Socio-Matrix

Chirag G

Department of Information Science Ramaiah Institute of Technology Bengaluru, India [1ms19is038@msrit.edu](mailto:1ms19is070@msrit.edu)

Gaurav Sood

Department of Information Science

Ramaiah Institute of Technology Bengaluru, India [1ms19is041@msrit.edu](mailto:1ms19is070@msrit.edu)

Jayesh Goyal

Department of Information Science Ramaiah Institute of Technology Bengaluru, India [1ms19is037@msrit.edu](mailto:1ms19is070@msrit.edu)

Eeshan Singh

Department of Information Science Ramaiah Institute of Technology Bengaluru, India [1ms19is052@msrit.edu](mailto:1ms19is070@msrit.edu)

***Abstract*—Cybercrime, Cyberbullying, malicious activities,Cybergrooming and spreading misinformation is working at its peak in this Era of the internet. Setting up tools to find out the people behind such acts with as little information that is available to us will come a long way in making sure that such people are caught. Within the world of investigations into missing, exploited, and trafficked children specifically, you often don’t start with much information. In many cases you can at least find a person with some combination of an email, a phone number, or a user name, which can be used as a starting step for any investigation process.**

**In this paper an attempt has been made to make Socio-Matrix to analyze multiple social media accounts and gather information such as profile data, followers, posts, similarity between different accounts, and various analysis from the metadata gathered. The analysis was targeted towards individuals of interest, and it was used to investigate malicious activities that occurred online, such as cyberbullying, spreading misinformation,and cyber crimes. Socio-Matrix was implemented in the form of a CLI or a web application, and users were able to choose from multiple analysis techniques based on their requirements. Overall, it was very powerful and useful for social media analysis of an individual. Socio-Matrix covers over 800+ websites.**

1. INTRODUCTION

As the number of people using social networking and social media platforms continues to rise, there has been a significant increase in the use of these platforms for performing various malicious acts. As a result, social media analysis has become an essential field to keep these crimes in check by providing valuable insights into user behavior, preferences, and trends. Socio-Matrix has been developed to facilitate easy analysis of social media, with all major platforms such as Instagram, Twitter, and Facebook included. SocioMatrix was mainly built using Python and offers a variety of analysis modules for users to choose from, including user profiling, network analysis, and hashtag analysis. Each module has an associated rating matrix with rate values ranging from 0 to 100, indicating the probability of a positive or negative match for the person being searched. The insights gained from Socio-Matrix can be used by businesses and individuals to make informed decisions, conduct research projects, or pursue personal interests. Socio-Matrix can be deployed as a CLI mainly as a Python or Nodejs script, but when deployed as a web application, it offers additional features beyond the functionality of the CLI.

II. Methodology

Socio-Matrix is written in Python and is designed to work with popular social media networks such as Instagram, Twitter, and Facebook. It includes various features and functions, such as username enumeration, hashtag analysis, and gathering post information 1 and other metadata. Socio-Matrix follows a three-phase approach to data collection and analysis: data collection, data processing, and data analysis

1. *Data Collection*

Socio-Matrix utilizes publicly available information from targeted social media platforms to analyze a person’s required information. For example, Socio-Matrix employs the Instagram Graph API to access multiple types of user data, such as user profiles, posts, and followers. Similarly, Socio-Matrix uses the Twitter API to access user profiles, tweets, and followers for Twitter data. After gathering data from these websites, it is stored in a database for further processing and analysis.

*B. Data Examination*

After completing the data collection process, Socio-Matrix employs various algorithms and statistical methods to perform data analysis and derive insights. Socio-Matrix offers several data analysis features, including username enumeration, hashtag analysis, and post metadata collection.

*C. Enumeneration of Usernames*

One of the features of Socio-Matrix is username enumeration, which enables users to gather information about an individual or a group of target audience using their usernames on various social media platforms. This method aims to collect the target user’s publicly available information, such as their full name, biography, and location. Additionally, it can also gather data on the user’s followers and those they follow, which is useful for personality analysis

*D. Analysis of Hashtags*

Socio-Matrix also includes a feature for hashtag analysis, which enables users to study the usage of hashtags related to a specific topic on a particular social media platform. This feature can assist in identifying popular or trending hashtags, as well as the users who are utilizing them. Additionally, it can be used to analyze the level of engagement associated with different hashtags.

