Machine Reasoning Project Report

Social Media Scrapper

Team Members

Sachin Varghese

Anton Cheriyan

Manalito Japlos

1 EXECUTIVE SUMMARY

GSS Certis is a leading advanced integrated security organization that develops and delivers multi-disciplinary security and integrated services. As a unique specialist operations-technology outsourcing partner, Certis integrates advanced security, technology, facilities management, customer service and talent to build and operate bespoke solutions for complex, critical operations that extend beyond security. Our design-for-purpose solutions are led by an extensive track record of running operations and design thinking to drive operational efficiencies and deliver business-critical outcomes for our customers.

Being the premium security provider in the market it is always expected to have social media presents which is by both the staff and competitors.

Current process of monitoring the social media presents is a manual process where staff would look at all social media like Instagram and Twitter and would create a report based on the same and submit daily basis which is a hectic process that is being done now.

2 BUSINESS PROBLEM BACKGROUND:

In GSS Certis current reports are manual and prone to human mistakes which we are looking forward to solve, and also the time and effort that is being spent on the report generation on a daily basis is significantly high.

We believe that there is time and effort that has been spent on preparing the reports and with the effective implementation of WEB scrapper we will be able to efficiently transform the process.

3 PROJECT OBJECTIVE

Having defined the business problem our group's aim was:

1. To come up with a solution to track the social media spams about the organization efficiently.

2. Save time and increase efficiency.

4 PROJECT SOLUTION

The first step in solutioning was to come up with a knowledge model. Knowledge modelling is a two-step process:

1. Knowledge Acquisition

2. Knowledge Representation

4.1 KNOWLEDGE ACQUISITION

Knowledge acquisition sets up the sources from where we would get the relevant data to build our intelligent system. The sources are enumerated in the below table:

S/N	Source of Information	Information elicited	Acquisition Technique used

1	Public Domain – Instagram Twitter	The bulk of information is provided from these sources e.g.- - Instagram Tags - Twitter tags	- Web Scraping to get the data from Twitter and Instagram

4.2 KNOWLEDGE REPRESENTATION

Considering the data that has been collected is for reporting and analytics we will be keeping the data in excel format and images to be saved in root folder.

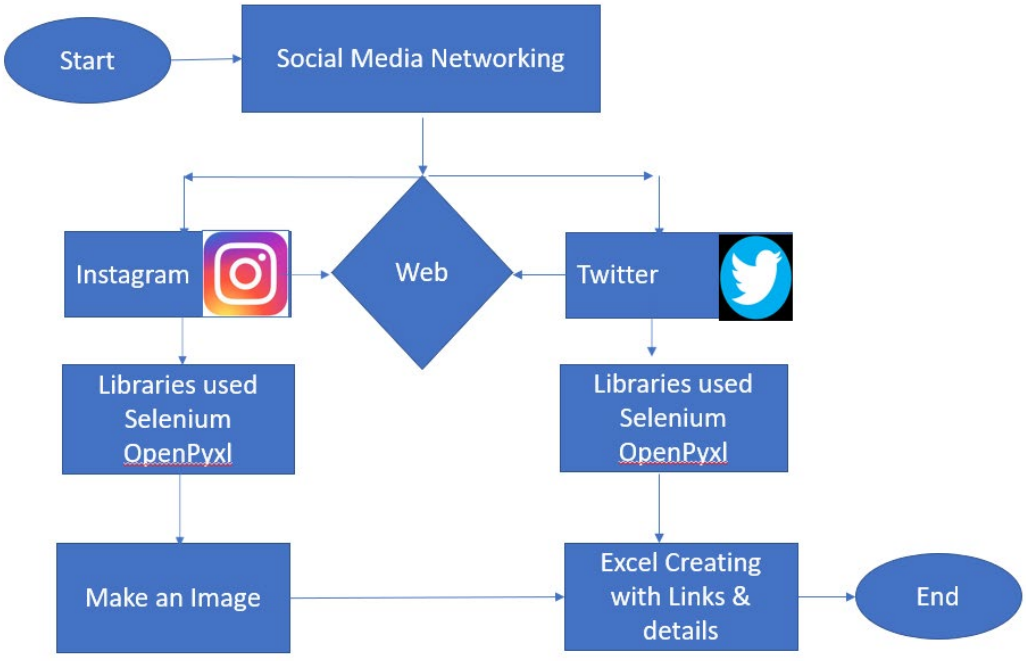
4.3 PROJECT SCOPE

While data mining can be performed continuously, in the context of this project, its scope is limited by the (i) keywords that will be used (ii) the amount of data that is available at the time of extraction. The system provides a collection of data from both Instagram and Twitter. However, the team believes that the project can be further developed to collect information from more locations.

5 BUSINESS PROCESS DESIGN

The Business Process is shown in the diagram below.

Data was crawled from Instagram and Twitter websites using libraries tweepy, openpyxl and selenium to Load all the data and push to Excel report and save images to root folder.



6 SYSTEM'S FEATURES

Despite the restricted scope and assumptions, the team went through a thorough thinking process to develop key features inside the Scraper system that may significantly increase the value of the business.

SCALABILITY

The system is developed to be easily accessible and can be executed by anyone at any point of time, the application would be relevant at any point of time and data can be extracted based on the user requirement.

7 LIMITATIONS

The current limitation would be the access to social media websites for offices with limited connectivity to social media websites. Another shortfall at this point would be that we are collecting data from Instagram and Twitter, the team is looking forward to extend the search.

8 CONCLUSION

Working on this project was a great experience for our team, and we learned a lot of essential skills along the way. The gathering of data was an important element of the entire procedure. We wouldn't have been able to construct a system based on all the different rules if we didn't have a solid knowledge basis from the lectures. The process of building the platform itself brought a completely new set of challenges. Working on the activity as a group allowed everyone to pick up on one another's skills. Overall, it was a tough start to the MTech program, and one which we are not complaining about.

9 IMPROVEMENTS

1) Web Based GUI : We would have loved to include a Web based GUI, so that the look and feel could have been better.

2) Certainty factors: add more platforms for search criteria.