

Penerapan Fungsi Metode *Rolling Hash* Pada Algoritma *Winnowing* Untuk Mendeteksi Kemiripan Teks Abstrak Berbasis Web

Yovi Apridiansyah¹, Ardi Wijaya² Ari Purjiawan³

^{1,2,3} Dosen Tetap Program Studi Teknik Informatika, Universitas Muhammadiyah Bengkulu
 Jl. Bali Kota Bengkulu, telp (0736) 22765/ fax (0736) 26161;
 e-mail: yoviapridiansyah@umb.ac.id, ardiwijaya@umb.ac.id, aripurjiawan12@gmail.com

(received: November 2021, revised : Februari 2022, accepted : April 2022)

Abstract— Plagiarisme merupakan pengambilan karangan atau pendapat orang lain dan menjadikannya seolah-olah karangan atau pendapat sendiri, misalnya menerbitkan karya tulis orang lain atas nama dirinya sendiri. Dalam hal mengantisipasi pendapat karangan orang lain yaitu dibutuhkan cara dalam menganalisis suatu dokumen, salah satunya yaitu dengan menerapkan fungsi metode *rolling hash* pada algoritma *winnowing* karena metode *winnowing* ini algoritma yang digunakan untuk melakukan proses pengecekan kesamaan kata (*document fingerprinting*) untuk mengidentifikasi plagiarisme atau penjiplakan. Dengan menerapkan fungsi metode *rolling hash* pada algoritma *winnowing* dapat menghasilkan perbandingan persentase plagiat atau kemiripan suatu dokumen atau teks abstrak. Dari Percobaan membandingkan teks abstrak dalam salah satu jurnal mendapatkan nilai persentase 50,06% dengan menggunakan persamaan *Jaccard Coefficient*.

Kata Kunci : *Plagiarisme, Winnowing, Rolling Hash, fingerprinting*

Abstract— Plagiarism is the taking of the essays or opinions of others and making it as if the essay or opinion itself, for example publishing the writings of others in the name of themselves. In terms of anticipating the opinions of others, it takes a way in analyzing a document, one of which is by applying the rolling hash method function to the winnowing algorithm because this winnowing method is an algorithm used to check the similarity of words (*document fingerprinting*) to identify plagiarism or plagiarism. By applying the rolling hash method function to the winnowing algorithm can result in a comparison of plagiarism percentages or similarities of a document or abstract text. From the experiment comparing abstract text in one of the journals got a percentage value of 50.06% using the *Jaccard Coefficient* equation.

Keyword : *Plagiarisme, Winnowing, Rolling Hash, fingerprinting*

I. PENDAHULUAN

Plagiarisme pada dasarnya selalu menjadi perhatian dibanyak sektor, terutama dalam sektor akademis. Kasus-kasus praktik menyalin beberapa bagian atau keseluruhan tulisan tanpa mencantumkan sumber menjadi hal biasa

yang sering ditemukan pada tugas harian, makalah, bahkan skripsi. Pencegahan dan pendeteksian merupakan cara yang dapat dilakukan untuk mengurangi *plagiarisme*. Pencegahan berarti menghalangi munculnya plagiarisme yang lebih ditekankan kepada moral masyarakat dan sistem pendidikan. Cara ini akan memberikan efek jangka panjang yang luar biasa. Dengan adanya sistem pendeteksian berarti cara untuk mengungkapkan *plagiarisme* [1], [2].

Untuk mengatasi tindakan plagiarisme tersebut, Kementerian Riset Teknologi Pendidikan Tinggi mengeluarkan peraturan tentang plagiarisme yang mengharuskan karya ilmiah melakukan pengecekan plagiasi menggunakan software tertentu dan hasilnya harus di bawah 20 persen. Akan tetapi, tidak sedikit akademik yang merasa kesulitan untuk memenuhi persyaratan tersebut, karena belum mengerti bagaimana sistem pengecekan plagiasi itu bekerja. Masing masing perguruan tinggi memiliki batas tersendiri untuk menentukan tingkat persentase plagiasi. Selain itu, semakin tinggi pendidikan yang ditempuh maka persyaratan tentang penggunaan jurnal di dalam sebuah penelitian semakin bertambah, yang kemudian tidak menutup kemungkinan meningkatkan tingkat plagiasi. Untuk mengatasi hal tersebut maka peneliti perlu teliti, cermat dan bertanggung jawab ketika akan mengutip dari sebuah jurnal [3],[4],[5].

