CMATA: Cyber Trafficking Monitoring and Tracking Prototype

Risty Acerado*, Mauee Czelsie Samaniego, and Roselia Morco

***Abstract*— With the evolution of the Internet, trafficking has broader and borderless operations targeting virtual spaces where individuals of different ages can be subjected to cyber trafficking. Despite the government's efforts to combat cyber trafficking, the growing number of cases is alarming. One apparent reason is that traffickers keep advancing their technical strategies, so it becomes more challenging for law enforcement to investigate complex and transnational organized crime. Thus, this paper presents the development of a cyber monitoring and trafficking web application integrated with a web scraper that identifies websites with a high probability of being used in cyber trafficking activities. The web scraper was written using Beautiful Soup. With the result of the sensitivity analysis conducted in the study, beautiful soup was determined to be the most suitable tool for developing web scraping algorithms based on performance, portability, and accuracy compared with the Scrapy and Selenium tools. Upon the initial testing of the developed prototyped in tracking and monitoring suspected cyber traffickers, it was able to scrape 45 potential trafficking websites. The initial testing result shows that this kind of application integrated with a scraper to identify cyber traffickers with a visualization dashboard can help law enforcement officers identify, track, and monitor online traffickers easily and quickly.**

***Index Terms*—Beautiful soup, selenium, scrapy, cyber trafficking, cyber trafficking monitoring application, scraping tools, selenium, web scraping.**

I. Introduction

For several years, human traffickers exploited victims from different countries. In the Philippines, a significant number of overseas workers became victims of sex trafficking or labor trafficking in numerous industries: industrial fishing, shipping, construction, manufacturing, education, health care, agriculture, domestic work, janitorial service, and other hospitality-related jobs. Most of these victims reported experiencing forced labor, illegal recruitment, and sex trafficking [1]. Many young and old are at high risk of trafficking at any given time and place. The incident worsened in the country with the advent of the COVID-19 pandemic, which increased sexual exploitation using the internet. Law enforcement reports indicate that the Philippines is one of the largest known online sexual exploitation sources of children. Traffickers sexually exploit children, individually and in groups. Sometimes these are conducted in live internet broadcasts in exchange for compensation wired through a money transfer agency by individuals, most often in another country, including the United States, Australia, Canada, and the United Kingdom [2].

Within just six months, according to [1], 1,434 females were victims of sex trafficking in the Philippines. As stated in [3], the ratio between the number of convicted traffickers and the victims is 1:5. Note that there are still many unreported and unvalidated cases as many traffickers brainwash their victims [4]. Many trafficking victims report the incident to their local police station, but the case is often not pursued for different reasons. These cause more challenges for the Philippine law enforcement agency to track and collect suspected traffickers.

The secretive nature of trafficking, especially done online, makes it nearly impossible to collect accurate data on the number of victims and, most importantly, to identify and catch the traffickers. The lack of a centralized database for tracking illegal recruitment and human trafficking continued to hamper the government's efforts to prevent trafficking and hold traffickers accountable.

The number of trafficking cases is still alarming despite the government's efforts. One apparent reason is that traffickers keep advancing their technical strategies, so it becomes more challenging for law enforcement to investigate complex and transnational organized crime. Indeed, in many countries, trafficking is still a new crime involving "new and untested laws" [3].

Trafficking is a serious issue that must be dealt with by the government and others who have the ability and care for society. Thus, research studies and other related project studies would be very relevant. Crime detection and monitoring is one new field where mobile technology is gaining great utility. Some recent studies have focused on identifying traffickers in advertisements and social media posts using semi-supervised algorithms, such as the studies of [6, 7].

In the Philippines, the first monitoring facility for trafficking, particularly cyber trafficking, was launched in late 2020. But according to the interview of Justice Guevara stated in this article [5], "expertise does not come overnight." Although the facility has the latest hardware and software, there are no proactive approaches yet in combating cyber trafficking, such as utilizing machine learning algorithms to determine the cyber trafficking websites quickly. Most of these current schemes and interventions are very reactive. In which early detection and tracking of potential victims and perpetrators are crucial. Therefore, this project piloted a proactive approach to anti-cyber trafficking by preventing the

Manuscript received November 23, 2022; January 10, 2023; accepted February 20, 2023. This work was part of the project funded and supported by United Nation Populations Fund (UNFPA).

