

### Тестовое задание

Используя фреймворк Scrapy необходимо написать парсер для получения информации о товарах интернет-магазина из списка категорий по заранее заданному шаблону, информация должна быть структурирована в виде словарей согласно формату выходных данных и сохранена в файл формата json.

На вход подается список категорий в виде ссылок (минимум 3), с количеством товаров от 100 штук на категорию

на сайте alkoteka.com (Например:  
https://alkoteka.com/catalog/slaboalkogolnye-napitki-2)

Обязательно осуществлять сбор данных с учетом выбранного региона для парсинга - Краснодар

По возможности для получения информации добавить возможность использовать подключение через прокси

### Входные данные

Изменяемый список ссылок на категории, можно выполнить в одном из двух вариантов:

1. В виде константной переменной формата: START\_URLS = ["url\_1", "url\_2", ..., "url\_3"]
2. В виде отдельного файла, который будет содержать список ссылок в произвольном формате

**\*\*\* Обратите внимание, что тестовое задание выполняется только на Scrapy (без использования Playwright, Headless, Selenium и тд)**

## **Выходные данные**

Формат выходных данных для одного товара:

*Это общий шаблон формата данных, и он не учитывает особенности сайта, выбранного для задания, если вы не смогли найти какие-то данные во время анализа сайта, все поля все равно должны быть собраны согласно стандартным значениям, так как являются обязательными для корректной проверки.*

```

1: {
2:   "timestamp": int, # Дата и время сбора товара в формате timestamp.
3:   "RPC": "str", # Уникальный код товара.
4:   "url": "str", # Ссылка на страницу товара.
5:   "title": "str", # Заголовок/название товара (! Если в карточке товара указан
   цвет или объем, но их нет в названии, необходимо добавить их в title в
   формате: "{Название}, {Цвет или Объем}").
6:   "marketing_tags": ["str"], # Список маркетинговых тэгов, например:
   ['Популярный', 'Акция', 'Подарок']. Если тэг представлен в виде
   изображения собирать его не нужно.
7:   "brand": "str", # Бренд товара.
8:   "section": ["str"], # Иерархия разделов, например: ['Игрушки',
   'Развивающие и интерактивные игрушки', 'Интерактивные игрушки'].
9:   "price_data": {
10:     "current": float, # Цена со скидкой, если скидки нет то = original.
11:     "original": float, # Оригинальная цена.
12:     "sale_tag": "str" # Если есть скидка на товар то необходимо вычислить
   процент скидки и записать формате: "Скидка {discount_percentage}%".
13:   },
14:   "stock": {
15:     "in_stock": bool, # Есть товар в наличии в магазине или нет.
16:     "count": int # Если есть возможность получить информацию о
   количестве оставшегося товара в наличии, иначе 0.
17:   },
18:   "assets": {
19:     "main_image": "str", # Ссылка на основное изображение товара.
20:     "set_images": ["str"], # Список ссылок на все изображения товара.
21:     "view360": ["str"], # Список ссылок на изображения в формате 360.
22:     "video": ["str"] # Список ссылок на видео/видеообложки товара.
23:   },
24:   "metadata": {
25:     "__description": "str", # Описание товара
26:     "KEY": "str",
27:     "KEY": "str",
28:     "KEY": "str"
29:     # Также в metadata необходимо добавить все характеристики товара
   которые могут быть на странице.
30:     # Например: Артикул, Код товара, Цвет, Объем, Страна производитель
   и т.д.
31:     # Где KEY - наименование характеристики.
32:   }
33:   "variants": int, # Кол-во вариантов у товара в карточке (За вариант
   считать только цвет или объем/масса. Размер у одежды или обуви
   вариантами не считаются).
34: }

```

## Ожидаемый результат

Результатом выполненного задания ожидается парсер который можно запустить командой

```
1: scrapy crawl spider_name -O result.json
```

После отработки парсера должен быть результирующий файл с собранными данными товаров по заданным категориям в формате, установленным согласно выходным данным.

Плюсом будет использование прокси (хотя бы прокидывание в запросы бесплатных прокси)

Плюсом будет сбор данных по заданному региону (могут отличаться цены, наличие и прочие данные для сайта)

Чем больше навыков Вы продемонстрируете в тестовом задании, тем лучше. (ООП/прокси/middleware)