

Домашние задания

Домашнее задание 1

1) Прочитать статью в Википедии:

<https://ru.wikipedia.org/wiki/Python>

2) Прочитать PEP 8

Кто дружит с английским, читаем тут:

<https://www.python.org/dev/peps/pep-0008/>

Кто не дружит с английским, подружитесь, но пока почитайте тут:

<https://pythonworld.ru/osnovy/pep-8-rukovodstvo-po-napisaniyu-koda-na-python.html>

3) Просмотреть все книги, которые лежат на Гугл Диске.

4) Все что Вам нужно знать о типах данных в Python написано на этой странице:

<https://docs.python.org/3/library/stdtypes.html>

Прочитайте эту страницу полностью. Пусть даже что-то будет непонятно, но все равно прочитайте. Проясняется будет дальше по ходу курса.

5) Также прочитайте книгу Python Advanced, книга лежит на Гугл Диске, до страницы 26.

6) Заходите в Google Play или App Store и пишите Sololearn Python. Устанавливаете приложение.

Находите там Курс по Python 3 и проходите первый модуль: "Начнем с Элементарного".

Прямо со смартфона. Скриншот о том что модуль пройден загружаем на Гугл Диск.

За загруженный скриншот с пройденным модулем - 10 баллов.

7) Скачиваете себе файл google.py из папки на ГуглДиске

https://drive.google.com/drive/u/0/folders/1selq3H2fv1jxNehP3tW8ge_1FnV2RYmO меняете в коде ключевое слово с online casino на купить вентилятор, запускаете, и сохраняете скриншот с результатом.

Скриншот загружаете на Гугл Диск. За загруженный скриншот - 10 баллов.

Итого, максимальная оценка за ДЗ - 20 баллов.

Домашнее задание 2

Читаем:

1) В книге Python Advanced страницы 39 (Контейнеры) - 53 (Функции)

2) <https://docs.python.org/3/library/stdtypes.html>

3) Проходим приложение Sololearn Python дальше.

Задание 1.

Распечатать документ Python Cheat Sheet из папки Гугл Диска.

Сфоткать и залить фотку в свою папку с ДЗ2
Оценка за задание - 5 баллов.

Задача 2.

Написать скрипт, в котором вручную задать 5 переменных url, title, description, h1, keyword.
Внутри этих переменных написать данные с какой-нибудь реальной html страницы.
Затем нужно определить количество символов и количество слов в url, title, description, h1.
Определить количество вхождений ключевого слова в title, description, h1.
Определить плотность ключевого слова в процентах в title, description, h1.
Всю информацию вывести в консоль в виде словаря в “красивом” виде.

Оценка за задачу 20 баллов.

Задача 3.

В папке на Гугл Диске (<https://drive.google.com/drive/u/0/folders/1EBn9m0M4IZ81lf0dCj-ks0bPNSXkWUKj>)
есть скрипт google_with_similar.py
Нужно определить каких типов данных все переменные в этом скрипте. Результат написать
комментариями (#) в самом скрипте.

Оценка за задачу 10 баллов.

Задача 4.

В папке на Гугл Диске - <https://drive.google.com/drive/u/0/folders/1EBn9m0M4IZ81lf0dCj-ks0bPNSXkWUKj>
лежит скрипт title.py
Внутри него сделать так, чтобы url можно было задавать из консоли.
Сделать так, чтобы в скрипте считалось: количество слов в Title, количество символов в Title,
количество уникальных слов, плотность слов, заспамленность и любые другие параметры title, которые
по вашему мнению влияют на SEO. Результат работы скрипта необходимо вывести в консоль в виде
словаря.

Оценка за ДЗ 20 баллов.

Итого, максимальная оценка за ДЗ - 55 баллов.

Домашнее задание 3

Задание 1

В книге Python Advanced читаем все до страницы 53 (Функции).

Задание 2

Проходим приложение Sololearn Python второй модуль “Управляющие конструкции”.
Загрузить скриншот на Гугл Диск. Оценка за скриншот - 5 баллов.

