Web Scraping and Social Media Scraping OTOMOTO(second-hand car selling website)

Names and ID's of all participants

397090 Wojciech Konarzewski

433886 Min-Jhen Wu

Note
✓ Names and IDs
✓ Topic
✓ Mechanics
✓ Output
✓ Data analysis
✓ Detailed description

Description

1. Short description of the topic and the web page.

   This web page is full of the information of cars. There are numbers of brands, types and price for each car. We picked 24 items up from the page and then scraped the pages. (eg. source, category, brand,…etc.)

2. Short description of your scraper mechanics.

   ### A. BEAUTIFUL SOUP

   • Searching for number of pages with offers

   • Creating list of links to all pages

   • Creating list of links to offers from all pages

   • Scraping desired data from every offer

   ### B. SCRAPY

   • Similar to BEAUTIFUL SOUP, there are two spiders

   • First scrapes links to offers and put them in csv

   • Second takes the csv file and then scrapes data from those links previously scraped

   ### C. SELENIUM

   • Open the website and accept the cookies

   • Collect every cars' link from all/100 pages

   • Use the link collected above enter each them

   • Scraping desired data from each page of car

3．Short technical description of the output you get.

    We scraped 23 items are following information.

1. Source – where was the car come form(Firm or Private person)
2. Category – we set as only Personal
3. Brand – we set as Opel
4. Model – we set as Astra
5. Generation – lifecycle of a design for that specific model
6. Year – year of production
7. Mileage – a number of miles car travelled or covered
8. Capacity – the maximum amount that something can contain
9. Fuel – kind of fuel for the car
10. HorsePower – indicate the power produced by a car's engine
11. Transmission – Manual or Automatic
12. Drive – the car is fronted wheel drive or backed wheel drive
13. Type – type of the car
14. Doors – number of doors for the car
15. Seats – number of seats in the car
16. Color – color of the car
17. Origin – where was the car origin from
18. First Own – if the car is new or not
19. Not_Crashed – if the car is crashed or not
20. ASO service – if the car has Administrative services or not
21. Used – if the car has been used or not
22. Price – price of the car
23. Currency – currency of the price of the car

4．Data analysis

We can use the data we scraped to see that how many offers are for diesel or how many for gasoline. To see that in the future it can be checked whether people get more ecological and tend to sell their diesels. Also, we can see what color or model of the car which people tend to purchase it.

5．Detailed description

397090 Wojciech Konarzewski - Beautiful Soup / Scrapy

433886 Min-Jhen Wu - Selenium / Description