DETECTING HIDDEN ILLEGAL ONLINE GAMBLING SITES ON .GO.ID WEBSITES USING WEB SCRAPER ALGORITHMS

A THESIS

A Partial Requirement to Fulfill for Master Degree in Computer Science

Muchlis Nurseno

NIM: 20210130036

Supervisor:

Haris Al Qodri Maarif, S.T., M.Sc., PhD

Umar Aditiawarman, S.T., M.Sc., PhD

SCHOOL OF COMPUTER SCIENCE

NUSA PUTRA UNIVERSITY

2024

STATEMENT OF AUTHENTICITY

The undersigned below:

Name	:	Muchlis Nurseno
ID of student	:	20210130036
Faculty	:	Computer Science
The Tittle of Thesis	:	Detecting Hidden Illegal Online Gambling Sites on Indonesia Government Websites using Web Scraper Algorithms

Stating truthfully that this thesis has nothing in common with other thesis. Thus this statement is made without coercion from any party. If this statement is not true, it will be sanctioned by the faculty leader.

Sukabumi, 14 January 2024

Writer,

SEPULUH RIBU RUPIAH 10000 TGL. 20 METERAI TEMPEL 7A6AKX783828409

Muchlis Nurseno

NIM. 20190130012

APPROVAL OF THESIS

Title : Detecting Hidden Illegal Online Gambling Sites on .GO.ID Websites using Web Scraper Algorithms

Name : Muchlis Nurseno

ID of student Faculty : 20210130036

This thesis has been reviewed and approved

Sukabumi, January 2024

Head of Study Program, Supervisor,

Prof. Ir. Teddy Mantoro, M.Sc., PhD
NIDN. 0323096491

Haris Al Qodri Maarif, S.T., M.Sc., PhD
NIDN. 0418068505

THESIS APPROVAL

Title : Detecting Hidden Illegal Online Gambling Sites on .GO.ID Websites using Web Scraper Algorithms

Name : Muchlis Nurseno

ID of student Faculty : 20210130036

This Thesis has been tested and defended in front of the Board of Examiners in Thesis session on 14 January 2024. In our review, this Thesis is adequate in terms of quality for the purpose of awarding the Master of Computer Degree.

Sukabumi, January 2024

Supervisor 1,	Examiner 1,
Haris Al Qodri Maarif, S.T., M.Sc., PhD	Deshinta Arrova Dewi, M.Sc.,PhD
NIDN. 0418068505	NIDN. 0416127102
Supervisor 2,	Examiner 2,
Umar Aditiawarman, S.T., M.Sc., PhD	Prof. Ir. Media Anugerah Ayu, M.Sc., PhD
NIDN. 0424068107	NIDN. 0315046903

PUBLICATION APPROVAL

As a member of the academic community of Nusa Putra University, i undersigned:

Name	:	Muchlis Nurseno
ID of student	:	20210130036
Study Program	:	Computer Science
Type of Work	:	Thesis

For the sake of scientific development, agree to grant to the University of Nusa Putra the Non-Exclusive Royality-Free Right for my scientific work entitled: **Detecting Hidden Illegal Online Gambling Sites on .GO.ID Websites using Web Scraping Algorithms**

Along with existing devices (if needed). With this non-exclusive royalty-free right, Nusa Putra University has the right to store, transfer media/formats, process in the form of a database, maintain and publish my thesis as long as I keep my name as the author/creator and as the copyright owner.

This statement I made in truth.

Made in : Sukabumi

At the Date of : January 2024

That States

(Muchlis Nurseno)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Detecting Hidden Illegal Online Gambling Sites on .GO.ID Websites using Web Scraper Algorithms

Widespread connectivity has driven the global surge of online gambling, posing adverse effects on Indonesian society and beyond. Online gambling websites in Indonesia are suspected of employing web defacement for SEO enhancement and as a promotional tactic to evade government restrictions. Online gambling operators collaborating with hackers have progressed beyond conventional black hat SEO tactics. They have now adopted Stealthy Defacement techniques to specifically target sections of web pages, rendering illicit content nearly invisible to both authorities and legitimate users. This manipulation enables the defaced content to achieve higher rankings on search engines. The "DESLOT - detecting slot" initiative identifies potentially compromised websites with .go.id domains which are traditionally reserved for governmental entities, commonly exploited by hackers to boost online gambling site ratings. This method involves analyzing keywords like 'slot,' 'judi,' 'gacor,' and 'togel' within webpage content. Research indicates DESLOT's impressive 99.1% detection accuracy in discerning compromised primary web pages of .go.id domains associated with online gambling. Additionally, the study highlights the success of intricate HTML coding techniques in concealing online gambling URLs, making them nearly imperceptible. The research underscores elaborate strategies by online gambling operators and emphasizes the significance of innovative approaches like DESLOT in countering digital space manipulation.

