

Deteksi Kemiripan Dokumen Pengajuan Proposal Menggunakan Algoritma Biword Winnowing Pada Sistem Informasi Penelitian Dan Pengabdian

Maskur*¹, Deny Qutara Putra², Nur Hayatin³

^{1,2,3}Teknik Informatika/Universitas Muhammadiyah Malang

maskur@umm.ac.id*¹, denysqp@gmail.com², noorhayatin@umm.ac.id³

Abstrak

SIMPPM atau Sistem Informasi Manajemen Penelitian dan Pengabdian Kepada Masyarakat merupakan sistem yang dikembangkan oleh DPPM Universitas Muhammadiyah Malang. Sistem ini menampung proposal penelitian dan pengabdian dosen seluruh universitas muhammadiyah malang. Reviewer dalam menilai suatu proposal masih kesulitan untuk mengetahui apakah proposal yang diajukan itu sudah atau pernah diajukan, sehingga reviewer tidak tahu nilai originalitas dari proposal yang diajukan. Algoritma Biword Winnowing digunakan untuk mencari fingerprint dari dokumen proposal yang ada dalam sistem. Penelitian ini menggunakan algoritma tersebut untuk mencari nilai kemiripan proposal satu dengan lainnya, sehingga setting algoritma Biword Winnowing yang baik perlu untuk dicari guna untuk meningkatkan nilai akurasi. Hasil akurasi dengan algoritma ini dibandingkan dengan aplikasi Plagiarism Checker X sebagai ukuran akurasi dari Algoritma Biword Winnowing. Hasil pengujian menunjukkan bahwa nilai window dan basis sangat mempengaruhi hasil akhir similarity.

Kata Kunci: *Biword Winnowing, Kemiripan Dokumen, Proposal SIMPPM, Plagiarism Checker X*

Abstract

SIMPPM or Management Information System Research and Community Service is a system developed by DPPM University of Muhammadiyah Malang. This system accommodates research proposal and dedication of all university lecturers muhammadiyah malang. The reviewer in assessing a proposal is still difficult to know whether the proposed proposal has been or has been submitted, so the reviewer does not know the originality of the proposed proposal. The Biword Winnowing algorithm is used to find the fingerprint of the proposed documents in the system. This research uses the algorithm to find the value of similarity of proposal with one another, so setting a good Biword Winnowing algorithm needs to be searched in order to increase accuracy value. The result of accuracy with this algorithm is compared with Plagiarism Checker X application as accuracy measure from Biword Winnowing Algorithm. The test result shows that the value of window and base greatly affect the final result of similarity.

Keywords: *Biword Winnowing, Similarity Documents, SIMPPM Proposal, Plagiarism Checker X*

1. Pendahuluan

UMM atau Universitas Muhammadiyah Malang adalah Perguruan Tinggi Muhammadiyah yang berdiri sejak tahun 1964 di Jalan Raya Tlogomas 246 Kota Malang, Jawa Timur. Sistem Informasi Manajemen Penelitian dan Pengabdian Kepada Masyarakat (SIMPPM) adalah sistem informasi yang dikembangkan oleh pihak DPPM Universitas Muhammadiyah Malang guna mendukung pelaksanaan penelitian dan pengabdian kepada masyarakat.

Dalam pengelolaan proposal penelitian dan pengabdian, untuk menghasilkan proposal yang layak didanai, proposal yang diajukan oleh peneliti dan pengabdi harus berkualitas. Salah satu yang menjadi faktornya adalah nilai originalitas dari proposal yang diajukan. Reviewer yang bertugas dalam mengkaji dan menilai proposal selama ini tidak bisa mengetahui apakah proposal yang diajukan itu sudah atau pernah diajukan, sehingga reviewer tidak tahu nilai originalitas dari proposal yang diajukan. originalitas dari dokumen terkait dengan plagiarisme dimana menggunakan ide, penulisan, hasil milik orang lain dan mempresentasikannya sebagai karya sendiri tanpa memberitahu atau mensitasi penulis yang sebenarnya [1], dalam kasus ini terkait dengan proposal penelitian dan pengabdian yang diajukan dalam SIMPPM. Maka dari itu

sangatlah penting bagi reviewer untuk mengetahui nilai originalitas dari proposal yang diajukan ke dalam sistem.

Untuk mengatasi masalah diatas, perlu dilakukan deteksi kemiripan proposal. Hal ini sulit jika dilakukan secara manual, sehingga perlu ditambahkan fitur deteksi kemiripan dokumen proposal ke dalam Sistem Informasi Penelitian dan Pengabdian (SIMPPM). Metode yang digunakan untuk melakukan deteksi kemiripan dokumen proposal adalah dengan menggunakan Biword Wnnowing. Biword Wnnowing merupakan algoritma pengembangan dari Wnnowing.

