

Akurasi dalam Mengidentifikasi Tingkat Similarity pada Artikel Ilmiah Menggunakan Algoritma Jaro Winkler

Firman Santosa¹✉

¹Sekolah Tinggi Keguruan dan Ilmu Pendidikan (STKIP) Rokania

firmansnts@gmail.com

Abstract

Plagiarism is an issue that often develops and always occurs, especially in universities. Sekolah Tinggi Keguruan dan Ilmu Pendidikan (STKIP) Rokania already has a scientific article recording system called E-Jurnal which is always used by lecturers to publish journals and search for relevant topics and literature. In making scientific articles, plagiarism is often not detected in scientific articles submitted by lecturers. This fraud is carried out by combining the available abstracts to form one abstract resulting from the merger. This means that the abstract is not the result made by the researcher himself. The biggest problem is when this fraudulent act is not detected, which is due to manual checking of documents. This of course can result in lowering the reputation of the accredited E-Journal. Of course, this problem must be immediately given the right solution to identify the level of similarity in scientific articles that already exist. Identification of the level of similarity in scientific articles is made through a structured development stage using the Jaro Winkler algorithm which is chosen to detect the similarity of abstract documents of scientific articles with abstracts that have been stored in the E-Jurnal database. The system will display the percentage level of similarity of the abstract of the scientific article so that the journal admin makes the right decision when accepting the scientific article or rejecting it. Through this research, lecturers can do an initial check of abstracts from prospective scientific articles to minimize plagiarism. Thus, it can minimize the actions of lecturers' fraud in making scientific articles and produce high-quality journals of higher value.

Keywords: Algorithm Jaro Winkler, Article, Plagiarism, Similarity, Open Journal System (OJS).

Abstrak

Plagiarisme merupakan isu yang kerap berkembang dan selalu terjadi khususnya di Perguruan Tinggi. Sekolah Tinggi Keguruan dan Ilmu Pendidikan (STKIP) Rokania telah memiliki sistem pencatatan artikel ilmiah yang bernama E-Jurnal yang selalu digunakan oleh dosen untuk mempublikasikan jurnal dan mencari topik dan literatur yang relevan. Dalam pembuatan artikel ilmiah, sering sekali tidak terdeteksinya tindakan plagiarisme terhadap artikel ilmiah yang diajukan oleh dosen. Kecurangan ini dilakukan dengan mengkombinasikan abstrak-abstrak yang telah tersedia sehingga membentuk satu abstrak hasil penggabungan. Artinya abstrak bukanlah hasil yang dibuat sendiri oleh peneliti. Permasalahan terbesarnya adalah ketika tidak terdeteksinya tindakan kecurangan ini yang disebabkan pengecekan dokumen dilakukan secara manual. Hal ini tentu saja dapat mengakibatkan menurunkan reputasi E-Jurnal yang sudah terakreditasi tersebut. Permasalahan ini tentu saja harus segera diberikan solusi yang tepat untuk mengidentifikasi tingkat similarity pada artikel ilmiah yang sudah ada sebelumnya. Identifikasi tingkat similarity pada artikel ilmiah dibuat melalui tahapan pengembangan terstruktur menggunakan algoritma Jaro Winkler yang dipilih untuk mendeteksi similaritas dokumen abstrak artikel ilmiah dengan abstrak-abstrak yang telah tersimpan pada basis data E-Jurnal. Sistem akan menampilkan berapa persen tingkat kesamaan abstrak artikel ilmiah tersebut sehingga admin jurnal memberikan keputusan yang tepat saat menerima artikel ilmiah tersebut atau menolaknya. Melalui penelitian ini, dosen dapat melakukan pengecekan awal abstrak dari calon artikel ilmiahnya untuk meminimalkan terjadinya plagiasi. Dengan demikian dapat meminimalisir tindakan kecurangan dosen dalam pembuatan artikel ilmiah dan menghasilkan jurnal yang berkualitas bernilai lebih tinggi.

Kata kunci: Algoritma Jaro Winkler, Artikel, Plagiarism, Similarity, Open Journal System (OJS).

JIdT is licensed under a Creative Commons 4.0 International License.



