

#W6PA: GitHub Spy

1. Описание задания	2
2. GitHub шпион (spy)	2
3. Литература для расширения кругозора	2



1. Описание задания

В этом задании вам нужно написать консольное приложение, которое будет шпионить за интересующим вас конкурентом на примере GitHub. Вам нужно написать инструмент, который позволяет скачивать страницы из Интернет, парсить содержимое HTML и проверять число платных и бесплатных предложений. Цель задания:

1. Научить делать запросы с помощью requests;
2. Научить разделять интеграционные/системные и юнит-тесты;
3. Парсить содержимое HTML.

2. GitHub шпион (spy)

Консольный интерфейс библиотеки:

```
$ github_spy.py analyze1  
free products: 10  
enterprise products: 25
```

Библиотека должна быть протестирована и предоставлять возможность тестировать функциональность парсинга без доступа в Интернет. Должна быть возможность запускать только юнит-тесты, только системные тесты, или и те и другие. Пример pytest marker'a slow можно взять с GitHub проекта купца (см. conftest.py):

- <https://github.com/big-data-team/python-course>

Системные тесты должны проверять, что число "free products" и "enterprise products" остались без изменений. Если же это не так, попробуйте в Assert вывести разницу между ожидаемыми и полученными ссылками.

3. Литература для расширения кругозора

В обсуждениях первого домашнего задания мы затронули тему "чистого кода" (Clean Code). Очень рекомендуется взять в домашнюю библиотеку одноименную книгу:

- Clean Code: A Handbook of Agile Software Craftsmanship
- автор: Robert C. Martin

¹ По умолчанию анализируется содержимое страницы <https://docs.github.com/>



- опубликовано: August 2008

Вероятно есть перевод, но поскольку автор курса читал эту книгу в оригинале, то за качество перевода не отвечает ;)

Полезные ссылки для погружения в материалы учебного модуля:

- <http://python-requests.org/> + <https://pypi.org/project/requests/>
- BeautifulSoup: <https://www.crummy.com/software/BeautifulSoup/bs4/doc/#>
- Soup Sieve: <https://facelessuser.github.io/soupsieve/>
- Scrapy | A Fast and Powerful Scraping and Web Crawling: <https://scrapy.org/>
- lxml: <https://lxml.de/> + <https://lxml.de/parsing.html>
- W3School XML & XPath: https://www.w3schools.com/xml/xml_xpath.asp
- XPath functions: <https://developer.mozilla.org/en-US/docs/Web/XPath/Functions>

Всем удачи!