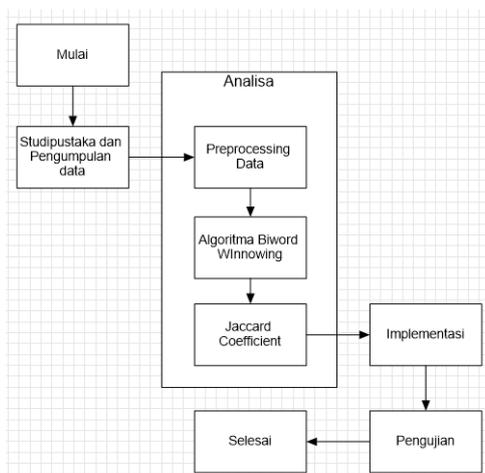
Ada dua jenis deteksi plagiarisme yaitu deteksi secara intrinsik dan deteksi secara ekstrinsik. Algoritma Wnnowing tergolong metode deteksi plagiarisme secara ekstrinsik karena Algoritma ini mendeteksi bagian kalimat atau kata yang memiliki kemiripan dari dokumen yang diuji plagiarismenya dengan dokumen sumber. Algoritma ini merupakan pengembangan dari metode Rabin-Karp Fingerprint. Perbedaannya adalah Algoritma Wnnowing menggunakan konsep window guna untuk mengoptimalkan hasil deteksi sedangkan Rabin-Karp tidak [2]. Pada pendeteksiannya, Algoritma Wnnowing harus memenuhi kebutuhan mendasar yaitu Whitespace insensitivity, Noise surpression dan Position independence [3]. Algoritma Wnnowing ini bertujuan agar dapat mengidentifikasi penjiplakan atau kemiripan dokumen, termasuk bagian-bagian kecil yang mirip dalam dokumen yang berjumlah banyak seperti yang sudah dilakukan oleh Nur Fadillah Ulfa yang membuat aplikasi deteksi kemiripan dokumen guna untuk mendeteksi plagiasi [4]. Namun, algoritma ini juga tidak terbatas untuk deteksi plagiasi saja, seperti yang sudah implementasikan oleh Sariyunti Astutik yang memanfaatkan algoritma ini untuk sistem penilaian esai otomatis yang dilakukannya pada tahun 2014 lalu [5].

Algoritma Wnnowing berfungsi untuk membuat memproses teks yang ada dalam dokumen menjadi suatu fingerprint. Algoritma tersebut akan menghitung nilai hash dari tiap k-gram dengan menggunakan fungsi rolling hash. Kemudian, dibentuk window window dari nilai hash yang terbentuk. nilai minimum hash akan ditunjuk dari setiap window yang terbentuk. nilai hash yang ditunjuk akan digunakan sebagai fingerprint dari dokumen. Fingerprint yang sudah dibentuk akan dijadikan sebagai dasar perbandingan antara file dokumen lainnya. [3].

Penelitian lainnya, menggunakan Algoritma Wnnowing yang sudah dikembangkan, yaitu dengan menggunakan pendekatan Biword (Algoritma Biword Wnnowing). Perbedaan dari algoritma yang dikembangkan oleh M. Ridho dengan Algoritma Wnnowing terletak pada teknik fingerprintnya. Algoritma Wnnowing menggunakan teknik fingerprint berbasis karakter sedangkan Algoritma Biword Wnnowing menggunakan teknik fingerprint berbasis 2 kata. hal ini bertujuan untuk menjaga arti kata atau frasa dalam teks dokumen. [6]

Hasil deteksi kemiripan proposal perlu dilakukan sebagai bahan pertimbangan bagi reivewer untuk mengkaji/menilai proposal. Sehingga diharapkan dengan menerapkan metode Biword Wnnowing untuk Deteksi Kemiripan Dokumen Proposal pada Sistem Informasi Penelitian dan Pengabdian dapat membantu pihak reviewer untuk mengetahui tingkat kemiripan proposal yang diajukan terhadap proposal-proposal yang sudah atau pernah diajukan sebelumnya.

2. Metode Penelitian



Gambar 1. Tahapan Penelitian

Untuk melakukan deteksi kemiripan dokumen diperlukan metode yang tepat, sehingga perlu dibuat sebuah Metode Penelitian yang dapat menjelaskan alur penggunaan Algoritma Biword Wnnowing dalam penelitian ini. sehingga tahapan penelitian yang dilakukan dalam penelitian ini seperti pada Gambar 1 diatas.

2.1 Studi Pustaka dan Pengumpulan Data

Pada tahapan ini dilakukan pengumpulan data dan informasi yang berhubungan dengan algoritma deteksi dokumen dan informasi terkait lainnya yang dpat menyelesaikan masalah yang diteliti pada tugas akhir ini melalui jurnal ilmiah, skripsi dan buku-buku lainnya.

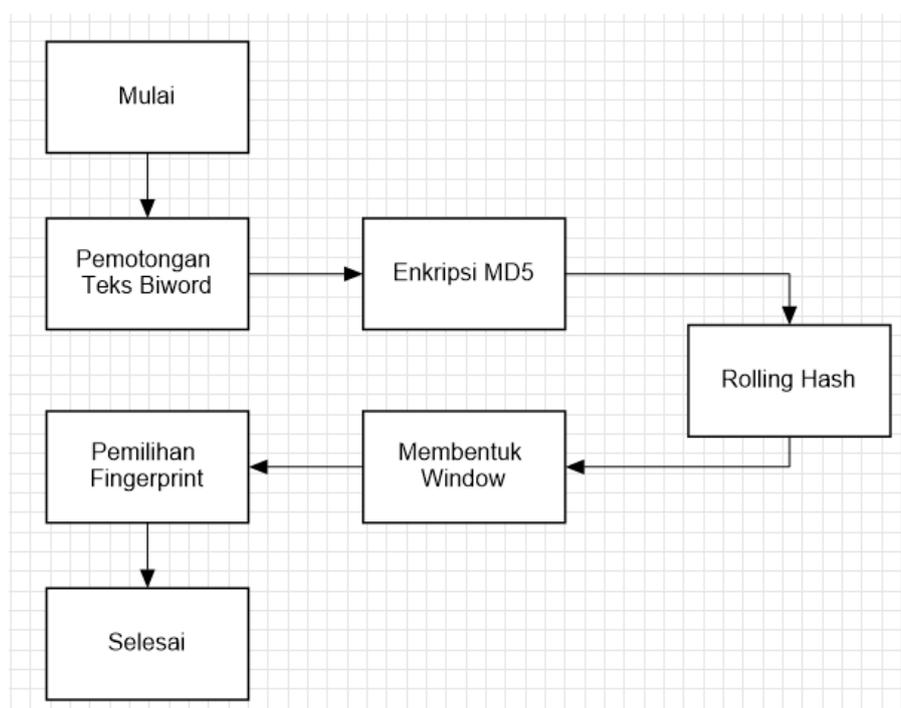
2.1.1 Analisa

1. Preprocessing Data

Pembersihan teks judul dan abstrak proposal yang ada dalam Sistem Informasi Penelitian dan Pengabdian melalui beberapa tahap,diantaranya [7] :

