Matthew Wolfgram

SI 206 003

3/25/2018

Project Proposal—Revised

I plan to crawl the PhoneArena website (https://www.phonearena.com/ ,) crawling the list of different phone manufacturers, then displaying a list of their available models once a user selects a brand on the command line. The HTML for the manufacturers and each phone are cached. After the caching has been executed, the HTML stored in it is parsed and compiled into two tables. The first table is generated as soon as the user selects a brand—it contains the brand name and the model ID affiliated with each specific model. Once a specific phone is chosen from the brand, the 100 listings are put into a second table that compiles data about the RAM, memory, battery, release date, pixel density, screen-to-body ratio, chip type, performance specifications, camera specifications, and screen size.The second table is linked to the first via the model ID associated with each phone. Once this displayable data is compiled, the user can choose to generate Plotly graphs of certain items in the table—if specified, a graph could be generated about the distribution of performance specifications, showing how certain phones compare to each other. The user could generate similar distribution graphs for camera size, pixel density, screen size, and battery size.

***Challenge Score:*** Crawling [and scraping] multiple pages in a site you haven’t used before ✣ (8) = ***8 points***