Banyak software yang ditawarkan menjadi solusi untuk deteksi plagiarisme seperti Turnitin, WCopyFind yang digunakan oleh University of Virginia, *Crosscheck* yang ditemukan oleh *Elsevier, Springer* yang diciptakan oleh

Massachusetts Institute of Technology pers, dan penerbit ternama lainnya [6],[7]. Hanya saja semua aplikasi yang tersebut itu nilai komersialnya sangat tinggi, oleh karena itu tujuan dari penelitian ini adalah untuk lebih menghemat biaya dengan tidak membeli software-software ternama tersebut.

Solusi yang ditawarkan adalah dengan mencoba menerapkan sebuah metode *rolling hash* pada algoritma *winnowing* yang akan diimplementasikan pada aplikasi sebagai deteksi perbandingan dokumen pada abstrak. Metode *rolling hash* itu sendiri merupakan salah satu metode *hashing* pada algoritma *winnowing* yang memberikan kemampuan untuk menghitung nilai hash tanpa mengulangi seluruh *string* per kata [8]. Oleh karena itu algoritma *winnowing* dapat digunakan untuk melakukan proses pengecekan perbandingan kesamaan kata pada teks abstrak untuk mendeteksi plagiarisme.

II. LANDASAN TEORI

A. Plagiarisme

Plagiarisme adalah mengambil ide-ide atau kata-kata orang lain dan mengakui sebagai milik sendiri. Plagiarisme adalah jenis pencurian intelektual. Plagiarisme dapat mengambil banyak bentuk, dari kecurangan yang disengaja untuk sengaja menyalin dari sumber tanpa pengakuan [9].

B. Metode Mendeteksi Plagiarisme

Dalam melakukan pendeteksian penjiplakan terdapat tiga metode [6] yaitu:

1. Perbandingan Teks Lengkap

Metode ini diterapkan dengan membandingkan semua isi dokumen. Namun pendekatan ini membutuhkan waktu yang lama tetapi cukup efektif.

2. Dokumen *Fingerprint*

Dokumen *fingerprint* merupakan metode yang digunakan untuk mendeteksi keakuratan kesamaan antar dokumen. Prinsip kerja dari metode dokumen *fingerprint* ini dengan menggunakan teknik *hashing*. Teknik *hashing* adalah sebuah fungsi yang menkonversi setiap *string* menjadi bilangan. Algoritma yang digunakan pada metode ini seperti algoritma *Winnowing*, algoritma *Manber* dan algoritma *Rabin-Karp*.

3. Kesamaan Kata Kunci

Prinsip dari metode kesamaan kata kunci adalah mencari kata kunci dari dokumen dan kemudian dibandingkan dengan kata kunci pada dokumen lain.

C. Algoritma *Winnowing* (*Biword*)

Winnowing adalah algoritma yang digunakan untuk melakukan proses pengecekan kesamaan kata (dokumen *fingerprinting*) untuk mendeteksi plagiarisme [6]. Secara teknis *winnowing* adalah ekstensi dari implementasi algoritma *rabin-karp fingerprint* dengan penambahan metode *window*.

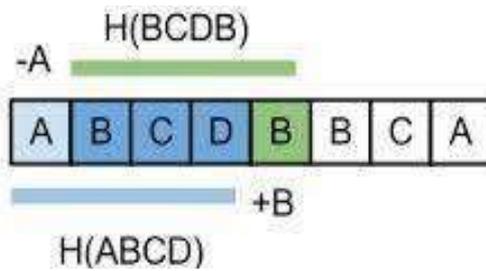
D. *N-Gram*

N-grams adalah rangkaian token dengan panjang *n*. Dalam *konteks komputasi linguistik*, token ini dapat berupa katakata, meskipun mereka dapat berupa karakter atau himpunan bagian dari karakter. Nilai *n* hanya mengacu pada jumlah token. Metode *n-grams* ini digunakan untuk mengambil potongan-potongan karakter huruf sejumlah *n* dari sebuah kata yang secara kontinuitas dibaca dari teks sumber hingga akhir dari dokumen.

