

Penerapan Algoritma Winnowing dalam Mendeteksi Kesamaan Judul Kerja Praktek Program Studi Ilmu Komputer Universitas Katolik Widya Mandira Kupang

Yunita Mildayani Putri Bria¹, Yulianti Paula Bria^{*2}, Yovinia Carmeneja Hoar Siki³

¹⁻³ Ilmu Komputer, Universitas Katolik Widya Mandira Kupang

E-mail: yunitampbria@gmail.com¹, yulianti.bria@unwira.ac.id^{*2}, yoviniacarmeneja@unwira.ac.id³

Abstrak. Pengajuan judul kerja praktek pada Program Studi Ilmu Komputer masih dilakukan secara manual sehingga dapat menyebabkan duplikasi judul. Untuk mencegah adanya unsur duplikasi judul kerja praktek maka perlu dikembangkan sebuah sistem deteksi kesamaan judul menggunakan algoritma Winnowing. Algoritma Winnowing merupakan algoritma berbasis *hashing-approach* yang menerapkan *hash-function* dan pembentukan *window* untuk memperoleh *fingerprints* pada saat mendeteksi judul. Berdasarkan *fingerprints* tersebut, tingkat kesamaan judul dapat dihitung. Penelitian ini bertujuan untuk mengukur tingkat kesamaan judul kerja praktek pada Program Studi Ilmu Komputer Universitas Katolik Widya Mandira. Hasil dari penelitian ini menunjukkan bahwa algoritma yang diterapkan dapat mendeteksi kesamaan judul kerja praktek dengan baik.

Kata kunci: algoritma *Winnowing*; deteksi kesamaan; judul kerja praktek; ilmu komputer.

Abstract. The submission of internship project titles in the Computer Science Study Program is still done manually, which can lead to title duplication. To prevent title duplication, a system needs to be developed using the Winnowing algorithm. The Winnowing algorithm is a hashing-based approach that applies a hash function and forms a window to obtain fingerprints when detecting titles. Based on these fingerprints, the degree of title similarity can be calculated. This study aims to measure the similarity of internship project titles in the Computer Science Department at Widya Mandira Catholic University. The results of this study show that the implemented algorithm can effectively detect title similarities.

Keywords: Winnowing algorithm; similarity detection; internship project title; computer science.

1. Pendahuluan

Perkembangan teknologi informasi memberikan dampak signifikan di berbagai bidang termasuk bidang pendidikan. Kecepatan, ketepatan, dan akurasi dalam penyediaan serta aksesibilitas informasi kini menjadi prasyarat utama untuk memenuhi kebutuhan informasi yang dinamis dan berguna. Di era digital yang semakin mengglobal, kemampuan untuk mengakses informasi secara cepat dan akurat tidak hanya menjadi kebutuhan, tetapi juga menjadi kunci penting dalam menunjang efektivitas dan efisiensi proses pembelajaran serta penelitian di lingkungan akademik. Informasi yang diperlukan harus dapat diakses

dengan cara yang efektif dan efisien oleh semua pihak yang membutuhkannya. Lembaga-lembaga akademik sepenuhnya menggunakan sistem informasi terkomputerisasi untuk meningkatkan kualitas pengelolaan data. Dengan penerapan sistem informasi yang terkomputerisasi, data dapat tersimpan secara terorganisir dengan baik, dan informasi dapat diakses secara cepat, tepat, dan akurat dibandingkan dengan sistem yang masih manual dan belum terkomputerisasi [1].

Seiring dengan kemajuan teknologi, kini sebagian besar pekerjaan dilakukan secara daring. Selain itu, tugas mahasiswa juga diserahkan dalam format elektronik karena lebih memudahkan dosen dan mahasiswa [2]. Namun, penggunaan format elektronik ini meningkatkan risiko terjadinya plagiarisme. Kemudahan akses internet bagi banyak pengguna memungkinkan mereka memperoleh berbagai hasil penelitian, makalah, artikel, dan buku, yang berpotensi mendorong terjadinya praktik plagiarisme. Dampak negatifnya juga dapat mempengaruhi proses belajar mahasiswa. Oleh karena itu, diperlukan metode yang efektif untuk mendeteksi plagiarisme pada data.

