Northeastern University

Course: CS6020

Assignment: Week 6 - Web Scraping through Toolkits

Total Points: 100

Date Due: Posted on Blackboard

Learning Objectives

In this assignment, you will learn how to:
- scrape data from HTML through a toolkit
- specify search parameters through a URL
- compare toolkits and write a report

Tasks

The objective of this assignment is to research web scraping toolkits and compare them through a case study.

A. (20 Points) Pick at least 3 web scraping toolkits and use them to extract the data from a web site of your choice (e.g., YellowPages.com, BoatTrader.com, CityOfBoston.gov).
B. (50 Points) Extract the data into a data file format of your choice and import that data file into a data frame in R. SHow an output of the data frame and put screen shots into your report for question C.
C. (30 Points) Write a report that compares the tools with a focus on cost, ease of use, features, and your recommendation. Discuss your experience with the tool chosen for question B.

Complete the tasks and attach your report as a PDF to your submission. Include pictures, screenshots, data file extracts, charts, and anything else that shows your analysis of the toolkits.