**Документация, содержащая описание функциональных характеристик модуля программного обеспечения**

**Модуль «Модерация контента»**

**г. Москва 2025**

**Содержание**

[1. Введение 3](#_heading=h.z7nixxlw7dmi)

[2. Назначение модуля 3](#_heading=h.gpk62jcsynam)

[3. Функциональные возможности 3](#_heading=h.wqq80x4vfv8d)

[4. Архитектура и принцип работы 3](#_heading=h.qx5myiguucxn)

[4.1. Принцип работы 3](#_heading=h.bukc3xp8kv2a)

[4.2. Структура входного и выходного файлов 3](#_heading=h.2penylrqmrip)

[5. Используемые языки программирования и фреймворки 5](#_heading=h.qhc22np3oh53)

[6. Возможные ошибки и решения 6](#_heading=h.xirqi35rhje0)

# **1. Введение**

Модуль «Модерация контента» предназначен для автоматизированного анализа и разметки видеоконтента (фильмы, сериалы) в онлайн-кинотеатре. Он выполняет модерацию контента в соответствии с установленными критериями, обеспечивая безопасную и здоровую среду для пользователей платформы.

# **2. Назначение модуля**

Модуль разработан для автоматизированной проверки видеоконтента на наличие нежелательных элементов, таких как:

* Пропаганда наркотиков, алкоголя и курения;
* Сцены с ЛГБТ-контентом;
* Контент, нарушающий политику платформы.

# **3. Функциональные возможности**

Модуль выполняет следующие задачи:

* Анализ видеофайлов и их метаданных, включая информацию из сервисов Кинопоиск и IMDB;
* Определение наличия ключевых признаков (наркотики, курение, алкоголь, ЛГБТ и др.);
* Формирование разметки контента и предоставление отчетов по результатам модерации;
* Хранение данных о модерации в стандартизированном формате.

# **4. Архитектура и принцип работы**

### 4.1. Принцип работы

Модуль загружает входной файл с данными о контенте.

Запрашивает информацию о контенте в сервисах Кинопоиск и IMDB.

Анализирует полученные данные, используя набор предустановленных критериев.

Формирует выходной файл с разметкой контента и классификацией по заданным категориям.

### 4.2. Структура входного и выходного файлов

Входной файл (XLSX):

| **Название столбца** | **Описание** |
| --- | --- |
| Название фильма | Название фильма на русском языке |
| Оригинальное название | Оригинальное название фильма |
| Год | Год выпуска фильма |
| Страна производства | Страна производства фильма |
| Кинопоиск | Ссылка на фильм в сервисе Кинопоиск |
| IMDB | Ссылка на фильм в сервисе IMDB |

Выходной файл (XLSX):

| **Название столбца** | **Описание** |
| --- | --- |
| Название фильма | Название фильма на русском языке |
| Оригинальное название | Оригинальное название фильма |
| Год | Год выпуска фильма |
| Страна производства | Страна производства фильма |
| Кинопоиск | Ссылка на фильм в сервисе Кинопоиск |
| IMDB | Ссылка на фильм в сервисе IMDB |
| Метки Кинопоиск | Полный список тегов фильма в сервисе Кинопоиск |
| Разметка курения | Цветовая разметка наличия или отсутствия признаков курения |
| Теги курения | Список тегов, связанных с курением (если есть) |
| Разметка алкоголя | Цветовая разметка наличия или отсутствия признаков алкоголя |
| Теги алкоголя | Список тегов, связанных с алкоголем (если есть) |
| Разметка наркотиков | Цветовая разметка наличия или отсутствия признаков наркотиков |
| Теги наркотиков | Список тегов, связанных с наркотиками (если есть) |
| Разметка ЛГБТ | Цветовая разметка наличия или отсутствия признаков ЛГБТ |
| Теги ЛГБТ | Список тегов, связанных с ЛГБТ (если есть) |

# 5. Используемые языки программирования и фреймворки

| Языки программирования | Python |
| --- | --- |
| Используемые библиотеки | Pandas, Selenium, Requests, OpenPyXL, Loguru, charset-normalizer, chardet, Levenshtein, random\_user\_agent, transliterate, undetected\_chromedriver, user\_agent, CatBoost |
| Контейнеризация | Docker, Yandex Managed Service for Kubernetes |
| Репозитории контейнеров | Docker Hub, mirror.gcr.io, Yandex Container Registry |
| СУБД | ClickHouse (Yandex Managed Service for ClickHouse) |

# **6. Возможные ошибки и решения**

Ошибка подключения к интернету – проверить соединение;

Нет данных в IMDB/Кинопоиске – контент не найден в базе, требуется ручная проверка;

Ошибка анализа файла – убедиться в корректности структуры входного файла.

В этом случае:

Проверьте наличие соединения с сетью Интернет;

Проверьте доступность сервиса Кинопоиск, открыв сайт https://www.kinopoisk.ru/ в браузере; Проверьте доступность сервиса IMDB, открыв сайт https://www.imdb.com в браузере.

Если все сервисы доступны, но ошибка все равно возникает, обратитесь к разработчику.