Plagiarisme dalam penulisan akademis menjadi permasalahan yang serius, karena tidak hanya merugikan pencipta karya asli, tetapi juga menurunkan kualitas akademik secara keseluruhan [3]. Untuk mengatasi hal ini, Kementerian Riset, Teknologi, dan Pendidikan Tinggi telah menetapkan kebijakan yang mewajibkan setiap karya ilmiah menjalani pemeriksaan plagiasi menggunakan perangkat lunak tertentu, dengan hasil yang harus berada di bawah ambang batas 20% [4]. Namun, banyak akademisi mengalami kesulitan dalam memenuhi persyaratan ini karena kurang memahami cara kerja sistem pengecekan plagiasi [5]. Setiap perguruan tinggi juga memiliki kebijakan yang berbeda terkait batas toleransi plagiasi yang diperbolehkan. Semakin tinggi jenjang pendidikan, semakin ketat pula persyaratan yang ditetapkan, termasuk penggunaan referensi dari jurnal ilmiah yang dapat meningkatkan tingkat kemiripan teks. Oleh karena itu, mahasiswa dan peneliti harus lebih teliti dan cermat dalam mengutip sumber agar dapat meminimalkan risiko plagiarisme [6]. Sebagai salah satu lembaga akademik, Universitas Katolik Widya Mandira (UNWIRA) Kupang telah mengambil langkah-langkah signifikan dalam mencegah dan mendeteksi plagiarisme di lingkungan akademik. UNWIRA mewajibkan setiap mahasiswa yang akan mengikuti seminar proposal atau ujian skripsi untuk menyerahkan salinan digital skripsi atau tesis mereka guna dilakukan pengecekan plagiarisme. Jika tingkat kesamaan dalam dokumen tersebut melebihi ambang batas yang ditetapkan, mahasiswa diwajibkan untuk melakukan revisi sebelum melanjutkan ke tahap berikutnya. Langkah ini bertujuan untuk memastikan integritas akademik serta meningkatkan kesadaran mahasiswa akan pentingnya orisinalitas dalam karya ilmiah mereka.

Berbagai perangkat lunak tersedia sebagai solusi untuk mendeteksi plagiarisme, seperti Turnitin, WCopyFind [7]. Namun, perangkat lunak tersebut umumnya memiliki biaya komersial yang cukup tinggi. Oleh karena itu, penelitian ini bertujuan untuk menyediakan alternatif yang lebih hemat biaya tanpa perlu membeli perangkat lunak komersial tersebut [8]. Solusi yang ditawarkan adalah dengan mencoba menerapkan sebuah metode *rolling hash* pada algoritma *Winnowing* yang akan diimplementasikan pada aplikasi deteksi kesamaan judul kerja praktek (KP). KP merupakan mata kuliah wajib yang berbasis praktik lapangan dengan bobot 3 SKS dilakukan secara mandiri dan harus ditempuh oleh setiap mahasiswa [9]. Tujuan utama KP ialah untuk meningkatkan pengetahuan yang dipelajari mahasiswa terhadap masalah yang terjadi di dunia nyata sehingga mahasiswa dapat menemukan solusi yang tepat untuk menghadapinya. 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. Algoritma *Winnowing* adalah salah satu algoritma yang dapat melakukan metode dokumen *fingerprinting* untuk mendeteksi kesamaan. Pendeteksian kesamaan *source code* pemrograman [10]. Algoritma *Winnowing* dapat digunakan untuk menghitung persentase kesamaan teks pada judul penelitian yang diajukan dengan judul penelitian yang terdapat pada basis data Google Scholar. Sebagai mesin pencari yang menjadi rujukan publikasi ilmiah, Google Scholar menyediakan data karya ilmiah yang valid dan dapat digunakan sebagai acuan dalam mendeteksi kesamaan judul proposal tugas akhir mahasiswa [11]. Jika dibandingkan dengan algoritma *fingerprint* lainnya, algoritma *Winnowing* mempunyai keunggulan dalam membandingkan suatu teks atau dokumen yaitu adanya penambahan konsep *window* yang dari *window* tersebut nilai *hash* yang dipilih menjadi nilai *fingerprint* minimum [12]. Pada tahun 2019, algoritma *Winnowing* digunakan untuk mendeteksi

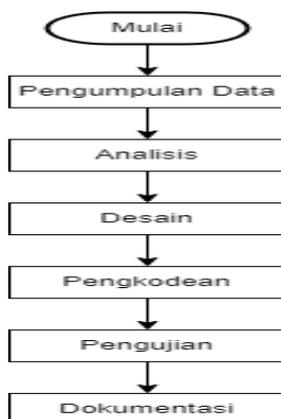
kemiripan teks skripsi dengan cara membandingkan kedua teks melalui teknik *fingerprinting*. Proses ini melibatkan pengubahan n-gram huruf menjadi angka yang disebut *hash* [13]. Algoritma *Winnowing* menggunakan Teknik pra-pemrosesan yang digunakan adalah *case folding, tokenizing, filtering, dan stemming* dengan *stemmer* Nazief-Andriani. *Hash function* menggunakan *rolling hash*. Pembentukan n-gram menggunakan *character-level trigrams*. Pengukuran similaritas menggunakan Jaccard Coefficient. Hasil penelitian menunjukkan bahwa algoritma *Winnowing* dengan *character-level trigrams* dapat mendeteksi kesamaan kata dengan hasil sebesar 23,53% [14].