*E. Collection of Post-Metadata*

Socio-Matrix feature known as post metadata collection enables users to gather information about posts made by specific social media accounts. This function can extract data such as the post’s date and time, the number of likes, comments, and shares, and the content of the post.

*F. Plugin Creation*

Socio-Matrix’s functionality can be extended through the use of plugins, which allow users to add their own custom features. Creating plugins is an integral part of Socio-Matrix’s development process because it enables users to tailor Socio-Matrix to their specific needs. To create a plugin, users need to write code that interacts with Socio-Matrix’s API and performs specific functions. For instance, a plugin could be designed to gather data on users who have liked a particular post or to identify the most engaged users on a specific social media platform. Once created, plugins can be shared with other users, allowing the wider Socio-Matrix community to benefit from the collective knowledge.

III. Design and Workflow

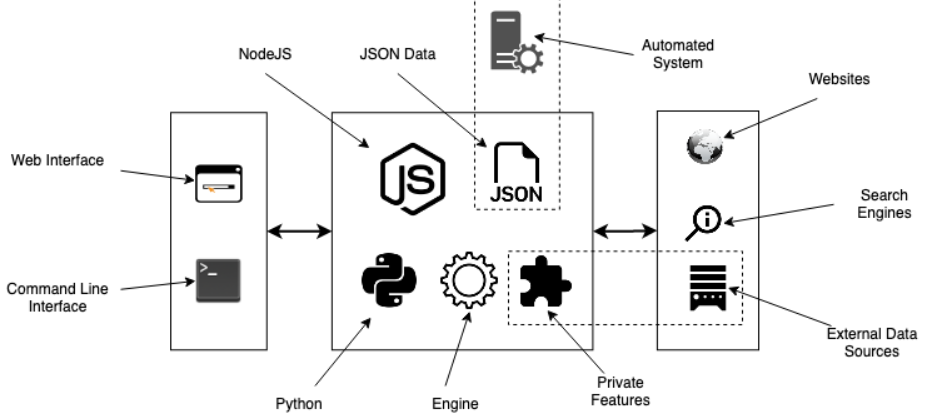


Fig 1. Design and Workflow

1. *Layer 1*

The Layer 1 consists of two components: Web Interface and Command Line Interface.

1. *Web Interface*

The web interface can provide additional features and benefits beyond the command-line interface (CLI) version of the tool.

* User-friendly interface: A web interface can be more intuitive and user-friendly than a CLI, making it easier for non-technical users to access the tool and perform social media analysis.
* Real-time updates: With a web-based interface, users can access real-time updates and analysis results as they become available, rather than waiting for batch processing to complete.
* Visualizations: A web interface can provide interactive visualizations that make it easier to understand and interpret the analysis results, which can be especially useful for complex data sets.
* Collaboration: With a web interface, users can easily share their analysis results and collaborate with others, which can be useful for research projects or investigations.
* Accessibility: A web interface can be accessed from anywhere with an internet connection, making it more accessible and convenient for users who may be working remotely or on-the-go.

1. *CLI*

While a web interface can offer several benefits, there are also some advantages to using a command-line interface (CLI).

* Automation: A CLI can be more easily automated than a web interface, making it useful for scripting and batch processing of large data sets.
* Lightweight: A CLI can be more lightweight and require fewer system resources than a web interface, which can be useful for users with limited computing power.
* Customizability: With a CLI, users can more easily customize the analysis parameters and settings to suit their specific needs and requirements.
* Speed: A CLI can be faster and more efficient than a web interface, particularly for performing analysis on large data sets or complex queries
* Security: A CLI can be more secure than a web interface, particularly if it is used on a local machine, as it is less vulnerable to network-based attacks and other security threats.