1. Casefolding (Mengubah huruf kapital menjadi huruf kecil)
2. Stopping (Penghapusan kata yang tidak memiliki informasi penting)
3. Stemming (Proses merubah kata berimbuhan dalam kalimat menjadi kata dasar)

2. Algoritma Biword Wnnowing



Gambar 2. Tahapan Algoritma Biword Wnnowing

Pembuatan fingerprint dari judul dan abstrak proposal target dan sumber yang sudah dibersihkan. Pada tahap ini akan dilakukan:

1. Melakukan pemotongan teks yang sudah dilakukan preprocessing pada tahap sebelumnya menjadi biword yang kemudian dienkripsi menggunakan MD5
2. Menghitung nilai hash (rolling hash). rumus dari rolling hash seperti pada Persamaan 1.

$$H(1, n) = c_1 * b^{(k-1)} + c_2 * b^{(k-2)} + \dots + c_k * b^0 \quad [8] \quad (1)$$

3. Membentuk window ukuran w
4. Pemilihan beberapa nilai hash menjadi dokumen fingerprint [6]
5. Jaccard Coefficient

Menghitung tingkat kemiripan dokumen proposal dengan persamaan jaccard coefficient [6]. Rumus dari Jaccard Coefficient seperti pada Persamaan 2.

$$\text{Similarity}(d_i, d_j) = \frac{W(d_i) \cap W(d_j)}{W(d_i) \cup W(d_j)} \times 100\% \quad [9] \quad (2)$$

2.1.2 Implementasi

Pada proses implementasi akan dilakukan pembuatan program berdasarkan algoritma yang telah dianalisa sebelumnya untuk diimplementasikan pada bahasa pemrograman (php) dan dilakukan pengujian untuk mengetahui keberhasilan dari deteksi kemiripan proposal.

2.1.3 Pengujian

Tujuan tahap pengujian ini adalah untuk mencapai nilai akurasi terbaik dari penggunaan algoritma biword winnowing. pengujian dilakukan dengan mencari nilai window (winnowing) dan nilai basis (rolling hash) terbaik unyuk studi kasus deteksi dokumen proposal berdasarkan judul dan abstrak. Skenario dari pengujian yang akan dilakukan adalah sebagai berikut:

1. Mengumpulkan koleksi data judul dan abstrak proposal yang ada pada Sistem Informasi Penelitian dan Pengabdian masing masing sebanyak 233 dengan 122 data penelitian dan 111 data pengabdian. Data yang digunakan dalam pengujian untuk mencari nilai setting terbaik adalah Data Training yang merupakan 80% dari keseluruhan data. Sisa 20% data akan digunakan sebagai Data Testing untuk menguji setting yang didapat dari hasil training. Akan dipilih dari dua kelompok masing masing satu data yang akan digunakan sebagai tolak ukur kemiripan proposal.
2. Melakukan proses deteksi kemiripan menggunakan biword winnowing dan aplikasi Plagiarism Checker X sesuai dengan masing masing kelompok data. Aplikasi tersebut diperlukan sebagai pembanding guna untuk mencari nilai margin dari window dan basis
3. Menghitung nilai akurasi dari biword winnowing terhadap aplikasi plagiarism checker dengan menggunakan nilai basis dan window yang sudah ditentukan.
4. Proses no 3 dan 4 terus dilakukan dengan menggunakan nilai basis dan window yang berbeda hingga nilai perbedaan akurasi dari biword winnowing terhadap aplikasi plagiarism checker mendekati 0%.

Plagiarism Checker X merupakan aplikasi yang dapat membantu pengguna untuk mendeteksi plagiarisme dari paper, blog, tugas atau yang lainnya. aplikasi ini memiliki fitur diantaranya *Online plagiarism checking, cross-comparison of multiple documents, keywords searching, bulk searching, and website scanning* [10].

3. Hasil Penelitian dan Pembahasan

Algoritma Biword Winnowing akan diterapkan terhadap data penelitian dan data pengabdian. Jumlah data yang digunakan untuk Training adalah 80% dari total data pada masing masing kelompok data (penelitian dan pengabdian). sedangkan sisanya yaitu 20%, akan digunakan sebagai data Testing. Hasil dari Tabel 1 algoritma ini berupa fingerprint dari masing masing data yang nantinya akan digunakan untuk mencari nilai kemiripan.

Tabel 1. Detail Penggunaan DataSet

Jenis Data	Penelitian	Pengabdian
Keseluruhan (100%)	122	111
Training (80%)	98	89
Testing (20%)	24	22

Namun nilai fingerprint yang didapat akan berbeda jika setting nilai basis (b) dan window (w) berbeda. untuk itu diperlukan training untuk menentukan nilai basis dan window yang sesuai. Setting nilai basis dan window yang akan diuji tertera pada Tabel 2.

Tabel 2. Nilai Basis dan Setting yang Digunakan Dalam Pengujian

BasisWindow	1	2	3	4	5	6	7	8
1	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8
2	2,1	2,2	2,3	2,4	2,5	2,6	2,7	2,8
3	3,1	3,2	3,3	3,4	3,5	3,6	3,7	3,8
4	4,1	4,2	4,3	4,4	4,5	4,6	4,7	4,8
5	5,1	5,2	5,3	5,4	5,5	5,6	5,7	5,8

6	6,1	6,2	6,3	6,4	6,5	6,6	6,7	6,8
7	7,1	7,2	7,3	7,4	7,5	7,6	7,7	7,8
8	8,1	8,2	8,3	8,4	8,5	8,6	8,7	8,8

3.1 Hasil Uji Plagiarism Checker X

Aplikasi ini akan digunakan sebagai tolok ukur dalam melakukan pengujian algoritma biword winnowing yang sudah diimplementasikan. maka dari itu diperlukan hasil prosentase similarity dari seluruh data. Data yang digunakan sebagai sumber dalam aplikasi ini harus sama dengan data sumber yang digunakan pada algoritma biword winnowing.

