



Создание рекомендательной системы по аниме на основе синопсиса и жанра

Бычкова Евгения, Кирилова Алиса, Б5122-45.03.03ликц

ЦЕЛЬ ПРОЕКТА

Разработка прототипа рекомендательной системы для аниме на основе семантического анализа **текстовых описаний (синопсисов) и жанров**, определяющего содержательное сходство между произведениями

ЗАДАЧИ ПРОЕКТА

01

Отобрать синопсисы
различных аниме различных жанров

03

Векторизовать синопсисы с
использованием **TF-IDF**

02

Провести **лингвистическую**
обработку текстов

04

Реализовать **алгоритм поиска**
в полученном векторном
пространстве

КОРПУС ОТОБРАННЫХ АНИМЕ



25 ПРОИЗВЕДЕНИЙ

Атака титанов; Бездомный Бог; Блич; Вайолет Эвергарден; Город, в котором меня нет; Корзинка фруктов; Клинок, рассекающий демонов; Магическая битва; Неудержимая юность; Наруто; Паразит; Последний серафим; Стальной алхимик; Тетрадь смерти; Токийские мстители; Токийский гуль; Твое имя; Убийца Акаме!; Форма голоса; Хантер x Хантер; Хоримия; Человек-бензопила; Черный клевер; Очень приятно, Бог; Торадора!

СИНОПСИСЫ

+ **ЖАНРЫ**

+ **ТИПЫ**


```
# жанры анимешек
```

```
genres = {  
    "Хантер х Хантер": "экшен приключения сёнэн",  
    "Токийский гуль": "экшен тёмный драма психология ужасы",  
    "Паразит": "ужасы драма философия фантастика",  
    "Человек-бензопила": "экшен тёмный ужасы сёнэн",  
    "Магическая битва": "экшен магия сёнэн",  
    "Стальной алхимик": "экшен магия драма война философия",  
    "Черный клевер": "экшен магия сёнэн",  
    "Блич": "экшен сверхъестественное сёнэн",  
    "Атака титанов": "экшен война драма постапокалипсис",  
    "Наруто": "ниндзя приключения дружба",  
    "Город, в котором меня нет": "драма психология триллер",  
    "Клинок, рассекающий демонов": "экшен тёмный демоны сёнэн",  
    "Форма голоса": "драма школа драма психология",  
    "Твое имя": "романтика фантастика драма романтика",  
    "Хоримия": "романтика школа повседневность",  
    "Корзинка фруктов": "драма семья романтика школа",  
    "Неудержимая юность": "романтика школа",  
    "Очень приятно, Бог": "романтика сверхъестественное",  
    "Тетрадь смерти": "драма психология триллер сверхъестественное",  
    "Вайолет Эвергарден": "драма эмоции повседневность",  
    "Убийца Акаме!": "экшен тёмный драма война",  
    "Бездомный Бог": "экшен сверхъестественное комедия",  
    "Торадора!": "романтика школа комедия",  
    "Последний серафим": "экшен тёмный вампиры постапокалипсис",  
    "Токийские мстители": "драма криминал экшен боевка"
```

```
# типы анимешек
```

```
types = {  
    "Хантер х Хантер": "приключенческий экшен",  
    "Токийский гуль": "тёмный хоррор",  
    "Паразит": "тёмная фантастика",  
    "Человек-бензопила": "тёмный экшен",  
    "Магическая битва": "боевой сёнэн",  
    "Стальной алхимик": "эпический экшен",  
    "Черный клевер": "боевой сёнэн",  
    "Блич": "боевой сёнэн",  
    "Атака титанов": "эпическая трагедия",  
    "Наруто": "приключенческий сёнэн",  
    "Город, в котором меня нет": "психологическая драма",  
    "Клинок, рассекающий демонов": "боевой сёнэн",  
    "Форма голоса": "повседневная драма",  
    "Твое имя": "романтическая драма",  
    "Хоримия": "школьная романтика",  
    "Корзинка фруктов": "семейная драма",  
    "Неудержимая юность": "школьная романтика",  
    "Очень приятно, Бог": "романтическое фэнтези",  
    "Тетрадь смерти": "психологический триллер",  
    "Вайолет Эвергарден": "эмоциональная драма",  
    "Убийца Акаме!": "трагический экшен",  
    "Бездомный Бог": "приключенческое фэнтези",  
    "Торадора!": "романтическая комедия",  
    "Последний серафим": "постапокалиптический экшен",  
    "Токийские мстители": "криминальная драма"
```

ПРОЦЕСС РАБОТЫ

ИСПОЛЬЗОВАНИЕ TF-IDF

- Семантическая значимость слов учитывается более точно;
- Малозначимые слова получают низкий вес;
- Подходит для описательных текстов (синопсисы).

ИСПОЛЬЗОВАНИЕ КОСИНУСНОЙ СХОЖЕСТИ - метод измерения близости двух векторов путем вычисления косинуса угла между ними

- Отсутствие зависимости от количества текста в синопсисе;
- Хорошая совместимость с TF-IDF для оценки сходства текстов;
- Повсеместно используется в рекомендательных системах для текстов



ПРОБЛЕМЫ, ВОЗНИКШИЕ В ПРОЦЕССЕ РАБОТЫ

- 1) Необходимость в добавлении жанров и типов аниме;
- 2) **Расширение** корпуса;
- 3) Неточность в результатах рекомендательной системы:
появление **неподходящих картин на 2-3 местах** в списке рекомендаций



КОНЕЧНЫЙ РЕЗУЛЬТАТ

Рекомендации по: Хоримия

- | | |
|-----------------------|----------|
| 1. Неудержимая юность | — 62.34% |
| 2. Торадора! | — 23.60% |
| 3. Корзинка фруктов | — 19.30% |
| 4. Вайолет Эвергарден | — 18.09% |
| 5. Твое имя | — 17.17% |

Рекомендации по: Последний серафим

- | | |
|--------------------------------|----------|
| 1. Атака титанов | — 15.34% |
| 2. Человек-бензопила | — 15.30% |
| 3. Убийца Акаме! | — 14.78% |
| 4. Клинок, рассекающий демонов | — 14.38% |
| 5. Токийский гуль | — 12.81% |

Рекомендации по: Паразит

- | | |
|----------------------|----------|
| 1. Токийский гуль | — 17.28% |
| 2. Стальной алхимик | — 16.76% |
| 3. Твое имя | — 16.55% |
| 4. Человек-бензопила | — 14.14% |
| 5. Форма голоса | — 9.57% |



Создание рекомендательной системы по аниме на основе синопсиса и жанра

Бычкова Евгения, Кирилова Алиса, Б5122-45.03.03ликц