

Programming: Text Classification
Nitya Mathur
Due Date: January 21, 24

Web Scraping

This was my first time attempting web scraping. Hence I have only uploaded what time allowed. I will try the rest of the text classification tasks on the data I have extracted.

Topic for search of articles: Stocks
News Website: New York Times
Number of articles scraped: 200

Approach: I was able to extract text from 1 html file using BeautifulSoup but was unsure of how to scrape multiple articles. Hence I also used Selenium to get the data. I formatted the code in such a way that I can change the search query from Stocks to whatever I want on the NYT website to scrape many different articles.

Tools used:

1. BeautifulSoup: Python library to pull data from HTML or XML files. Useful for web scraping and easy to use once familiar with data.
2. Selenium: Python tool to mainly perform automated test cases for browsers or web applications but can also be used for web scraping

GitHub: <https://github.com/nitya0812/Web-Scraping/>