Keywords: Black Hat SEO, Stealthy Defacement, Web Scraper, Government Website, Online Gambling

CHAPTER I
INTRODUCTION

1.1 Background

The rapid growth of internet penetration in Indonesia has revolutionized various aspects of society and the economy. With a staggering 212.9 million internet users at the beginning of 2023, Indonesia has positioned itself as one of the most connected nations globally. This remarkable figure represents a 77.0 percent internet penetration rate(Kemp, 2023), indicating the widespread access to digital media across Indonesian society. The exponential growth of internet connectivity has created a range of opportunities and challenges, including the emergence of online gambling offenses that exploit the digital landscape.

This widespread connectivity has not only facilitated the growth and popularity of online gambling games within Indonesia but has also transcended geographical boundaries, evolving into a global predicament (Setiawati & Ayu Sri Daulat, 2022). The unparalleled ease of 24/7 accessibility, allowing individuals to indulge in online gambling from any location, has contributed significantly to the escalating appeal of these games, particularly among the younger generation and various social circles. As a consequence, online gambling has emerged as a pressing and intricate challenge, transcending national borders and necessitating comprehensive interventions to mitigate its adverse impact on society.

The development of information technology and electronic media has brought about new forms of criminal activities, including online gambling offenses. The borderless nature of the internet allows criminal actors to exploit the medium for transnational criminal activities, with online gambling being a significant aspect. These offenses involve the promotion and facilitation of gambling activities through digital platforms. The accessibility and anonymity provided by the internet have contributed to the proliferation of online gambling, making it a pressing concern for law enforcement agencies and policymakers.

Nowadays, web defacements have become a cost-effective and prospering practice to inject fraudulent content into compromised (high-ranking) websites for promoting illicit goods and services such as unlicensed drugs, adult content, and illegal gambling operations. Web defacements, one of the major promotional channels for online underground economies, inflict significant harm to websites reputations and revenues and may lead to legal ramifications. Criminal actors utilize black hat SEO techniques and hacking of Indonesian websites to promote online gambling activities. In response to these cybercrimes, the Indonesian government has implemented the Information and Electronic Transaction (ITE) Law to address digital offenses and safeguard the integrity of the digital sphere. However, viewed from the standpoint of the sociology of law, the functioning of online gambling laws as a tool for social control is not yet perfect within society. In other words, online gambling laws have not been effective (Kuasa & Jaya, 2022).

In the context of online gambling in Indonesia, the Ministry of Communication and Information (Kemenkominfo) has taken decisive measures to address this matter from 2018 to September 12, 2023. During this period, Kemenkominfo carried out extensive actions, targeting a significant total of 949,388 websites and associated content related to online gambling. This comprehensive effort involved the identification and removal of 898,112 websites and internet protocol (IP) addresses directly linked to online gambling, along with 44,196 instances on file-sharing platforms and an additional 7,080 instances on various social media platforms (CNN Indonesia, 2023).

The proliferation of gambling-related content was widespread, permeating a diverse range of websites, even infiltrating the realm of social media. Additionally, Kemenkominfo remained vigilant by monitoring government websites that had fallen victim to such illicit content. Between January 1, 2022, and September 6, 2023, a concerning total of 9,052 government websites were found to have been infiltrated with online gambling material, highlighting the significance of the issue within the country (Prasetyo, 2023).

Many countries have released regulations and required the involved parties (e.g., search engines) to detect web defacements to thwart the spread of illegal goods and services. Unfortunately, web defacements remain fugitive and challenging to eradicate even though search engines, security service providers (e.g., Google Safe Browsing and other URL safety-checking tools), and law-enforcement officials have fought against them for years (Zhao, 2023).