Files	penelitian 19.txt	penelitian 20.txt	penelitian 21.txt	penelitian 24.txt	penelitian 25.txt	penelitian 26.txt	penelitian 27.
penelitian 13 sumber.txt	0 % Duplicate	0 % Duplicate					

Gambar 3. Screenshot Hasil Plagiarism Checker X menggunakan data Proposal Penelitian

Files	pengabdian 10.	pengabdian 11.	pengabdian 17.	pengabdian 18.	pengabdian 19.	pengabdian 20.	pengabdian 2
pengabdian 77 sumber.txt	0 % Duplicate	2,37 % Duplicate	0 % Duplicate				

Gambar 4. Screenshot Hasil Plagiarism Checker X menggunakan data Proposal Pengabdian

Pada Gambar 3 dan Gambar 4 sudah ditampilkan hasil dari nilai kemiripan yang dilakukan dengan menggunakan aplikasi Plagiarism Checker X. dengan sumber proposal yang sudah ditentukan.

3.2 Hasil Uji Algoritma Biword Winnowing

Pada tahap ini akan dilakukan pengujian algoritma biword winnowing menggunakan Matrik setting yang sudah disediakan pada tabel 2. Jika dilihat dari matrik pengujian pada Tabel 2, akan ada 64 kombinasi setting basis dan window yang akan diuji. Pengujian akan dibagi menjadi 2 tahap, yaitu tahap Training dan Tahap Testing dengan menggunakan data yang sudah tertera pada Tabel 1.

3.2.1 Tahap Training

Tahap ini bertujuan untuk mengetahui nilai setting basis dan window yang terbaik yang didapat dengan menggunakan Data Training yang sudah ditentukan sebelumnya yaitu sejumlah 98 data untuk penelitian dan 89 data untuk pengabdian. Pengujian Training melibatkan keseluruhan setting karena tahap ini akan mencari satu setting terbaik untuk masing masing kategori (penelitian dan pengabdian) dari 64 setting yang ada pada matrix Tabel 2.

Total data yang didapat dari hasil training ialah sebanyak 6272 (64×98) untuk penelitian dan 5696 (64×89) untuk pengabdian. selisih persentase yang didapat akan cari nilai rata ratanya berdasarkan set window dan set basis masing masing, hal ini dilakukan untuk mendapatkan nilai rata rata selisih persentase per setting.

Nilai rata rata selisih akan digunakan sebagai patokan untuk menentukan nilai setting basis dan window yang baik, semakin nilai rata rata selisih mendekati 0 maka akan semakin baik nilai window dan basis yang digunakan.

Tabel 3. Hasil Evaluasi Training Detail Menggunakan Data Proposal Penelitian

No	Window	Basis	Rata Rata Hasil	Rata Rata Hasil	Rata Rata Selisih (%)
			Biword Wnnowing (%)	Plagiarism Checker X (%)	
25	2	4	0,49361772	0,279285714	0,463892119

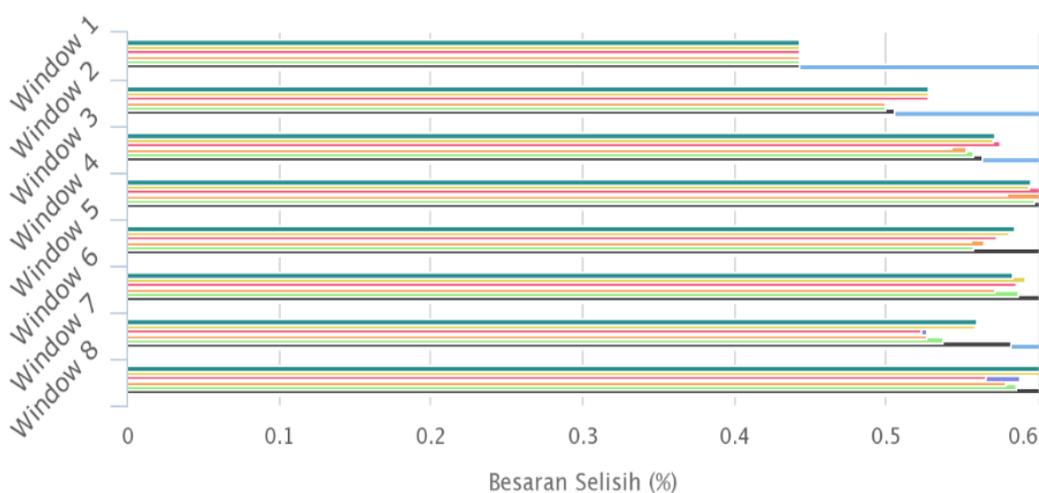
Tabel 4. Hasil Evaluasi Training Detail Menggunakan Data Proposal Pengabdian

No	Window	Basis	Rata Rata Hasil	Rata Rata Hasil	Rata Rata Selisih (%)
			Biword Wnnowing (%)	Plagiarism Checker X (%)	
9	1	2	0,425786255	0,064044944	0,44343891
17	1	3	0,425786255	0,064044944	0,44343891
25	1	4	0,425786255	0,064044944	0,44343891
33	1	5	0,425786255	0,064044944	0,44343891
41	1	6	0,425786255	0,064044944	0,44343891
49	1	7	0,425786255	0,064044944	0,44343891
57	1	8	0,425786255	0,064044944	0,44343891