Beberapa jenis *N-Gram*, yaitu unigram yaitu token data teks twitter yang hanya terdiri dari satu kata, kemudian bigram yaitu token data teks twitter yang terdiri dari dua kata dan trigram yaitu token data teks twitter yang terdiri dari tiga kata. Penerapan *N-gram* dapat dilihat seperti berikut :Berikut ini adalah contoh *n grams* dengan $n=5$: “Teks: A do run run run, a do run run” dari teks tersebut dilakukan penghilangan spasi menjadi. “Teks: adorunrunrunadorunrun” dari teks tersebut dihasilkan rangkaian 5-grams yang diturunkan menjadi “Teks: adoru dorun orunr runru unrun nrunr runru unrun nruna runad unado nador adoru dorun orunr runru unrun”. Tujuan pemakaian *N-gram* dilakukan pada penelitian ini karena dalam bahasa Indonesia banyak frase yang tidak hanya terdiri dari satu kata [10].

E. Rolling Hash

Rolling hash merupakan salah satu metode *hashing* yang memberikan kemampuan untuk menghitung nilai hash tanpa mengulangi seluruh *string*. Berikut ini contoh *string* “abcdbbca” dan anda ingin menemukan pola “bcd b” dalam *string* ini.

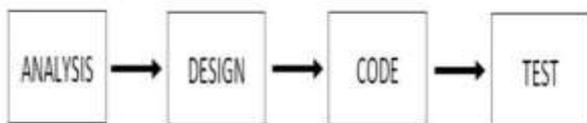


Gambar 1. Window Rolling Hash

Berikut penjelasan metode dari algoritma *winning* : *Preprocessing*, proses awal dari metode dimana input file akan melalui berbagai macam proses manipulasi text. Macam – macam metode *preprocessing* meliputi *case folding, filtering, stemming, dan tokenizing*. *Case folding* adalah proses manipulasi case-sensitive. Sedangkan *filtering* atau sering dikenal dengan istilah *stopword removal* adalah proses penghapusan kata yang tidak relevan dalam teks. *Stemming* ialah proses pemisahan kata menjadi kata dasar. *Tokenizing* adalah proses pemisahan kata berdasarkan susunan kata [11]. Hasil dari proses ini akan menghasilkan dokumen teks yang relevan untuk diproses dan dicari kecocokannya. Pada preprocess file teks akan dibentuk menjadi rangkaian substring senilai *k* atau *k-gram*.

III. METODE PENELITIAN

Dalam penelitian ini model pengembangan sistem yang digunakan yaitu model incremental. incremental model adalah model pengembangan sistem pada rekayasa perangkat lunak berdasarkan perangkat lunak persyaratan yang dipecah menjadi beberapa fungsi atau bagian sehingga model pengembangannya secara bertahap [12],[13],[14].

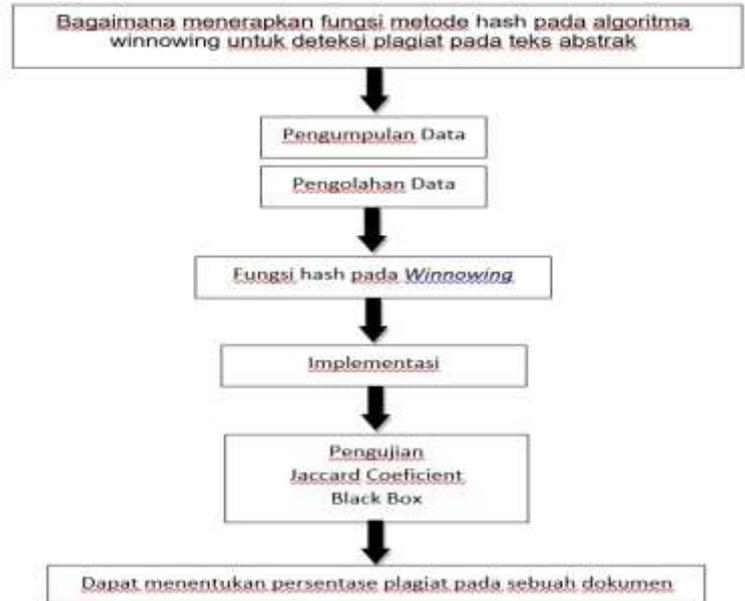


Gambar 2. Model Incremental

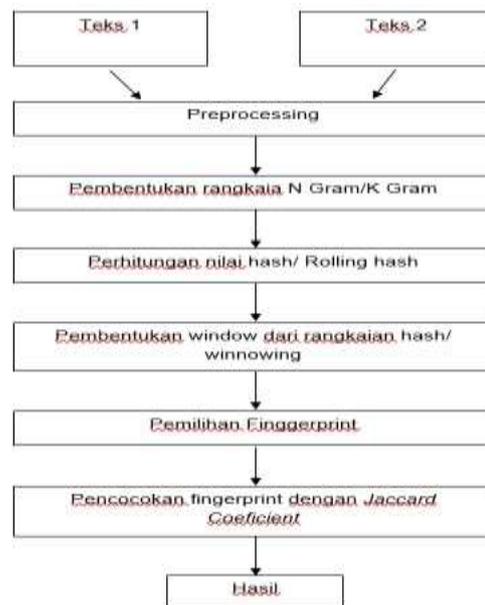
1. Analisa, Analisis data diartikan sebagai upaya mengolah data menjadi informasi, sehingga karakteristik atau sifat-sifat data tersebut dapat dengan mudah dipahami dan bermanfaat untuk menjawab