Recognizing the persistent issue of online gambling in Indonesia, often employing .go.id domains to bypass government restrictions and, in some instances, to enhance their search engine visibility, we embarked on the development of a specialized web scraping algorithm. Constructed using the Python programming language, our web scraper tool is proficient in identifying concealed online gambling websites within webpage content. Following the acquisition of results, an analysis is conducted to investigate how .go.id domain websites are manipulated for black hat SEO purposes in the online gambling industry. The research was conducted from July 27 to November 8, 2023, with the web scraper implementation taking place from October 25 to 26, 2023. We anticipate that this research will yield significant advantages for the public in mitigating hacking activities associated with online gambling.

1.2 Problem Statement

The total accumulated turnover of funds in 2023 related to online gambling, as identified by the Financial Transaction Reports and Analysis Center (PPATK) of the Republic of Indonesia (Kompas, 2024), amounted to Rp 327 trillion in 168 million transactions. Within this total, 3,295,310 individuals were found to engage in online gambling, contributing deposits to online gambling sites totaling Rp 34.51 trillion. The overall cumulative turnover of online gambling funds from 2017 to 2023 reached Rp 517 trillion. This implies that the turnover in 2023 alone accounts for 63 percent of the total turnover of online gambling funds since 2017. Based on these findings, it can be assessed how massive online gambling activities are in Indonesia. Therefore, this research can be considered significant as it unveils how they operate in Indonesia by hacking a website, especially those with the .go.id domain, without being detected by administrators and regular website visitors.

In the realm of Indonesia's digital governance, this issue is crucial and requires immediate attention and investigation. The core problem revolves around the surreptitious infiltration of online gambling activities into websites operating under the esteemed .go.id domain. Traditionally reserved for governmental entities, these platforms serve as vital conduits for essential services, information dissemination, and the promotion of transparency.

The issue at hand is not merely an incursion; it represents a profound threat to the very fabric of governmental digital presence. What intensifies the urgency of this inquiry is the covert manner in which these online gambling phenomena operate. Malicious actors deftly navigate through the digital landscape, utilizing .go.id domains to promote, advertise, or market online gambling activities, thus evading detection by both unsuspecting visitors and vigilant site administrators.

This infiltration poses a dual challenge. Firstly, it compromises the credibility of these platforms, introducing illicit content that contradicts the intended use of .go.id domains. Secondly, it raises significant cybersecurity risks, as these clandestine activities not only challenge the integrity of governmental websites but also amplify concerns about the potential exploitation of sensitive information residing within these platforms.

The gravity of the situation extends to broader societal impacts, as the rise of online gambling activities has been associated with various issues affecting the well-being of individuals and communities across Indonesia. The unregulated and illicit nature of these operations contributes to problems such as addiction, financial distress, and social disruption, magnifying the negative impact on Indonesian society. The research questions guiding this study are as follows:

a. To what extent have websites under the .go.id domain been infiltrated by online gambling activities, operating discreetly without detection?
b. How do malicious actors execute their activities within websites under the .go.id domain without being detected by visitors or administrators?
c. Can the developed web scraper algorithm effectively identify and reveal instances of online gambling infiltration on sites under the .go.id domain without relying on traditional detection methods?

To unravel the depths of this critical issue, this research employs a sophisticated web scraper algorithm, meticulously developed using the Python programming language. The overarching aim is to illuminate the extent of the infiltration and dissect the methodologies employed by these malicious actors. Such insights are imperative, not only for comprehending the scale of the problem but also for devising strategic countermeasures to fortify the resilience and security of governmental online platforms.

As we delve into this investigation, the narrative unfolds with an unwavering focus on the paramount importance of safeguarding these digital bastions of governance from the insidious incursion of online gambling activities. The revelations from this research are poised to not only inform our understanding of the issue but also to pave the way for proactive measures that will shield government websites from the shadows cast by illicit online activities.

1.3 Research Objectives

a. To identify and detect illegal online gambling sites on “.go.id” websites.

b. To develop and implement web scraper algorithms for effectively detecting hidden illegal online gambling websites on ".go.id" domains.

c. To assess the prevalence and characteristics of illegal online gambling sites in the ".go.id" website landscape.