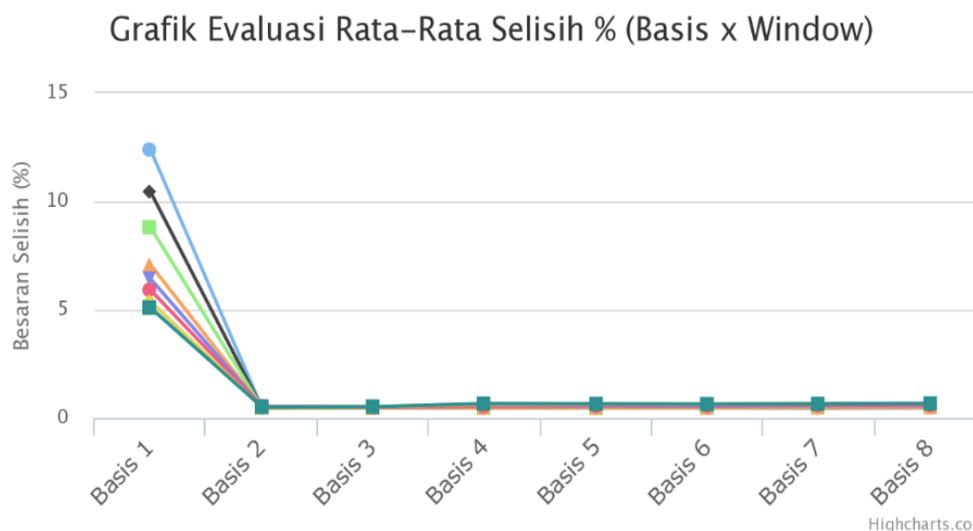
Pada Tabel 3 dan Tabel 4 merupakan hasil evaluasi dari data training penelitian dan pengabdian yang sudah di ambil nilai rata rata selisih yang paling kecil. Setting yang memiliki nilai rata rata selisih yang paling kecil terdapat pada nilai 0,463892119% dengan setting window = 2 dan basis = 4. sedangkan untuk pengabdian terdapat pada nilai 0,44343891% dengan setting window = 1 dan basis = 2, 3, 4, 5, 6, 7, 8. terdapat banyak nilai setting dikarenakan basis yang memiliki window 1 kecuali basis = 1, akan memiliki hasil selisih yang sama. Sehingga basis tidak memiliki pengaruh terhadap hasil selisih jika nilai window adalah 1 kecuali basis 1. hal ini dapat dilihat pada Gambar 5.

Sesuai dengan Gambar 5 dan Gambar 6, nilai dari basis dan window dapat mempengaruhi hasil persentase kemiripan dokumen, hasil rata rata selisih akurasi dari basis 1 memiliki nilai selisih yang lebih besar dari basis 2 sampai 8 sehingga nilai basis 1 tidak cocok digunakan sebagai setting basis daripada basis yang lainnya.

Grafik Evaluasi Rata-Rata Selisih % (Window x Basis)



Gambar 5. Diagram Evaluasi Rata-Rata Selisih % Data Proposal Pengabdian (Window x Basis)



Gambar 6. Diagram Evaluasi Rata-Rata Selisih % Data Proposal Pengabdian (Basis x Window)

3.2.1 Tahap Testing

Pada Tabel 5, tahap ini bertujuan untuk menguji setting yang sudah didapatkan melalui tahap training dengan menggunakan data testing sehingga diharapkan hasil dari testing tidak memiliki perbedaan yang jauh dari hasil training.

Tabel 5. Setting Terbaik hasil Training Data Penelitian dan Pengabdian

Kategori	Window	Basis	Hasil Selisih	Keterangan
Penelitian	2	4	0,463892119	-
Pengabdian	1	2	0,44343891	dipilih basis 2 untuk meminimalisir proses perhitungan

Pada tahap training sudah didapatkan setting yang terbaik menurut hasil training. Proses yang dilakukan untuk tahap ini sama dengan tahap Training, hanya saja jumlah data yang digunakan adalah 20% dari data keseluruhan dan hanya menggunakan 1 setting saja pada tiap kategori.

Tabel 6. Hasil Evaluasi Testing Penelitian dan Pengabdian dengan 2 Algoritma

No	Kategori	Algoritma	W	B	Rata Rata Hasil Biword Wnnowing (%)	Rata Rata Hasil Plagiarism Checker X (%)	Rata Rata Selisih (%)
1	Penelitian	Biword Wnnowing	2	4	4,7733665854	4,257916667	0,634635956
2	Pengabdian	Biword Wnnowing	1	2	5,142701624	4,782727273	0,586314628
3	Penelitian	Wnnowing	2	4	5,042889450	4,257916667	0,784972784
4	Pengabdian	Wnnowing	1	2	7,18766646	4,782727273	2,404939188

Tabel 6 menunjukkan hasil evaluasi testing secara keseluruhan dengan menggunakan Algoritma yang berbeda. Hasil menunjukkan bahwa dengan menggunakan Algoritma Biword Wnnowing dapat menghasilkan selisih rata-rata yang lebih kecil daripada menggunakan algoritma winnowing dengan kondisi Setting W (Window) dan B (Basis) yang sama.

Hasil rata rata selisih dengan menggunakan Algoritma Biword Wnnowing menunjukkan untuk penelitian memiliki nilai rata rata selisih 0,634635956% sedangkan untuk pengabdian bernilai 0,586314628%. jika dibandingkan dengan hasil dari training ditunjukkan pada Tabel 7.

