



CLUB OF DATA ENGINEERS
SCHOOL OF INFORMATION SCIENCES

Project 001 | 05/04/2021 (Web scraping)

Goal :

This project's main goal is to let you familiarize with the **Web scraping process** and know more about the different technologies used in that process.

Task :

Omar is an investor who is interested in the world of crypto, He started his first Facebook group in 2015 where he wrote more than 700 Articles about crypto, how to invest in crypto and alot more.

Omar thought of migrating all of his Facebook group articles to his own website and needs your help.

Task : Create a script that extracts all of his Facebook group posts(Post content, images links, Date, number of likes and comments, extraction date & time) to an .csv file.

post_id	post_date	post_txt	post_imgs_link	post_likes	post_comments	extraction_date
1	5/4/2021	Lorem ipsum dolor Lorem ipsum dolor Lorem ipsum dolor	https://shil3ba.com/img.png	3400	590	5/4/2021
2	5/4/2021	Lorem ipsum dolor Lorem ipsum dolor Lorem ipsum dolor	https://shil3ba.com/img.png	3400	590	5/4/2021
3	5/4/2021	Lorem ipsum dolor Lorem ipsum dolor Lorem ipsum dolor	https://shil3ba.com/img.png	3400	590	5/4/2021
4	5/4/2021	Lorem ipsum dolor Lorem ipsum dolor Lorem ipsum dolor	https://shil3ba.com/img.png	3400	590	5/4/2021
5	5/4/2021	Lorem ipsum dolor Lorem ipsum dolor Lorem ipsum dolor	https://shil3ba.com/img.png	3400	590	5/4/2021
6	5/4/2021	Lorem ipsum dolor Lorem ipsum dolor Lorem ipsum dolor	https://shil3ba.com/img.png	3400	590	5/4/2021
7	5/4/2021	Lorem ipsum dolor Lorem ipsum dolor Lorem ipsum dolor	https://shil3ba.com/img.png	3400	590	5/4/2021
8	5/4/2021	Lorem ipsum dolor Lorem ipsum dolor Lorem ipsum dolor	https://shil3ba.com/img.png	3400	590	5/4/2021
9	5/4/2021	Lorem ipsum dolor Lorem ipsum dolor Lorem ipsum dolor	https://shil3ba.com/img.png	3400	590	5/4/2021
10	5/4/2021	Lorem ipsum dolor Lorem ipsum dolor Lorem ipsum dolor	https://shil3ba.com/img.png	3400	590	5/4/2021
11	5/4/2021	Lorem ipsum dolor Lorem ipsum dolor Lorem ipsum dolor	https://shil3ba.com/img.png	3400	590	5/4/2021
12	5/4/2021	Lorem ipsum dolor Lorem ipsum dolor Lorem ipsum dolor	https://shil3ba.com/img.png	3400	590	5/4/2021

** You can use the **CODE Projects** Facebook group as an example of Omar's one.

** Please consider submitting your code in this [github](#) repo.

** The project discussion will be in the CODE Projects Facebook group.

Khalid KASSI
C.O.D.E President
School of Information Sciences

Deadline:20/04/2021