

Homework #1

Task:

In this homework, you will perform a webscrape on the requested site and output the .csv in the format requested. In this task, you will create a python script that exports the data into a single .csv files. The definition of the csv output is defined below. For submission, you only need to submit your .py file. For this scenario, you will be scraping phone related data into a spreadsheet.

Submission Requirements:

- Python file used to scrape and create the necessary .csv files
- Site to scrape: <http://drd.ba.ttu.edu/2020c/isqs6339/hw1/index.php>
- You also need to scrape the associated mobcards that are available on urls related to each mob listed on the URL above.
 - ✎ Your scraper should dynamically acquire those URLs to scrape further. i.e. If I added another phone, your scraper should work with no code changes.

Grading:

- You will be graded based upon:
 - ✎ Quality of your code
 - ✦ Note, I am not looking for the most efficient code, though I do want to see how few of web request you can perform to gather the data. I am looking for code that is well documented (i.e. commented) and follows a logical progression. Your goal is to write code that another developer could pick up and know what you are doing.
 - ✎ Adhering to best practices listed in the lectures
 - ✦ Example: Variables that I can change to run your code in my environment. I should not have to look through your code for items to change. These should be listed at the top of the file. i.e. filepaths, urls, etc.
 - ✎ Correctly generating .csv files by the requested standards (displayed on page 2).

Hints

- In many cases, I have reused classes on the webpage for spans. I would suggest you look at using a find_all and using array notation to access as it may make the problem much easier.

CSV Definitions: Your code should produce the following file with fields in this order.

- File #1:
 - id – *hint*, this is stored in the href. There are many ways to get this value, but I suggest an split on “=” might be useful, once you access the href.
 - Model
 - Product_Size
 - Color
 - Battery
 - Storage
 - Network
 - OS
- File #2
 - id – Value repeated as necessary
 - Model – Value repeated as necessary
 - Camera_Front (1 bullet per line in csv file)
- File #3
 - id – Value repeated as necessary
 - Model – Value repeated as necessary
 - Camera_Back (1 bullet per line in csv file)

Hint Your code should only scrape URLs that are visible (displayed on the website).