1.4 Significance of Study

a. Enhancing Cyber Security: The identification and detection of illegal online gambling sites on Indonesian government websites will contribute to strengthening cyber security measures within the government's digital infrastructure. By identifying unauthorized content or elements, the research can aid in implementing robust security protocols to safeguard sensitive information and prevent further cyber attacks.

b. Safeguarding Government Reputation: The presence of illegal online gambling activities on government websites can severely impact the reputation and trustworthiness of government platforms. By developing and implementing web scraper algorithms for efficient detection, the research aims to mitigate potential damage to the government's reputation and credibility.

c. Protecting the Public: The prevalence of illegal online gambling sites on government websites poses risks to the public, especially to vulnerable groups like the younger generation. The research seeks to protect citizens from exposure to such illicit activities and raises awareness about the dangers of online gambling.

d. Policy and Regulation: Assessing the prevalence and characteristics of illegal online gambling sites on government websites can inform policymakers and regulators about the extent of the issue. The findings can contribute to the formulation of effective policies and regulations to address the problem at a national level.

e. Developing Web Scraper Technology: The development and implementation of web scraper algorithms specialized in detecting illegal online gambling activities can advance the field of web scraping and content detection. The research can serve as a foundation for future advancements in web scraper technology to address various forms of online illicit activities.

f. Data-Driven Decision Making: By gathering empirical data on the prevalence and characteristics of illegal online gambling sites, the research enables data-driven decision making. This information can guide the allocation of resources and the development of targeted interventions to combat the issue effectively.

g. International Implications: As illegal online gambling is not confined to national borders, the research can have international implications. Understanding the landscape of illegal online gambling activities on government websites can contribute to international collaborations and information sharing to combat transnational cyber threats.

1.5 Limitation of Problems and Assumptions

a. Limited Representativeness: This research is focused on the .go.id domain, which has traditionally been affiliated or represented as government-owned websites. While this approach allows for a more specific investigation, it may exclude other government-related websites or sources that could also potentially contain illegal online gambling sites for black hat SEO. The data source used to compile a list of .go.id domain websites is not limited to the data published by the Ministry of Communication and Information (domain.go.id), which amounted to 3,410 as of September 20, 2023.

b. Limited Scraping Scope: The web scraper algorithm is designed to scrap only the main pages of the government websites. Recursive crawling to follow links within the websites is not implemented. Consequently, this may result in incomplete data collection, as deeper website content with potential illegal gambling promotions or sites might not be explored.

c. Language and Content Diversity**:** The diversity of languages and content used in online gambling promotions poses a challenge for detection. Some government websites may contain content in local languages or dialects specific to the Indonesian audience. Additionally, websites involved in black hat SEO or suspected of being compromised may employ specialized advertising language catered to their Indonesian audience, using terms such as ‘slot,’ ‘judi,’ ‘togel,’ or ‘gacor.’ The web scraper algorithm's effectiveness in identifying these specific terms and understanding their context.

d. Dynamic Website Content: The research data was collected on July 27 and 28, 2023, while the implementation of the developed algorithm took place from October 23 to November 8, 2023. As government websites are subject to continuous updates, the content observed during the research may change over time. The study's snapshot of the websites' content on specific dates may not fully represent the current state of hidden online gambling sites on government websites. Frequent updates and changes to website content may lead to variations in the prevalence of illegal gambling sites on these websites.

e. False Positives and Negatives: The utilization of a web scraper algorithm in this study could yield erroneous outcomes, such as false positives, which inaccurately detect .go.id websites manipulated for black hat SEO in support of online gambling activities, and false negatives, which overlook specific instances of concealed online gambling websites. These inaccuracies may impact the reliability and precision of the results, potentially leading to misleading conclusions.

CHAPTER V
CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

To identify and detect illegal online gambling sites on ".go.id" websites, our approach commenced with a comprehensive process that began with Google Dork searches, enabling us to examine commonly used keywords in online gambling and observe how webpages under the ".go.id" domain have been infiltrated with online gambling content. This strategic method not only facilitated the identification of prevalent keywords associated with online gambling but also provided insights into the specific techniques employed by malicious actors to embed such content into ".go.id" web pages.

This initial step in the identification process laid the foundation for the development and implementation of our web scraper algorithms, ensuring a targeted and effective exploration of the ".go.id" website landscape. Upon reflection of the outcomes derived from our extensive research, it is apparent that our journey has successfully fulfilled the established objectives. This accomplishment has resulted in a significant improvement in the accuracy of our algorithm and a more profound understanding of the online gambling landscape within ".go.id" websites.