Задача 3

В прошлом ДЗ была задача 4 в которой вы считали количество уникальных слов, плотность слов,
заспамленность и любые другие параметры title, которые по вашему мнению влияют на SEO.
Развиваем эту задачу дальше. Вводим из консоли url и keyword. Дальше парсим со страницы title,
description, h1. Вычисляем суммарное качество оптимизации страницы в баллах от 0 до 100 на основе

анализа title, description, h1. Необходимо задействовать все новые для вас операторы: if, for, while, try.
Оценка за задачу 40 баллов.

Оценка за ДЗ - 45 баллов.

Домашнее задание 4

Задание 1

Прочитать <https://docs.python.org/3/library/functions.html> и особое внимание уделить функции **open**.

Задача 2.

Скачать базу Букварикс.

<https://www.bukvarix.com/english-keywords.html>

И написать скрипт выбора ключевиков из базы, по аналогии с тем, как это делалось на занятии.

Повторить код с занятия. Оценка за задачу - 30 баллов.

Задача 3.

Вспоминаем скрипт google.py с которым мы неоднократно работали.

<https://drive.google.com/drive/u/0/folders/1HR4zhQUIY-2iLxrlDb284HNzIDIKY5y>

Написать на базе него чекалку позиций сайта. Запросы получаем из внешнего файла.

Результат (запрос, url, позиция) сохраняем тоже в файл. Оценка за задачу - 40 баллов.

Оценка за ДЗ - 70 баллов.

Домашнее задание 5

Задание 1

Прочитать <https://wiki.python.org/moin/TimeComplexity>

Задание 2

Загрузить скрин приложения Sololearn с пройденным разделом Функции.

Оценка за задание - 5 баллов.

Задача 3.

Скрипт выборки ключевиков из базы Букварикс переписать с использованием функции.

Оценка за задачу - 30 баллов.

Оценка за ДЗ - 35 баллов.

Домашнее задание 6

Задание 1

Прочитать Python Advanced Book страницы 53 - 85.

Задание 2

Загрузить скрин приложения Sololearn с пройденными разделами до ООП (Объектно-ориентированное программирование). Оценка за задание - 10 баллов.

Задача 3

Написать упрощенный парсер, аналог программы Screaming Frog SEO Spider.

На вход программы задается адрес сайта, домен.

Программа начинает сканировать страницы и доставать из них title и h1.

Данные в результате записываются в файл scan.csv со структурой:

url, title, h1, url_level, response_server_time

В результате парсер должен пройти по всем страницам сайта, сканируя новые обнаруженные.

Позаботиться о том, чтобы одни и те же страницы не сканировались по несколько раз.

Оценка за задачу: 50 баллов.

Дополнительно:

Тем, кому будет легко, реализовать проверку разрешения сканирования урлов в robots.txt

Оценка за дополнение + 20 баллов.

Максимальная оценка за ДЗ - 80 баллов.

Домашнее задание 7

Задание 1

Прочитать Python Advanced Book до конца.

Задание 2

Загрузить скрин приложения Sololearn где пройдено всё до конца.

Оценка 5 баллов.

Задача 3

Написать 8 классов, с атрибутами и методами, которые вы считаете наиболее подходящими для данных классов: Website, Page, Link, Text, Book, Image, Title, Description. Некоторые классы могут наследоваться один от другого, некоторые могут не наследоваться ни от кого. Создать по 2 экземпляра каждого класса. Организовать взаимодействие между экземплярами классов. Например, из текстов можно создать книгу, из книги вытащить текст, страница (Page) - это элемент сайта, тайтл и дескрипшн это тоже текст с методами текста.

Оценка за задачу - 30 баллов.

Максимальная оценка за ДЗ - 35 баллов.

Домашнее задание 8

Прочитать:

1. Python Advanced Book до конца.

2. <https://habr.com/ru/post/72757/>

3. <https://habr.com/ru/post/337314/>

4. <https://habr.com/ru/post/140581/>

5. <https://habr.com/ru/post/455796/>

6. <https://habr.com/ru/post/145835/>
7. <https://habr.com/ru/post/421993/>
8. <https://habr.com/ru/post/422311/>
9. <https://habr.com/ru/post/415829/>
10. <https://habr.com/ru/post/339272/>

Задача:

Переписать парсер, аналог Screaming Frog SEO Spider из домашнего задания 6. Сейчас он написан в функциональном стиле. Необходимо переписать его взяв за основу ООП парадигму.