Pemilihan algoritma yang tepat dalam pencarian kesamaan dokumen teks memiliki pengaruh yang signifikan terhadap akurasi perhitungan. Kesalahan dalam pemilihan algoritma dapat menyebabkan penurunan keakuratan dalam mendeteksi tingkat kemiripan dokumen [15]. Oleh karena itu, metode yang digunakan harus mampu mengidentifikasi kesamaan dalam dokumen berjumlah besar secara lebih efisien dibandingkan metode tradisional. Pengujian algoritma *Winnowing* dalam sistem informasi pengajuan judul skripsi menunjukkan bahwa algoritma ini efektif dalam menghitung persentase kemiripan dengan membandingkan *fingerprint* dari kalimat yang diajukan. Nilai n-gram yang lebih kecil cenderung menghasilkan tingkat kemiripan yang lebih tinggi, sedangkan n-gram yang lebih besar menurunkan nilai kemiripan. Dengan demikian, algoritma ini dapat memberikan rekomendasi penerimaan judul berdasarkan batas kemiripan yang telah ditetapkan dalam sistem [16].

Penelitian terkait pengecekan kesamaan pada judul dan abstrak skripsi telah berhasil dilakukan [17] dimana metode yang diterapkan dalam deteksi kemiripan judul memberikan hasil yang cukup akurat. Penelitian tersebut menggunakan algoritma *Winnowing* untuk mendeteksi kata atau kalimat dalam teks, serta menerapkan metode Jaccard Similarity untuk mengukur persentase kesamaan. Hasil penelitian menunjukkan bahwa algoritma *Winnowing* memiliki tingkat akurasi lebih tinggi dibandingkan algoritma Rabin-Karp dan Knuth-Morris-Pratt dalam mendeteksi plagiarisme dalam dokumen teks [18]. Pemilihan Algoritma *Winnowing* dibandingkan Algoritma Rabin-Karp dalam penelitian ini didasarkan pada beberapa faktor utama, yaitu akurasi, efisiensi, dan stabilitas dalam mendeteksi kesamaan teks, khususnya pada judul kerja praktek atau skripsi. Algoritma *Winnowing* memiliki tingkat akurasi yang lebih tinggi karena menggunakan teknik *fingerprinting* dengan konsep *windowing*, yang memungkinkan algoritma menangkap pola kesamaan lebih baik dibandingkan metode pencocokan langsung [19]. Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengembangkan sistem deteksi kesamaan judul kerja praktek menggunakan algoritma *Winnowing* yang diharapkan dapat membantu perguruan tinggi dalam memastikan orisinalitas judul tugas akhir mahasiswa serta mendukung integritas akademik secara lebih optimal.

2. Metode Penelitian

Gambar 1 menunjukkan tahapan sistematis untuk mengembangkan aplikasi deteksi judul KP menggunakan algoritma *Winnowing*. Tahap penelitian diawali dengan pengumpulan data judul KP, diikuti dengan tahap analisis kebutuhan sistem, desain, pengkodean, pengujian dan dokumentasi.



Gambar 1. Tahapan Penelitian

2.1 Analisis Kebutuhan Sistem

2.1.1 Analisis Peran Sistem

Sistem ini memiliki beberapa peran penting dalam proses pengelolaan kerja praktek:

- Sistem yang dibangun dapat menerima dan menyimpan judul kerja praktek yang diajukan mahasiswa.
- Sistem yang dibangun dapat mendeteksi kesamaan judul kerja praktek yang diajukan oleh mahasiswa menggunakan algoritma.

2.1.2 Analisis Peran Pengguna

Pembuatan sistem ini mempunyai tiga kategori pengguna yaitu:

a. Admin

Admin bertanggung jawab untuk mengelola data mahasiswa dan dosen, memperbarui data judul kerja praktek secara berkala termasuk menambahkan, menghapus dan menyimpan data yang sudah tidak relevan.

b. Mahasiswa

Mahasiswa berperan sebagai pengguna sistem untuk mencari judul kerja praktek yang sudah diajukan oleh mahasiswa lain. Mahasiswa dapat mengecek tingkat kesamaan judul kerja praktek melalui sistem dan membantu mahasiswa dalam bimbingan laporan KP.

c. Dosen

Dosen berperan sebagai pengguna sistem untuk mengecek judul-judul kerja praktek yang diajukan oleh mahasiswa, memastikan bahwa judul yang diajukan tidak mirip dengan judul yang sudah pernah diajukan sebelumnya dan melihat hasil akhir laporan KP mahasiswa bimbingan secara online.