The strategic refinement of our algorithm signifies a pivotal milestone in our research journey. Through collaborative efforts, we seamlessly integrated an additional layer of automated validation, conducting a meticulous analysis of the entire HTML structure containing URLs associated with online gambling. This augmentation played a crucial role in refining the precision of our algorithm, ensuring a thorough examination of websites entwined with online gambling activities.

Our algorithm follows a two-fold approach, initially scrutinizing the entire HTML content to identify elements containing URLs related to online gambling. It then enhances its analysis by specifically examining the 'Title' segment. This comprehensive strategy enables us to determine with heightened accuracy whether a website has been infiltrated by malicious actors involved in online gambling. Notably, the incorporation of strategic keywords such as 'slot,' 'judi,' 'gacor,' and 'togel' has propelled the algorithm's ability to recognize concealed online gambling-related content.

In evaluating the effectiveness of our algorithm, a meticulous validation process unfolded. Out of 183,927 web-scraped sites, 1,482 were identified with keywords like 'slot,' 'judi,' 'gacor,' or 'togel.' However, at this stage, definitive confirmation of whether these keywords indicated infection remained elusive. Python was once again employed to validate whether these sites were genuinely infected with online gambling-related elements. Manual verification ensued to assess accuracy after obtaining results. Continuous improvement was pursued by introducing an additional layer of automated validation, specifically targeting the 'Title' segment after the initial validation of the 'URL' section. This approach elevated accuracy to 99.1%, showcasing the algorithm's proficiency in effectively identifying websites related to online gambling.

Our observations delved into the darker recesses of the Indonesian online landscape, laying bare the widespread use of advanced HTML coding techniques employed by websites housing numerous URLs related to online gambling. The clandestine operations were characterized by methods such as transparency, hidden attributes, and background-matching colors, all strategically orchestrated to cloak online gambling links from casual visitors. This revelation shed a stark light on the prevalence of black hat SEO practices, particularly within the ".go.id" domains, unraveling a complex web of illicit activities.

The interconnected domains observed in our study illuminated a coordinated effort to exploit compromised ".go.id" websites for bolstering the ratings and SEO of online gambling sites. The revelation of interconnecting subdomains with shared IP addresses hinted at a network of compromised websites nestled within the ".go.id" domain space. This network's actions prompted serious concerns regarding the security and integrity of ".go.id" domains, urging the imperative need for a comprehensive investigation into the extent of this infiltration.

The strategic insertion of hidden links within the content and design elements of online gambling websites emerged as a pivotal discovery. These intentionally obscured links, accessible to search engine crawlers while remaining invisible to regular users, mirrored tactics reminiscent of Black Hat SEO practices in China.

This finding underscores the urgent necessity for vigilance in identifying and countering such tactics to preserve the integrity of online platforms and search engine algorithms.

Understanding the significance of these hidden links involves recognizing their role as a form of backlink manipulation. In the realm of search engine optimization (SEO), backlinks serve as essential elements that contribute to a website's authority and visibility in search results. A backlink is essentially a hyperlink from one website to another, acting as a vote of confidence or recommendation. When a website with a strong reputation and relevance provides a backlink to another site, it can positively impact the latter's search engine rankings.

In the context of the discovered hidden links on online gambling websites, it becomes evident that they were strategically placed to manipulate search engine algorithms. These covert links likely aimed to boost the visibility and rankings of the implicated websites within search results, employing tactics associated with Black Hat SEO. Such practices go against the principles of ethical SEO, as they attempt to deceive search engines and artificially inflate a site's perceived authority.

In summation, our research transcends mere algorithmic refinement; it unravels a tapestry of intricacies characterizing the nexus between online gambling, black hat SEO practices, and compromised ".go.id" domains. The implications of our findings extend beyond academic realms, beckoning industry stakeholders, regulators, and cybersecurity experts to collaboratively address the multifaceted challenges posed by the evolving landscape of cyber threats within the Indonesian online sphere.

5.2 Recommendations

In light of our findings and the complexity of online gambling promotion, several crucial recommendations emerge to mitigate the risks associated with black hat SEO practices and enhance the security and integrity of online domains.