Оценка за задачу - 50 баллов.

Максимальная оценка за ДЗ - 50 баллов.

Домашнее задание 9

Прочитать:

1. Python Advanced Book до конца.
2. <https://requests-html.kennethreitz.org/>
3. <https://requests.readthedocs.io/en/master/>
4. <https://ru.wikipedia.org/wiki/XPath>

Задача 1:

Переписать парсер из прошлого ДЗ. Берем файл dz8_2.py из папки:

<https://drive.google.com/drive/u/0/folders/1K9BxcAWw6IPaubuSzkf5nrtT3O-UI-hM>

Расширить его функционал. Добавить сканирование тегов h2, h3, h4, h5, h6, p, img.

Оценка за задачу - 50 баллов.

Максимальная оценка за ДЗ - 50 баллов.

Задача 2:

Переписать парсер из прошлого ДЗ. Берем файл dz8_2.py из папки:

<https://drive.google.com/drive/u/0/folders/1K9BxcAWw6IPaubuSzkf5nrtT3O-UI-hM>

Расширить его функционал. Доработать его под кастомный парсер olx.ua.

Так, чтобы данный парсер можно было натравить на olx и в результате получить файл с данными обо всех объявлениях. Оценка за задачу - 50 баллов.

Максимальная оценка за ДЗ - 100 баллов.

Домашнее задание 10

Прочитать:

1. Python Advanced Book до конца.
2. <https://selenium-python.readthedocs.io/>
3. <https://habr.com/ru/post/464261/>

Задача 1:

Написать скрипт автоперевода на селениуме.

Перевод получаем тут - <https://www.bing.com/translator>

Скрипт заходит на страницу, вставляет фразу "hello world" и переводит ее на немецкий язык.

Результат перевода вытаскиваем из соседнего окна, принтим в консоль и записываем в файл.

Оценка за задачу - 80 баллов.

Задача 2:

На сервисе - <https://jsonplaceholder.typicode.com/>

Сделать запрос в API и получить все ссылки фотографий.

Записать их в файл txt.

Оценка за задачу - 20 баллов.

Максимальная оценка за ДЗ - 100 баллов.

Домашнее задание 11

Прочитать:

1. <https://developers.google.com/webmaster-tools/search-console-api-original/v3/quickstart/quickstart-python>
2. <https://best-proxies.ru/api/>
3. <https://developers.google.com/analytics/devguides/reporting/core/v4/quickstart/service-py?hl=ru>
4. <https://python-wordpress-xmlrpc.readthedocs.io/en/latest/>

Задача 1:

Написать скрипт, который делает запрос в Гугл по указанному ключу. Каждый запрос в Гугл делается со случайным юзер-агентом. Далее скрипт заходит на первые 3 сайта. Набирает по одному абзацу текста с каждого сайта. Далее переводит текст с русского на английский.

Используя библиотеку: <https://pypi.org/project/googletrans/>

Формирует из автоперевода статью и заливает на вордпресс (py4seo.com, admin, 123456), через xmlrpc.

Оценка за задачу: 120 баллов.

Задача 2:

Скачать данные из Google Search Console через API для любого своего сайта.

Результат решения задачи - скриншот, где видно, что это ваш сайт, и что это его данные.

Оценка за задачу: 30 баллов.

Оценка за ДЗ - 150 баллов.

Домашнее задание 12

Прочитать:

1. <https://www.nltk.org/>
2. <https://spacy.io/>
3. <https://talktotransformer.com/>
4. <https://pillow.readthedocs.io/en/stable/index.html>
5. <https://stackoverflow.com/questions/30227466/combine-several-images-horizontally-with-python>
6. <https://docs.opencv.org/master/>

7. <https://tproger.ru/translations/opencv-python-guide/>
8. <https://opencv-python-tutroals.readthedocs.io/en/latest/index.html>

Задача 1:

Скачать словарь синонимов.

http://usyn.ru/blog.php?id_blog=11

На основании него написать функцию синонимизации русского текста.