2.2 Desain

Perancangan *interface* adalah tahap penting dalam pengembangan sistem yang fokus pada bagaimana pengguna akan berinteraksi dengan sistem secara visual. Dalam konteks sistem pengelolaan dan deteksi kesamaan judul kerja praktek, perancangan *interface* bertujuan untuk menciptakan antarmuka yang intuitif, mudah digunakan dan efisien dalam menyajikan informasi yang dibutuhkan oleh pengguna seperti mahasiswa, dosen, dan *admin* dapat dilihat pada gambar 2.

The image shows a web interface for detecting similarity in titles and abstracts. It features a sidebar menu with options like Home, Edit User Data, Change Password, View My Proposal, and Logout. The main area contains input fields for 'Judul' and 'Abstrak', a 'Periksa' button, a file upload section for PDFs, and a results section for similarity percentages. A 'Ajukan Proposal KP' button is at the bottom.

Gambar 2. Desain deteksi kesamaan judul

2.3 Pengkodean/Implementasi

Pengembangan sistem diimplementasikan menggunakan bahasa pemrograman PHP dan database MySQL.

2.4 Pengujian -User Acceptance Test

User Acceptance Test (UAT) adalah pengujian aplikasi yang telah dibangun sebelum aplikasi dirilis dan digunakan langsung oleh pengguna. UAT dapat menyakinkan pengguna aplikasi yang telah dibangun apakah aplikasi dapat diterima pengguna dengan baik atau tidak. Dalam pengujian ini, peneliti memberikan beberapa pernyataan kepada beberapa responden dimana responden ini adalah dosen, mahasiswa, dan tata usaha. Pertanyaan yang akan disajikan berkaitan dengan kelayakan dari sistem yang telah dikembangkan oleh peneliti.

3. Hasil dan Pembahasan

3.1 Hasil Implementasi

Pada penelitian ini, tahap pra-pemrosesan data meliputi *case folding* untuk mengubah seluruh judul menjadi huruf kecil (*lowercase*) dan *filtering* dengan menghilangkan tanda simbol [+ -] yang terdapat dalam teks. Tahap *filtering* pada kata-kata tidak dilakukan dengan tujuan untuk mempermudah proses pendeteksian kesamaan berdasarkan makna kalimatnya. Dengan demikian tanda spasi dibiarkan tetap ada (*case sensitivity ignored*). Proses *stemming* atau pengembalian kata ke akar katanya juga tidak dilakukan sehingga kedua teks dibiarkan sesuai makna aslinya.

3.1.1 Menu Deteksi Kesamaan Judul

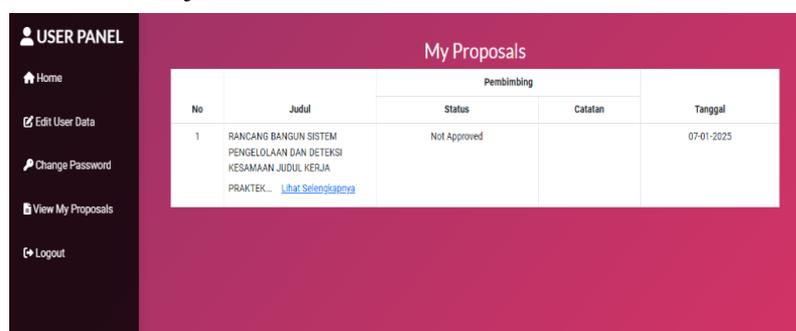
Gambar 3 menunjukkan tampilan *user interface* untuk mendeteksi kesamaan judul. Pada menu ini ditampilkan judul yang telah diajukan. Kita juga dapat mengakses hasil pengecekan kesamaan judul serta melakukan revisi atau pengajuan ulang judul KP.



Gambar 3. Deteksi kesamaan judul

3.1.2 Menu View My Proposal

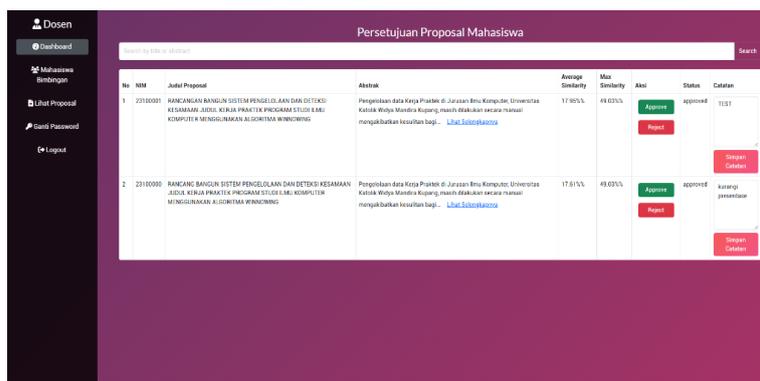
Gambar 4 menunjukkan menu View My Proposal dimana mahasiswa dapat melihat presentasi kesamaan dan mengakses catatan dosen terkait judul KP.



