

Designing an algorithm for web scraping using python

Task given by the placement cell

WEB SCRAPING: Web scraping refers to the extraction of data from web, In machine learning language making un-structured data to a very meaningful structured data.

When I am looking your task, I was really excited to do this, because I am very interested in doing projects on web development, so I do this web project as a by-product of my knowledge. In the beginning of this project, I imported the required python standard libraries then I created a .csv file to store the data which is scraped (collected) from the amazon web page. After I defined a function to handle the scraping process by various statements such as requesting the server to get necessary data.

As per the given ASIN code, I created a URL for each ASIN code to scrape the data likewise name of product, description, images link, rating of manufacturer and total rating of the product.

Problem faced:

- Most of the ASIN code are outdated.
- All ASIN code doesn't have same HTML.
- Many products don't contain required data.

Advantage:

- This algorithm can handle the outdated ASIN code.

Limitations:

- Algorithm can only work for ASIN code.
- User cannot give the URL as an input.

Algorithm is successfully satisfying the condition that completion upon a single execution and output of the algorithm is in .csv file.