

In-class Participation Assignment

Group 2

MIE 1624H

Due: January 23, 2017

Web scraping or *web data extraction* is an automatic software technique for extracting information from web-sites. As not all e-commerce web-sites provide APIs, one of the most popular areas for web scrapping is *online retail*.

Find an online retailer web-site where web scrapping is allowed. Parse HTML pages with products of the same type to extract product name, product price, indication if a product is on sale, etc. You can use `BeautifulSoup` library in Python for pulling data out of HTML and XML files.

Explain how web scrapping works based on identifying a structure of HTML tags in web-pages. Provide IPython example of your web scrapping algorithm for a web-site of your choice. In your implementation, loop over 3-4 web-pages with products sold on that web-site. Print a table with your scrapping results and export those as a `csv` file.

Prepare 12-15 minute presentation of your results. After the presentation, send your PowerPoint slides and IPython Nootebook `ipynb` file to the course instructor or upload via Blackboard, such that those can be posted on the course web-page and re-used by your colleagues for assignments and a course project.