Nanyang Technological University



**Web Crawling and Web Scraping of sg.carousell.com**

H6752 Data Extraction Techniques

Individual Assignment

|  |  |
| --- | --- |
| **Name** | **Matriculation No.** |
| ALBERT LENG CHAN SENG | G1801172E |

March 2020

## **TASK**

To use Selenium and BeautifulSoup to crawl and scrap a website of your choice (have to be legal).

**PROBLEM STATEMENT**

My family needs to buy a baby chair for our baby boy. We have decided to purchase it from carousell.com. The key criteria which we use to decide which baby chair to buy are its price, its condition, the seller’s ratings and the seller’s location. Searching from carousell.com directly often returns a page full of unnecessary and distracting details like advertisements and sponsored items.

Our goal is to collect only the items which match our criteria efficiently and effectively. We need tools which automate the whole process and return the result datasets in a clear and concise manner. By analyzing the datasets, we can decide which baby chair to purchase. We will use Selenium and Webcrawler to achieve this goal.

Below are summary of item and the key criteria:

Website: <https://sg.carousell.com/>

Item: Baby Chair

Condition: New

Seller’s location: Within 5km from Woodlands

Price Range: $5 to $150

Seller’s Ratings: The more stars the better. The maximum number of stars is five.

The results should be returned from the most recent first as we prefer to buy latest posted item.

**COLLECTION OF DATA**

**ANALYSIS OF RESULT**

## **REFERENCES**