Gambar 4. Menu View My Proposal

3.1.3 Menu Dosen Persetujuan Proposal

Gambar 5 menunjukkan tampilan persetujuan proposal. Pada halaman ini dapat dilihat persetujuan atau penolakan dari dosen pembimbing serta catatan revisi dari dosen.



Gambar 5. Persetujuan Proposal

3.2 Hasil Pengujian

3.2.1 Hasil Pengujian Black Box

Setelah melalui tahap pra-pemrosesan, judul asli maupun judul perbandingan kemudian diubah ke dalam potongan-potongan gram. Setelah diubah ke dalam bentuk grams, fungsi *rolling hash* diterapkan kepada setiap gram untuk mendapatkan nilai *hash* (*fingerprints*) dari kedua judul yang akan dibandingkan. Sebelum mengidentifikasi nilai-nilai *hash* yang sama di antara kedua judul, seluruh *fingerprints* pada masing-masing judul dikelompokkan ke dalam *window* untuk mendapatkan nilai *hash* minimum. Pada penelitian ini lebar *window* yang digunakan adalah $w=4$.

1. Perbandingan Judul

Berikut adalah 2 judul yang dibandingkan:

- sistem pakar pendeteksian kerusakan pada laptop menggunakan metode naïve bayes.
- sistem pakar pendeteksi penyakit tanaman dan hama buah naga menggunakan naïve bayes.

2. Test Processing

Dari dua judul yang telah dipilih, maka akan dilakukan *text preprocessing* menggunakan *case folding*, *remove punctuation* dan *remove space* sehingga terbentuk seperti Tabel 1.

Tabel 1. Preprocessing

	Text Preprocessing judul 1	Text Preprocessing judul 2
Judul	Sistempakarpendeteksikerusakanpada laptopmenggunakanmetodenaivebayes	Sistempakarpendeteksipenyakittanaman danhamabuahnagamenggunakannaivebayes

3. Pembentukan N-Gram

N-gram adalah sebuah metode yang dilakukan dengan mengambil suatu rangkaian substring dari string yang berjumlah n. pada penelitian ini, peneliti menggunakan n-gram=4 dan 5. Dari dua judul tersebut terbentuk n-gram dengan nilai 4 seperti Tabel 2.

Tabel 2. Pembentukan N-Gram

N-Gram Judul 1	N-Gram Judul 2
siste, istem, stem , tem p, em pa, m pak, paka, pakar, akar , kar p, ar pe, r pen, pend, pende, endet, ndete, detek, eteks, teks, eksi , ksi k, si ke, i ker, keru, kerus, erusa, rusak, usaka, sakan, akan , kan p, an pa, n pad, pada, pada , ada l, da la, a lap, lapt, lapto, aptop, ptop , top m, op me,	siste, istem, stem , tem p, em pa, m pak, paka, pakar, akar , kar p, ar pe, r pen, pend, pende, endet, ndete, detek, eteks, teks, eksi , ksi p, si pe, i pen, peny, penya, enyak, nyaki, yakit, akit , kit t, it ta, t tan, tana, tanam, anama, naman, aman , man d, an da, n dan, dan , dan h, an ha, n ham, hama, hama ,

N-Gram Judul 1	N-Gram Judul 2
p men, meng, mengg, enggu, nggun, gguna, gunak, unaka, nakan, akan, kan m, an me, n met, meto, metod, etode, tode, ode n, de na, e nai, naiv, naive, aiveb, iveba, vebay, ebaye, bayes, ayes	ama b, ma bu, a bua, buah, buah, uah n, ah na, h nag, naga, naga, aga m, ga me, a men, meng, mengg, enggu, nggun, gguna, gunak, unaka, nakan, akan, kan n, an na, n nav, navi, navie, avie, vie b, ie ba, e bay, baye, bayes, ayes

N-gram juga dilihat pada Gambar 6.

Step 1: Preprocessing

sistem pakar pendeteksi kerusakan pada laptop menggunakan naive bayes

Step 2: N-Gram Generation

siste, istem, stem, tem p, em pa, m pak, paka, pakar, akar, kar p, ar pe, r pen, pend, pende, endet, ndete, detek, eteks, teks, eksi, ksi p, si pe, i pen, peny, penya, enyak, nyaki, yakit, akit, kit t, it ta, t tan, tana, tanam, anama, naman, aman, man d, an da, n dan, dan, dan h, an ha, n ham, hama, hama, ama b, ma bu, a bua, buah, buah, uah n, ah na, h nag, naga, naga, aga m, ga me, a men, meng, mengg, enggu, nggun, gguna, gunak, unaka, nakan, akan, kan n, an na, n nai, naiv, naive, aive, ive b, ve ba, e bay, baye, bayes, ayes, yes m, es me, s men, meng, mengg, enggu, nggun, gguna, gunak, unaka, nakan, akan, kan n, an na, n nai, naiv, naive, aive, ive b, ve ba, e bay, baye, bayes, ayes