*B. Layer 2 and 3*

The Layers 2 and 3 showcase the following steps

* The system comprises a Python-based web server that serves the command-line interface (CLI) client, and a JavaScript-based Express.js web server that serves the web interface client.
* Additionally, a search engine written in JavaScript allows for the exploration of websites, as well as external data sources. The search engine crawls these sources and collects relevant data, which is then converted to JSON and passed through an automated system.
* The automated system incorporates various analysis logics to analyze and visualize the collected data.

IV. Implementation

The social media analysis tool can be implemented using a combination of Python and JavaScript technologies, including Flask or Django for the Python web server, Express.js for the JavaScript-based web server, Cheerio or Puppeteer for the search engine, JSON for data conversion, and various analysis libraries for the automated system.

1. *Python and Web Server*

The CLI client can be served through a Python web server using a framework such as Flask or Django. The server can listen for requests from the CLI client and respond with the appropriate data.

1. *JavaScript-based Express.js Web Server*

The web interface client can be served through a JavaScript-based Express.js web server. The server can listen for requests from the web interface client and respond with the appropriate data. The web interface can be developed using a front-end framework such as React, Angular, or Vue.js.

1. *Search Engine*

To explore websites and external data sources, a search engine can be written in JavaScript using a library such as Cheerio or Puppeteer. The search engine can crawl various data sources and collect the data. The collected data can be cleaned and processed to remove any irrelevant information

D*. Data Conversion*

The collected data can be converted to JSON using a library such as json.dumps() in Python or JSON.stringify() in JavaScript. This ensures that the data can be easily passed through the automated system for analysis.

*E. Automated System*

The automated system can be written in Python or JavaScript and contains various analysis logics to analyze and visualize the data. The system can perform various analysis modules, such as user profiling, network analysis, and hashtag analysis, by using libraries such as NetworkX or Matplotlib for Python, or D3.js or Plotly for JavaScript. The analysis results can be presented to the user in a user-friendly format, such as a dashboard or a report.

V. Results

The outcomes of the tool are dependent on the work at hand. We show you how to use the tool to analyse social media data in the examples below.

1. *Hashtag Investigation*

In social media analysis, hashtag analysis is a typical task. It entails examining how hashtags are used by users on a specific platform. the tool can be used to collect and analyse hashtag data from sites such as Twitter and Instagram. the tool, for example, can detect the most widely used hashtags, the most important people, and the themes connected with specific hashtags. Businesses and individuals can utilise this information to better understand their target audience.

1. *Profiling of users*

Another common duty in social media analysis is user profiling. It entails analysing user data in order to comprehend user behaviour and preferences. the tool can be used to collect and analyse user data from platforms such as Twitter and Instagram in order to develop user profiles. the tool, for example, may determine the most prevalent forms of information provided by users, the times of day when users are most active, and the topics most commonly linked with certain people. Businesses and individuals can use this data to construct targeted marketing efforts or to personalise interactions with their target audience.

1. *Network examination*

In social media analysis, network analysis is a more difficult task. It entails examining the relationships between users on a specific platform. The the tool can be used to collect data on user activities such as retweets, mentions, or comments, and then analyse this data to uncover patterns of user relationships. For example, the tool may discover user groups that frequently communicate with one another or major influencers who have a big impact on the behaviour of other users. Businesses and individuals can utilise this information to better understand the social dynamics of a specific platform and modify their social media strategy accordingly.

E. Output

1. *Socio-Matrix in CLI*

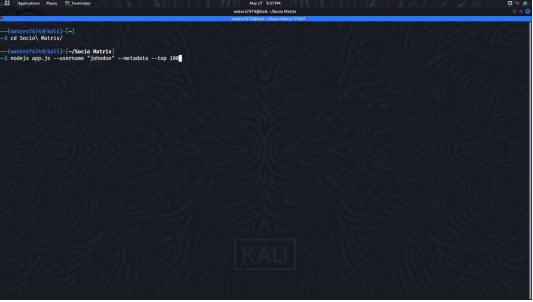


Fig 2. Socio-Matrix in CLI

Socio-Matrix offers various commands depending on the usage of the user. Depending on the functionality and the feature requirement the commands of Socio-Matrix has been setup .Socio-matrix cli is efficient and easy to use.

