

Тестовое задание для Senior Python Developer

Написать программу на Python 3, которая называется `scan.py` и которая в качестве аргумента командной строки принимает адрес произвольной веб-страницы.

На заданной веб-странице программа должна найти весь текст, который содержится непосредственно внутри тегов `<div>`. В результате должен быть чистый текст без html внутри себя. Далее этот текст нужно разбить на слова и подсчитать, сколько раз каждое слово встречается. Регистр необходимо игнорировать, то есть, например, слова `python` и `Python` считаются одинаковыми. Также не нужно считать стоп-слова, список которых можно найти здесь: <https://www.ranks.nl/stopwords> (работаем с английским языком). Знаки пунктуации тоже игнорируются.

Программа должна распечатать на экран (вывести в `stdout`) заголовок этой веб-страницы (содержимое тега `<title>`), и, после пустой строки, вывести все слова в порядке от самых часто встречающихся к самым редким, с указанием количества, сколько раз встретилось каждое слово.

Пример вызова программы:

```
./scan.py https://www.poetryfoundation.org/poems/48860/the-raven
```

Вывод программы (для краткости показаны только первые 7 слов):

The Raven by Edgar Allan Poe | Poetry Foundation

door	14
nevermore	11
chamber	11
raven	10
bird	10
lenore	8
thy	7

Что будет оцениваться:

- Корректность и работоспособность программы
- Какие сторонние библиотеки были выбраны
- Насколько эффективно применяются эти библиотеки (знание их API)
- Организация исходного кода (на какие функции разделен код, какие названия даны функциям и переменным, и т. п.)
- Читаемость, модульность исходного кода
- Стиль написания исходного кода (PEP8 приветствуется)
- Наличие юнит-тестов к функциям
- Наличие и содержание комментариев или docstrings

Таким образом, исходный код не менее важен, чем итоговая программа!

Приветствуется аккуратность и внимание к деталям:

- Выравнивание вывода программы (колонки слов и чисел, как в примере)
- Преобразование escape-последовательностей html в нормальный текст (например, `—`; это тире)
- Валидация аргумента командной строки
- и т. п.

На этих веб-страницах можно тестировать программу:

<https://www.poetryfoundation.org/poems/48860/the-raven>

<https://www.azlyrics.com/lyrics/ironmaiden/rimeoftheancientmariner.html>

Результат рекомендуется опубликовать с помощью Github Gist.