

Web Scraping Project

The goal of this assessment exercise is to design and execute a data acquisition process that involves web scraping using at least one of the technologies introduced in the course.

You should write a report that describes the process performed and accompany it with an R code that allows to reproduce the process (if possible from beginning till end, if not, explain).

The project should emulate as much as possible a real scraping process. Following the proposal you have elaborated in class it should include:

- Your goal
- The data source and web technology it is based upon
- The approach/technology that you will use for downloading
- How you have scraped the data (be detailed, but not unnecessarily exhaustive).
- If it makes sense, show (or provide) the results of the scraping process.
- If it is reasonably short you can do an analysis of the data. If you think it adds too much extra work, explain how you would perform the analysis.
- Finish the report with a brief discussion of the limitations that you have found to do the scraping and what do you think that you should have done in a real situation

Remember

- Accompany your instructions with reproducible code, preferably in an appendix and/or an additional file.
- Name your exercise (all the files) with your last names and your first name using the following pattern:
LastName1_FirstName1.LastName2_FirstName2_Module3_Assessment.* (*= .R, .pdf, .txt, ...) (You may use "Report" instead of "Module3_Assessment")
- Put everything in a folder in your github repo and submit (push) it and send me a message after you have done this.

The final date for submission is July the 8th

Let me know if you need extra time or if have any doubts or need help

Alex