R. Acerado and R. Morco are with Technological Institute of the Philippines, Quezon City, Philippines

M. C. Samaniego was with Technological Institute of the Philippines, Quezon City, Philippines

*Correspondence: (e-mail: racerado.is@tip.edu.ph).

crime right before it started. This paper presents a Cyber Monitoring and Tracking Application that would greatly help government agencies to combat cyber trafficking crimes. In addition, the result of the web scraping tool's sensitivity analysis in portability, accuracy, and performance conducted in this study could be referenced by other researchers in their future related research.

A. Objectives

The project aims to develop a website tracking application for the suspected cyber trafficking platform.

Specifically, the project aims to

- Determine the scraping tool appropriate for the project in terms of portability, accuracy, and performance;
- Develop a scraping algorithm that will scrape websites based on the identified keywords;
- Develop an application integrated with the scraping algorithm that identifies the suspected websites.

B. Significance

As proposed in this paper, an application that integrates a predictive classification analysis will help promote parents' awareness of these hidden transactions in cyberspace, especially since even small children are widely known targets of human trafficking. With this proposed project, early detection of possible or suspected websites can help save the lives of children by preventing transactions from those websites.

In addition, this could be useful to law enforcement agencies. It is an excellent line of defense to act on early signs of human trafficking before it has taken a victim. It will be a great tool for law enforcement agencies to analyze transactions on the web, which is the most notorious place to execute illegal transactions faster. In this way, law enforcement can develop a task force to watch over red flags identified by the algorithm. Scrutinizing each online transaction would be much easier for law enforcement agencies, allowing them to act on the concerns of a larger community in need of their assistance.

II. Tradeoffs and Sensitivity Analysis

A. Web Scraping Tools

Beautiful Soup, Scrapy, and Selenium web scraping tools were analyzed in this study to determine the most reliable web scraper for the project. Beautiful Soup is a library that allows you to quickly and easily extract information from HTML. It is commonly used for web scraping projects in practice. Scrapy is a web scraping API that allows you to scrape the web without getting blocked. It was designed to scrape the web, but it can also be used to extract data via APIs or as a general-purpose web crawler. Selenium is an open-source web-based automation tool. It is primarily used for testing in the industry, but it can also be used for web scraping. The tools were tested and compared the portability in terms of memory consumption, accuracy, and performance to determine how the tools would function on the application.

B. Summary of Web Scraping Constraints

Table I summarizes the results of evaluating the three web scraping tools' portability, accuracy, and performance.

TABLE I: Summary of Constraints and Tradeoffs

Web Scraping Tool	Response Performance	Portability	links scrape
Beautiful Soup	5.416	0.001 MB	241
Scrapy	0.433	69.3 MB	213
Selenium	3.966	19.3 MB	279

Three trials were conducted to determine the response time performance of the tools, and the result was averaged. For the portability criterion, it was based on the memory consumption of each web scraping tool. Meanwhile, the accuracy was based on the number of data gathered from each web scraping tool. Scrapy takes the shortest time to execute compared to other web scraping tools in terms of performance

Eqs. (1-2) were used in the sensitivity analysis of the tools based on the set criteria. HV indicates the higher value rating and LV for the lower value rating. The first step is to get the percentage difference of the value rating result of the criterion; the tools then, the last step is to determine the subordinating rank.

$$\%difference = (HV - LV)/HV \quad (1)$$

$$Subordinate\ rank = Governing\ rank - [(\%difference) \times 10] \quad (2)$$

The ranking value of eight (8), nine (9), and ten (10) was based on their ratings in each specified criterion, with ten as the highest. These values were used in Eq. (2) as the governing rank to validate the subordinate rank further.

Meanwhile, Table I shows that BeautifilSoup has the most minimum memory consumption than the other two web scraping tools. Thus, in this criterion, BeautifulSoup got the highest rank of ten (10).

In terms of accuracy, based on the number of more accurate links scraped by the tools, Selenium got the highest number of scraped data compared with the two scraping tools. This time, Selenium got the highest ranking in this criterion.