Tabel 7. Hasil Evaluasi Perbandingan Testing dan Training Penelitian dan Pengabdian dengan Menggunakan Algoritma Biword Wnnowing

Kategori	Rata Rata Selisih (%)		Selisih
	Training	Testing	
Penelitian	0,463892119	0,634635956	0,17074384
Pengabdian	0,44343891	0,586314628	0,14287572

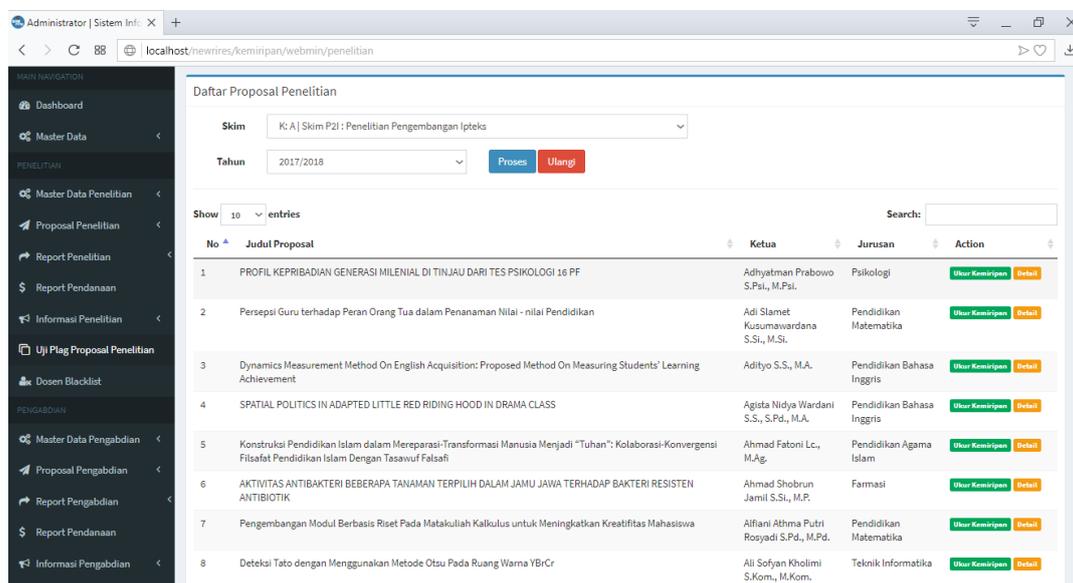
Dari nilai selisih dapat disimpulkan bahwa untuk setting basis = 4 dan window = 2 pada kategori penelitian memiliki akurasi error sebesar 0,17% sedangkan untuk kategori pengabdian dengan menggunakan setting window 1 dan basis 2 memiliki akurasi error sebesar 0,14%. semakin kecil nilai akurasi error maka semakin baik setting yang digunakan untuk algoritma biword winnowing, sehingga setting setting yang sudah disebutkan diatas dapat digunakan sebagai setting terbaik untuk diimplementasikan ke dalam algoritma biword winnowing yang nantinya akan digunakan untuk deteksi kemiripan dokumen proposal pada Sistem Informasi Penelitian dan Pengabdian.

3.3 Implementasi Algoritma Biword Wnnowing ke Sistem Informasi Penelitian dan Pengabdian

Hasil Setting terbaik yang sudah ditemukan pada tahap sebelumnya akan gunakan Algoritma Biword Wnnowing untuk diimplementasikan ke dalam Sistem. Setting tersebut dapat dilihat pada tabel 5. Untuk Hasil pengimplementasian Deteksi Kemiripan ke dalam sistem adalah sebagai berikut:

3.3.1 User Admin

Admin dapat mencari nilai kemiripan proposal penelitian dan pengabdian dengan seluruh proposal yang ada pada sistem. Admin bebas untuk memilih proposal yang ingin dicari nilai kemiripannya. Berikut adalah tampilan menu untuk deteksi kemiripan proposal pada bagian admin.



Gambar 7. Tampilan Menu Uji Plagiarisme Proposal Penelitian User Admin

Pada Gambar 7 terdapat form select skim dan tahun, yang artinya Admin harus memilih Skim mana dan tahun ajaran mana data yang ingin dicari nilai kemiripannya. Tabel dibawahnya akan menampilkan data proposal dari skim dan tahun yang dipilih sebelumnya. dalam tabel terdapat row Action yang didalamnya berisi tombol ukur kemiripan dan detail. Tombol ukur kemiripan berfungsi untuk mencari nilai kemiripan proposal yang diuji dengan proposal lainnya dalam bentuk persentase. sedangkan tombol detail berfungsi untuk menampilkan detail dari proposal itu sendiri.