1. *Socio-Matrix results generated in command line*

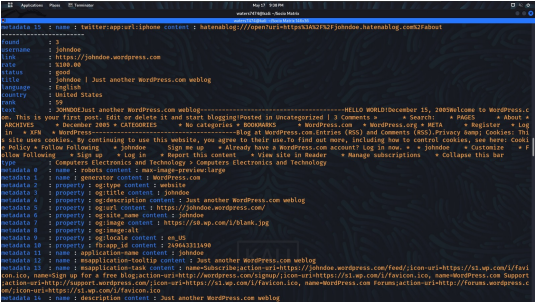


Fig 3. Socio-Matrix results generated in command line

The Figure represents the results generated by Socio-Matrix when used in command line, Common Information such as the Title, Description, Rank, status of the social media sites along with data of specific social media websites is shown one below the other making it easy for the user to go through the results.

1. *Socio-Matrix web interface*

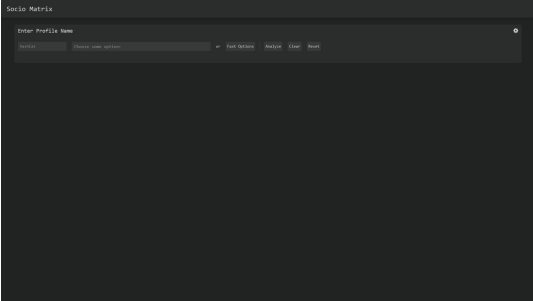


Fig 4. Socio-Matrix web interface

This figure shows Socio-Media deployed as a web-app, The user interface is simple and easy to toggle around. Socio-Matrix when deployed as a web-app makes sure that everyone including general public can use all of the features provided.

*d.Socio-Matrix filter selection*

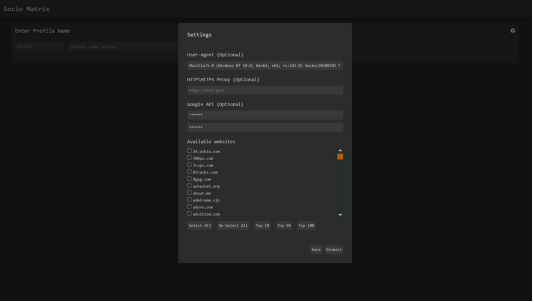


Fig 5. Socio-Matrix filter selection

Figure depicts the various websites that haven been covered by Socio-Matrix for analysis and result generation . A total of 900+ websites have been included . The user can choose the social media sites from the list provided , select all the sites in the list or just select the top 10/50/100 sites.

e. *Socio-Matrix’s execution based on provided filters*

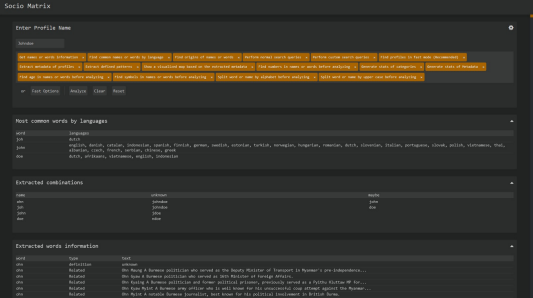


Fig 6. Socio-Matrix’s execution based on provided filters

This figure shows a demonstration of the working of the tool, various filters have been included in the search option along with the list of websites. This demonstration was performed on the Web-app.