C. Sensitivity Analysis

After accumulating three requirements of the three web scraping tools on the various constraints, the total rank of each criterion was calculated to determine which is considered the most reliable and most effective web scraping tool for the proposed application.

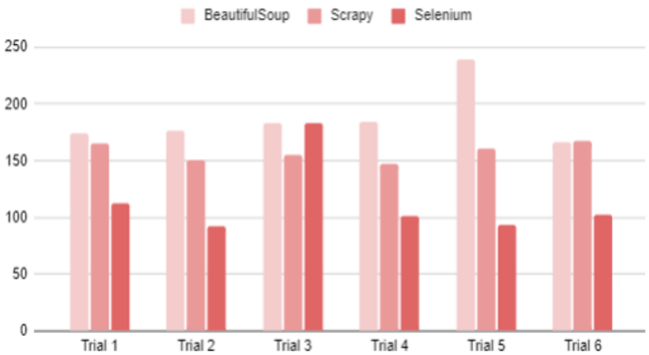

Fig. 1. Summary of Sensitivity Analysis

TABLE I: SUMMARY OF CONSTRAINTS AND TRADEOFFS

Criteria	Criterion's Importance (On scale 1-10)	Beautiful Soup		Scrapy		Selenium	
Performance	*9*	0.8	**7.20**	10.00	**90.00**	1.09	**9.81**
Portability	*8*	10.00	**80.00**	-0.89	**-7.12**	0.3	**2.40**
Accuracy	*10*	8.64	**86.40**	7.63	**76.00**	10	**100.00**
TOTAL RATING			**173.60**		**158.88**		**112.21**

In evaluating how one criterion varies from the others under a specific set of assumptions, a sensitivity analysis was employed using the 10-9-8 Criterion Rank of Web Scraping Tools. Table II shows the summary of 6 combinations of sensitivity analysis for each criterion.

As we can see in Fig. 1 and in Table II, Beautiful Soup emerges as the best suitable web scraping tool for this project, winning most of the complexity analysis.

III. PROTOTYPE DEVELOPMENT

There were 23 keywords used for scraping the suspected trafficking websites. Cybersecurity consultants suggested the initial keywords used for tracking the websites. Other keywords were picked from the first few websites that were scraped using the initial keywords.

The scraper was developed using Python programming language in the environment of Visual Studio (VS) Code. The scraper works by having a supplied set of keywords used as the identifier of the different cyber trafficking websites online. These keywords are stored in a google sheet file, which the scraper will access through Google API once integrated into a cloud server. But for this prototyping project, the database to store the suspected trafficking websites was set up on a local server for development.

The scraping process starts by getting the keywords from the google sheet file. The main function then uses these keywords to scrape the web for websites that contain these keywords. The scraped websites were then saved with the following column labels: URL, keyword, classification, keyword_count, and date.

The tagging of website classification was done manually, a team member was assigned to visit and validate the website if it was a trafficker.

Twice within a week of testing, the prototype of the proposed application was able to scrape 64 websites.

However, some websites were duplicated. The total distinct websites were trimmed into 45.

The project's prototype was developed using VueJS Framework. The database was integrated into the proposed cyber trafficking application.

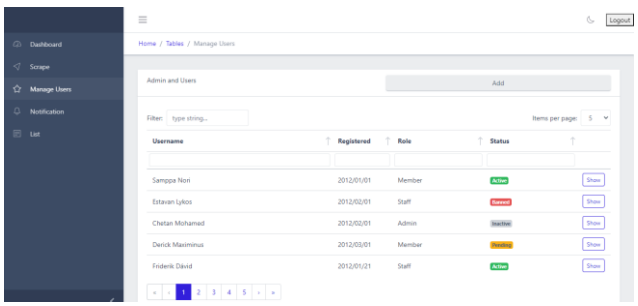

Fig. 2. Manage users.

An interactive and user-friendly dashboard contains the main menu of the application. The other parts or pages of the application are the tables that show the list of websites stored in the database subject for monitoring and validation and a list of keywords used for scraping.

The following images are some of the screenshots of the proposed application.

