

Разработка информационной системы по рекомендациям культурных событий на основе языковых моделей

Токарев Дмитрий Сергеевич

Студент (магистр)

Удмуртский государственный университет, Ижевск, Россия

E-mail: tokareal.dima@gmail.com

Современные информационные технологии и методы машинного обучения открывают новые возможности для автоматизации процессов рекомендации культурных событий. В декабре 2024 года в приложении «Госуслуги.Культура» был запущен чат-бот, который помогает пользователям подбирать мероприятия. Однако текущая версия бота ограничена в функциональности: он не взаимодействует с пользователями в диалоговом режиме, а лишь предлагает выбрать из предопределенных фильтров. Это снижает качество рекомендаций и не позволяет учитывать индивидуальные предпочтения пользователей.

Целью данной работы является разработка информационной системы, которая сможет предоставлять персонализированные рекомендации культурных событий на основе диалогового взаимодействия с пользователем. Для этого предлагается дообучить языковую модель на данных с сайта <https://www.culture.ru/afisha/russia>, где представлено более 120 000 мероприятий.

Основные этапы работы:

1. **Сбор и обработка данных:** Еженедельный парсинг данных с сайта <https://www.culture.ru/afisha/russia> для получения актуальной информации о мероприятиях. Данные будут включать описание, дату, время, регион, возрастные ограничения, стоимость, теги (жанры) и другие параметры. Для обработки данных будет использоваться библиотека BeautifulSoup или Scrapy.

2. **Дообучение языковой модели:** В качестве базовой модели будет выбрана одна из современных языковых моделей, таких как YandexGPT, GPT-3 или их аналоги. Модель будет дообучена на собранных данных, что позволит ей лучше понимать контекст и предпочтения пользователей, а также генерировать более точные рекомендации.

3. **Разработка телеграмм-бота:** Интеграция дообученной языковой модели в телеграмм-бота. Бот сможет задавать уточняющие вопросы (например, о временных рамках, составе группы и предпочтениях) и предлагать мероприятия на основе полученных ответов. Для разработки бота будет использоваться библиотека python-telegram-bot.

4. **Тестирование и оценка:** Проведение тестирования бота с реальными пользователями для оценки качества рекомендаций. Анализ результатов тестирования и выявление возможных улучшений.

Достоинства предлагаемого подхода:

· **Персонализация:** Возможность учитывать индивидуальные предпочтения пользователей, такие как временные рамки, состав группы и жанровые предпочтения.

· **Практическая значимость:** Разработанная система может быть интегрирована в существующие платформы, такие как «Госуслуги.Культура», для повышения качества взаимодействия с пользователями и увеличения посещаемости культурных мероприятий.

Источники и литература

- 1) YandexGPT-5-Lite-Pretrain: - <https://huggingface.co/yandex/YandexGPT-5-Lite-8B-pretrain>

- 2) Python Telegram Bot's documentation - <https://pythontelegrambot.readthedocs.io/en/latest/>
- 3) Культура РФ: <https://www.culture.ru/afisha/russia>