1. Pendahuluan

Perkembangan teknologi informasi yang semakin pesat membawa dampak baik dan buruk bagi kehidupan manusia, dampak buruk yang ditimbulkan salah satunya adalah masalah plagiarisme [1]. Tindakan meniru karya ilmiah penulis lain tanpa mencantumkan sumber rujukan atau dengan kata lain mengakui karya tersebut merupakan buah pemikiran sendiri telah berlangsung lama di berbagai perguruan tinggi baik negeri maupun swasta di Indonesia [2]. Pelaku

plagiarisme bertindak dengan mengambil dan menuliskan karya penulis lain pada karya yang diakui sebagai miliknya tanpa menuliskan asal dari tulisan yang ditulis tersebut. Kesamaan suatu dokumen dengan dokumen lainnya atau suatu *text* dengan *text* lainnya dapat diketahui dengan menggunakan pendekatan *string metric* yaitu dengan melakukan perbandingan dua *string*. Kedua *string* tersebut dimasukkan ke dalam fungsi matematis tertentu untuk mengetahui jarak antara keduanya [3]. Menghitung jumlah kemiripan kata adalah tugas umum, namun memberikan peran yang sangat penting dalam

berbagai aplikasi *Natural Language Processing* (NLP), seperti mesin pencari, detektor plagiarisme, sistem penjawab pertanyaan, dan lainnya [4]. Tentu saja menghitung jumlah kemiripan kata sangat bisa digunakan untuk mendeteksi tingkat plagiat pada artikel ilmiah, mengingat karya ilmiah merupakan hasil pemikiran seseorang yang dituangkan dalam bentuk teks. Algoritma yang dapat digunakan untuk mencari kesamaan string dan kemiripan kata yaitu *Jaro-Winkler*, *Hamming*, *Damerau Levenshtein*, dan *Levenshtein Distance*. Algoritma *Jaro-Winkler* dipilih karena sangat cocok dan sesuai dalam memberikan hasil terbaik pada pencocokan dua *string* pendek. Penggunaan algoritma *Jaro-Winkler* telah dilakukan oleh peneliti sebelumnya yaitu, koreksi otomatis dan saran ejaan, fitur, mendeteksi dokumen plagiarisme, *stemming* kata kompensasi tidak standar bahasa Inggris [5]. Maka dari itu pemerintah, dosen pembimbing, dan pengelola jurnal selalu cermat dan berhati-hati dalam menerima dan memeriksa suatu karya ilmiah, bahkan pemerintah mengeluarkan kebijakan untuk setiap karya ilmiah yang diunggah ke *repository* kampus wajib dicek plagiasinya [6].

Meninjau apa yang disebut jurnal predator mungkin membuang-buang waktu dan tenaga. Sebagai salah satu dari banyak kekurangan, jurnal predator tidak menjamin pengarsipan dan akses jangka panjang ke isinya. Untuk menyoroti beberapa masalah jurnal predator, seorang peneliti mengirimkan manuskrip yang cacat secara sengaja dan pengulas melihat permasalahan pada karya ilmiah tersebut, sementara editor menerima karya ilmiah itu. Sama seperti penerbitan penelitian di jurnal predator cenderung membuang-buang sumber daya, meninjau untuk outlet ini harus dilihat dalam cahaya kritis yang sama, dan upaya diperlukan untuk menghentikan praktik ini [7]. Seseorang mengutip atau menyalin sebagian atau semua hasil karyanya sendiri yang sudah dipublikasikan. *Self plagiarism* dapat dihindari atau diatasi dengan cara menulis ulang karya tersebut dengan konsep yang baru [8].

Sejak diberlakukannya Surat Edaran Dirjen-DIKTI No.2050/E/T/2011 tentang Kebijakan Unggah Karya Ilmiah dan Jurnal; Surat Edaran Dirjen-DIKTI No. 152/E/T/2012 tentang Publikasi Karya Ilmiah S1 di Jurnal Nasional, S2 di Jurnal Nasional Terakreditasi, dan S3 di Jurnal Internasional. Kewajiban publikasi yang sudah diatur bagi dosen dan peneliti diharapkan dapat mendorong peningkatan jumlah dan mutu publikasi ilmiah pada tingkat nasional dan internasional sehingga Indonesia mampu bersaing dengan bangsa-bangsa lain. Kemudian diikuti oleh Surat Edaran Dirjen-DIKTI No. 212/E/T/2012 tentang Panduan Pengelolaan Jurnal Terbitan Berkala Ilmiah Elektronik; dan Surat Edaran Dirjen Risbang-DIKTI No.93/E/SE/XII/2015 tentang Akreditasi Jurnal Ilmiah Secara Elektronik, pemerintah menetapkan kebijakan bahwa mulai 1 April 2016, akreditasi jurnal ilmiah nasional hanya berlaku untuk jurnal yang telah

dikelola secara elektronik dengan *platform Open Journal System* (OJS) agar proses penilaian lebih mudah, cepat, akurat dan transparan. Jurnal yang telah diterbitkan secara elektronik kemudian didaftarkan pada database Akreditasi Jurnal Nasional (ARJUNA). Adanya kebijakan tersebut, maka setelah 1 April 2016, para pengelola/penerbit jurnal di Indonesia harus mematuhi segala persyaratan/ketentuan dalam penerbitan e-journal agar terbitannya dapat terakreditasi secara nasional [9].