Fig. 2 shows the interface for managing users of the application such as setting the roles of the users. The admin can add, delete or edit users listed in the table. Only those with login credentials can access the application.

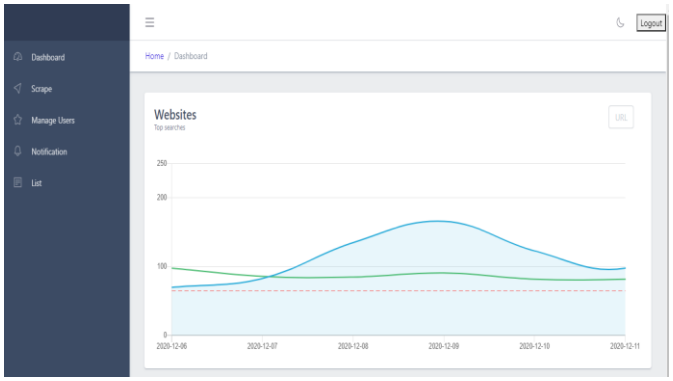

Fig. 3. CMATA dashboard.

Fig. 3. Shows the dashboard of the application. Once the user successfully logs into the application, the default homepage would be the dashboard, which will show some of the visualizations of the database.

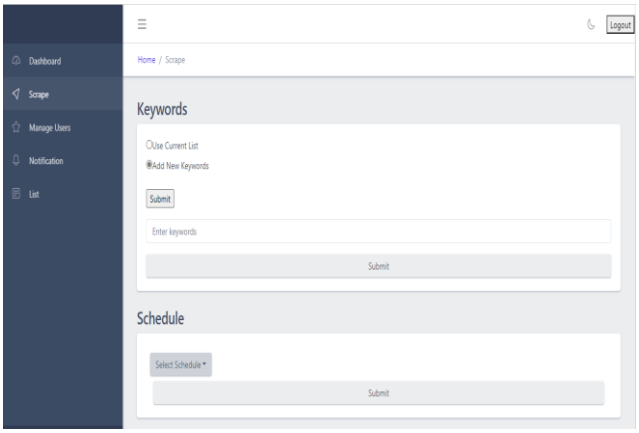

Fig. 4. Manage keywords.

Fig. 4 shows the interface for managing keywords used to scrape websites. On this page, the user can decide whether to use the current list of keywords available in the database or add new keywords for scraping new websites. The user may also select when the application scrapes from the internet. Options can be daily, weekly, or monthly.

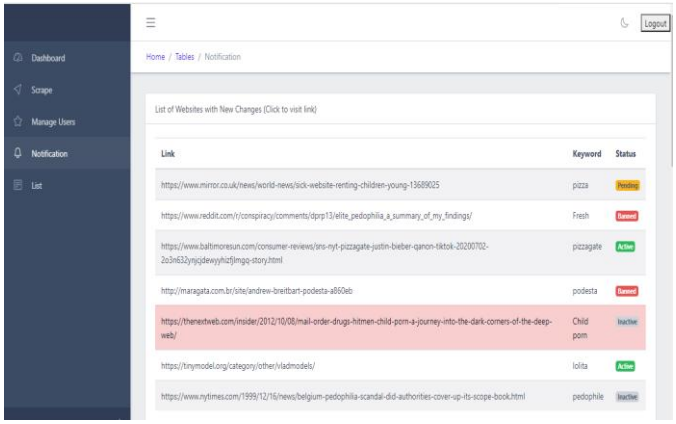

Fig. 5. Notification page.

Fig. 6 shows the notification page containing the website's status and keywords list. Links are also clickable, so the user has the option to redirect to that website.

IV. Conclusion

Technical constraints analysis and tradeoffs helped determine the best scraping tool suitable for identifying suspected cyber trafficking websites.

Based on the analysis result and the overall performance concerning speed, portability, and accuracy, Beautiful Soup is a powerful tool specifically for identifying trafficking websites.

The prototype developed using the vueJS framework integrated with the web scraper using Beautiful soup was able to scrape 45 suspected trafficking websites.