Gambar 6. Pembentukan n-gram

4. Perhitungan Fungsi Hash

Hashing merupakan metode untuk mengkonversi sebuah string menjadi suatu nilai unik dengan penentuan panjang tertentu yang berfungsi sebagai penanda string tersebut. Rumus fungsi hash dapat dilihat pada Persamaan (1).

$$H(x_k) = C_1 * b^{(k-1)} + C_2 * b^{(k-2)} + \dots + C_k * b^{(k-k)} \quad (1)$$

Keterangan:

C₁ = nilai ASCII dari setiap huruf

b = basis bilangan prima (2)

k = jumlah n-gram

Berikut merupakan penerapan rumus *rolling hash* untuk menghitung n-gram dari 2 judul tersebut.

$$H(\text{pend}) = \text{ASCII}(p) * 2^{(3-1)} + \text{ASCII(e)} * 2^{(2-1)} + \text{ASCII(d)} * 2^{(1-1)}$$

$$H(\text{pend}) = 122.2^{(4-1)} + 101.2^{(4-2)} + 110.2^{(4-3)} + 100.2^{(4-4)}$$

$$H(\text{pend}) = 896 + 404 + 220 + 100 = 1620$$

Jadi, berdasarkan perhitungan fungsi *hash* dari huruf Pend menghasilkan nilai *hash* yaitu 1620. Perhitungan tersebut dilakukan pada seluruh rangkaian n-gram sehingga menghasilkan nilai seperti terlihat pada Tabel 3.

Tabel 3. Pembentukan *rolling hash*

N-gram judul 1	N-gram Judul 2
12076312652, 11046020346,	12076312652, 11046020346,
12087502373, 12176182170,	12176182170, 10622741151,
10622741151, 11376705757,	3446326945, 11755805527,
3446326945, 11755805527,	11235568362, 10211651056,
10205102147, 11235568362,	3446368055, 11759957624,
10211651056, 11897008169,	11550716328, 10511294125,
3446368055, 11759957624,	12176170144, 10621526460,
10624464028, 11550716328,	12075465565, 10960464560,
10511294125, 10630656640,	11759959741, 10624677836,
12176170144, 10621526460,	12692349946, 10205183957,
11254021891, 12075457491,	11046195266, 12105169371,
10959649077, 3364004286, 3441216971,	12172077729, 10208192610,
11239698154, 10628760044,	10207162345, 11443648348,
11984613950, 12294551933,	11480643749, 3433963571,
12067986726, 10205101743,	10207529040, 11480684552,
11235527558, 10207529848,	10923342639, 10207154533,
11480766151, 3446326238,	10127840141, 3432105748,
11755734038, 10197881834,	12276070360, 10201347840,
10506316734, 10127940146,	3444265939, 11547643839,
3442206265, 11339616844,	10818498042, 10127950749,
10210447248, 11775423483,	11447776726, 10624494834,
12186515780, 11666435765,	10825547030, 10839898534,
11688856764, 3443277155,	11547684721, 10205101743,
11447776726, 10624494834,	10207529646, 11480745767,
11553827743, 10825547030,	11547797731, 10216515753,
10839898534, 12289400428,	11030738933, 10544079749,
11547684721, 10205101743,	10299103132, 10219567266
11235527555, 10207529549,	
11480735968, 3443277769,	
11447838737, 10630757929,	
12186400260, 11654768176,	
10510430275, 10543407772,	
3364034476, 3444266162, 11547666431,	
10203254453, 11048951254,	
12383524146, 10544079749,	
3431904149, 10299103132, 10219567266	

5. Pembentukan Window

Proses pembentukan *window* sama dengan pembentukan *n-gram*. Pada penelitian ini, peneliti menggunakan *window* =4 yang dapat dilihat pada Tabel 4.

Tabel 4. Pembentukan *window*

<i>Window</i> judul 1	<i>Window</i> judul 2

W1= 12076312652, 11046020346, 12087502373, 12176182170	W1= 12076312652, 11046020346, 12087502373, 12176182170
W2= 11046020346, 12087502373, 12176182170, 10622741151	W2= 11046020346, 12087502373, 12176182170, 10622741151
W3 = 12087502373, 12176182170, 10622741151, 11376705757	W3= 12087502373, 12176182170, 10622741151, 11376705757,
W5 =12176182170, 10622741151, 11376705757, 3446326945.....	W5= 12176182170, 10622741151, 11376705757, 3446326945.....

6. Pencarian Fingerprint

Dokumen *fingerprint* merupakan metode yang digunakan untuk mendeteksi dari salinan antara teks atau antar dokumen baik dari seluruh teks dokumen atau sebagian dari teks dokumen yang dapat dilihat pada Tabel 5.