No	Judul Proposal	Ketua	Skim	Persentase Kemiripan
1	UPAYA MENINGKATKAN LITERASI SAINTIK MAHASISWA DENGAN MENGGUNAKAN PEMBELAJARAN BERBASIS PENELITIAN	Agung Deddiliawan Ismail S.Pd., M.Pd.	Skim PPKPPT : Penelitian Peningkatan Kualitas Pembelajaran di Perguruan Tinggi	9.38 %
2	Persepsi Legiun Veteran Republik Indonesia (LVRI) Terhadap Jepang Pasca Diplomasai Geisha di Indonesia (Studi Deskriptif Kuantitatif Pada Anggota LVRI Kota Malang)	M. Syaprin Zahidi S.I.P., M.A.	Skim P2I : Penelitian Pengembangan Ipteks	7.5 %
3	VALIDASI ALAT UKUR KEPERIBADIAN BIG FIVE (MINI – INTERNATIONAL PERSONALITY ITEM POOL)	Aifah Nabilah Madurah S.Pd., M.A.	Skim P2I : Penelitian Pengembangan Ipteks	6.25 %
4	PENGEMBANGAN MODUL BAHASA PEMROGRAMAN GUI MATLAB UNTUK MAHASISWA PENDIDIKAN MATEMATIKA	Octavina Rizky Utami Putri M.Pd.	Skim P2I : Penelitian Pengembangan Ipteks	5.96 %
5	Analisis Pembelajaran Tematik Berbasis Kearifan Lokal di SD Muhammadiyah 1 Kota Malang	Ima Wahyu Putri Utami S.Pd., M.Pd	Skim P2I : Penelitian Pengembangan Ipteks	5.13 %
6	ANALISIS PENERAPAN EDMODO PADA PEMBERIAN TES FORMATIF MAHASISWA PENDIDIKAN PROFESI GURU MATEMATIKA	Reni Dwi Susanti M.Pd	Skim P2I : Penelitian Pengembangan Ipteks	5 %
7	REPRESENTASI KONSEP EKOBUDAYA DALAM NOVEL PRESIDEN KARYA WISNAN HADI	Dr. Hj. Sugarti M.Si.	Skim PDK : Penelitian Dasar Keilmuan	4.55 %
8	MODEL MANAJEMEN STRATEGI PADA PERUSAHAAN MULTI NASIONAL DI INDONESIA	Dr. Uci Yuliaty M.M.	Skim PDK : Penelitian Dasar Keilmuan	4.26 %
9	PENGEMBANGAN PRODUKSI CALF MILK REPLACER UNTUK OPTIMALISASI PEMELIHARAAN PEDDET PERAH JANTAN SEBAGAI SUMBER 'BEEF-FRISIEH'	Dr. Ir. Lestari Hendraningsih MP.	Skim PBP : Penelitian Berorientasi Produk	4.26 %
10	EFEKTIFITAS EKSTRAK DAUN KELOR (Moringa oleifera lam) DALAM MEMPERBAIKI PROFIL SERUM IRON DAN TOTAL IRON BINDING CAPACITY TIKUS PUTHY JANTAN DIET RENDAH ZAT BESI	dr. Diah Hermayanti Sp.Pk.	Skim PDK : Penelitian Dasar Keilmuan	4.26 %

Gambar 8. Tampilan Menu Hasil Uji Plagiarisme Proposal Penelitian User Admin

Hasil uji kemiripan dokumen yang dipilih akan tampil seperti pada Gambar 8. Akan tampil tabel hasil yang menyajikan judul proposal mana saja yang memiliki kemiripan dengan proposal yang diuji. tombol plus yang ada di sebelah field No. akan menampilkan kata-kata apa saja yang memiliki kemiripan terhadap proposal yang dituju, seperti pada Gambar 9.

No	Judul Proposal	Ketua	Skim	Persentase Kemiripan
1	UPAYA MENINGKATKAN LITERASI SAINTIK MAHASISWA DENGAN MENGGUNAKAN PEMBELAJARAN BERBASIS PENELITIAN	Agung Deddiliawan Ismail S.Pd., M.Pd.	Skim PPKPPT : Penelitian Peningkatan Kualitas Pembelajaran di Perguruan Tinggi	9.38 %

Kata Dasar dominan yang sama:
salah satu tuju teliti guna dekat

Abstrak Proposal:
Rendahnya minat baca yang ditunjukkan dari hasil analisis PISA dan UNESCO serta didukung dengan hasil penelitian menyebutkan bahwa minat baca serta literasi saintifik mahasiswa yang masih rendah. Hasil observasi yang telah dilaksanakan pada 149 mahasiswa angkatan 2017 Prodi Pendidikan Matematika UMM menyebutkan bahwa hanya 9 dari 149 mahasiswa yang gemar membaca. Membaca merupakan salah satu unsur dalam literasi saintifik. Berdasarkan permasalahan tersebut maka tujuan penelitian yang akan dilakukan adalah meningkatkan literasi saintifik mahasiswa dengan menggunakan pembelajaran berbasis penelitian (research). Pendekatan penelitian yang digunakan adalah pendekatan kuantitatif. Jenis penelitian yang digunakan adalah Classroom Action Research atau Penelitian Tindakan Kelas (PTK). Luaran yang diharapkan pada penelitian ini adalah artikel ilmiah yang akan dipublikasikan.

Gambar 9. Tampilan Detail Proposal Hasil Uji Plagiarisme User Admin

3.3.2 User Dosen Sebagai Reviewer

Sistem Informasi Manajemen Penelitian dan Pengabdian Kepada Masyarakat Reviewer

Daftar Proposal Pengajuan Penelitian

Filter: 2017/2018 [Proses]

#	Judul	Skim	Ketua	DE	EP	Status	Cek	Aksi
1	Politik Lingkungan: Model Kebijakan Network Governance dalam Konservasi Sumber Mata Air di Kota Batu	Penelitian Pengembangan Ipteks (P2I)	Muhammad Kamil S.I.P., MA.	500	540	DE: Sudah / EP: Sudah	Uji Kemiripan	Review / Uraiah PPP
2	Persepsi Legiun Veteran Republik Indonesia (LVRI) Terhadap Jepang Pasca Diplomasai Geisha di Indonesia (Studi Deskriptif Kuantitatif Pada Anggota LVRI Kota Malang)	Penelitian Pengembangan Ipteks (P2I)	M. Syaprin Zahidi S.I.P., M.A.	480	-	DE: Sudah	Uji Kemiripan	Review
3	Senyapan dan Kilir Lidah dalam Komunikasi Verbal Pembawa Acara dan Narasumber Acara Televisi: Kajian Psikoinguistik	Penelitian Pengembangan Ipteks (P2I)	Fida Pangesti S.Pd., M.A.	515	-	DE: Sudah	Uji Kemiripan	Review
4	Dynamics Measurement Method On English Acquistation: Proposed Method On Measuring Students' Learning Achievement	Penelitian Pengembangan Ipteks (P2I)	Adityo S.S., M.A.	450	-	DE: Sudah	Uji Kemiripan	Review
5	Penerimaan Khalayak Terhadap Tayangan Infotainment Tentang Fenomena Artis Lepas Hijab (Studi Pada Mahasiswa UMM Berdasarkan Kelompok Sosial)	Penelitian Pengembangan Ipteks (P2I)	Winda Hardyanti S.Sos., M.Si.	500	520	DE: Sudah / EP: Sudah	Uji Kemiripan	Review / Uraiah PPP

Gambar 10. Tampilan Menu Uji Plagiarisme Proposal Penelitian User Dosen

Tampilan Gambar 10 merupakan Pengujian Proposal dari reviewer sama dengan yang ada pada Admin, hanya saja Reviewer hanya dapat menguji proposal yang direview, sedangkan Admin dapat menguji semua proposal yang ada.