*f. Socio-Matrix detected profile results*

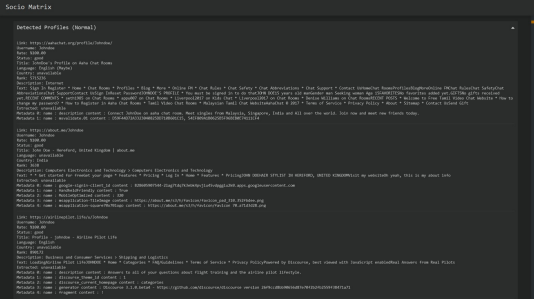


Fig 7. Detected profile results

The Results generated for the search haven been shown here. Information gathered from various sites are shown one below the other. The user is also given the option to download the information gathered

g*. Socio-Matrix’s visual graph of detected metadata*



Fig 8. Socio-Matrix’s visual graph of detected metadata

This figures draws out a graph of the metadata gathered from the internet and shows various connections among them . Visualisation of the data provides easy and faster understanding of the results.

VI. Conclusion

Socio-Matrix is a simple way to examine social media data from various platforms. In this section, some essential features, limitations, and future research possibilities will be discussed.

Analysis that was conducted, highlighted that social media site metadata can provide valuable insights into user behaviour and preferences. For instance, study of Instagram data revealed that posts with photos or videos tend to receive more engagement than those with only text. Additionally, posts with more hashtags or tags tend to have greater visibility and reach. Similarly, examination of Twitter data found that users are more active at certain times of the day and that including mentions or hashtags can enhance interaction and visibility.

Overall, the analysis emphasises the importance of social media analysis in understanding user behaviour and preferences and creating effective social media strategies for businesses and individuals.

However, it is important to note that social media analysis has limitations, such as Socio-Matrix’s dependence on social media network APIs, which limits the amount of data that can be collected. It is also essential to ensure that all data analysed follows ethical standards and best practices.

Future research in this area should focus on creating more advanced analytical methods to provide users with more insights into the topic or target individual. Using Socio-Matrix as a base, further advancements in the field of social media analysis can be made.

VII. References

1. D. Ghimire, “Comparative study on python web frameworks: Flask and django,” 2020.
2. A. Yim, C. Chung, and A. Yu, “Matplotlib for python developers: Effective techniques for data visualization with python.” Packt Publishing Ltd, 2018.
3. R. Mitchell, Web scraping with Python: Collecting more data from the modern web. ” O’Reilly Media, Inc.”, 2018.
4. A. Mardan and A. Mardan, “Using express. js to create node. js web apps.” Springer, 2018, pp. 51–87.
5. N. Q. Zhu, “Data visualization with d3. js cookbook.” Packt Publishing Ltd, 2013
6. E. Persson, “Evaluating tools and techniques for web scraping,” 2019.
7. D. S. Sirisuriya et al., “A comparative study on web scraping,” 2015.
8. Z. Tan, C. He, Y. Fang, B. Ge, and W. Xiao, “-based extraction of news contents for text mining,” IEEE Access, vol. 6, pp. 64 085–64 095, 2018
9. J. Liu, X. Kong, X. Zhou, L. Wang, D. Zhang, I. Lee, B. Xu, and F. Xia, “Data mining and information retrieval in the 21st century: A bibliographic review,” Computer science review, vol. 34, p. 100193, 2019.
10. D. Canali, M. Cova, G. Vigna, and C. Kruegel, “Prophiler: a fast filter for the largescale detection of malicious web pages,” in Proceedings of the 20th international conference on World wide web, 2011, pp. 197–206.
11. O. Peled, M. Fire, L. Rokach, and Y. Elovici, “Entity matching in online social networks,” in 2013 International Conference on Social Computing. IEEE, 2013, pp. 339–344.
12. G. Stringhini, C. Kruegel, and G. Vigna, “Detecting spammers on social networks,” in Proceedings of the 26th annual computer security applications conference, 2010, pp. 1–9.
13. Z. Zhang, Q. Gu, T. Yue, and S. Su, “Identifying the same person across two similar social networks in a unified way: Globally and locally,” Information Sciences, vol. 394, pp. 53–67, 2017.
14. A. Gupta and R. Bhatia, “Ensemble approach for web page classification,” Multimedia Tools and Applications, vol. 80, pp. 25 219–25 240, 2021.
15. S. R. Eddy, “Accelerated profile hmm searches,” PLoS computational biology, vol. 7, no. 10, p. e1002195, 2011.