Pengembangan *Engine Web Crawler* Sebagai Pencari Jejak Serangan *Cyber Stored Cross-Site Scripting*

Ilham Yoga Prabhaswara[a1], I Made Agus Dwi Suarjaya[a2], Ni Kadek Dwi Rusjayanthi[a3]
[a]Program Studi Teknologi Informasi, Fakultas Teknik, Universitas Udayana
Bukit Jimbaran, Bali, Indonesia, Telp. (0361) 701806
e-mail: [1]ilhamprabhaswara@gmail.com, [2]agussuarjaya@it.unud.ac.id,
[3]dwi.rusjayanthi@unud.ac.id

Abstrak

Kerentanan Cross-site Scripting (XSS) telah lama menjadi perhatian dalam keamanan aplikasi web dan termasuk dalam daftar OWASP Top 10. Pada tahun 2017 XSS menduduki peringkat 6, namun pada tahun 2021 kerentanan ini naik ke peringkat 4 dalam kategori Injection. Kerentanan ini memanfaatkan form input yang tidak tervalidasi dengan baik. Penelitian ini bertujuan untuk mengidentifikasi laman web yang rentan terhadap serangan Stored Cross-site Scripting. Metode penelitian dilakukan dengan melakukan pencarian pada tiga tingkat kedalaman. Web scraping digunakan untuk mengekstrak data dari laman web, dan source code laman web dibandingkan dengan pola serangan Stored Cross-site Scripting menggunakan Algoritma Knuth-Morris-Pratt. Hasil penelitian menunjukkan bahwa beberapa laman web memiliki pola serangan dan jejak serangan yang terdeteksi, sementara beberapa lainnya hanya memiliki pola serangan tanpa jejak serangan yang terlihat. Berdasarkan analisis manual dari 56 data yang diambil secara acak dari penelitian, ditemukan bahwa 5 laman web memiliki nilai true positive, di mana pola serangan ditemukan dan terdapat jejak serangan di dalamnya. Sementara itu, 49 laman web lainnya memiliki nilai true negative, di mana pola serangan ditemukan namun tidak terdapat jejak serangan di dalamnya. Penelitian ini memberikan wawasan tentang laman web yang rentan terhadap serangan Stored Cross-site Scripting. Hasil penelitian ini dapat digunakan untuk meningkatkan keamanan aplikasi web dan mengurangi potensi serangan Cross-site Scripitng di masa depan..

Kata kunci: *Stored Cross-site Scripting, Kerentanan Web, OWASP Top 10, Web Scraping, Algoritma Knuth-Morris-Pratt.*

Abstract

Cross-site Scripting (XSS) vulnerability has long been a concern in web application security and is included in the OWASP Top 10 list. In 2017, XSS ranked 6th, but in 2021, it rose to the 4th position in the Injection category. This vulnerability exploits poorly validated input forms. This study aims to identify web pages that are vulnerable to Stored Cross-site Scripting attacks. The research is conducted by performing a search at three levels of depth. Web scraping is used to extract data from web pages, and the source code of the web pages is compared to Stored Cross-site Scripting attack patterns using the Knuth-Morris-Pratt algorithm. The results of the study indicate that some web pages exhibit detected attack patterns and traces of attacks, while others only show attack patterns without visible traces of attacks. Based on manual analysis of 56 randomly selected data from the research, it was found that 5 web pages had true positive values, indicating the presence of attack patterns and traces of attacks. Meanwhile, 49 other web pages had true negative values, where attack patterns were detected but no traces of attacks were found. This research provides insights into web pages vulnerable to Stored Cross-site Scripting attacks. The findings can be used to enhance web application security and reduce the potential for Cross-site Scripting attacks in the future.