4. Kesimpulan

Berdasarkan implementasi yang sudah dilakukan sebelumnya, dapat disimpulkan bahwa :

1. Algoritma Biword Winnowing merupakan algoritma yang dapat digunakan untuk mencari nilai similarity dengan bantuan rumus jaccads coefficient
2. Algoritma Biword Winnowing memiliki setting nilai basis pada rolling hash dan nilai window yang dapat dirubah. perubahan dari dua nilai tersebut dapat mempengaruhi nilai akhir similarity yang akan dihasilkan.
3. Setting nilai basis = 1 tidak cocok digunakan untuk mencari nilai similarity karena nilai similarity yang dihasilkan terlalu tinggi. hal ini telah dibuktikan dengan besarnya nilai selisih yang didapat pada tahap pengujian.
4. Algoritma Biword Winnowing dapat digunakan untuk mencari nilai similarity menggunakan data Penelitian dan Pengabdian pada Sistem Informasi Penelitian dan Pengabdian UMM.
5. Hasil Pengujian menunjukkan bahwa nilai basis dan window yang terbaik untuk proses similarity adalah:
 - a. basis = 4 dan window = 2 untuk data Penelitian
 - b. basis = 2 dan window = 1 untuk data Pengabdian

Disamping itu Penelitian ini masih memiliki banyak kekurangan. sehingga diharapkan peneliti selanjutnya dapat menutupi kekurangan dari penelitian ini, diantaranya:

1. Perlu dilakukan pengujian lebih lanjut untuk nilai basis dan window diatas 8. hal ini perlu dilakukan untuk mengetahui pola yang jelas karena pada basis dan window dengan jarak nilai pengujian 1-8 masih tidak ditemukan pola yang jelas kecuali basis nilai satu yang memiliki nilai selisih tertinggi dan nilai tersebut turun seiring dengan naiknya nilai window.
2. Perlu penanganan untuk data selain yang berbahasa Indonesia, hal ini dikarenakan pada Sistem Informasi Penelitian dan Pengabdian terdapat proposal yang berbahasa selain bahasa indonesia, contohnya Bahasa Inggris.

Daftar Notasi

n	:	karakter ke n
c	:	nilai ascii dari karakter
b	:	basis
k	:	panjang string
di	:	dokumen i
dj	:	dokumen j
w(di)	:	frekuensi fingerprint pada dokumen i
w(dj)	:	frekuensi fingerprint pada dokumen j
$H_{(1,n)}$:	nilai rolling hash
Similarity(di,dj)	:	nilai kemiripan dokumen i dengan dokumen j

Referensi

- [1] Pandey A, Kaur M, Goyal P. The Menace of Egalitarianism. *Int Symp Emerg Trends Technol Libr Inf Serv.* 2015;(1):1–6.
- [2] Wibowo RK, Hastuti K. Penerapan Algoritma Winnowing Untuk Mendeteksi Kemiripan Teks pada Tugas Akhir Manusia. *TechnoCOM.* 2016;15(4):303–11.
- [3] Setiawan A. Implementasi Algoritma Winnowing Untuk Deteksi Kemiripan Judul Skripsi Studi Kasus STIMK Budidarma. *Maj Ilm INTI.* 2017;7(1):1–9.
- [4] Ulfa NF, Mustikasari M. Pembuatan Aplikasi Pengukuran Tingkat Kemiripan Dokumen Berbasis Web Menggunakan Algoritma Winnowing. *J Inform dan Komput.* 21(3):61–8.
- [5] Astutik S, Cahyani AD, Sophan MK. Sistem Penilaian Esai Otomatis Pada E-Learning Dengan Algoritma Winnowing. *J Inform [Internet].* 2014;12(2):47–52. Tersedia pada: <http://jurnal.informatika.petra.ac.id/index.php/inf/article/view/19138>
- [6] Ridho M. Rancang Bangun Aplikasi Pendeteksi Penjiplakan Dokumen Menggunakan Algoritma Biword Winnowing. *Teknik Informatika Universitas Islam Negeri SLTAN Syarif Kasim Pekanbaru Riau.* niversitas Islam Negeri SLTAN Syarif Kasim Pekanbaru Riau; 2013.

- [7] Putera Utama Siahaan A, Mesran, Rahim R, Siregar D. K-Gram As A Determinant Of Plagiarism Level In Rabin-Karp Algorithm. *Int J Sci Technol Res.* 2017;6(7):350–3.
- [8] Jarwati, Cahya Prihandoko A, Eka Yulia R W. Penerapan Algoritma Winnowing pada Sistem Rekomendasi Penentuan Dosen Pembimbing Skripsi (Studi Kasus: Prodi Sistem Informasi Universitas Jember Jember). 2016;(1):11–20.
- [9] Oktapriansyah Lubis F, Zul MI, Rachmawati H. Rancang Bangun Aplikasi Pendeteksi Kemiripan File Berbasis Web Menggunakan Jaccard Coefficient dan Algoritma Winnowing. *J Aksara Komput Terap Politek Caltex Riau.* 2016;5(1):260–7.
- [10] Rathore FA, Farooq F. Plagiarism detection softwares: Useful tools for me dical writers and editors. *Learn Res [Internet].* 2018;64(11):1329–31. Tersedia pada: <http://www.jpma.org.pk/PdfDownload/7088.pdf>

