



Social Scrapers

Siddhi Jain, Srishti Jain, Yashita Maltare ,Prof. Amrish Shrivastav

Department of Computer Science and Engineering, Acropolis Institute of Technology and Research, Indore ,India

ABSTRACT :-

Internet is a collection of information and nowadays, much information can be easily accessed through the internet. Some social media web applications, such as “Meta”, “Twitter” and even “Instagram”, provide users with easy information sharing features. However, the information that is presented is in the form of a timeline or a feed, which is at many times not relevant to some users that need that for a specific purpose or it is quite hard to access by the user because of the redundancy of the information. This situation can be resolved with a Web Scraping tool, that is able to search information, combine and present it in a better structured way according to user preferences. A system is developed to implement the proposed method.

1.INTRODUCTION :-

Social Scraper is a system that can search the relevant information, filter this kind of information, and present it to the user. Social Scraper uses Web scraping which is one of the most popular techniques to extract web data or content. Getting accurate information from the pool of the web to analyze a person's profile to generate an idea about the overall aspect and personality is difficult. Web Scraping provides a way to search information, combine and present it in a better way according to user preferences.

2.PROBLEM FORMULATION :-

As of 2018, more than 2.5 quintillion bytes of data were being generated daily. That is a mind-boggling amount of data. Thinking of how to process it is enough to give you a migraine. But what if it didn't have to be that hard? What if there was a method you could use to access data wherever you want at any quantity you want? Thus, we want to provide a tool to search social profiles of professionals, businesses, and organizations quickly, effortlessly, and efficiently.

Our research process contained several steps:

Awareness of the problem –

We decided to start by performing a literature review in order to understand previous research into the topic as well as to understand some of the issues that can arise when scraping websites. Most importantly, we wanted to know how other researchers had designed their web scrapers and crawlers.

Suggestion for solving the problem - The results of the literature review as well as creative input from the researchers are used for this step. The creative input introduces a measure of non-repeatability that the reader should be aware of.

Implementation of our suggestion - The implementation process was iterative, and contained several stages: i) requirements analysis, ii) architectural design, iii) coding, and iv) testing in order to develop a prototype.

The evaluation of our implementation - The prototype implementation of the previous step is used in order to evaluate the design and see if the proposed system could be used to solve the problems identified at the start of the process. The evaluation answers the question of whether the proposed design and proof-of-concept implementation solves the problems identified previously.

Concluding the results of our implementation - The research ends with a consolidation of what has been learned throughout the process as well as identifying areas that need further research in order to communicate the results to the research community.