Keywords : *Stored Cross-site Scripting, Web Vulnerability, OWASP Top 10, Web Scraping, Knuth-Morris-Pratt algorithm.*

1. Pendahuluan

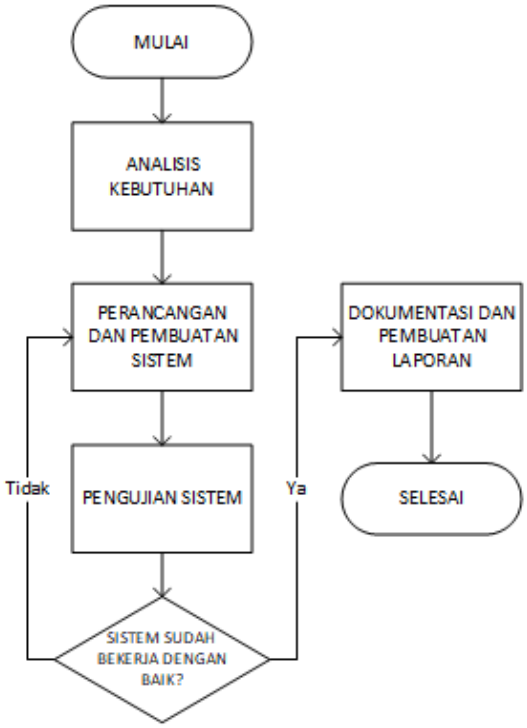
Cross-site Scripting merupakan salah satu jenis serangan pada aplikasi website melalui sebuah *form input* pada browser yang tidak tervalidasi dengan baik. *Cross-site Scripting* menyebabkan penyerang dapat melewati celah keamanan di sisi klien dengan tujuan untuk mendapatkan informasi sensitif milik korban atau menyisipkan aplikasi berbahaya yang dapat menyebabkan kerugian pada korban [1]. *Cross-site Scripting* yang telah menargetkan sebuah website dapat membuat website menyimpan banyak *string* yang dijadikan *payload* untuk melakukan serangan *cross-site scripting*. Risiko yang terjadi dapat dikurangi dengan metode *web scraping* yang mengekstraksi serta mengumpulkan data dari sebuah website. Data hasil ekstraksi dibandingkan dengan pola serangan *Cross-site Scripting* menggunakan Algoritma KMP (Knuth-Morris-Pratt) sehingga diharapkan adanya pola serangan *Stored Cross-site Scripting* pada suatu website diteliti.

Penelitian terhadap *cross-site scripting* telah dilakukan oleh sejumlah peneliti. Penelitian mengenai *cross-site scripting* dengan memanfaatkan framework DVWA (*Damn Vulnerable Web Application*) yang merupakan framework untuk melakukan penetration testing milik OWASP (*Open Web Application Security Project*). Penelitian ini bertujuan untuk mengetahui cara kerja jenis serangan *Reflected* XSS dan *Stored* XSS serta menyediakan pencegahan atau tindakan preventif terhadap jenis serangan ini [2].

Penelitian mengenai implementasi *web scraping* untuk mengakuisisi data pada jurnal SINTA berupa judul jurnal, judul penelitian, author 2 dan afiliasi [3]. Hasil analisis berupa tren topik penelitian kesehatan di Indonesia khususnya pada Jurnal SINTA. Implementasi Algoritma Knuth-Morris-Pratt [4] dilakukan dengan memodifikasi fitur pencarian dengan menggunakan bahasa pemrograman PHP. Sistem dikembangkan untuk mengoptimalisasi pencarian definisi istilah Standar Operasional Prosedur (SOP) pada Lembaga Pnjamin Mutu UIN Ar-Raniry. Hasil penelitian menunjukkan bahwa algoritma Knuth-Morris-Pratt (KMP) mampu mengoptimalkan fitur pencarian. Tujuan penulis adalah untuk menganalisis laman website yang memiliki pola serangan dan jejak serangan, dengan menggunakan web scraping untuk mengekstraksi data dan membandingkannya dengan pola serangan menggunakan algoritma KMP. Pengujian dilakukan secara luas dengan tiga level kedalaman menggunakan metode web crawler.

2. Metode Penelitian

Penelitian terkait pengembangan *engine web crawler* sebagai pencari jejak serangan *stored cross-site scripting* dilakukan dalam empat langkah yang dijabarkan melalui diagram alir dapat dilihat pada Gambar 1 berikut.



Gambar 1. Alur Penelitian

Gambar 1 merupakan tahapan dari metodologi penelitian yang digunakan dalam pembuatan penelitian. Pelaksanaan terhadap pembuatan penelitian terdiri dari empat tahapan yaitu, analisa kebutuhan perancangan dan pembuatan sistem, pengujian sistem dan melakukan dokumentasi terhadap sistem yang telah dibuat dalam bentuk laporan. Berikut rincian detail dari masing-masing tahapan metodologi penelitian.

