

Кейс “Поисковик фильмов”

Команда **BackSpace:**

Короткова Анна Владимировна
Седов Михаил Юрьевич
Горюнов Платон Витальевич

Наставник:

Томилин Никита Васильевич

ГАОУ СО Дворец молодежи ДТ “Кванториум”

Цель и задачи



Цель: создать веб-сервис с функцией поиска фильмов по тегам или текстовому описанию из заготовленной базы данных

Задачи:

- 1) Составить список функций для веб-сервиса
- 2) Определиться с формой веб-платформой
- 3) Собрать базу данных
- 4) Написать веб-сервис
- 5) Провести тестирования и отладку

BackSpace



Короткова Анна

- разработка бота,
тимлид



Седов Михаил

- написание парсера



Горюнов Платон

- работа с базами
данных

Список функций и выбор платформы

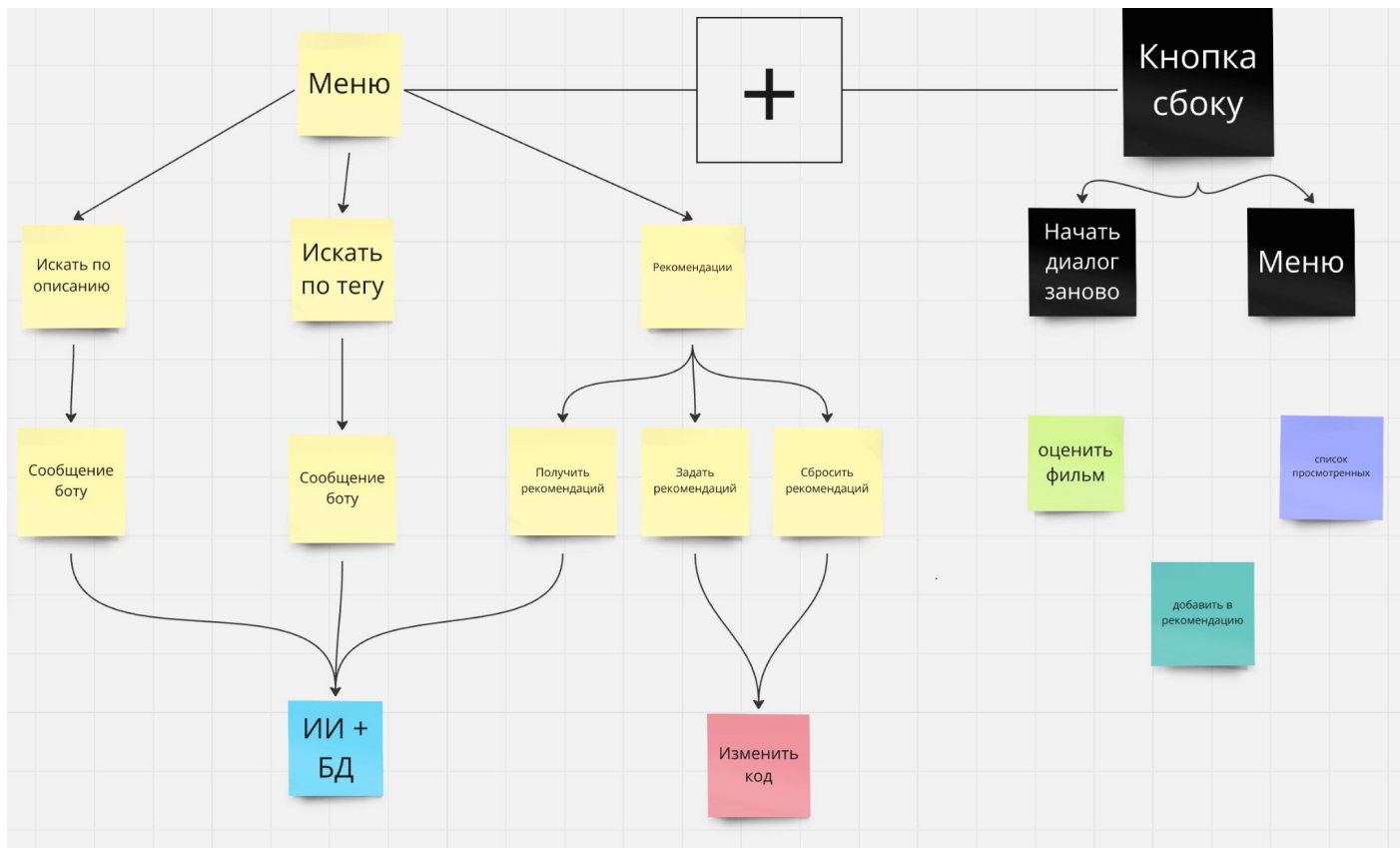
Функции:

- 1) Поиск фильмов по тегам
- 2) Поиск фильмов по описанию
- 3) Рекомендательная система
- 4) Пользовательская статистика



В качестве веб-сервиса нами был выбран телеграмм-бот, так как эта площадка является наиболее распространенной и мобильной, а также удобной в реализации взаимодействия между пользователем и программой

Иерархия функций



Сбор базы данных



Для сборки базы данных мы использовали парсинг данных. Для этого нами был выбран сайт с 1000 лучших фильмов ([Коллекции фильмов - 1000 фильмов, которые нужно посмотреть, прежде чем умереть \(The Guardian\)](#)). Он оказался одним из самых информативных, по нашему мнению.

Парсинг происходил по следующему принципу: парсилась ссылка на страницу фильма, после чего второй парсер заходил на неё и сохранял всю нужную нам информацию. Сама база данных сохранялась в формате “sqlitedb”, чтобы при дальнейшей работе с ней было удобно взаимодействовать через встроенные библиотеки.

Итоговая база данных с фильмами состояла из 250 строк и таких колонок, как название, рейтинг, жанр, теги, описание, режиссёр, актёрский состав, длительность, а также эмбединг описания и эмбединг тегов, которые формировались с помощью готового алгоритма.

Алгоритм



Для нахождения наиболее подходящих по запросу пользователя фильмов мы использовали специализированную на русском языке семантическую модель - “sberbank-ai/sbert_large_nlu_ru”. Она кодирует текст запроса в векторное представление, после чего происходит вычисление косинусного сходства с эмбедингами фильмов, полученного ранее в процессе сборки базы данных.

Данный способ позволяет реализовать поиск фильмов по смыслу, а не по конкретным словам. Однако модель теряет свою точность при работе с объемной базой данных. Именно по этой причине нами было принято решение уменьшить размер базы данных с 1000 до 250 строк. Данное количество является наиболее оптимальным как для точности работы модели, так и полноты предоставляемых данных.

Итоговый функционал



Основные функции бота:

- Рекомендации фильмов
- Поиск наиболее подходящих фильмов по запросам
- Семантический поиск
- Работа с базами данных

Дополнительные функции:

- Отметка фильма как просмотренного
- Оценка фильма пользователем
- Добавление фильма в список рекомендаций
- Просмотр списка просмотренных фильмов и фильмов с оценками
- Интерактивное управление через inline-клавиатуры

Демонстрация работы



@KINOGO_TECHNOSTR
ELKABOT