

Министерство образования и науки Российской Федерации
Санкт-Петербургский политехнический университет Петра Великого

Институт компьютерных наук и технологий

Кафедра «Программная инженерия»

Лабораторная работа №058

«Скачивание картинок с сайта»

**по дисциплине «Технологии разработки
качественного программного обеспечения»**

Выполнил:

Студент группы в3530904/90322

Демченко К.Т.

Преподаватель:

Смирнов Н.Г.

Санкт-Петербург

2022

Цель работы

Скачивание картинок с сайта. На входе - URL страницы.
Необходимо скачать все изображения на странице, а также на всех дочерних страницах, на которые можно перейти по гиперссылкам. Скачивать изображения только > X КБ.
Параметры приложения (URL, каталог для сохранения, размер X и др.) задавать в property-файле.

Описание работы приложения

На рисунке 1 изображена диаграмма последовательностей, на которой показано взаимодействие акторов: компьютера и сервера в рамках работы программы.

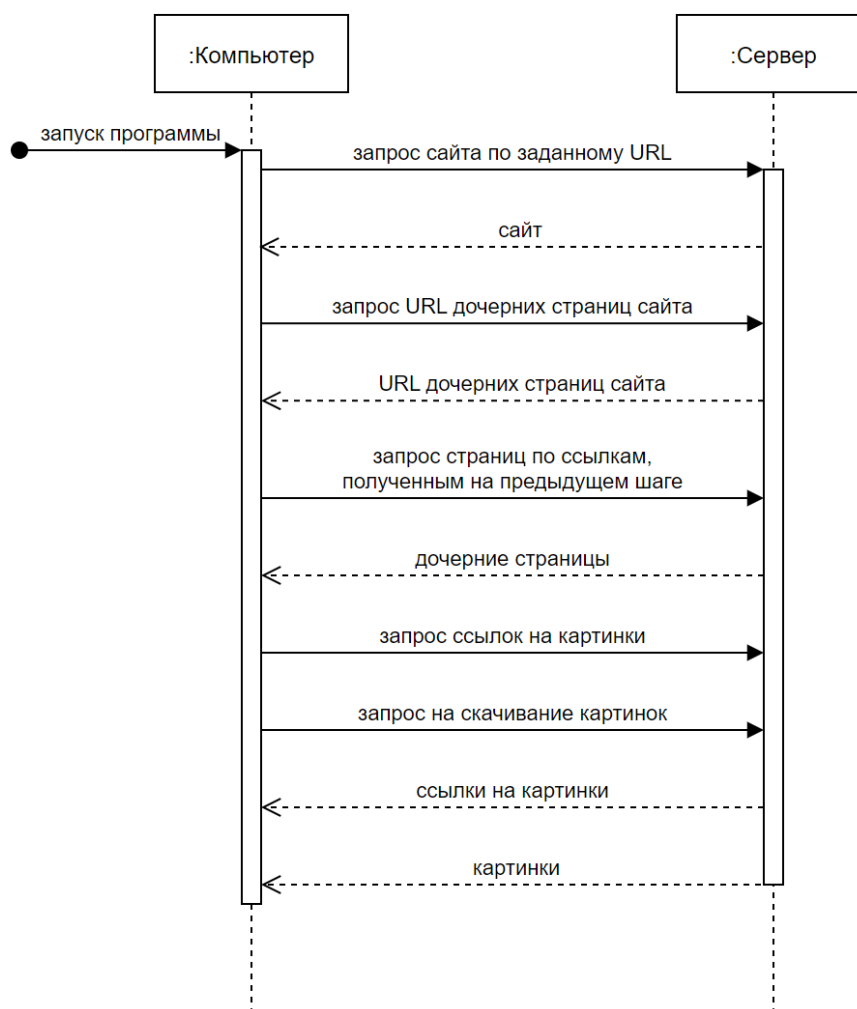
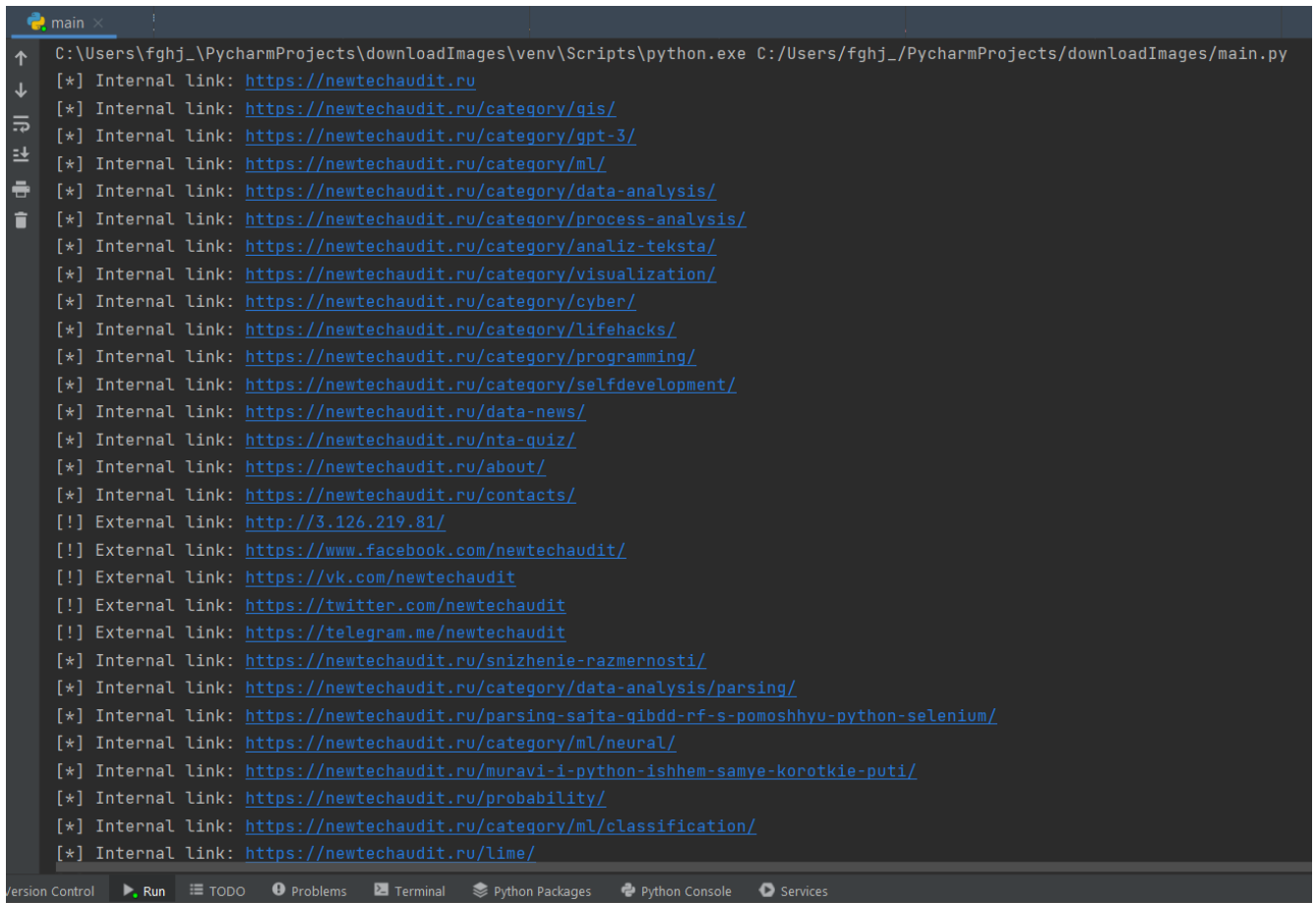