3. Kajian Pustaka

Kajian pustaka adalah konsep-konsep penopang sebagai penuntun dalam pengembangan *engine web crawling* sebagai pencari jejak serangan *cyber Stored Cross-Site Scripting* (XSS).

3.1 Cross-site Scripting

Serangan Cross-site Scripting (XSS) merupakan salah satu jenis serangan pada aplikasi website melalui sebuah form input pada browser yang tidak tervalidasi dengan baik. Serangan XSS dilakukan oleh penyerang dengan cara Cross-site Scripting memasukkan kode HTML, JavaScript atau client side script berbahaya lainnya ke dalam aplikasi website yang bertujuan sebagai upaya untuk mencuri kredensial Login pengguna dan informasi sensitif lainnya seperti session token atau informasi akun keuangan [1], [5] . Pola serangan *Cross-site scripting* yang digunakan pada penelitian ini terdapat dalam cheat sheet milik OWASP [6], [7].

3.2 Python

Python adalah bahasa pemrograman tingkat tinggi yang sangat populer saat ini. Hal ini tidak lepas dari bahasa Python yang dianggap powerful dan dekat dengan bahasa manusia. Python adalah bahasa pemrograman "intrepeter". Artinya kode dieksekusi langsung sesuai dengan instruksi yang ditulis dalam bahasa pemrograman atau bahasa scripting tanpa terlebih dahulu dikonversi ke kode objek, seperti *compiler* [8].

3.3 Web Crawling

Web crawling adalah teknik pengumpulan data otomatis yang dilakukan oleh program komputer untuk mengekstrak informasi dari situs web secara sistematis [9].

3.4 Web Scraping

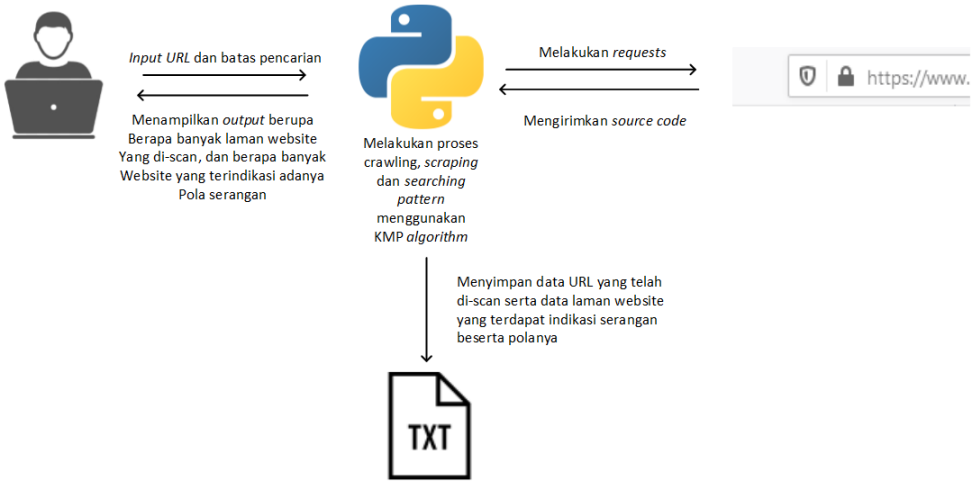
Web scraping adalah proses mengekstrak informasi dari website secara otomatis dengan cara mengurai tag *hypertext* dan mengambil informasi berupa teks, gambar, dan video yang disematkan di dalamnya dari sejumlah besar data dari halaman website [10].

3.5 Algoritma Knuth-Morris-Pratt

Algoritma Knuth-Morris-Pratt (KMP) merupakan suatu algoritma pencarian string untuk mencari teks berdasarkan urutan dari kiri ke kanan yang dikembangkan oleh D. E. Knuth, J. H. Morris dan V. R. Pratt [11]. Algoritma KMP bekerja dengan cara mencocokkan *pattern* atau susunan kata yang dicari dari kiri ke kanan pada awal teks, lalu menggeser susunan kata sampai susunan kata tersebut berada di ujung teks.