First and foremost, there is a clear need for the development of more advanced web scraping tools. These tools should be designed to adapt to the diverse text formats employed by malicious actors in the online gambling sector. Continuous research and development in this area are essential to ensure data collection remains accurate and comprehensive, especially as cybercriminals evolve their tactics.

Secondly, fostering collaborative efforts among stakeholders is paramount. Given the transnational nature of online gambling promotion, the exchange of information, expertise, and best practices among researchers, cybersecurity experts, law enforcement agencies, and domain administrators is critical. Establishing a coordinated network for sharing threat intelligence will enhance the collective ability to combat illicit online activities that transcend national borders.

Furthermore, the development and enforcement of policies and regulations specifically targeting black hat SEO practices in online gambling are imperative. Policymakers and regulatory bodies should work in tandem with industry experts to formulate clear legal frameworks and regulations. Such regulations act as a deterrent to these activities and provide a legal basis for taking action against offenders. Law enforcement agencies should be well-equipped and adequately resourced to investigate and prosecute those involved.

To maintain the effectiveness of security measures and regulatory frameworks, the establishment of continuous monitoring and evaluation mechanisms is crucial. These mechanisms involve tracking the impact of interventions over time, assessing the reduction in compromised websites, and analyzing any changes in the tactics employed by cybercriminals. Ongoing evaluation ensures that strategies remain adaptive and effective in combating evolving threats.

Finally, the adoption of robust cybersecurity best practices by website owners and administrators is vital. This includes regular security assessments, timely software updates, strong password policies, and network segmentation to protect websites against potential threats. Remove unused subdomains on website. Neglecting supervision over inactive subdomains can create opportunities for attacks. Remember, the more doors open, the greater the potential for vulnerabilities that can be exploited by criminals. Avoid disclosing sensitive information about the server, such as SSH, FTP, database credentials, and others, to unauthorized individuals or publicly. Safeguarding this information carefully can help protect system's security from potential threats.

Collaboration with cybersecurity experts and the implementation of advanced intrusion detection systems will help identify and mitigate security breaches promptly. Administrators can also implement measures to address incidents of online gambling web defacement by referring to documents published by the National Cyber and Crypto Agency (Badan Siber dan Sandi Negara, 2023). Proactively safeguarding websites is key to reducing the risk of falling victim to IP address compromises and strengthening the overall cybersecurity posture.

REFERENCE

Akarsh, I., & Regulagedda, R. M. (2021). Parallelization of Web Crawler with Multithreading and Natural Language Processing. *International Research Journal of Engineering and Technology*. www.irjet.net

Albalawi, M., Aloufi, R., Alamrani, N., Albalawi, N., Aljaedi, A., & Alharbi, A. R. (2022). Website Defacement Detection and Monitoring Methods: A Review. In *Electronics (Switzerland)* (Vol. 11, Issue 21). MDPI. https://doi.org/10.3390/electronics11213573

Alfarisy, G. A. F., & Bahtiar, F. A. (2017). Focused Web Crawler for Indonesian Recipes. *International Conference on Sustainable Information Engineering and Technology (SIET)*.

Amudha, S., Sc, M., & M Phil. (2017). Web Crawler for Mining Web Data. *International Research Journal of Engineering and Technology*. www.irjet.net

Arifanto, R., Asnar, Y. D. W., & Liem, M. M. I. (2018). Domain Specific Language for Web Scraper Development. *Proceedings of 2018 5th International Conference on Data and Software Engineering (ICoDSE) : Jayakarta Hotel, Senggigi Beach, Lombok, Indonesia, November 7th-8th, .*

Badan Siber dan Sandi Negara. (2023). *Web Defacement : Judi Online*.

Bricongne, J. C., Meunier, B., & Pouget, S. (2023). Web-scraping housing prices in real-time: The Covid-19 crisis in the UK. *Journal of Housing Economics*, *59*. https://doi.org/10.1016/j.jhe.2022.101906

Canali. Davide, Cova, M., Kruegel, C., & Vigna, G. (2011). Prophiler: A Fast Filter for the Large-Scale Detection of Malicious Web Pages. *Proceedings of the 20th International Conference on World Wide Web, March, Hyderabad, India*, 197–206.