Menurut pandangan pendapat tokoh ahli di atas, bisa ditarik kesimpulan adalah hasil sebuah artikel ilmiah yang berbentuk sebuah pemaparan tulisan hasil dari sebuah penelitian yaitu membahas suatu permasalahan dalam suatu bidang tertentu dengan menggunakan kaidah-kaidah yang berlaku. Akan tetapi seiring dengan berjalannya waktu perkembangan informasi serta teknologi, tak hanya memberikan efek positif tetapi memberikan efek negatif. Hal ini diakibatkan banyaknya bertebaran sebuah informasi yang sering dimanfaatkan sebagai rujukan di dalam pembuatan artikel ilmiah. Sehingga tak jarang ditemukan kecurangan seperti meniru ide orang lain maupun menjiplak artikel ilmiah yang juga diistilahkan dengan *plagiarisme* [10].

Atas dasar itu maka perlu dibuat suatu sistem yang mampu mendeteksi tingkat *similarity* pada artikel ilmiah di kampus STKIP Rokania. Penelitian ini diharapkan mampu membangun sistem pendeteksi tingkat *similarity* pada artikel ilmiah berbasis web terintegrasi dengan E-Jurnal STKIP Rokania dan mengetahui tingkat akurasi kegiatan plagiarisme pada dokumen artikel ilmiah. Hasil dari penelitian ini diharapkan dapat digunakan sebagai tambahan pengetahuan bagi pembaca atau bagi peneliti lain, serta dapat menjadi referensi untuk penelitian sejenis.

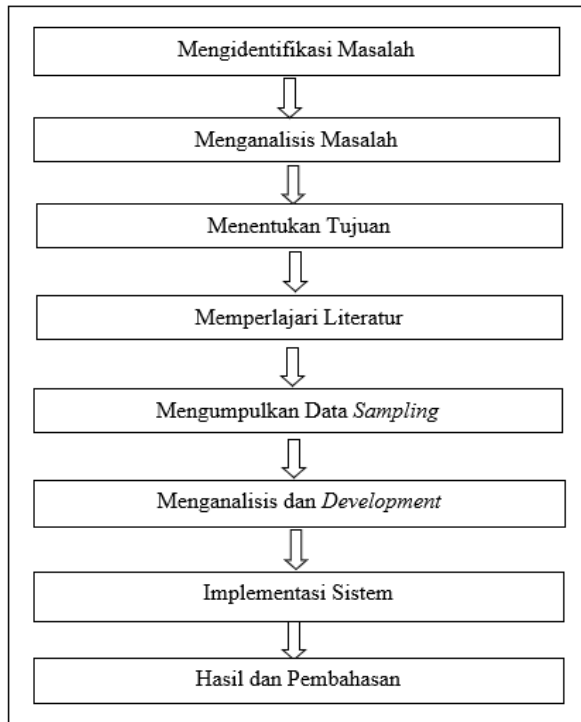
2. Metodologi Penelitian

Metodologi penelitian ini juga digunakan untuk mengidentifikasi permasalahan yang ditemukan, kemudian membuat analisa dari permasalahan tersebut yang pada akhirnya akan menemukan penyelesaian masalah. Dalam penelitian ini diperlukan suatu penyelesaian permasalahan dalam mendeteksi tingkat plagiarisme pada artikel ilmiah yang akan dipublikasikan di E-Jurnal Sekolah Tinggi Keguruan dan Ilmu Pendidikan Rokania.

2.1. Kerangka Kerja

Dalam metodologi penelitian ini menggunakan urutan kerangka kerja yang harus dipatuhi. Beberapa tahapan yang terdapat dalam kerangka kerja ini adalah identifikasi masalah, analisis masalah, tujuan penelitian, pengumpulan data, analisa dan implementasi metode, pengujian data dan pengambilan keputusan. Dengan mengikuti kepada tiap-tiap tahapan yang terdapat pada kerangka kerja dalam penelitian ini yang diharapkan hasil akhir berupa pengambilan keputusan yang sesuai dengan

tujuan. Berikut adalah Kerangka Kerja yang digunakan pada penelitian ini.



