Assignment 3 – Web Scraping

INDIVIDUAL ASSIGNMENT

(1) Do the following IN YOUR BROWSER, no programming required. Please write down your answers (concise written answers please).

Use your browsers development tools. Open the network tab and analyze the network for the following:

a) go to https://www.ebay.com and search for "lg phone"

b) what type of search request is eBay using, GET or POST?

c) which URL variable represents the search term?

d) click on “Auction”. Which URL variable represents auction searches?

d) can you come up with a shorter URL that produces the same search result page?

e) click on the next search result page and observe how the URL changes. What variable in the URL identifies the page number?

f) what is the feature common to each item in the search results page? I.e., what item do we need to select to obtain each item among the search results?

g) identify the number of bids for each item in search results. What do they have in common? How does it look in the HTML source code?

(2) Let's program!

a) Use the URL identified above and write code that loads eBay’s search result page for "**lg phone**". Save the result to file. (Please give it a meaningful filename. E.g., "ebay_lg_phone_01.htm".)

b) Take your code in (a) and write a loop that will download the first 10 pages of search results (or the maximum result page number, whichever is less). Save each of these pages. IMPORTANT: Each page request needs to be followed by at least a 10 second pause! Remember, you want your program to mimic your behavior as a human and help you make good purchasing decisions.

c) Write a separate piece of code that loops through the pages you downloaded in (b) and opens and parses them into a Python or Java xxxxsoup-object. Next find the **number of bids** of each item on each search result page and **print them to screen along with each item's URL**.

What to Turn In:

- a single PDF with your written solutions
- please submit your code snippets along with the console output in the PDF

Late work policy:

The deadline is strict: homework submitted after 9am will be considered late.