Tabel 5. Hasil *Fingerprint*

<i>Fingerprint</i> judul 1				<i>Fingerprint</i> Judul 2			
110460	110460	120875	120875	110460	110460	120875	121761
20346	20346	02373	02373	20346	20346	02373	82170

Jumlah <i>fingerprint</i> judul 1	=4
Jumlah <i>fingerprint</i> judul 2	=4
Gabungan semua	=8
Jumlah yang sama	=3

Perhitungan Jaccard coefficient untuk menghitung *similarity* dapat dilihat pada Persamaan (2).

$$Similarity = \frac{intersection}{union - intersection} \times 100\% \tag{2}$$

Keterangan :

intersection = *fingerprint* yang sama
union = gabungan *fingerprint*

$$Similarity = \frac{3}{(8 - 3)} \times 100\%$$

$$= 60\%$$

Berdasarkan hasil perhitungan persentase dari kemiripan 2 judul yang dibandingkan diperoleh hasil *similarity* sebesar 60%.

3.2.2 Hasil Pengujian UAT

User Acceptance Testing (UAT) dilakukan untuk menguji sejauh mana sistem yang telah dikembangkan dapat diterima oleh pengguna. UAT dilakukan untuk menguji sejauh mana aplikasi yang telah dikembangkan dapat diterima oleh pengguna. UAT ini melibatkan 7 responden yang terdiri dari 5 orang mahasiswa, 1 dosen dan 1 pegawai tata usaha. Setiap responden memasukkan judul, melakukan validasi dan menginput judul di dalam sistem. Setelah pengujian, responden diminta untuk memberikan penilaian terhadap aspek-aspek sistem dengan menggunakan skala Likert (1-5), 1: sangat tidak setuju, 2: tidak setuju, 3: netral, 4: setuju, 5: Sangat Setuju. Hasil pengujian UAT dapat dilihat pada Tabel 6

Tabel 6. Pengujian Sistem

Responden	Kemudahan penggunaan	Kejelasan informasi	Kecepatan sistem	Kepuasan pengguna	Rata-Rata skor
Mahasiswa 1	5	5	4	5	4.75
Mahasiswa 2	3	4	5	4	4.0
Mahasiswa 3	4	4	5	5	4.25
Mahasiswa 4	5	5	3	4	4.25
Mahasiswa 5	5	5	5	5	5.0
Tata Usaha	5	5	5	5	5.0
Dosen	4	5	4	4	4.25

Hasil pengujian UAT menunjukkan rata-rata skor dari semua responden adalah sebesar 4,5 yang menunjukkan bahwa responden secara umum merasa puas dengan kemudahan, kejelasan dan kecepatan sistem yang dikembangkan dalam membantu proses deteksi judul.

4. Kesimpulan dan Saran

4.1 Kesimpulan

Aplikasi pengelolaan dan deteksi kesamaan judul KP pada Program Studi Ilmu Komputer berbasis web telah berhasil dibangun. Aplikasi ini memungkinkan pengelolaan dan deteksi kesamaan judul KP sehingga mempermudah mahasiswa dalam mengecek kesamaan judul, membantu pegawai tata usaha dalam mengelola data judul, serta mendukung dosen dalam proses bimbingan secara *online*. Selain itu, informasi mengenai judul KP dapat diakses dengan mudah dan dalam waktu yang relatif singkat.

4.2 Saran

Aplikasi ini dapat dikembangkan lagi dengan menambahkan fitur rekomendasi judul KP berdasarkan topik yang belum banyak dieksplorasi sehingga membantu mahasiswa yang kesulitan mencari dan mengurangi duplikasi.

5. Referensi

- [1] M. P. Rihi, A. A. Pekuwali, and D. A. Sitaniapessy, "Algoritma WInnowing untuk Mendeteksi Kesamaan Judul Skripsi Teknik Informatika: WInnowing Algorithm for Detecting the Similarity of Informatic Engineering Undergraduate Thesis Title," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 2, no. 2, pp. 42-52, 2022.
- [2] S. Shrestha, S. Shakya, and S. Gautam, "WInnowing vs Extended-WInnowing: A Comparative Analysis of Plagiarism Detection Algorithms," *Journal of Trends in Computer Science and Smart Technology*, vol. 5, no. 3, pp. 213-232, 2023.
- [3] M. Nurudin, W. Jayanti, R. D. Saputro, M. P. Saputra, and Y. Yulianti, "Pengujian Black Box Pada Aplikasi Penjualan Berbasis Web Menggunakan Teknik Boundary Value Analysis." *Jurnal Informatika Universitas Pamulang*, vol. 4, no. 4, 2019.
- [4] A. Rahmatulloh, N. I. Kurniati, I. Darmawan, A. Z. Asyikin, and D. Witarsyah, "Comparison between the stemmer porter effect and nazief-adriani on the performance of winnowing algorithms for measuring plagiarism," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 9, no. 4, pp. 1124-1128, 2019.
- [5] S. Taerungruang, and W. Aroonmanakun, "Constructing an Academic Thai Plagiarism Corpus for Benchmarking Plagiarism Detection Systems," *GEMA Online Journal of Language Studies*, vol. 18, no. 3, 2018.
- [6] J. Jarwati, and A. C. Prihandoko, "Penerapan Algoritma WInnowing Pada Sistem Rekomendasi Penentuan Dosen Pembimbing Skripsi (Studi Kasus Program Studi Sistem Informasi)," *BERKALA SAINSTEK*, vol. 5, no. 1, pp. 11-20, 2017.
- [7] S. Schleimer, D. S. Wilkerson, and A. Aiken, "WInnowing: local algorithms for document fingerprinting," In *Proceedings of the 2003 ACM SIGMOD international conference on Management of data*, pp. 76-85, Juni 2003.