masalah-masalah yang berkaitan dengan kegiatan penelitian.

2. *Design*, Dalam tahapan *design* ini menghasilkan rancangan yang memenuhi kebutuhan yang ditentukan selama tahapan analisis. Hasil akhirnya berupa spesifikasi rancangan yang sangat rinci sehingga mudah diwujudkan pada saat pemrograman.



Gambar 3. Kerangka Penelitian



Gambar 4. Blok dDiagram Proses Winnowing

3. Code, Pengkodean yang mengimplementasikan hasil desain kedalam kode atau bahasa yang dimengerti oleh mesin komputer dengan menggunakan bahasa matlab. Pada tahap pengkodean system penulisan kode

program merupakan tahap penerjemahan desain sistem yang telah dibuat ke dalam bentuk perintah-perintah yang dimengerti komputer dengan mempergunakan bahasa pemrograman.

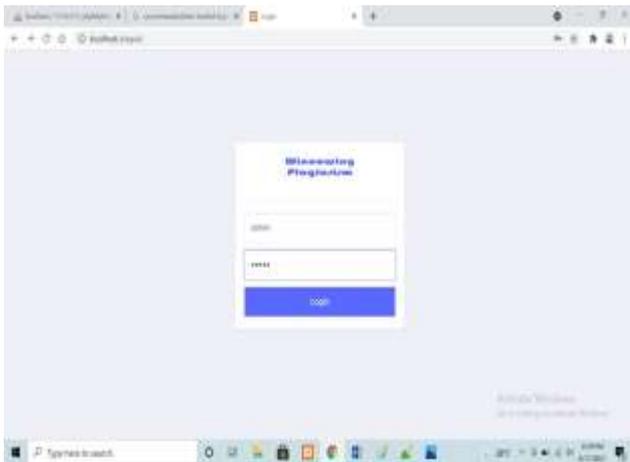
4. Test, Metode pengujian yang dipakai adalah dengan menggunakan pengujian aplikasi *blackbox testing* atau test fungsional yaitu pengujian program dengan memberikan input tertentu dan melihat hasil yang didapatkan dari input tersebut. dengan kata lain, *black box testing* terfokus pada *fungsionalitas* sistem [15], [16] dan pengujian *Jaccard Coeficient*.

Persamaan *Jaccard Coeficient* [15].

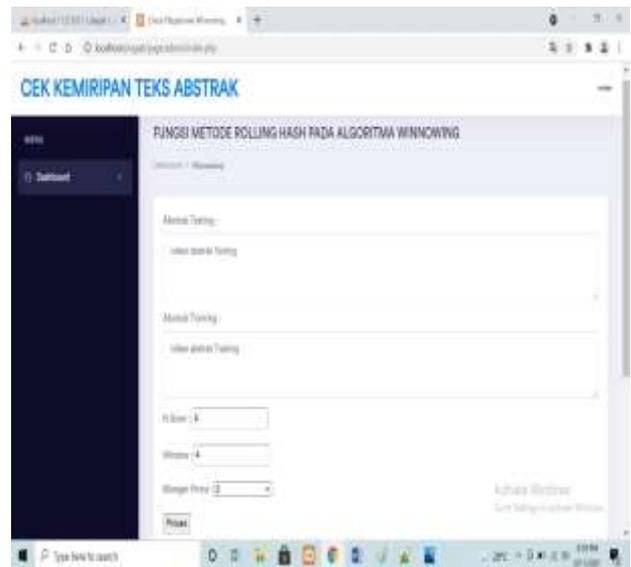
$$Similarity = \frac{Jumlah\ Fingerprint\ sama}{Total\ Fingerprint} \times 100\% \quad (1)$$

