помощью косинусной меры.
- Выбор топ 10 похожих документов.

# Технологии

- Серверная часть:
  - Java
  - Spring Boot
  - Spring Jdbc
- Клиентская часть:
  - Vaadin Framework
- База данных:
  - PostgreSQL



# Результаты



username	article_url
Akson87	455624
...	...



# Результаты (продолжение)

Использование мыши и клавиатуры на консолях — это читерс

[Найти](#)

Многие игроки считают, что XIM Apex — это необнаруживаемое читерское устройство, которое портит видеоигры. Но для людей с нарушениями моторики это единственная возможность играть. Мультиплеер игры Call of Duty получил репутацию «мясорубки». Игрок спаунится, делает один-два убийства, если он достаточно хорош, потом умирает, респаунится, и всё повторяется заново. Я уже очень давно не играл в Call of Duty на консолях, поэтому должен был стать лёгкой мишенью, но на самом деле я тащу. Я играю в Call of Duty: WWII на Xbox One, и мой kill/death ratio (отношение убийств к смертям) выше среднего. В режимах Team Deathmatch и Kill Confirmed я занимаю второе место с K/D ratio 21/14, первое место с 18/15 и пятое место с 14/11. Я впервые играю в мультиплеер Call of Duty: WWII, то есть не знаю карту, у меня нет мышечной памяти оружия, его темпа стрельбы и отдачи. Но мне с лёгкостью удаётся вести бегущие к укрытиям цели. Когда враг подбирается ко мне, я способен легко направить на него прицел и первым убить его. Я бы с радостью сказал, что это просто моё мастерство, но правда в том, что я пользуюсь устройством, позволяющим играть вместо стандартного контроллера с мышью и клавиатурой (и другими устройствами ввода, которые официально не поддерживаются). Обычно на консолях такое невозможно. Устройство под названием XIM Apex стоит 100 долларов и выглядит как USB-стик и концентратор, позволяющий подключить к Xbox One (или PlayStation 4) мои мышь и клавиатуру от PC, в то время как мои противники и напарники скорее всего играют обычными контроллерами Xbox One. Многие игроки могут заявить, что это читерство, ведь мышь позволяет мне целиться быстрее и точнее, чем аналоговые стики стандартных контроллеров. Общеизвестно, что игроки с мышью и клавиатурой всегда будут иметь преимущество перед игроками с контроллером, но разработчики игр и производители консолей не имеют или желания, или возможностей блокировать устройства наподобие XIM. Благодаря XIM я, в детстве игравший в Call of Duty на PC с клавиатурой и мышью, начинаю

Краткий обзор консольного рынка

Моддер смог объединить PSP и GameCube

Основные принципы разработки игр

Демотиватор для геймера

Альтернативные аркады на GDC: галерея сумасбродных контроллеров домашнего производства

Идеальный tutorial в игре — советы от PopCap

Пять проблем при разработке мобильных free-to-play игр

Игровая механика: давайте разберём ядро игры по косточкам

Интервью с Дином Холлом о процессе разработки DayZ

NVIDIA Shield — лучшая портативная консоль для гика

# Результаты (конец)

SadJoker

Найти

А как же всё-таки работает многопоточность? Часть I: синхронизация

Обзор `java.util.concurrent.*`

Руководство по оформлению HTML/CSS кода от Google

Книги, которые должен прочитать Java программист: от новичка до профессионала

Как правильно сортировать контент на основе оценок пользователей

Основные принципы настройки Garbage Collection с нуля

Золотые правила успешной кнопки

Знакомство с АОП

ORM в Android с помощью ORMLite

31 метод эффективного программирования под Android

1	Рецепт нагрузочного тестирования на JMeter
2	Tips & tricks for MySQL Developers. Работа с SQL
3	JDK concurrent package
4	Реализация Singleton в JAVA
5	Почему стоит изучить Clojure?
6	Книги для тимлидов и руководителей проектов
7	Gradle и решение задач автоматизации
8	Concurrency: 6 способов жить с shared state
9	У вас здесь ошибка... или о практике инспекций кода в мобильной разработке
10	Мой отчет по годовому изучению платформы Java EE
11	Делаем релизы с помощью Maven в Java
12	Введение в JMeter
13	Модель памяти в примерах и не только
14	Впечатления от работы с Play! Framework 2.1 + Java
15	Учимся проектировать на основе предметной области (DDD: Domain Driven Design)
16	Десять смертных грехов в оценке трудоёмкости разработки программного обеспечения
17	Захабренный договор на разработку сайта, дизайна, софта.
18	Введение в анализ сложности алгоритмов (часть 4)
19	Улучшаем опыт взаимодействия с формами
20	Вся правда о Chrome (и не только) Web Inspector. Часть 1

# Производительность

Тестовый набор данных, состоящий из 20826 статей, 19747 пользователей и 492694 пользовательских оценок.

Действие	Время (мс)
Поиск статьи и формирование списка похожих статей	197
Добавление новой статьи к уже имеющимся	6847
Формирование матрицы user–user, с учетом того, что в эту матрицу попали 9385 пользователей (количество пользователей, у которых количество оценок более или равно 9)	168453
Формирование списка рекомендаций для пользователя	144

# Выводы

- Рекомендательный сервис может быть использован как научными цифровыми библиотеками, так и любыми пользователями.
- Быстрое формирование персональных рекомендаций.
- Быстрое формирование списка похожих статей.

# Дальнейшее развитие

- Повышенная интеграция сервиса с научными цифровыми библиотеками.
- Улучшение точности контентной фильтрации.



Казанский федеральный  
УНИВЕРСИТЕТ  
ВЫСШАЯ ШКОЛА  
информационных технологий  
и интеллектуальных систем

# Разработка рекомендательного сервиса для научных цифровых библиотек

---

Мустаев Артур Рустемович 11-502

Научный руководитель:

Елизаров Александр Михайлович