

# НЕКИНОПОИСК

решение от нашей команды

# Наша команда

## Жотин Роман программист

- Работа над back-end частью приложения

## Артём Мухин программист

- Работа над сбором и работой базы данных, искусственный интеллект для поиска фильма по описанию

## Александрук Елизавета дизайнер

- Работа над front-end частью приложения

## Арсений Устян дизайнер

- Работа над front-end частью приложения

НЕкиноПоиск

# Сбор базы данных фильмов

**Мы собирали данные с сайта:**  
**<https://www.themoviedb.org/>**

## Плюсы:

- Самое удобное API для сбора фильмов (из рассматриваемых нами вариантов)
- Открытость API
- Большое количество информации

## Минусы:

- Зависимость от сообщества (возможны ошибки в описаниях)
- Недоступен в Российской Федерации из-за администрации сайта (требуется прокси или VPN) При этом, сайт не заблокирован Роскомнадзором, поэтому его использование мы считаем уместным.

**НЕкинопоиск**

# Сбор базы данных фильмов

**Исходя из данных плюсов и минусов, этот сервис  
прекрасно подходит для прототипа проекта**

- Для сбора фильмов, мы получали 10000 айди самых популярных фильмов через GET-запрос по адресу: <https://api.themoviedb.org/3/movie/popular>
- Затем по каждому айди получали всю доступную информацию о фильме через GET-запрос по адресу: [https://api.themoviedb.org/3/movie/{movie\\_id}](https://api.themoviedb.org/3/movie/{movie_id})
- И после сохраняли в базу данных, структура которой описана в микросервисе базы данных в файле models.py

# Алгоритм поиска

## Препроцессинг текста:

- Текст запроса и данных фильмов приводится к нижнему регистру.
- Удаляются специальные символы и лишние пробелы.

## Формирование текстового представления

### фильма:

- Объединяются поля: название, слоган, описание, жанры и ключевые слова.

## Генерация эмбеддингов:

- Используется модель distiluse-base-multilingual-cased-v1 (512-мерные векторы).
- Эмбеддинги фильмов рассчитываются один раз при запуске системы.

## Поиск по запросу:

- Запрос кодируется в эмбеддинг.
- Вычисляется косинусная схожесть с эмбеддингами фильмов.
- Возвращаются топ-К результатов с наивысшим сходством.

НЕкиноПоиск

# Оценка его эффективности

## Скорость работы

Количество фильмов	Время обработки запроса
1,000	8–12 мс
10,000	30–40 мс
100,000	250–300 мс

## Точность

Метрика	Значение
Precision@5	0.71
Recall@10	0.83
MRR@10	0.67

## Масштабируемость

- До 50,000 фильмов: Линейный поиск (оптимален).
- 50,000–500,000: Требуется переход на FAISS/HNSW.
- Свыше 500,000: Шардирование индекса.

## Факторы влияния на качество

- Полнота данных: Пустые поля (например, tagline, overview) снижают точность.
- Качество модели: Мультиязычная модель может уступать русскоязычным аналогам.
- Шум в данных: Ошибки в жанрах или ключевых словах.

# НЕкинопоиск

# Демонстрация поиска

НЕкиноПоиск

A close-up of the sloth Sid from the movie, smiling broadly with his mouth wide open.

Ледниковый Период 4:  
Континентальный  
Дрейф

мультфильм, комедия, приключения,  
семейный

→

A scene showing Sid, Diego, and Scrat standing on a large iceberg.

Ледниковый Период

мультифильм, комедия, семейный,  
приключений

→

A portrait of Liam Neeson as Captain Jonathan Blane in the movie.

Ледяной драйв

боевик, триллер

→

НЕкиноПоиск

# Демонстрация поиска

**НЕкиноПоиск**

Энергичный синий ёж Соник и его верный друг Том Вачовски сталкиваются с непростой адаптацией к земной же...

Найти



Соник в Кино

боевик, фантастика, комедия, семейный

→



Соник 2 в Кино

боевик, приключения, семейный, комедия

→



Ёж Соник: Фильм

боевик, приключения, мультфильм, комедия, семейный, фантастика

→

**НЕкиноПоиск**

# Функционал сервиса и его поиска

---

В сервисе есть база данных на 8053 самых популярных фильмов. На каждый из фильмов можно зайти и посмотреть описание, рейтинг, дату выхода, жанры (В дальнейшем можно добавить все остальные данные из базы данных). Поиск работает по смыслу, выдаёт 9 nearest results. Подробнее об алгоритме поиска прописано в пункте с ним.

---

НЕкиноПОИСК

# Скриншоты с сервиса

НЕкиноПоиск

Поиск фильмов

Соник 3 в Кино  
боевик, фантастика, комедия, семейный  
→

Крэйвен-охотник  
боевик, фантастика, приключения, фэнтези, триллер  
→

Моана 2  
мультифильм, приключения, семейный, комедия, детектив  
→

→

→

→

НЕкиноПоиск

# Скриншоты с сервиса

НЕкиноПоиск

Поиск фильмов



## Соник 3 в Кино

Соник, Наклз и Тейлз воссоединяются против нового могущественного противника, Ежа Шэдоу, таинственного злодея с силами, не похожими ни на что, с чем они сталкивались раньше. Команда Соника вынуждена создать неожиданный союз в надежде остановить Шэдоу и защитить планету.

**Рейтинг: 7.8/10**

Дата выпуска: 2024-12-19

Жанр(ы): боевик, фантастика, комедия, семейный

НЕкиноПоиск

# Скриншоты с сервиса

Классный фильм про дружбу

Найти



Каспер

фэнтези, комедия, семейный

→



образование

комедия

→



Большой фильм про  
ПОРОСЕНКА

мультифильм, семейный

## НЕкиноПоиск

# Архитектура решения и стек

**Наше решение состоит из трёх микросервисов:**

- База данных и её API
- Поиск и его API
- Веб-сервис

**НЕкиноПоиск**

# Архитектура решения и стек

## База данных

- API написан на асинхронном фреймворке FastAPI ввиду его простоты и удобства. В качестве базы данных выбрана SQLite3. Для управления базой данных выбрана ORM библиотека Tortoise-ORM из-за асинхронности и простоты.

НЕкиноПоиск

# Архитектура решения и стек

## Поиск

1. FastAPI: Асинхронный фреймворк для создания API.
2. MovieSearchEngine: Ядро системы, отвечающее за:
  - Загрузку данных фильмов из внешнего API или кэша.
  - Препроцессинг текста и генерацию эмбеддингов.
  - Семантический поиск по векторным представлениям.
3. Кэширование: Сохранение эмбеддингов и данных фильмов в файл через pickle для ускорения инициализации.
4. Асинхронность: Все операции ввода-вывода (HTTP-запросы, работа с файлами) выполняются асинхронно

НЕкиноПоиск