Gambar 1. Kerangka Kerja

Berdasarkan Gambar 1 di atas, maka dapat dijelaskan tahapan-tahapan kerangka kerja sebagai berikut:

- a. Mengidentifikasi masalah
Masalah yang akan diidentifikasi dalam penelitian ini adalah belum adanya system pendeteksi tingkat plagiarisme untuk artikel ilmiah pada E-Journal STKIP Rokania.
- b. Menganalisa Masalah
Analisis masalah dalam penelitian ini dilakukan dengan cara memahami alur proses jurnal yang sudah publish sebelumnya dengan melakukan wawancara langsung kepada Kepala Unit Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) STKIP Rokania.
- c. Menentukan Tujuan
Setelah melakukan analisis dalam menentukan masalah, tahap berikutnya akan menentukan tujuan yang akan dicapai dari penelitian ini. Hal ini dimaksudkan agar hasil yang diinginkan sesuai dengan yang diharapkan. Yaitu menyediakan wadah pengecekan tingkat plagiarisme pada artikel ilmiah secara *online* sebelum dipublikasikan ke *Open Journal System*.
- d. Mempelajari Literatur
Mempelajari tentang literatur yang akan digunakan sebagai bahan referensi dalam penelitian ini. Adapun literatur yang digunakan berasal dari jurnal-jurnal ilmiah yang berkaitan dengan *Algorithm Jaro-Winkler* dan *Plagiarism Checker*. Literatur-literatur ini akan menjadi acuan dalam

melakukan penelitian ini dengan tujuan membantu dan memudahkan proses penelitian.

- e. Mengumpulkan Data *Sampling*
Data *Sampling* ditarik langsung dari *Open Journal System* (OJS) Sekolah Tinggi Keguruan dan Ilmu Pendidikan (STKIP) Rokania berbentuk JSON. Kemudian seluruh data *sampling* akan disimpan ke database yang baru sebagai data perbandingan terhadap teks yang ditambahkan untuk pengujian nantinya.
- f. Menganalisis dan *Development*
Pada tahapan ini, peneliti akan menganalisis sistem yang baru untuk mendapatkan solusi dari permasalahan yang ditemukan. Tahapan analisis ini meliputi pembuatan *Data Flow Diagram*, *Entity Relationship Diagram*, Rancangan Tabel, Rancangan *Input* dan Rancangan *Output*. Kemudian akan dilanjutkan dengan pengkodean menggunakan bahasa pemrograman PHP dengan memanfaatkan *framework CodeIgniter* berdasarkan rancangan yang sudah dibuat sebelumnya.
- g. Implementasi Sistem
Setelah dilakukan tahapan analisis dan *development*, langkah berikutnya adalah implementasi sistem yang selesai dibuat, yaitu dengan mengunggah sistem ke dalam hosting dan mengkaitkannya dengan domain. Sehingga sistem dapat diakses oleh pengguna secara langsung dan digunakan untuk mendeteksi tingkat plagiarisme pada artikel ilmiah yang akan mereka ajukan di E-Journal STKIP Rokania.
- h. Menguji Hasil
Setelah system diimplementasikan, tahapan selanjutnya adalah melakukan pengujian terhadap hasil, apakah sudah sesuai dengan yang diharapkan dan dirancang sebelumnya. Pengujian ini menggunakan *Black Box Testing*.

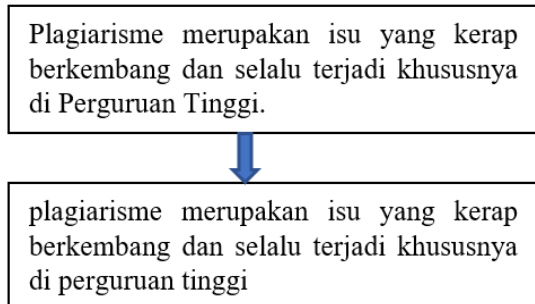
2.2. Text Pre-processing

Text Pre-processing adalah suatu proses untuk menyeleksi data *text* agar menjadi lebih. Tidak ada aturan pasti tentang setiap tahapan dalam *Text Pre-processing*. Semua itu tergantung dengan jenis serta kondisi data yang dimiliki [11]. *Text Pre-processing* merupakan salah satu implementasi dari text mining. *Text mining* sendiri adalah suatu kegiatan menambang data, dimana data yang biasanya diambil berupa *text* yang bersumber dari dokumen-dokumen yang memiliki *goals* untuk mencari kata kunci yang mewakili dari sekumpulan dokumen tersebut sehingga nantinya dapat dilakukan analisa hubungan antara dokumen-dokumen tersebut [12]. *Text Pre-processing* mempersiapkan teks yang tidak terstruktur menjadi data yang baik dan siap untuk diolah. *Text Pre-processing* memiliki tahapan-tahapan yang meliputi proses *case folding*, *cleaning*, tokenisasi, *stopword removal* dan *stemming* [13].