IV. HASIL DAN PEMBAHASAN

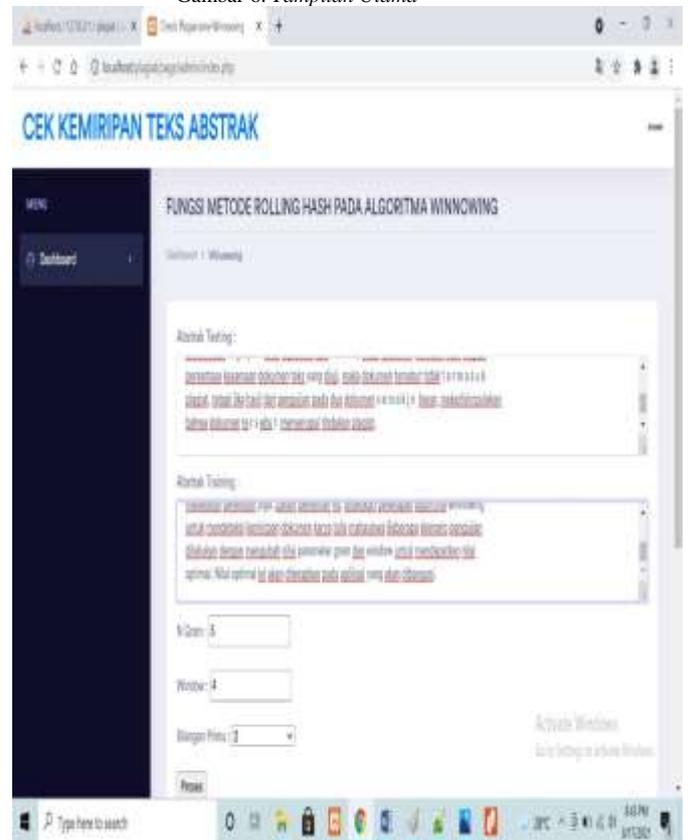
Bagian hasil dan pembahasan ini merupakan pembahasan tentang rancangan antar muka aplikasi yang dibuat.



Gambar 5. Login



Gambar 6. Tampilan Utama



Gambar 7. Tampilan Proses Perbandingan Kemiripan Abstrak



Gambar 8. Tampilan Hasil Proses

Dari hasil percobaan diatas persentase plagiarisme yang dihasilkan adalah 50,06% menggunakan persamaan *Jaccard Coefficient*.

$$\begin{aligned}
 \text{koefisien jaccard} &= \frac{\text{Intersection}}{\text{Union Intersection}} \times 100\% \\
 &= \frac{443}{885} \times 100\% \\
 &= 50,06\%
 \end{aligned}$$

Tabel 1. Karakteristik Basis Data

| Pengujian | Tes | Hasil | Kesimpulan |
|-----------|-------------|---------|------------|
| Login | Input | Bisa | Valid |
| | Login | Diakses | |
| Proses | Proses | Bisa | Valid |
| | Cek Plagiat | Diakses | |
| Logout | Proses | Bisa | Valid |
| | Logout | Diakses | |

V. PENUTUP

A. Kesimpulan

1. Dengan menerapkan fungsi metode *rolling hash* pada algoritma *winning* dapat menghasilkan

perbandingan persentase plagiat suatu dokumen atau abstrak.

2. Dari Percobaan membandingkan teks abstrak dalam salah satu jurnal mendapatkan nilai persentase 50,06% dengan menggunakan persamaan *Jaccard Coefficient*.

B. Saran

Untuk kedepannya bias dijadikan literatur dalam penelitian selanjutnya dengan metode *rolling hash* pada algoritma *winning* dapat menghasilkan perbandingan persentase plagiat suatu dokumen yang lebih baik