Рисунок 1 – Диаграмма последовательностей

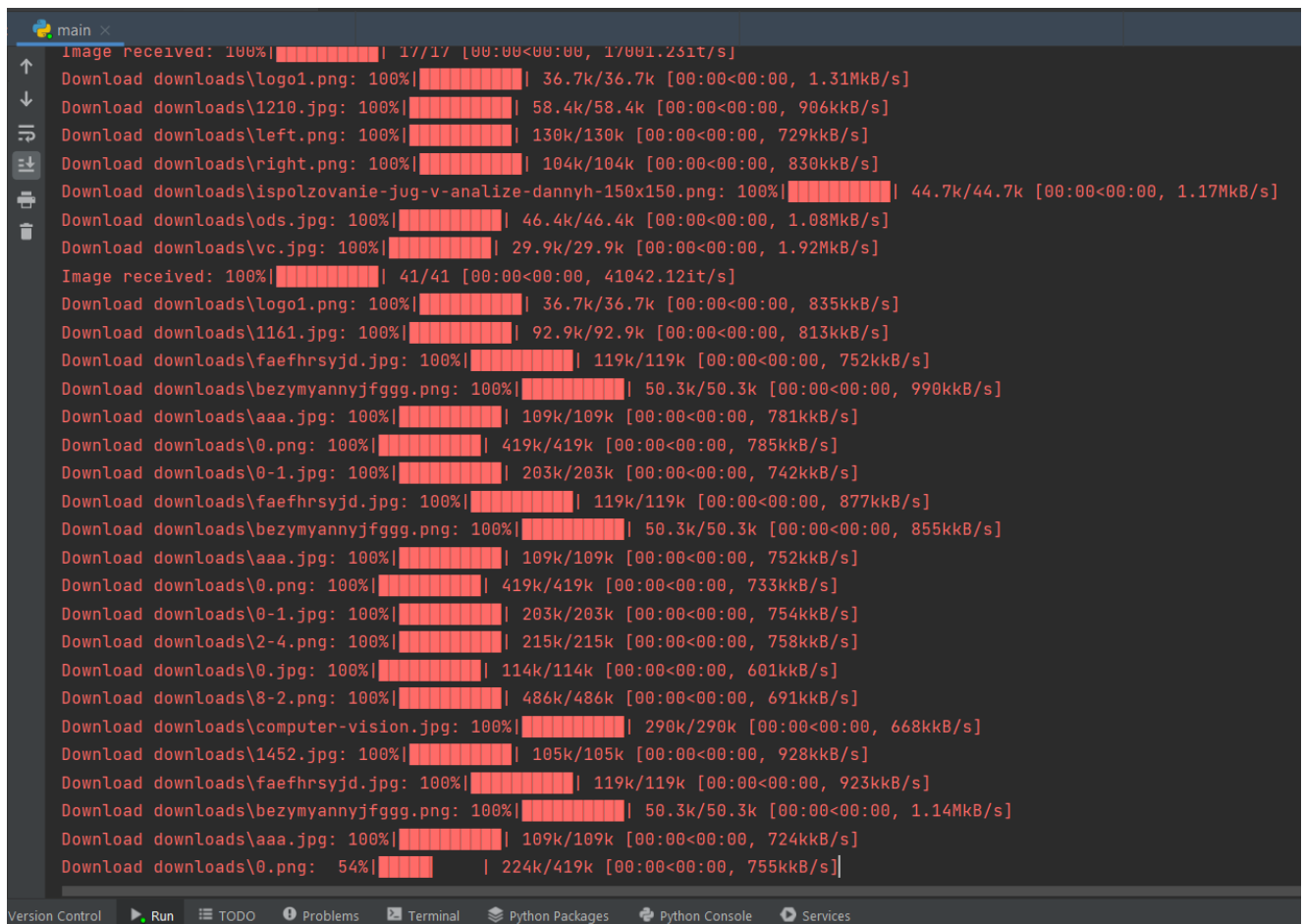
Скриншоты приложения

На следующих рисунках показан процесс работы приложения.



```
main x
C:\Users\fgjhj_\PycharmProjects\downloadImages\venv\Scripts\python.exe C:/Users/fgjhj_/PycharmProjects/downloadImages/main.py
[*] Internal link: https://newtechaudit.ru
[*] Internal link: https://newtechaudit.ru/category/gis/
[*] Internal link: https://newtechaudit.ru/category/gpt-3/
[*] Internal link: https://newtechaudit.ru/category/ml/
[*] Internal link: https://newtechaudit.ru/category/data-analysis/
[*] Internal link: https://newtechaudit.ru/category/process-analysis/
[*] Internal link: https://newtechaudit.ru/category/analiz-teksta/
[*] Internal link: https://newtechaudit.ru/category/visualization/
[*] Internal link: https://newtechaudit.ru/category/cyber/
[*] Internal link: https://newtechaudit.ru/category/lifehacks/
[*] Internal link: https://newtechaudit.ru/category/programming/
[*] Internal link: https://newtechaudit.ru/category/selfdevelopment/
[*] Internal link: https://newtechaudit.ru/data-news/
[*] Internal link: https://newtechaudit.ru/nta-quiz/
[*] Internal link: https://newtechaudit.ru/about/
[*] Internal link: https://newtechaudit.ru/contacts/
[!] External link: http://3.126.219.81/
[!] External link: https://www.facebook.com/newtechaudit/
[!] External link: https://vk.com/newtechaudit
[!] External link: https://twitter.com/newtechaudit
[!] External link: https://telegram.me/newtechaudit
[*] Internal link: https://newtechaudit.ru/snizhenie-razmernosti/
[*] Internal link: https://newtechaudit.ru/category/data-analysis/parsing/
[*] Internal link: https://newtechaudit.ru/parsing-sajta-qibdd-rf-s-pomoshhyu-python-selenium/
[*] Internal link: https://newtechaudit.ru/category/ml/neural/
[*] Internal link: https://newtechaudit.ru/muravi-i-python-ishhem-samye-korotkie-puti/
[*] Internal link: https://newtechaudit.ru/probability/
[*] Internal link: https://newtechaudit.ru/category/ml/classification/
[*] Internal link: https://newtechaudit.ru/lime/
```

Рисунок 2 – Вывод гиперссылок с заданного сайта



```
main x
Image received: 100%|██████████| 17/17 [00:00<00:00, 17001.23it/s]
Download downloads\logo1.png: 100%|██████████| 36.7k/36.7k [00:00<00:00, 1.31MkB/s]