- [8] R. K. Wibowo, and K. Hastuti, “Penerapan Algoritma WInnowing Untuk Mendeteksi Kemiripan Teks pada Tugas Akhir Mahasiswa,” *Techno. com*, vol. 15, no. 4, pp. 303-311, 2016.
- [9] R. A. Putra, F. P. Utama, and A. Erlansari, “Penerapan Algoritma WInnowing pada Sistem Pengelolaan Kerja Praktik Dengan Pendekatan Human-Centered Design (Studi Kasus: Program Studi S-1 Informatika Universitas Bengkulu),” *Pseudocode*, vol. 10, no. 1, pp. 30-44, 2023.
- [10] R. Ramdhani, A. Fadlil, and S. Sunardi, “Penerapan Algoritma WInnowing dan Word-Level Trigrams Untuk Mengidentifikasi Kesamaan Kata,” *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 2, pp. 427-435, 2022.
- [11] A. Yudhana, and I.A. Mukaromah, “Implementation of WInnowing Algorithm with Dictionary English-Indonesia Technique to Detect Plagiarism,” *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 5, 2018.
- [12] Normah, B. Rifai, S. Vambudi, and R. Maulana, “Analisa Sentimen Perkembangan Vtuber dengan Metode Support Vector Machine Berbasis SMOTE,” *J. Tek. Komput. AMIK BSI*, vol. 8, no. 2, pp. 174–180, 2022.
- [13] N. Alamsyah, and M. Rasyidan, “Deteksi Plagiarisme Tingkat Kemiripan Judul Skripsi Pada Fakultas Teknologi Informasi Menggunakan Algoritma WInnowing,” *Technol. J. Ilm.*, vol. 10, no. 4, p. 197, 2019.
- [14] S. Sunardi, A. Yudhana, and I. A. Mukaromah, “Indonesia Words Detection Using Fingerprint WInnowing Algorithm,” *Jurnal Informatika*, vol. 13, no. 1, p. 7, Jan. 2019.
- [15] R. A. Putra, F.P. Utama, and A. Erlansari, “Penerapan Algoritma WInnowing pada Sistem Pengelolaan Kerja Praktik dengan Pendekatan Human-Centered Design (Studi Kasus: Program Studi S-1 Informatika Universitas Bengkulu),” *Pseudocode*, vol. 10, no. 1, pp. 30-44, 2023.
- [16] N. P. B. Pratama, M. Mustaqiem, and M. Minarni, “Pengembangan Sistem Informasi Pengelolaan Judul Skripsi dan Tugas Akhir dengan Fitur Deteksi Kemiripan Menggunakan Algoritma WInnowing,” *TIN: Terapan Informatika Nusantara*, vol. 2, no. 5, pp. 271-278, 2021.
- [17] H. Sucipto, M. F. Rizal, A. H. Mujiyanto, M. Ali, T. Kistofer, C. Mashuri, and A. D. Indriyanti, “Artikel Rancang Bangun Sistem Pengukuran Tingkat Kemiripan Judul dan Abstrak Skripsi menggunakan Algoritma WInnowing dan Dice Similarity,” In *Prosiding Seminar Nasional Sains, Teknologi, Ekonomi, Pendidikan dan Keagamaan (SAINSTEKNOPAK)*, vol. 8, pp. 229-238, Desember 2024.
- [18] E. L. Amalia, V. A. Lestari, V. N. Wijayaningrum, and A. A. Ridla, “Automatic essay assessment in e-learning using winnowing algorithm,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 29, no. 1, pp. 572-582, 2023.
- [19] A. M. Siahaan, and D. Aryanto, “Implementasi Algoritma Rabin Karp Dan Optimasi Dengan Algoritma Stemmer Sastrawi Dalam Deteksi Plagiat Pada Jurnal Skripsi Mahasiswa,” *Jurnal Armada Informatika*, vol. 8, no. 1, pp. 23-31, 2024.