Castillo, C., Donato, D., Gionis, A., Murdock, V., & Silvestri, F. (n.d.). Know your Neighbors: Web Spam Detection using the Web Topology. *SIGIR '07 : 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*.

Chung, young-joo, Toyoda, M., & Kitsuregawa, M. (2009). A Study of Link Farm Distribution and Evolution using a Time Series of Web Snapshots. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 9–16.

CNN Indonesia. (2023, September 13). Cara Kominfo Adu Siasat dengan Situs Judi Online yang Terus Muncul. *Https://Www.Cnnindonesia.Com/Teknologi/20230913145007-192-*

998526/Cara-Kominfo-Adu-Siasat-Dengan-Situs-Judi-Online-Yang-Terus-Muncul.

FOFA. (n.d.). *FOFA: Foresee & Find it All*. Https://En.Fofa.Info/about/En.

FOFA. (2023). *FOFA Python SDK Documentation*. Https://Github.Com/Fofapro/Fofa-Py.

Invernizzi, L., Comparetti, P. M., Benvenuti, S., Kruegel, C., Cova, M., & Vigna, G. (2012). EvilSeed: A guided approach to finding malicious web pages. *Proceedings - IEEE Symposium on Security and Privacy*, 428–442. https://doi.org/10.1109/SP.2012.33

John, J. P., Yu, F., Xie, Y., Krishnamurthy, A., & Abadi, M. (2011). deSEO: Combating Search-Result Poisoning. *20th USENIX Security Symposium (USENIX Security 11)*.

Kemp, S. (2023, February). *DIGITAL 2023: INDONESIA*. Https://Datareportal.Com/Reports/Digital-2023-Indonesia.

Kuasa, D. A., & Jaya, F. (2022). Fenomena Judi Online: Hukum & Masyarakat. *Jurnal Hukum*, *5*(2). http://publishing-widyagama.ac.id/ejournal-v2/index.php/yuridika/

Ma, X., & Yan, M. (2021). Design and implementation of crawler program based on python. *Journal of Physics: Conference Series*, *2033*(1). https://doi.org/10.1088/1742-6596/2033/1/012205

Maggi, F., Balduzzi, M., Flores, R., Gu, L., & Ciancaglini, V. (2018). Investigating web defacement campaigns at large. *ASIACCS 2018 - Proceedings of the 2018 ACM Asia Conference on Computer and Communications Security*, 443–456. https://doi.org/10.1145/3196494.3196542

Mao, B. M., & Bagolibe, K. D. (2019). A contribution to detect and prevent a website defacement. *Proceedings - 2019 International Conference on Cyberworlds, CW 2019*, 344–347. https://doi.org/10.1109/CW.2019.00062

Min, M., Lee, J. J., & Lee, K. (2022). Detecting Illegal Online Gambling (IOG) Services in the Mobile Environment. *Security and Communication Networks*, *2022*. https://doi.org/10.1155/2022/3286623

Mitchell, R. (2018). *Web Scraping with Python: COLLECTING MORE DATA FROM THE MODERN WEB*. www.allitebooks.com

Muehlethaler, C., & Albert, R. (2021). Collecting data on textiles from the internet using web crawling and web scraping tools. *Forensic Science International*, *322*. https://doi.org/10.1016/j.forsciint.2021.110753

Ntoulas, A., Najork, M., Manasse, M., & Fetterly, D. (2006). Detecting Spam Web Pages through Content Analysis. *WWW '06: Proceedings of the 15th International Conference on World Wide Web*, 83–92.

Park, A. J., Quadari, R. N., & Tsang, H. H. (2017). Phishing Website Detection Framework Through Web Scraping and Data Mining. *2017 8th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 680–684.

Pramudita, Y. D., Anamisa, D. R., Putro, S. S., & Rahmawanto, M. A. (2020). Extraction System Web Content Sports New Based on Web Crawler Multi Thread. *Journal of Physics: Conference Series*, *1569*(2). https://doi.org/10.1088/1742-6596/1569/2/022077

Prasetyo, A. (2023, September 14). Menkominfo: Ruang Digital bakal Bersih dari Judi Olnline dalam Sepekan. *Https://Mediaindonesia.Com/Politik-Dan-Hukum/613296/Menkominfo-Ruang-Digital-Bakal-Bersih-Dari-Judi-Online-Dalam-Sepekan*.