Download downloads\1210.jpg: 100%|██████████| 58.4k/58.4k [00:00<00:00, 906kkB/s]
Download downloads\left.png: 100%|██████████| 130k/130k [00:00<00:00, 729kkB/s]
Download downloads\right.png: 100%|██████████| 104k/104k [00:00<00:00, 830kkB/s]
Download downloads\ispolzovanie-jug-v-analize-dannyh-150x150.png: 100%|██████████| 44.7k/44.7k [00:00<00:00, 1.17MkB/s]
Download downloads\ods.jpg: 100%|██████████| 46.4k/46.4k [00:00<00:00, 1.08MkB/s]
Download downloads\vc.jpg: 100%|██████████| 29.9k/29.9k [00:00<00:00, 1.92MkB/s]
Image received: 100%|██████████| 41/41 [00:00<00:00, 41042.12it/s]
Download downloads\logo1.png: 100%|██████████| 36.7k/36.7k [00:00<00:00, 835kkB/s]
Download downloads\1161.jpg: 100%|██████████| 92.9k/92.9k [00:00<00:00, 813kkB/s]
Download downloads\faefhrsyyd.jpg: 100%|██████████| 119k/119k [00:00<00:00, 752kkB/s]
Download downloads\bezmyannyjfggg.png: 100%|██████████| 50.3k/50.3k [00:00<00:00, 990kkB/s]
Download downloads\aaa.jpg: 100%|██████████| 109k/109k [00:00<00:00, 781kkB/s]
Download downloads\0.png: 100%|██████████| 419k/419k [00:00<00:00, 785kkB/s]
Download downloads\0-1.jpg: 100%|██████████| 203k/203k [00:00<00:00, 742kkB/s]
Download downloads\faefhrsyyd.jpg: 100%|██████████| 119k/119k [00:00<00:00, 877kkB/s]
Download downloads\bezmyannyjfggg.png: 100%|██████████| 50.3k/50.3k [00:00<00:00, 855kkB/s]
Download downloads\aaa.jpg: 100%|██████████| 109k/109k [00:00<00:00, 752kkB/s]
Download downloads\0.png: 100%|██████████| 419k/419k [00:00<00:00, 733kkB/s]
Download downloads\0-1.jpg: 100%|██████████| 203k/203k [00:00<00:00, 754kkB/s]
Download downloads\2-4.png: 100%|██████████| 215k/215k [00:00<00:00, 758kkB/s]
Download downloads\0.jpg: 100%|██████████| 114k/114k [00:00<00:00, 601kkB/s]
Download downloads\8-2.png: 100%|██████████| 486k/486k [00:00<00:00, 691kkB/s]
Download downloads\computer-vision.jpg: 100%|██████████| 290k/290k [00:00<00:00, 668kkB/s]
Download downloads\1452.jpg: 100%|██████████| 105k/105k [00:00<00:00, 928kkB/s]
Download downloads\faefhrsyyd.jpg: 100%|██████████| 119k/119k [00:00<00:00, 923kkB/s]
Download downloads\bezmyannyjfggg.png: 100%|██████████| 50.3k/50.3k [00:00<00:00, 1.14MkB/s]
Download downloads\aaa.jpg: 100%|██████████| 109k/109k [00:00<00:00, 724kkB/s]
Download downloads\0.png: 54%|██████████| 224k/419k [00:00<00:00, 755kkB/s]
```