DAFTAR PUSTAKA

- [1] A. Menggunakan, / Ratcliff, Y. Obershelp, J. Lady, A. Sinsuw, and A. Jacobus, "Rancang Bangun Aplikasi Deteksi Kemiripan Dokumen Teks," *J. Tek. Inform.*, vol. 11, no. 1, 2017.
- [2] D. W. Wicaksono and M. I. Irawan, "Sistem Deteksi Kemiripan Antar Dokumen Teks Menggunakan Model Bayesians Pada Term Latent Semantic Analysis (LSA)," *J. Sains dan Seni POMITS*, vol. 3, no. 2, pp. 1–6, 2014.
- [3] S. Taerungruang and W. Aroonmanakun, "Constructing an academic thai plagiarism corpus for benchmarking plagiarism detection systems," *GEMA Online J. Lang. Stud.*, vol. 18, no. 3, pp. 186–202, 2018, doi: 10.17576/gema-2018-1803-11.
- [4] J. Priambodo, "Pendeteksian Plagiarisme Menggunakan Algoritma Rabin-Karp dengan Metode Rolling Hash," *J. Inform. Univ. Pamulang*, vol. 3, no. 1, p. 39, 2018, doi: 10.32493/informatika.v3i1.1518.
- [5] W. E. Y. R. Jarwati, Antonius Cahya Prihandoko, "Penerapan Algoritma Wining pada Sistem Rekomendasi Penentuan Dosen Pembimbing Skripsi (Studi Kasus: Prodi Sistem Informasi Universitas Jember Jember)," no. 1, pp. 11–20, 2016.
- [6] S. Schleimer, D. S. Wilkerson, and A. Aiken, "Wining," no. April 2003, p. 76, 2003, doi: 10.1145/872757.872770.
- [7] R. K. Wibowo and K. Hastuti, "Penerapan Algoritma Wining Untuk Mendeteksi Kemiripan Teks pada Tugas Akhir Mahasiswa," *Techno.com*, vol. 15, no. 4, pp. 303–311, 2016, [Online]. Available: <https://publikasi.dinus.ac.id/index.php/technoc/article/view/1271%0Ahttps://core.ac.uk/download/pdf/289792296.pdf>.
- [8] A. Kurniawati and I. W. S. Wicaksana, "Perbandingan Pendekatan Deteksi Plagiarism Dokumen Dalam Bahasa Inggris," *Semin. Ilm. Nas. Komput. dan Sist. Intelijen (KOMMIT 2008)*, no. Kommit, pp. 284–291, 2008.
- [9] Anisah Hasan and Irwan, "J urnal E quilibrium J urnal E quilibrium," *Fenom. Plagiarisme Mhs.*, vol. IV, no. 1, pp. 1–10, 2016, [Online]. Available: <https://media.neliti.com/media/publications/70308-ID-fenomena-plagiarisme-mahasiswa.pdf>.
- [10] A. Nurfalah and A. Ardiyanti Suryani, "Analisis Sentimen Berbahasa Indonesia Dengan Pendekatan Lexicon-Based Pada Media Sosial," *J. Masy. Inform. Indones.*, vol. 2, no. 1, pp. 1–8, 2017.
- [11] S. Sunardi, A. Yudhana, and I. A. Mukaromah, "Implementasi Deteksi Plagiarisme Menggunakan Metode N-Gram Dan Jaccard Similarity Terhadap Algoritma Wining," *Transmisi*, vol. 20, no. 3, p. 105, 2018, doi: 10.14710/transmisi.20.3.105-110.
- [12] M. Utami and Y. Apridiansyah, "Implementasi Algoritma Sequential Searching Pada Sistem Pelayanan Puskesmas Menggunakan Bootstrap (Studi Kasus Puskesmas Kampung Bali Bengkulu)," *JSAI (Journal Sci. Appl. Informatics)*, vol. 2, no. 1, pp. 81–86, 2019, doi: 10.36085/jasai.v2i1.166.
- [13] M. Syarif and W. Nugraha, "Metode Incremental Dalam Membangun Aplikasi Identifikasi Gaya Belajar Untuk Meningkatkan Hasil Belajar Siswa," *Jusikom J. Sist. Komput. Musirawas*, vol. 4, no. 1, pp. 42–49, 2019, doi: 10.32767/jusikom.v4i1.441.

- [14] M. F. Fanani, Sholiq, and F. A. Muqtadiroh, "Implementasi Metode Incremental Dalam Membangun Aplikasi Use Case Point Pada Perusahaan DTS," *Seminar Nasional Sistem Informasi Indonesia*. pp. 169–174, 2015.
- [15] H. B. I. Alfaris, C. Anam, and A. Masy'an, "Implementasi Black Box Testing Pada Sistem Informasi Pendaftaran Santri Berbasis Web Dengan Menggunakan PHP Dan MYSQL," *J. Sains dan Teknol.*, vol. 6, no. 1, pp. 23–38, 2013, [Online]. Available: <http://ejournal.unwaha.ac.id/index.php/saintek/article/download/64/64>.
- [16] A. S. Fadhila Cahya Ningrum, Dandi Suherman, Sita Aryati, Handika Angga Prasetya, "Pengujian Black Box pada Aplikasi Perpustakaan Menggunakan Teknik Equivalence Partitions," *J. Inform. Univ. Pamulang*, vol. 4, no. 4, 2020, doi: 10.32493/jtsi.v3i3.5343.