Saputra Hasibuan, E. (2023). The Police are Indecisive: Online Gambling is Rising. Facts About The Eradication of Online Gambling in The Field. *Journal of Social Research*, *2*. http://ijsr.internationaljournallabs.com/index.php/ijsr

Schedlbauer, J., Raptis, G., & Ludwig, B. (2021). Medical informatics labor market analysis using web crawling, web scraping, and text mining. *International Journal of Medical Informatics*, *150*. https://doi.org/10.1016/j.ijmedinf.2021.104453

Search Engine Optimization (SEO) Starter Guide. (n.d.). Https://Developers.Google.Com/Search/Docs/Fundamentals/Seo-Starter-Guide.

Setiawati, S., & Ayu Sri Daulat, P. (2022). The Urgency of Special Regulations for online Gambling in Indonesia. *International Journal of Arts and Social Science*, *Volume 5*(Issue 7). www.ijassjournal.com

Sharma, A. K., Shrivastava, V., & Singh, H. (2020). Experimental performance analysis of web crawlers using single and Multi-Threaded web crawling and indexing algorithm for the application of smart web contents. *Materials Today: Proceedings*, *37*(Part 2), 1403–1408. https://doi.org/10.1016/j.matpr.2020.06.596

Shrivastava, V., & Saxena, S. (2019). Comparative Analysis of Web Crawling Algorithms for Improvement in Web Crawler. *International Conference on Sustainable Computing in Science, Technology & Management (SUSCOM-2019)*, 1531–1539. https://ssrn.com/abstract=3356275

Sun, G., Xiang, H., & Li, S. (2019). On Multi-Thread Crawler Optimization for Scalable Text Searching. *Journal on Big Data*, *1*(2), 89–106. https://doi.org/10.32604/jbd.2019.07235

Tong, S., Zhang, H., Shen, B., Zhong, H., Wang, Y., & Jin, B. (2016). *Detecting Gambling Sites From Post Behaviors*. http://hadoop.apache.org/

Kompas. (2024, January 10). *PPATK: Perputaran Dana Judi “Online” Capai Rp 327 Triliun Sepanjang 2023*. Https://Nasional.Kompas.Com/Read/2024/01/10/12150571/Ppatk-Perputaran-Dana-Judi-Online-Capai-Rp-327-Triliun-Sepanjang-2023.

Urvoy, T., Chauveau, E., Filoche, P., & Lavergne, T. (2008). Tracking Web spam with HTML style similarities. *ACM Transactions on the Web*, *2*(1). https://doi.org/10.1145/1326561.1326564

vanden Broucke, S., & Baesens, B. (2018). Practical Web Scraping for Data Science. In *Practical Web Scraping for Data Science*. Apress. https://doi.org/10.1007/978-1-4842-3582-9

Wu, B., & Davison, B. D. (2005a). Cloaking and Redirection: A Preliminary Study. *In Proceedings of the First International Workshop on Adversarial Information Retrieval on the Web (AIRWeb),* 7–16. http://www.googlebot.com/bot.html

Wu, B., & Davison, B. D. (2005b). Identifying Link Farm Spam Pages. *WWW ’05: Special Interest Tracks and Posters of the 14th International Conference on World Wide Web*, 820–829. http://www.lowcostwireless4u.com/

Wu, B., & Davison, B. D. (2006). Detecting Semantic Cloaking on the Web. *WWW ’06: Proceedings of the 15th International Conference on World Wide Web*, 819–828.

Yang, R., Chi, C., Wang, D., He, J., Pang, S., Wang, X., & Lau, W. C. (2021). Scalable Detection of Promotional Website Defacements in Black Hat SEO Campaigns. *Proceedings of the 30th USENIX Security Symposium*, 3703–3720.

Yang, R., Liu, J., Gu, L., & Chen, Y. (2020). Search & Catch: Detecting Promotion Infection in the Underground through Search Engines. *IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications*, 1566–1571. https://doi.org/10.1109/TrustCom50675.2020.00216/20/$31.00

Zhao, R. (2023). The Chameleon on the Web: an Empirical Study of the Insidious Proactive Web Defacements. *ACM Web Conference 2023 - Proceedings of the World Wide Web Conference, WWW 2023*, 2241–2251. https://doi.org/10.1145/3543507.3583377