Пермский филиал федерального государственного автономного   
образовательного учреждения высшего образования   
«Национальный исследовательский университет   
«Высшая школа экономики»

*Факультет экономики, менеджмента и бизнес-информатики*

Перминов Николай Александрович

**Лабораторная работа №6**

**«Web-разработка с использованием Flask»**

*Отчёт*

студента образовательной программы «Программная инженерия»   
по направлению подготовки *09.03.04 Программная инженерия*

Руководитель   
Доцент каф. ИТБ, к.п.н.   
И. В. Ильин

Пермь, 2020 год

Оглавление

[1 Постановка задачи 3](#_Toc73892268)

[2 Диаграмма классов 4](#_Toc73892269)

[3 Код программы 5](#_Toc73892270)

[4 Код unit-тестов 14](#_Toc73892271)

[5 Анализ покрытия тестов 16](#_Toc73892272)

# Постановка задачи

1. Спарсить данные с целевого сайта (myanimelist.com) (лабораторная работа №5).
2. Провести анализ (лабораторная работа №5).
3. Представить полученные данные на сайте
4. Представить анализ на сайте

# Анализ задачи

## Первая часть

Для поставленной задачи требуется спарсить сайт.

Парсинг сайта проводится в несколько этапов:

1. Обработка ссылок, которые потребуется парсить. Для этого парсер проходится по списку топ аниме до самого конца, если ему не удается докачать всё, то что он не может докачать попадает в файл ошибок
2. Обработка файла ошибок (докачивание страниц, где были ошибки).
3. Скачивание страниц для парсинга, адреса которых получены на первом этапе.
4. Обработка ошибок скачивания (докачивание страниц с ошибками)
5. Обработка всех страниц для получения с них данных.

Поскольку я парсил около 35 тысяч страниц мне пришлось (чтобы сай не блокировал мой ip) воспользоваться GitHub Actions, сервиса, который предоставляет виртуальные машины на время. Также его использование позволило распараллелить скачивание 35 тысяч страниц.

## Вторая часть

Для агрегации данных на сайте я использовал связку рендеринга html таблицы библиотекой pandas и генератор страниц Flask.

Для реализации поиска использовались формы, которые изменяли url-параметры, по которым производится поиск, они же указывают требуется выдача данных или графиков.

Для отрисовки графиков используется библиотека matplotlib, а на сайте уже выдаётся png изображение.

# Код программы

Код первой части <https://github.com/k0perX-X/mal-database>

Код второй части <https://github.com/k0perX-X/pythonHSE/tree/master/Laba%206>