

Поиск Art-объектов

Программное решение для поиска наиболее релевантных произведений искусства согласно запросу пользователя

Овечкин Евгений

Андрей Ефимов

Глазков Тимур

Киселев Виктор

Килин Георгий

1

Исследование

Исследование предметной области и поиск темы для работы

2

Реализация

Анализ задач и проектирование компонентов системы, а также непосредственное написание кода.

3

Презентация

Подготовка материалов для презентации программного решения

Решаемая задача

Аудитория приложения - клиенты интересующиеся искусством

Задача клиента - найти интересные произведения искусства по нечеткому текстовому запросу о своих предпочтениях

Оценка качества:

1. Оценка релевантности выбора через обратную связь от клиента
2. Оценка косинусного расстояния между описаниями выбранных арт-объектов и запросом клиента

Клиентский путь

Клиент вводит
поисковой запрос

to touch your soul
the art's search engine

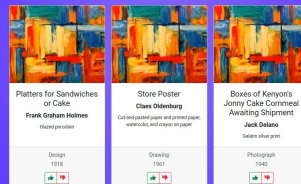
Something you keen on...

Для улучшения качества ответа
клиент может оценить качество
выдачи.

Система анализирует текст запроса по LLM
модели и находит в базе объектов искусства
наиболее близкие по смыслу запроса объекту
из базы Нью-Йоркского музея The Museum of
Modern Arts.

to touch your soul
the art's search engine

calla



Print



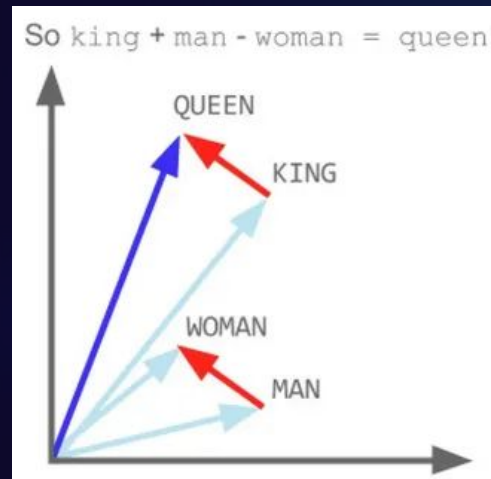
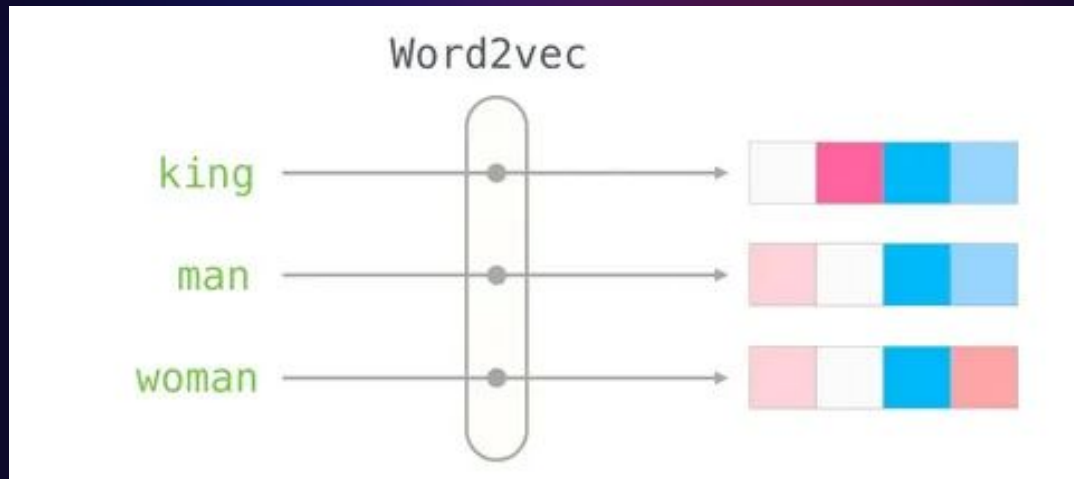
1. Первичное исследование данных
2. Объединение датасетов с произведениями искусства и с авторами в один
3. Предобработка данных и feature engineering
4. Очистка данных
5. Создание результирующего признака для NLP модели
6. Визуализация данных

A box plot showing the distribution of release dates for three size categories: big, medium, and small. The x-axis represents the release date, ranging from 800 to 2000. The y-axis lists the size categories. The 'big' category (blue) has a median release date around 1950, with a box from approximately 1900 to 1980. The 'medium' category (red) has a median release date around 1900, with a box from approximately 1850 to 1950. The 'small' category (green) has a median release date around 1920, with a box from approximately 1880 to 1960. Outliers are present for all categories, with 'big' having one outlier near 1800, 'medium' having one near 1700, and 'small' having several outliers between 800 and 1600.

[illegible]

Подготовка модели

Мы используем предварительно обученные на части набора данных Google News векторы (приблизительно 100 миллиардов слов). Модель содержит 300-мерные векторы для 3 миллионов слов и предложений. Векторы были получены с использованием подхода Word2Vec.



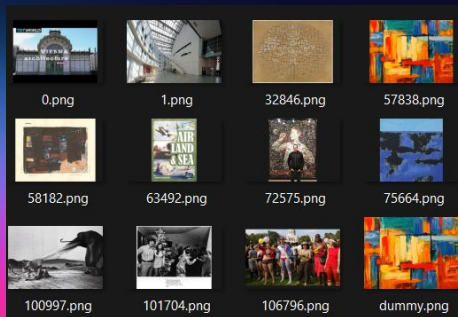
Задачи сервера

```
Waiting for application startup.
Application startup complete.
127.0.0.1:61249 - "GET / HTTP/1.1" 304 Not Modified
127.0.0.1:61249 - "GET /style.css HTTP/1.1" 304 Not Modified
127.0.0.1:61250 - "GET /script.js?version=4 HTTP/1.1" 304 Not Modified
127.0.0.1:61252 - "GET /api?q=red HTTP/1.1" 200 OK
127.0.0.1:61252 - "GET /images/57838.png HTTP/1.1" 304 Not Modified
127.0.0.1:61253 - "GET /images/58130.png HTTP/1.1" 304 Not Modified
127.0.0.1:61254 - "GET /images/58182.png HTTP/1.1" 304 Not Modified
ed', 'id': '58130', 'feedback': 1}
тот коммита

тот коммита
127.0.0.1:61259 - "POST /api/feedback HTTP/1.1" 200 OK
127.0.0.1:61277 - "GET /api?q=red HTTP/1.1" 200 OK
```

Обработка запросов

При получении запроса с фронта, бэк выполняет его обработку и возвращает результат на фронт.



Выгрузка картинок

После обработки запроса моделью, по полученным результатам происходит автоматизированный подбор и поиск картинок.

Filter objects: 1 • SELECT * FROM sql11672622.Data;

Result Grid: Filter Rows: Edit:

idData	q	artwork_id	result
1	dark	58832	1
2	dark	97509	1
3	fantasy	131330	0
4	clouds	73314	1
5	clouds	85287	1
6	clouds	43615	0
9	clouds	73314	1
20	hey	106796	0
21	red	58130	1

Запись фидбека

После получения поисковой выдачи, пользователь может оценить её результат, путём проставления лайка либо дизлайка.

Задел на будущее

- Руками собрать набор данных запросов и экспонатов, которые им соответствуют. На этом наборе посчитать метрики ранжирования: map@K , nDCG@K
- Использовать подход к сравнению текстов, учитывающий порядок слов в них

Спасибо за внимание!