Even with just an initial rapid test, the prototype exhibits a significant tool to identify and monitor websites used by cyber criminals to lure and exploit their targeted victims, especially the young and vulnerable ones. With the simple, easy-to-use interface of the application and interactive dashboard, law enforcement agencies can act quickly to halt transactions of the suspected websites and take immediate action against the traffickers.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

R. Acerado and M. C. Samaniego have written the paper. M. C. Samaniego developed the prototype and worked with another member to develop the web scraper. All authors collaborated in the requirements identification, data gathering, project finalization, and paper's final version.

Acknowledgment

The researchers would like to express their gratitude to the UNFPA Philippines for the support in this project, to ISSA-PH for the consultations, and to other data privacy experts from private and government organizations who selflessly provided a lot of insights to the researchers and to the entire development team.

References

[1] US Department of State. (2020). Trafficking in persons report: Philippines. [Online]. Available: https://www.state.gov/reports/2020-trafficking-in-persons-report/philippines/

[2] International Justice Mission with 14 Implementing and Advisory Partners. (2020). [Online]. Available: https://ijmstoragelive.blob.core.windows.net/ijmna/documents/Final_OSEC-Public-Summary_05_20_2020.pdf

[3] United Nations Office on Drugs *on* Crime, "Global report on trafficking in persons," *In Global Report on Trafficking in Persons*, 2016.

[4] Courtney Desilet. (2019). [Online]. Available: https://portal.ct.gov/-/media/DCF/HumanTrafficking/pdf/Stop-Human-Trafficking Article-v021419.pdf

[5] A. Gallagher and P. Holmes, *Developing an Effective Criminal Justice Response to Human Trafficking*, p. 320, 2008.

[6] H. Alvari, P. Shakarian, and J. E. K. Snyder, "Semi-supervised learning for detecting human trafficking," *Secur Inform*, vol. 6, no. 1, 2017.

[7] A.V olodko, E. Cockbain, and B. Kleinberg, "Spotting the signs" of trafficking recruitment online: exploring the characteristics of advertisements targeted at migrant job-seekers," *Trends Organ Crim*, vol. 23, pp. 7–35, 2020.

Risty Acerado finished her bachelor of science in computer science at Divine Word College of Legazpi (DWCL) Legazpi City, her master's in information management at the University of the East (U.E.), Metro Manila, and her doctor in information technology (DIT) at Technological Institute of the Philippines, Quezon City, Philippines.

She's a faculty member of the Technological Institute of the Philippines, Quezon City, Philippines.

Dr. Acerado is a member of the Information Systems Security Association (ISSA) P.H., the Philippine Society for Information Technology (PSITE), and the Analytics Association of the Philippines.

Mauee Czelsie Samaniego is a bachelor of science in information systems (BSIS) graduate from the Technological Institute of the Philippines. She graduated "with honors."

She was one of the members of the development team of the project funded by UNFPA last 2020 to combat cyber trafficking.

She was one of the participating students in the seameo-tvet Virtual Student Exchange Program between the Universitas Pembangunan Nasional Veteran Yogyakarta and the Technological Institute of the Philippines during the last school year, 2022-2023.

Her team was one of the teams who was granted funding by the United Nations Population Fund (UNFPA)- Philippines and supported by the Coalition Against Trafficking in Women – Asia Pacific (CATWAP) during the 2022 Hackathon Competition that aims to end Sexual Gender-based Violence (SGBV).

Currently, she is software programmer trainee at Fujitsu Engineering Technologies Philippines (FETP), Inc.

Roselia Morco is a graduate of bachelor of science in commerce majoring in computer science at Saint Mary's University, Nueva, Vizcaya. She took her Master of Science in Information Technology at the University of La Salette, Santiago City, and her Doctor in Information Technology at St. Paul University Philippines, Tuguegarao, Cagayan Valley.

She was a full-time faculty member in the College of Information Technology Education (CITE) from 2004-2012. Currently, she is the program chair of the Information Systems department of the College of Computer Studies (CCS), formerly known as CITE, Technological Institute of the Philippines, Quezon City, Philippines.

Dr. Morco is a member of the Information Systems Security Association (ISSA) PH, the Philippine Society for Information Technology (PSITE), and the Analytics Association of the Philippines.