На вход получает исходный любой текст. На выходе возвращает синонимизированный текст.

Оценка за задачу: 70 баллов.

Задача 2:

Написать функцию, которая получает на вход ключевое слово.

Дальше делает запрос в Гугл Картинки, вытаскивает 4 картинки.

Скачивает их и соединяет в одну картинку. Сохраняет полученную составную картинку на диске.

Оценка за задачу: 30 баллов.

Оценка за ДЗ - 100 баллов.

Домашнее задание 13

Прочитать:

1. <http://docs.peewee-orm.com/en/latest/>

Задача 1

В домашнем задании номер 9 в задаче 2 заменить сохранение данных в файл, на сохранение данных в базу данных Postgres.

```
connection = {  
    'user': 'py4seo',  
    'database': 'library',  
    'host': '46.30.164.249',  
    'password': 'PY1111forSEO'  
}
```

Ваша таблица должна называться "имя_фамилия".

Оценка за задачу 100 баллов.

Максимальная оценка за ДЗ - 100 баллов.

Домашнее задание 14

Прочитать:

1. <http://docs.peewee-orm.com/en/latest/peewee/relationships.html>
2. <https://database.guide/the-3-types-of-relationships-in-database-design/>
3. [https://en.wikipedia.org/wiki/Thread_\(computing\)](https://en.wikipedia.org/wiki/Thread_(computing))
4. <https://docs.google.com/presentation/d/1vNVTTI2hXWtMJ4MFrEkkYgWmbpYK1dBJHxpHWjJQGZo>

Задача 1

Написать парсер любого интернет-магазина, в котором парсятся: товары и категории в отдельные таблицы и связываются между собой методом ManyToMany. Результат записываем в таблицы базы. Таблицы называем с привязкой к своим ФИО.

```
connection = {  
    'user': 'py4seo',  
    'database': 'library',  
    'host': '46.30.164.249',  
    'password': 'PY1111forSEO'  
}
```

Оценка за задачу 100 баллов.

Задача 2

Написать 2 функции. Первая функция реализует получение ссылок из выдачи Google по ключевому слову. Вторая функция получает ссылки из выдачи bing.com по тому же ключевому слову. Обе функции используют селениум (headless). Когда запускается скрипт, пользователь вводит в консоль ключевое слово и нажимает Enter. Скрипт запускает оба парсера и Google и Bing в двух разных потоках.

Результаты и Бинга и Гугла принтытся и в консоль, и записываются в один общий файл results.txt.

Оценка за задачу 100 баллов.

Максимальная оценка за ДЗ - 200 баллов.

Домашнее задание 15

Прочитать:

1. <https://docs.python.org/3/library/concurrent.futures.html>
2. <https://habr.com/ru/post/229767/>
3. <https://docs.python.org/3/library/asyncio.html>
4. <https://habr.com/ru/post/282972/>
5. <https://github.com/aio-lib>
6. <https://habr.com/ru/post/337420/>

Задача 1

На диске в папке <https://drive.google.com/drive/u/0/folders/1UvubgSa5s827kQhIRKvg8o-SVBdvWTZ>

Лежат файлы case1-6. Взять любой из этих файлов и реализовать сохранение спаршеных данных (href, name) в многопоточном парсере в базу данных. Используем Peewee.

Оценка за задачу 100 баллов.

Задача 2

На диске в папке <https://drive.google.com/drive/u/0/folders/1UvubgSa5s827kQhIRKvg8o-SVBdvWTZ>

Лежит файл as_parser2.py. Взять этот файл и реализовать сохранение спаршеных данных (href, name) в этом асинхронном парсере в базу данных. Используем Aiopg (+ SQLAlchemy)

<https://aiopg.readthedocs.io/en/stable/>.

Оценка за задачу 200 баллов.

Максимальная оценка за ДЗ - 300 баллов.