4. Hasil dan Pembahasan

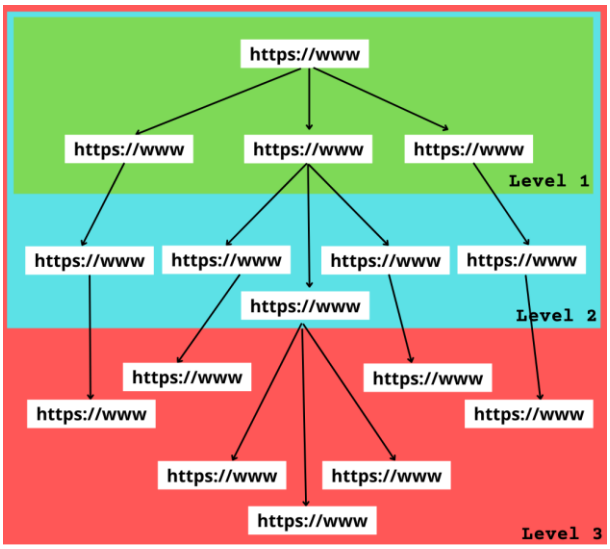
4.1 Gambaran Umum Sistem



Gambar 2. Gambaran umum

Gambar 2. merupakan gambaran umum sistem. Alur yang dapat dijelaskan pada gambaran umum sistem yaitu mulanya pengguna menjalan sistem yang telah dibuat. Saat sistem telah dijalankan, pengguna memasukkan URL yang ingin di-*scan* serta menginputkan batas tingkatan pencarian antara 1 - 3. Sistem akan mengirimkan *requests* ke laman tersebut lalu sistem akan menerima *feedback* berupa *source code* dari laman tersebut. Setelah menerima *feedback*, sistem akan melakukan proses *crawling*, *scraping* dan melakukan pencarian *pattern* dengan menggunakan algoritma yang sudah ditentukan. Hasil pemrosesan tersebut disimpan pada file txt yang sudah dibuat secara otomatis saat proses dijalankan.

4.2 Hasil Penelitian



Gambar 3. Kedalaman Pengujian

Gambar 3 merupakan ilustrasi dari pengujian yang telah dilakukan. Dapat dijelaskan bahwa kedalaman level satu merupakan pengujian pada satu level turunan. Sebagai contoh sebuah laman web memiliki 30 link yang terdapat di dalamnya, pengujian level satu adalah melakukan proses pencarian pola pada laman web tersebut serta pada 30 link yang terdapat di dalamnya. Pengujian level dua dilakukan pada laman web awal dan 30 link yang tersedia, namun dari 30 link yang tersedia juga akan dilakukan proses scrapping satu per satu untuk mendapatkan link yang terdapat pada masing-masing laman. Bila rata-rata pada suatu laman terdapat 30 link, maka total laman yang dijadikan target proses pencarian pola adalah sekitar 900-1000 laman. Konsep pengujian pada level tiga mengikuti konsep turunan pada konsep pengujian level sebelumnya.

4.2.1 Pengujian pada blog.devtiersoftware.com/posts

Berikut merupakan tabel hasil pengujian pada laman blog.devtiersoftware.com/posts dengan kedalaman pengujian dari level 1 sampai 3. Proses awal adalah menginputkan url dan batasan pencarian di dalam program. Selanjutnya dilakukan scraping data url sesuai kedalaman level pengujian yang sudah ditentukan.

Tabel 1. Tabel hasil pengujian level satu

URL	https://blog.devtiersoftware.com/posts
Level	1
Laman di-scan	28 laman
Pattern ditemukan pada laman	10 laman
Lama Waktu Eksekusi	6 menit 37 detik

Tabel 1 merupakan hasil pengujian sistem dengan kedalaman pencarian level satu. Pengujian sistem yang dilakukan dengan target URL https://blog.devtiersoftware.com/posts telah berhasil dilakukan dan berhasil melakukan scanning pada 28 laman dan menemukan pola serangan pada 21 laman tersebut. Pengujian sistem dengan kedalaman pencarian level satu memerlukan waktu selama 6 menit 37 detik untuk melakukan proses pengujian.