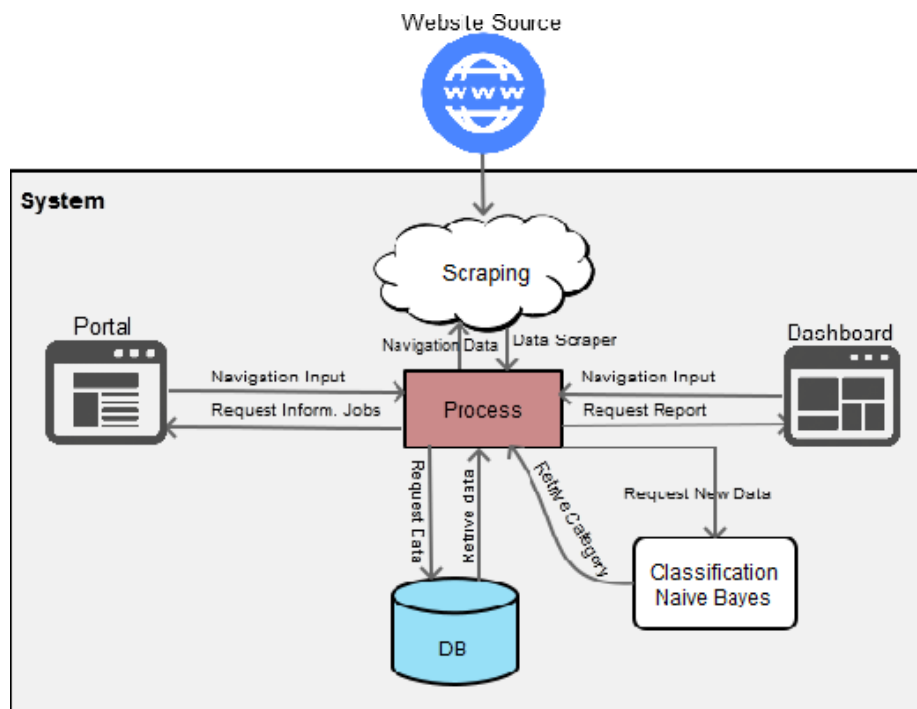
LITERATURE REVIEW:-

There are some existing tools present that provide method to scrape: Dexi.io –It demands users to be skilled in programming. Scrapinghub – It is not an effortless tool as it might be a complicated yet powerful tool for businesses. Parsehub –It becomes controversial due to its pricing policies. There are tools like “Scraping Robot” and “TransparenCEE” that work on similar concepts. But their target markets are the business organizations and people with political agendas. They either search the comments or do the sentiment analysis of those comments.

METHODOLOGY

Social Scrapers works through the following methods to get an outcome:-

- Request the web page using the URL.
- Parse the structure of the web page so your programming language can work with its contents.
- Extract the information we are interested in.
- Write this information to a file for future use.



RESULT AND DISCUSSIONS

Web scraping has been employed for several different purposes. Perhaps the most widely known applications of scrapers are the ones employed by search engines such as Google. Scraping systems employed by search engines fill the task of indexing and categorizing web pages. Scrapers and crawlers are also used for archival purposes such as the crawlers employed by the Internet Archive. Another use is for sentiment analysis by businesses interested in brand opinions. Extracting data and information from the web is of use in research and learning activities as well. Web scraping systems have also been shown to be able to extract information from social networks and forums.

Through Social Scrapers, we expect the following outcomes:

- Time saved by finding all sources.
- Better understanding of someone’s profile and/or work.
- Effortless marketing research and analysis.

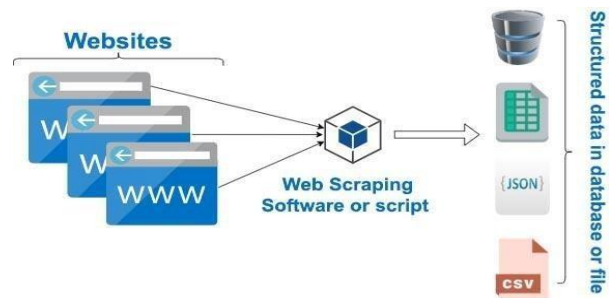


IMAGE SOURCE: webharvy.com

CONCLUSION

It's no secret that social media is changing the way people interact with each other. It's even changing the way we think about the world around us. The world of social media is constantly changing, so it can be a challenge for businesses and individuals to keep tabs on changes. Now that you know how social media scraping can positively impact you and your organization, I hope you'll choose Social Scrapers to help you continue on that journey.

REFERENCES

-
- "Scrapy", scrapy.org, 2019. [Online]. Available: <https://scrapy.org>. [Accessed May 17, 2019].
 - "Selenium", seleniumhq.org, 2019. [Online]. Available: <https://www.seleniumhq.org>. [Accessed May 17, 2019].
 - Y. Sun, I. G. Councill and C. L. Giles, "The Ethicality of Web Crawlers," in 2010 IEEE/WIC/ACM International Conference on Web Intelligent Agent Technology, Toronto, ON, 2010, pp. 668-675.
 - S. S R and S. Chaudhari, "An Effective Forum Crawler," in 2014 International Conference on Circuits, Systems, Communication and Information Technology Applications (CSCITA), Mumbai, 2014, pp. 230-234