Version Control Run TODO Problems Terminal Python Packages Python Console Services

Рисунок 4 – Вывод названий, размера загруженных изображений и времени их скачивания

Тестирование приложения

Для модульного тестирования приложения был использован модуль unittest. Всего было написано 8 тестов, разделённых на 4 класса, тестирующих 4 основных метода приложения.

На рисунке 5 показано успешное прохождение методами приложения всех тестов.

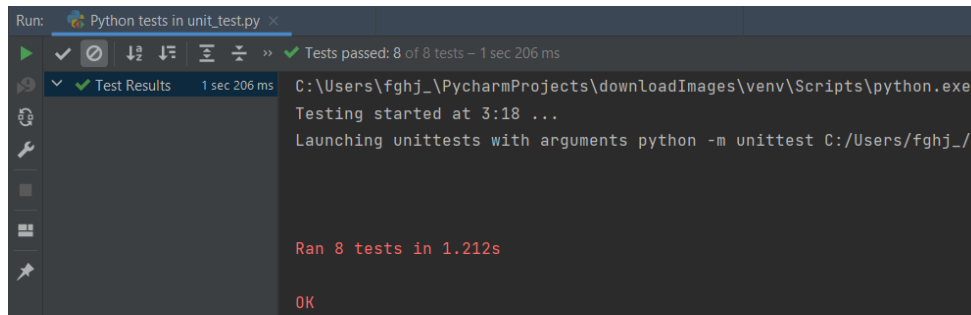


Рисунок 5 – Прохождение тестов

Использованные библиотеки и модули

1. Beautiful Soup — это библиотека-парсер для синтаксического разбора файлов HTML/XML, преобразовывает разметку в дерево синтаксического разбора. Он поддерживает простые и естественные способы навигации, поиска и модификации дерева синтаксического разбора.

2. Библиотека Requests открывает возможность использования API в Python, позволяет легко и с минимальным количеством кода взаимодействовать с веб-приложениями.

3. Модуль os предоставляет множество функций для работы с операционной системой, причём их поведение, как правило, не зависит от ОС, поэтому программы остаются переносимыми. Модуль os позволяет работать с файловой системой, с окружением, управлять процессами.

4. Библиотека tqdm позволяет создавать консольные линейные индикаторы прогресса и индикаторы прогресса с графическим интерфейсом.

5. Библиотека YAML довольно простая, поскольку единственная необходимая функциональность — это возможность анализировать файлы в формате YAML. Используется для хранения данных в виде сериализованных файлов конфигурации.

YAML (YAML Ain't Markup Language) - ещё один текстовый формат для записи данных. YAML более приятен для восприятия человеком, чем JSON, поэтому часто используется для описания сценариев в ПО.

6. Модуль urllib.parse определяет стандартный интерфейс для разбора URL-адреса на компоненты: протокол, порт, домен, путь и т. д., чтобы можно было объединять компоненты обратно в строку URL-адреса и преобразовать относительный URL-адрес в абсолютный URL-адрес с учетом базового URL-адреса.

7. Модуль unittest поддерживает автоматизацию тестов, использование общего кода для настройки и завершения тестов, объединение тестов в группы, а также позволяет отделять тесты от фреймворка для вывода информации. С его помощью можно собирать тесты в группы, собирать результаты выполнения тестов. ООП стиль позволяет уменьшить дублирование кода при схожих объектах тестирования.

Вывод

В результате была написана программа, позволяющая скачивать все изображения не менее заданного размера со страницы сайта и всех её дочерних страниц, на которые можно перейти по гиперссылкам. Код программы был покрыт модульными тестами.