Tabel 2. Tabel hasil pengujian level dua

URL	https://blog.devtiersoftware.com/posts
Level	2
Laman di-scan	76 laman
Pattern ditemukan pada laman	38 laman

Lama Waktu Eksekusi	18 menit 9 detik

Tabel 2 merupakan hasil pengujian sistem dengan kedalaman pencarian level dua. Pengujian sistem yang dilakukan dengan target URL https://blog.devtiersoftware.com/posts telah berhasil dilakukan dan berhasil melakukan scanning pada 76 laman dan menemukan pola serangan pada 38 laman tersebut. Pengujian sistem dengan kedalaman pencarian level dua memerlukan waktu selama 18 menit 9 detik untuk melakukan proses pengujian.

Tabel 3. Tabel hasil pengujian level tiga

URL	https://blog.devtiersoftware.com/posts
Level	3
Laman di-scan	167 laman
Pattern ditemukan pada laman	53 laman
Lama Waktu Eksekusi	1 jam 47 menit 29 detik

Tabel 3 merupakan hasil pengujian sistem dengan kedalaman pencarian level tiga. Pengujian sistem yang dilakukan dengan target URL https://blog.devtiersoftware.com/posts telah berhasil dilakukan dan berhasil melakukan scanning pada 167 laman dan menemukan pola serangan pada 53 laman tersebut. Pengujian sistem dengan kedalaman pencarian level tiga memerlukan waktu selama 1 jam 47 menit 29 detik untuk melakukan proses pengujian.

4.2.2 Pengujian pada jdih.klungkungkab.go.id

Berikut merupakan tabel hasil pengujian pada laman jdih.klungkungkab.go.id dengan kedalaman pengujian dari level 1 sampai 3. Proses awal adalah menginputkan url dan batasan pencarian di dalam program. Selanjutnya dilakukan scraping data url sesuai kedalaman level pengujian yang sudah ditentukan.

Tabel 4. Hasil pengujian level satu

URL	https:// jdih.klungkungkab.go.id
Level	1
Laman di-scan	95 laman
Pattern ditemukan pada laman	21 laman
Lama Waktu Eksekusi	49 menit 57 detik

Tabel 4 merupakan hasil pengujian sistem dengan kedalaman pencarian level satu. Pengujian sistem yang dilakukan dengan target URL https:// jdih.klungkungkab.go.id telah berhasil dilakukan dan berhasil melakukan scanning pada 95 laman dan menemukan pola serangan pada 21 laman tersebut. Pengujian sistem dengan kedalaman pencarian level satu memerlukan waktu selama 49 menit 57 detik untuk melakukan proses pengujian.

Tabel 5. Hasil pengujian level dua

URL	https://jdih.klungkungkab.go.id
Level	2
Laman di-scan	882 laman
Pattern ditemukan pada laman	118 laman
Lama Waktu Eksekusi	12 jam 35 menit 56 detik

Tabel 5 merupakan hasil pengujian sistem dengan kedalaman pencarian level dua. Pengujian sistem yang dilakukan dengan target URL https:// jdih.klungkungkab.go.id telah berhasil dilakukan dan berhasil melakukan scanning pada 882 laman dan menemukan pola serangan pada 118 laman tersebut. Pengujian sistem dengan kedalaman pencarian level dua memerlukan waktu selama 12 jam 35 menit 56 detik untuk melakukan proses pengujian.

Tabel 6. Hasil pengujian level tiga

URL	https:// jdih.klungkungkab.go.id
Level	3
Laman di-scan	4723 laman
Pattern ditemukan pada laman	883 laman

Lama Waktu Eksekusi	3 hari 15 jam 30 menit 6 detik

Tabel 6 merupakan hasil pengujian sistem dengan kedalaman pencarian level tiga. Pengujian sistem yang dilakukan dengan target URL https:// jdih.klungkungkab.go.id telah berhasil dilakukan dan berhasil melakukan scanning pada 4723 laman dan menemukan pola serangan pada 883 laman tersebut. Pengujian sistem dengan kedalaman pencarian level tiga memerlukan waktu selama 3 hari 15 jam 30 menit 6 detik untuk melakukan proses pengujian.

