

BLDUP, LLC
Junior Python Developer
Test Task

Привет! Это команда BLDUP. Очень надеемся, что тебе будет интересно выполнять наше тестовое задание. Сейчас нам важно увидеть структурированный, понятный код и огромное желание, поэтому ничего страшного, если будут ошибки. Так же мы готовы ответить на все твои вопросы в ходе выполнения задания (даже если они покажутся тебе глупыми). Удачи!

Необходимо получить данные с веб-сайта:

<http://www.tauntondeeds.com/>

Конкретно для этого задания требуется собрать данные со страницы (раздел Land Court):

<http://www.tauntondeeds.com/Searches/ImageSearch.aspx>

Мы ценим твое время и не хотим растягивать выполнение задания на несколько дней, поэтому финальное задание звучит следующим образом:

Получить данные по всем DEED документам за текущий год (2020).

Достаточно парсить только данные с preview (то, что появляется при нажатии на кнопку view, парсить не нужно)

Вывести в output данные в формате json (выводить можно по одному, не нужно накапливать все).

Список необходимых полей и их формат:

date: datetime
type: str
book: str
page_num: str
doc_num: str
city: str
description: str
cost: float
street_address: str
state: str
zip: str

Для некоторых полей нет отдельных колонок, однако вся информация есть в других колонках, значения которых и придется распарсить. Значения некоторых полей могут отсутствовать, поэтому нужно использовать None.

Данное приложение должно быть написано на Python (версии ≥ 3.4) с использованием Scrapy.

Все остальные пакеты по вашему усмотрению (однако не стоит этим увлекаться, например, устанавливать Java, чтобы парсить текст).

Есть несколько рекомендаций:

- * Реализуйте базовый класс на основе scrapy.spiders.Spider и наследуйте от него ваш скрапер для указанного выше сайта
- * Называйте правильно функции и переменные (плохо: Var1, хорошо: crawled_deeds)
- * Используйте type hints

Мы не устанавливаем временных рамок на выполнение задания, однако просим указать, сколько времени (суммарно, за вычетом перерывов) ты потратил на реализацию.