Tahapan *Text Preprocessing* bertujuan mempersiapkan teks menjadi data yang akan diolah sebelum masuk ke tahapan Perhitungan Kemiripan Data. Sekumpulan karakter dalam teks dipecah-pecah menjadi unsur yang lebih berarti [14]. *Text Preprocessing* pada penelitian ini melewati tiga proses, yaitu:

a. Case Folding

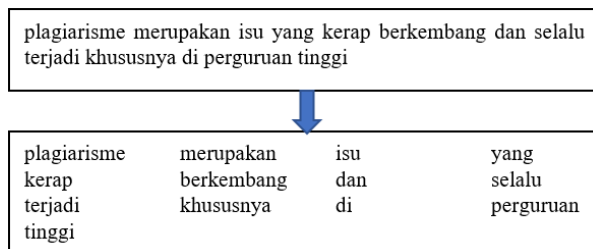
Tahapan ini merupakan proses mengubah teks menjadi huruf kecil, yaitu hanya huruf “a” sampai dengan “z” dan angka yang diterima [15]. Kata yang tidak mengandung huruf dan angka akan dihapus. Tahapan ini disajikan pada Gambar 2.



Gambar 2. Tahapan *Case Folding* pada teks

b. Tokenizing

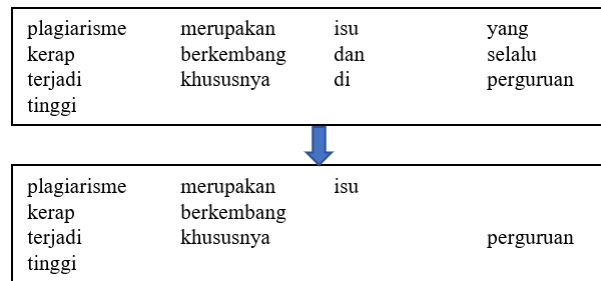
Tahapan *Tokenizing* merupakan proses pemotongan teks atau kalimat berdasarkan kata-kata penyusunannya [16]. Tahapan ini disajikan pada Gambar 3.



Gambar 3. Tahapan *Tokenizing* pada teks

c. Stopwords Removal

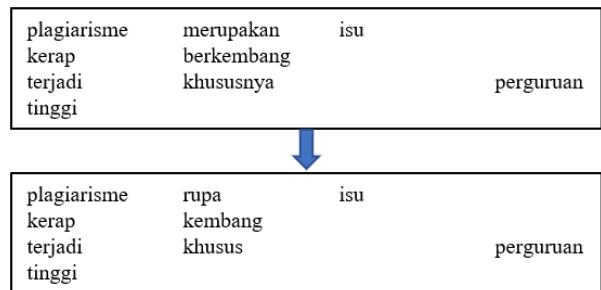
Pada tahapan *Stopwords Removal* mengeliminasi kata-kata yang dianggap tidak memiliki relevansi yang erat dengan isi dari teks yang diuji seperti kata: “dan”, “yang” dan “supaya” [17]. Gambar 4.4 menunjukkan contoh *Stopwords Removal*. Tahapan ini disajikan pada Gambar 4.



Gambar 4. Contoh *Stopwords Removal* pada teks

d. Stemming

Tahapan *Stemming* menghasilkan kata dasar dari setiap kata hasil proses *Tokenizing* dan *Stopwords Removal* [18]. Pada tahapan ini dilakukan proses pengembalian berbagai bentuk kata ke dalam suatu representasi yang sama. Pemeriksaan semua kemungkinan bentuk kata diasumsikan memiliki 2 awalan (prefix) dan 3 akhiran (sufiks). Pada setiap pemotongan diikuti dengan pemeriksaan pada kamus untuk memastikan hasil pemotongan itu sudah berada dalam bentuk dasar. Bila pemeriksaan berhasil maka proses dinyatakan selesai dan tidak perlu melanjutkan proses pemotongan imbuhan selanjutnya. Tahapan ini disajikan pada Gambar 5.



Gambar 5. Contoh *Stemming* pada teks

2.3. Perhitungan Kemiripan Data

Setelah tahapan *Text Preprocessing* dilakukan, tahapan selanjutnya adalah menghitung tingkat kemiripan data. Algoritma *Jaro-Winkler* adalah algoritma untuk mengukur kesamaan antara dua string, biasanya algoritma ini digunakan untuk mendeteksi duplikat [19]. Algoritma *Jaro Winkler* merupakan varian dari *Jaro Distance*, yaitu sebuah algoritma untuk mengukur kesamaan antara dua string [20]. Algoritma ini dapat digunakan pada pendeteksian duplikasi data. Semakin tinggi nilai *Jaro Winkler distance* untuk dua string, maka semakin mirip kedua string tersebut.