4.3 Pembahasan Hasil Penelitian

Analisis akurasi pengujian dilakukan dengan cara pengujian manual yang dilakukan dengan memilih 10 data secara acak. Pengujian manual dilakukan dengan mengunjungi laman web yang terkait dan membandingkan secara manual apakah terdeteksi adanya pola serangan serta jejak serangan atau tidak.

4.3.1 Hasil pada blog.devtiersoftware.com/posts

Tabel 7. Hasil akhir pengujian level satu

No.	URL	True Positive	True Negative	False Positive	False Negative
1.	https://blog.devtiersoftware.com/posts/et-odio-illo-distinctio		✓		
2.	https://blog.devtiersoftware.com/posts/exercitationem-perspiciatis-voluptate-sint-saepe		✓		
3.	https://blog.devtiersoftware.com/posts/aperiam-consequatur-eligendi-nihil-dignissimos-similique-voluptas-aut		✓		
4.	https://blog.devtiersoftware.com/posts/ut-velit-veritatis-asperiores-veritatis-laboriosam-voluptatum-dolorem		✓		
5.	https://blog.devtiersoftware.com/posts/ipsum-doler		✓		
6.	https://blog.devtiersoftware.com/posts/sad-sed-sad-sed		✓		
7.	https://blog.devtiersoftware.com/posts/consequatur-debitis-eius-a		✓		
8.	https://blog.devtiersoftware.com/posts/blanditiis-nihil-consequatur-rerum-aliquam		✓		
9.	https://blog.devtiersoftware.com/posts/donet-ue		✓		
10.	https://blog.devtiersoftware.com/posts/ipsam-voluptatum-nisi-magni-et	✓			

Tabel 7 merupakan tabel hasil akhir dari pengujian sistem level satu pada blog.devtiersoftware.com/posts. Pengujian sistem level satu menghasilkan 9 *True Negative* dan 1 *True Positive* dari 10 laman yang dipilih untuk proses analisis akurasi pengujian.

Tabel 8. Hasil akhir pengujian level dua

No.	URL	True Positive	True Negative	False Positive	False Negative
1.	https://blog.devtiersoftware.com/posts/nostrum-est-sit-omnis-aliquam		✓		
2.	https://blog.devtiersoftware.com/posts/ipsam-voluptatum-nisi-magni-et	✓			
3.	https://blog.devtiersoftware.com/posts/consequatur-aut-perspiciatis-nobis-minus-expedita-minus-eveniet		✓		
4.	https://blog.devtiersoftware.com/posts/saepe-provident-cumque-quaerat-architecto-aut		✓		

5.	https://blog.devtiersoftware.com/posts/consectetur-et-ex-facere		✓		
6.	https://blog.devtiersoftware.com/posts/et-voluptatem-ratione-ab		✓		
7.	https://blog.devtiersoftware.com/posts/ea-ut-consequatur-aut-voluptas-dolore-soluta-voluptatem		✓		
8.	https://blog.devtiersoftware.com/posts/enim-aut-consectetur-quia		✓		
9.	https://blog.devtiersoftware.com/posts/quis-molestias-necessitatibus-repellat-et	✓			
10.	https://blog.devtiersoftware.com/posts/praesentium-aut-culpa-autem-a-et-quia-nemo		✓		

Tabel 8 merupakan tabel hasil akhir dari pengujian sistem level dua pada blog.devtiersoftware.com/posts. Pengujian sistem level dua menghasilkan 8 *True Negative* dan 2 *True Positive* dari 10 laman yang dipilih untuk proses analisis akurasi pengujian.

