



Тестовое задание на стажера тех отдела

1. Задание по парсингу

Сейчас активно развивается направление парсинга и скрапинга данных. Кратно растет количество заказов на подобные услуги, а также собираются датасеты для шоурума в Датамаркет. В связи с этим, первое тестовое задание будет в этом направлении:

1. Написать парсер магазина auto.ru (<https://auto.ru>). Собрать до 5-и фото (в максимальном качестве) из одного объявления. Объявлений брать до 10-и штук. Спарсить нужно первые 10 марок из списка на главной странице:

Легковые автомобили

376 987 предложений, из них 79 458 размещены только на Авто.ру

Все Новые С пробегом ☐ В кредит

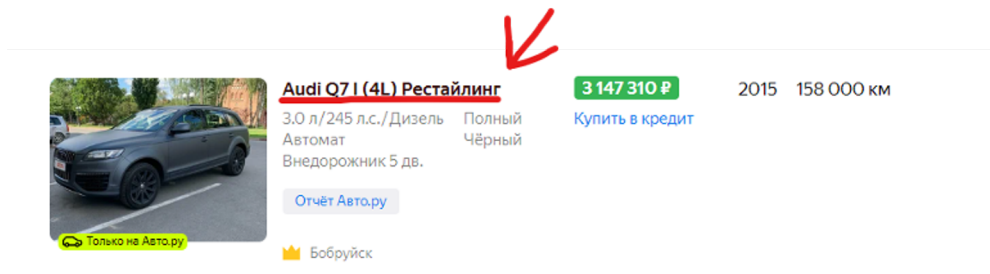
Марки Помощник

Американская классика

1 LADA (BAZ) 54323	8 EXEED 6157	Kia 20350	OMODA 7100	Toyota 25498
2 Audi 9373	9 Ford 9876	Land Rover 3941	Opel 5927	Volkswagen 15952
3 BMW 17273	10 Geely 5622	Lexus 4900	Peugeot 2991	Volvo 2881
4 Changan 7334	Haval 9120	Mazda 7012	Porsche 2411	ГАЗ 2683
5 Chery 13400	Honda 6174	Mercedes-Benz 17132	Renault 8269	Москвич 4643
6 Chevrolet 9004	Hyundai 19678	Mitsubishi 9381	Skoda 7152	YAZ 3660



Сортировка: папка с названием марки автомобиля, внутри папки с фото, с названием объявления, в каждой из которых фото:



Нужно прислать архив с фото и код парсера.

2. Задание по разметке

Не пугайтесь названия. Суть задания сугубо в работе с данными, в их обработке и анализе.

Прикреплено 3 xml файла: annotations.xml, annotations-2.xml, annotations-3.xml

Ваша задача получить количество:

1. Всего изображений
2. Всего изображений размечено
3. (если есть) Неразмеченных изображений
4. Статистика по классам:

Например:

класс1 : 100 фигур (бокс, полигон и тд - всё вместе)

класс2 : 200 фигур (бокс, полигон и тд - всё вместе)

5. Количество фигур (всего)

5.* Статистика по фигурам:

Например:

полигоны: 4444 шт

боксы: 1000 шт

6. Название и параметры (широта, высота) самого большого изображения и самого маленького. Если их будет несколько, то вывести на экран количество таких изображений и выдать инфу одного любого изображения.
7. Изменить id-шники изображений - сделать их в обратном порядке.
8. Изменить пате изображений - поменять расширение на 'png'.
9. Изменить пате изображений - удалить путь к файлу, оставить только название самого файла.

Написать 4 скрипта, скрипт должен работать для всех разметок

1 скрипт:

1, 2, 3, 5, 6 - общая статистика

2 скрипт:

4 - Статистика по классам

3 скрипт:

5* - Статистика по фигурам

4 скрипт:

7, 8, 9 - изменение xml файла

Пояснения:

фигуры: бокс, полигон, точка

класс: ignore, fur, item

Файлы задания

annotations.xml 170.0KB

annotations-2.xml 897.2KB

annotations-3.xml 458.7KB

3. Задание по маскам

Ваша задача сделать маски по заданным данным.

Прикреплен архив.

В папке «Задание2» лежит masks.xml и папка с 8 фотографиями.

В файле разметка 8 фотографий, задача сделать 2 вида масок. 1) наложить маску на оригинал, 2) Сделать маску на черном фоне. Класс Ignore должен быть черным цветом на маске и прозрачным на фото. Пример ниже.



Пояснения:

Для отрисовки масок следует использовать библиотеку работы с изображениями Pillow.

Обратите внимание на последовательность наложения слоев при рисовании маски.

Файлы задания

Задание2.zip 3112.3KB

Время выполнения заданий: 7 дней.

Форма сдачи: ссылка на репозиторий GitHub.

