Screen scraping with Python

## Python code to extract data from website and a macro to combine

Extracting website information is a common task which we come across daily. This could become a tedious work when we want to extract a lot of contact or other information from multiple pages. So, this python code will extract the website information using the “class”. This class information varies based on websites, so I suggest you check (using F12) and use the corresponding node or div details.

Python code is shared: Extract\_To\_Excel.py

This code will extract the context from Yellow pages website and save it to an excel. If there are multiple pages, as a search result, for loop range must be modified accordingly.

Macro file is created to merge all the output workbooks into a single file and then to combine all into one sheet. These will further help in reducing time to combine multiple files copy & paste tasks.

These are the modules created to perform the above-mentioned operations.

Module1: MergeExcelFiles

This will merge all the files in a folder into a workbook (different sheets)

Module2: Combine

This will combine all sheets to a sheet

Module3: DeleteSheet1

This will delete the sheets which has name other than “sheet1”