Saat membandingkan kata SAYUR dengan UAYSR, bisa dilihat seksama maka dapat dikatakan semua karakter yang ada di *s1* ada di *s2* dan sama dengan karakter yang ada di *s2*, tetapi dengan urutan yang berbeda. Dengan mengganti S dan U, dapat dilihat perubahan kata SAYUR menjadi UAYSR, pertukaran dua elemen string inilah contoh nyata dari transposisi yang dijelaskan.

Jaro Winkler Distance menggunakan *prefix scale* (p) yang memberikan tingkat penilaian yang lebih, dan *prefix length* (l) yang menyatakan panjang awalan yaitu panjang karakter yang sama dari *string* yang dibandingkan sampai ditemukannya ketidaksamaan. Maka *Jaro Winkler Distance* nya (d_w) disajikan pada Persamaan (1).

$$d_w = d_j + (l_p(1 - d_w)) \quad (1)$$

Dimana d_j merupakan *Jaro Distance* untuk *string* S1 dan S2. l adalah panjang *prefix* umum di awal *string* nilai maksimalnya 4 karakter (Panjang karakter yang sama sebelum ditemukan ketidaksamaan \max 4). p adalah konstanta *scalling factor*. Nilai standar untuk konstanta ini menurut *Winkler* adalah 0,1.

Dalam beberapa implementasi dari *Jaro Winkler*, *prefix* $l_p(1 - d_j)$ hanya ditambahkan bila perbandingan *string* *Jaro Distance* di bawah “*boost threshold*” atau ambang batas. Ambang batas atas dari implementasi *Winkler* adalah 0,7. Berikut ini adalah contoh perhitungan *Jaro Winkler Distance*. Jika *string*

$S1 = \text{FIRMAN}$ dan
 $S2 = \text{FIRAMN}$ maka:

$$m = 6$$
$$s1 = 6$$
$$s_2 = 6$$

karakter yang tertukar hanyalah M dan A, maka

$$t = 1$$

maka nilai *Jaro Distance* disajikan pada Persamaan (2).

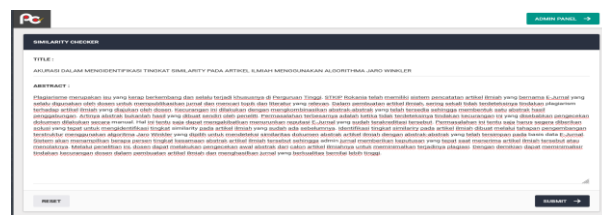
$$d_j = \frac{1}{3} \times \binom{6}{6} + \binom{6}{6} \left(\frac{6-1}{6} \right) = \mathbf{0,9444} \quad (2)$$

Kemudian bila diperhatikan susunan $s1$ dan $s2$ dapat diketahui nilai $1 = 3$, dan dengan nilai konstan $p = 0,1$ maka nilai *Jaro Winkler distance* disajikan pada Persamaan (3).

$$d_w = 0,9444 + (3 \times 0,1(1 - 0,9444)) = 0,961 \quad (3)$$

3. Hasil dan Pembahasan

Aplikasi pendeteksi tingkat *similarity* yang dibuat dibangun menggunakan bahasa pemrograman PHP dan database MySQL. Gambar 5 menunjukkan halaman antarmuka utama untuk mengecek data.



Gambar 6. Halaman Antarmuka

Setelah melakukan submit judul dan abstrak, sistem akan mengecek otomatis tingkat *similarity* data yang

disubmit berdasarkan data yang ada di dalam OJS dengan memanfaatkan algoritma *Jaro Winkler*.

Berdasarkan hasil pengecekan terhadap 67 data jurnal yang tersubmit pada OJS STKIP Rokania, didapatkan hasil dengan nilai tertinggi pada *record* data ke 52 sebagaimana dapat dilihat pada Gambar 7.