Tabel 9. Hasil akhir pengujian level tiga

No.	URL	True Positive	True Negative	False Positive	False Negative
1.	https://help.twitter.com/using-twitter/twitter-supported-browsers		✓		
2.	https://business.twitter.com/en/help/troubleshooting/how-twitter-ads-work.html?ref=web-twc-ao-gbl-adsinfo&utm_source=twc&utm_medium=web&utm_campaign=ao&utm_content=adsinfo		✓		
3.	https://blog.devtiersoftware.com/posts/accusamus-molestiae-reprehenderit-architecto-est-voluptatem-qui-repellat		✓		
4.	https://blog.devtiersoftware.com/posts/iste-perferendis-nostrum-ab-natus-ut-sint-ea	✓			
5.	https://blog.devtiersoftware.com/posts/debitis-suscipit-consequatur-doloribus-natus-nostrum		✓		
6.	https://support.twitter.com/articles/20170514		✓		
7.	https://blog.devtiersoftware.com/posts/natus-deleniti-necessitatibus-a-excepturi-velit		✓		
8.	https://blog.devtiersoftware.com/posts/aut-provident-a-delectus-deserunt-nesciunt-ut-nisi		✓		
9.	https://blog.devtiersoftware.com/posts/exercitationem-perspiciatis-voluptate-sint-saepe	✓			
10.	https://apps.apple.com/us/app/whatsapp-messenger/id310633997		✓		

Tabel 9 merupakan tabel hasil akhir dari pengujian sistem level tiga pada blog.devtiersoftware.com/posts. Pengujian sistem level tiga menghasilkan 8 *True Negative* dan 2 *True Positive* dari 10 laman yang dipilih untuk proses analisis akurasi pengujian.

4.3.2 Hasil pada klungkungkab.go.id

Tabel 10. Hasil akhir pengujian level satu

No.	URL	True Positive	True Negative	False Positive	False Negative
1.	http://jdihn.go.id		✓		
2.	https://jdihn.go.id		✓		

3.	https://jdih.banglikab.go.id		✓		
4.	https://jdih.gianyarkab.go.id		✓		
5.	https://www.youtube.com/watch?v=CFUbfuIGYls		✓		
6.	https://youtu.be/wVdt99Fvg1Q		✓		

Tabel 10 merupakan tabel hasil akhir dari pengujian sistem level satu pada jdih.klungkungkab.go.id. Pengujian sistem level satu menghasilkan 6 *True Negative* dari 6 laman yang dipilih untuk proses analisis akurasi pengujian.

Tabel 11. Hasil akhir pengujian level dua

No.	URL	True Positive	True Negative	False Positive	False Negative
1.	http://jdih.gianyarkab.go.id		✓		
2.	http://jdih.banglikab.go.id		✓		
3.	http://www.bphn.go.id		✓		
4.	https://jdih.gianyarkab.go.id/pengumuman/detail/8		✓		
5.	http://karangasemkab.go.id		✓		
6.	https://gianyarkab.go.id		✓		
7.	https://jdihn.go.id/pencarian/detail/1604432		✓		
8.	https://rechtsvinding.bphn.go.id/ejournal/index.php/jrv		✓		
9.	https://ejournal.balitbangham.go.id/index.php/dejure		✓		
10.	https://jdih.baliprov.go.id/profil/standar-operasional-prosedur		✓		

Tabel 11 merupakan tabel hasil akhir dari pengujian sistem level dua pada jdih.klungkungkab.go.id. Pengujian sistem level dua menghasilkan 10 True Negative dari 10 laman yang dipilih untuk proses analisis akurasi pengujian.

Tabel 12. Hasil akhir pengujian level tiga

No.	URL	True Positive	True Negative	False Positive	False Negative
1.	https://ejournal.balitbangham.go.id/index.php/kebijakan/about/submissions#authorGuidelines		✓		
2.	https://ejournal.balitbangham.go.id/index.php/ham/about		✓		
3.	http://pkp.sfu.ca/ojs		✓		
4.	https://portal.issn.org/resource/ISSN/2614-2414		✓		
5.	https://jdih.kpu.go.id/surat-edaran		✓		
6.	https://ejournal.balitbangham.go.id/index.php/ham/search?subject=bantuan%20hukum		✓		
7.	https://ejournal.balitbangham.go.id/index.php/dejure/about/editorialTeam		✓		
8.	https://jdihn.go.id/pencarian/detail/1604423		✓		
9.	https://ejournal.balitbangham.go.id/index.php/dejure/search?subject=criminal		✓		
10.	https://jdih.gianyarkab.go.id/produk/artikel/detail/35		✓		

Tabel 12 merupakan tabel hasil akhir dari pengujian sistem level tiga pada jdih.klungkungkab.go.id. Pengujian sistem level tiga menghasilkan 10 True Negative dari 10 laman yang dipilih untuk proses analisis akurasi pengujian.

5. Kesimpulan

Penelitian dilakukan dengan memanfaatkan *web scraping* untuk mendapatkan data berupa url-url yang terdapat pada sebuah laman website. Proses pencarian pola serangan pada laman website dilakukan dengan memanfaatkan Algoritma Knuth-Morris-Pratt yang membandingkan pola serangan yang sudah ditentukan dengan *source code* laman website yang menjadi target pengujian. Pengujian sistem dilakukan pada level pencarian 1 sampai 3. Data acak hasil dari masing-masing pengujian digunakan sebagai bahan pengujian secara manual. 56 data yang digunakan 6 diantaranya memiliki nilai *True Positive* yaitu pola serangan ditemukan dan terdapat jejak serangan pada laman tersebut, sedangkan 49 sisanya memiliki nilai *True Negative* yang berarti pola serangan ditemukan namun tidak ditemukan jejak serangan pada laman tersebut.

Daftar Pustaka

[1] A. Fadlil, I. Riadi, and A. Nugrahantoro, "Data Security for School Service Top-Up Transactions Based on AES Combination Blockchain Technology," *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 11, no. 3, p. 155, 2020, doi: 10.24843/lkjiti.2020.v11.i03.p04.

[2] I. O. Laleb, "Analisis Cross-Site Scripting (XSS) Injection-Reflected XSS And Stored XSS Mengggunakan Framework OWASP 10," *J. Ilm. Flash*, vol. 8, pp. 36–42, 2022.

[3] Y. Sahria, "Implementasi Teknik Web Scraping pada Jurnal SINTA Untuk Analisis Topik Penelitian Kesehatan Indonesia," *URECOL (Unversity Res. Colloqium)*, pp. 297–306, 2020, [Online]. Available: http://repository.urecol.org/index.php/proceeding/article/view/1079

[4] K. A. Khairan and H. Ahmadian, "Penerapan Algoritma Knuth-Morris-Pratt Pada Fitur Pencarian Definisi Istilah Standar Operasional Prosedur (Sop) Pada Lembaga Penjaminan Mutu Uin Ar-Raniry," *Cybersp. J. Pendidik. Teknol. Inf.*, vol. 3, no. 1, p. 26, 2019, doi: 10.22373/cj.v3i1.4723.

[5] F. B. Alataz, "Peningkatan Keamanan Website dari Serangan Cross Site Scripting (XSS) Dengan Metode Metacharacter dan Form Validation," 2021, [Online]. Available: https://repository.upnvj.ac.id/11185

[6] OWASP, "XSS Filter Evasion," *OWASP/Cheatsheetseries*, 2021. https://cheatsheetseries.owasp.org/cheatsheets/XSS_Filter_Evasion_Cheat_Sheet.html (accessed Feb. 21, 2023).

[7] P. Thapa, "XSS Payload," 2020. https://github.com/pgaijin66/XSS-Payloads/blob/master/payload/payload.txt

[8] Y. M. Rangkuti, S. I. Al Idrus, and D. D. Tarigan, *Pengantar Pemrograman Python*. Media Sains Indonesia, 2021.

[9] A. Surahman, A. F. Octaviansyah, and D. Darwis, "Ekstraksi Data Produk E-Marketplace Sebagai Strategi Pengolahan Segmentasi Pasar Menggunakan Web Crawler," *Sistemasi*, vol. 9, no. 1, p. 73, 2020, doi: 10.32520/stmsi.v9i1.580.

[10] G. Indrawan, A. Asroni, L. Joni Erawati Dewi, I. G. A. Gunadi, and I. K. Paramarta, "Balinese Script Recognition Using Tesseract Mobile Framework," *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 13, no. 3, p. 160, 2022, doi: 10.24843/lkjiti.2022.v13.i03.p03.

[11] D. S. Islamiyati and A. Fikri, "Penerapan Algoritma Knuth-Morris-Pratt dalam Mendeteksi Tingkat Kemiripan Judul Skripsi Berbasis Web," *J. Inf. Syst. Res.*, vol. 3, no. 2, pp. 58–63, 2022, doi: 10.47065/josh.v3i2.1168.