50.	PENDIDIKAN KARAKTER MELALUI MENDONGENI DI SD NEGERI 003 RAMBAH	0.80809206639962	80.81%	DETAIL →
51.	<p>Penguatan Kapasitas SDM Lembaga Pendidikan Dalam Sekolah Percontohan Inkusip Berpola Asrama Anak 3T Sman Plus Satap 1 Wasur Merauke Terhadap Penanganan Kedaruratan Adiksi Napza Melalui Perencanaan Pedoman Program Pemulihan Holistik-Komprehensif Berbasis: Penguatan Kapasitas SDM Lembaga Pendidikan Dalam Sekolah Percontohan Inkusip Berpola Asrama Anak 3T Sman Plus Satap 1 Wasur Merauke Terhadap Penanganan Kedaruratan Adiksi Napza Melalui Perencanaan Pedoman Program Pemulihan Holistik-Komprehensif Berbasis Kearifan Lokal</p>	0.83924738161258	83.92%	DETAIL →
52.	PEMANFAATAN HURUF SEBAGAI MEDIA PEMBELAJARAN PENGENALAN KUNIZ DI TK DIAN PRATAMA, JAKARTA TIMUR	0.82154867327661	82.15%	DETAIL →

Gambar 7. Hasil Pengecekan Data

Pada Gambar 7 dapat dilihat bahwa nilai kemiripan tertinggi pada seluruh jurnal yang di uji yaitu 0.83924738161258 dengan tingkat kemiripan adalah 83.92%.

Detail dari setiap hasil akan menampilkan bagaimana proses *Text Processing* dilakukan dan menghasilkan berapa nilai *Jaro Winkler* per masing-masing kata yang disajikan pada Gambar 8.

[illegible]

Gambar 8. Tampilan *Text Processing*

Selanjutnya sistem juga akan menampilkan *detail* dari hasil perhitungan *Jaro Winkler* per masing-masing kata sebagaimana dapat dilihat pada Gambar 9.

[illegible]

Gambar 8. *Detail Nilai Jaro Winkler Per-kata*

Hasil akhir sistem akan menampilkan *Final Result* yang merupakan nilai akhir dari deteksi tingkat *similarity* melalui data yang disubmit.



Gambar 9. Final Result

Gambar 9 menampilkan nilai akhir berdasarkan nilai tertinggi yang dihasilkan dari perhitungan *Jaro Winkler* yaitu 83.92 %.

4. Kesimpulan

Sebagaimana yang telah ditampilkan, algoritma *Jaro Winkler* mampu mendeteksi tingkat *similarity* satu jurnal dengan jurnal yang sudah ada pada *Open Journal System* (OJS) STKIP Rokania. Setiap data yang disubmit akan dikalkulasi berdasarkan data uji yang didapatkan dari API OJS STKIP Rokania. Semakin banyak data yang ada pada OJS STKIP Rokania, maka data uji mengikuti jumlah data tersebut. Hasil akhir yang dikeluarkan oleh sistem dapat dijadikan acuan untuk menentukan tingkat kesamaan data terhadap data yang diuji dengan menentukan batas nilai maksimal data tersebut dianggap mirip.

Daftar Rujukan

- [1] Nurdin, N., Rizal, R., & Rizwan, R. (2019). Pendeteksian Dokumen Plagiarisme dengan Menggunakan Metode Weight Tree. *Telematika*, 12(1), 31. <https://doi.org/10.35671/telematika.v12i1.775>
- [2] Swari, M. H. P., Putra, C. A., & Handika, I. P. S. (2021). Plagiarism Checker pada Sistem Manajemen Data Tugas Akhir. *Jurnal Sains dan Informatika*, 7(2), 192–201. <https://doi.org/10.34128/jsi.v7i2.338>
- [3] Rouf, H. A., Wijayanto, A., & Aziz, A. (2020). Deteksi Plagiarisme Skripsi Mahasiswa dengan Metode Single-link Clustering dan Jaro-Winkler Distance. *JURNAL PILAR TEKNOLOGI: Jurnal Ilmiah Ilmu Teknik*, 5(1), 26–31. <https://doi.org/10.33319/piltek.v5i1.50>
- [4] Vatolin, A. S., Smirnova, E. Y., & Shkarin, S. S. (2021). Russian News Similarity Detection with SBERT: Pre-training and fine-tuning. *Komp'juternaja Lingvistika i Intellektual'nye Tehnologii*, 2021-June(20), 692–697. <https://doi.org/10.28995/2075-7182-2021-20-692-697>
- [5] Yulianto, M. A., & Nurhasanah, N. (2021). The Hybrid of Jaro-Winkler and Rabin-Karp Algorithm in Detecting Indonesian Text Similarity. *Jurnal Online Informatika*, 6(1), 88. <https://doi.org/10.15575/join.v6i1.640>
- [6] Syaharuddin, S., Mandailina, V., Saddam, S., Perwira Negara, H. R., Satriawan, R., & Ibrahim, M. (2021). Minimalisasi Tingkat Plagiat Karya Ilmiah Mahasiswa Melalui Workshop Software Plagiarism Checker X. *Bakti Cendana*, 4(1), 62–67. <https://doi.org/10.32938/bc.v4i1.902>
- [7] Severin, A., Strinzel, M., Egger, M., Domingo, M., & Barros, T. (2021). Characteristics of scholars who review for predatory and legitimate journals: Linkage study of Cabells Scholarly Analytics and Publons data. *BMJ Open*, 11(7). <https://doi.org/10.1136/bmjopen-2021-050270>
- [8] Zahroh, F., & Wahyudin, N. (2020). Peran Open Access Institutional Repository (Oair) Dalam Mengurangi Tingkat Plagiarisme Pada Karya. *BIBLIOTIKA: Jurnal Kajian Perpustakaan dan Informasi*, 4(1), 121–129. <http://journal2.um.ac.id/index.php/bibliotika%0APERAN>
- [9] Nashihuddin, W. (2020). Mediatisasi Kebijakan Penerbitan E-Journal dengan Open Journal System di Indonesia. *Jurnal Ilmu Perpustakaan dan Informasi*, 4(1). <https://doi.org/10.29240/tik.v4i1.1176>
- [10] Rosalia, R., & Fuad, A. J. (2019). Peran dosen dalam meminimalisasi perilaku plagiasi mahasiswa. *Indonesian Journal of Islamic Education Studies (IJIES)*, 2(1), 61–77. <https://doi.org/10.33367/ijies.v2i1.882>
- [11] Wulandari, D., Indriati, & Dewi, C. (2019). Analisis Sentimen pada Ulasan “Lazada” Berbahasa Indonesia Menggunakan BM25 dan K-Nearest Neighbor (K-NN) dengan Perbaikan. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3(5), 5017–5023.
- [12] Khairunnisa, S., Adiwijaya, A., & Al Faraby, S. (2021). Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19). *Jurnal Media Informatika Budidarma*, 5(2), 406–414. <http://dx.doi.org/10.30865/mib.v5i2.2835>
- [13] Agustina, D. A., Subanti, S., & Zukhronah, E. (2021). Implementasi Text Mining Pada Analisis Sentimen Pengguna Twitter Terhadap Marketplace di Indonesia Menggunakan Algoritma Support Vector Machine. *Indonesian Journal of Applied Statistics*, 3(2), 109–122. <https://doi.org/10.13057/ijas.v3i2.44337>
- [14] Hanani, H. N., Jayadianti, H., & Rustamaji, H. C. (2021). Fuzzy String Matching for Semi-Automation of Words with Jaro Winkler Distance Algorithm on Microsoft Word Documents. *Seminar Nasional Informatika (SEMNASIF)*, 18(1), 1–10. <http://jurnal.upnyk.ac.id/index.php/semnasif/article/view/6067>
- [15] Hickman, L., Thapa, S., Tay, L., Cao, M., & Srinivasan, P. (2022). Text preprocessing for text mining in organizational research: Review and recommendations. *Organizational Research Methods*, 25(1), 114–146. <https://doi.org/10.1177/1094428120971683>
- [16] Nofiyani, N., & Wulandari, W. (2022). Implementasi Electronic Data Processing Untuk meningkatkan Efektifitas dan Efisiensi Pada Text Mining. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 6(3), 1621–1629. <http://dx.doi.org/10.30865/mib.v6i3.4332>
- [17] Rozi, I. F., & Putri, I. K. (2022). Peringkasan Teks Otomatis pada Portal Berita Olahraga menggunakan metode Maximum Marginal Relevance. *Jurnal Informatika Polinema*, 8(3), 21–30. <https://doi.org/10.33795/jip.v8i3.519>
- [18] Dalimunthe, K., & Hayadi, B. H. (2022). Information Text Retrieval Untuk Pencarian Data Penilaian Mengacu Pada Saran Dari Pengunjung Menggunakan Vector Space Modelimplementasi. *Journal Computer Science and Information Technology (JCoInt)*, 3(2), 73–79. <http://jurnal.ulb.ac.id/index.php/JCoInt/index>
- [19] Ikasari, D., & Widiastuti, W. (2021, January). Sentiment Analysis Review Novel “Goodreads” Berbahasa Indonesia Menggunakan Naïve Bayes Classifier. In *Semnas Ristek (Seminar Nasional Riset dan Inovasi Teknologi)* (Vol. 5, No. 1). <https://doi.org/10.30998/semnasristek.v5i1.5040>
- [20] Sanjaya, A. (2020). Optimasi Pencarian Data Menggunakan Text Filtering dan Algoritma Jaro Winkler. *Network Engineering Research Operation*, 5(1), 24–29. <http://dx.doi.org/10.21107/nero.v5i1.149>