

Implementasi Perbaikan Ejaan Pada Web Semantik Ruang Baca Fakultas Teknik Dengan Menggunakan Algoritma Jaro-Winkler Distance

Fahrim Irhamna Rachman^{*1}, Nur Alam², Titin Wahyuni³, Lukman Anas⁴

Universitas Muhammadiyah Makassar

fachrim141020@unismuh.ac.id, alam.nur@student.unismuh.ac.id, titinwahyuni@unismuh.ac.id,
lukmananas@unismuh.ac.id

Abstrak - Penelitian ini bertujuan untuk meningkatkan efisiensi dan akurasi sistem pencarian judul skripsi di Ruang Baca Fakultas Teknik dengan menerapkan algoritma Jaro-Winkler Distance. Algoritma ini digunakan untuk mengukur tingkat kesamaan antara dua string atau kata, sehingga sangat cocok untuk digunakan dalam pencarian judul skripsi yang dimasukkan secara tidak lengkap atau salah eja. Sistem yang dikembangkan juga mengintegrasikan proses perbaikan ejaan dan pencarian sinonim untuk meningkatkan hasil pencarian, yang membuatnya lebih tanggap terhadap variasi dalam pengetikan pengguna. Dalam penelitian ini, rata-rata waktu proses pencarian dengan menggunakan judul skripsi yang tidak lengkap adalah 9926,75 ms per kata, termasuk proses perbaikan ejaan dan pencarian sinonim yang dilakukan oleh algoritma Jaro-Winkler Distance. Hasil penelitian ini menunjukkan bahwa penerapan algoritma ini secara signifikan efektif dalam menyelesaikan masalah kesalahan ejaan, sekaligus meningkatkan akurasi pencarian judul skripsi. Dengan demikian, sistem pencarian yang dikembangkan mampu mempermudah pengguna dalam menemukan judul yang mereka cari, meskipun kata kunci yang dimasukkan tidak lengkap atau kurang tepat.

Kata Kunci : perbaikan ejaan; web semantik; algoritma jaro-winkler distance; jaro-distance; jaro-winkler

Abstract - This study aims to improve the efficiency and accuracy of the thesis title search system in the Faculty of Engineering Reading Room by implementing the Jaro-Winkler Distance algorithm. This algorithm is used to measure the similarity between two strings or words, making it highly suitable for use in searching thesis titles that are entered incompletely or with misspellings. The system developed also integrates spelling correction and synonym search processes to enhance search results, making it more responsive to variations in user input. In this study, the average search process time using incomplete thesis titles was 9926.75 ms per word, including the spelling correction and synonym search processes performed by the Jaro-Winkler Distance algorithm. The results of this study indicate that the application of this algorithm is significantly effective in addressing spelling errors while also improving the accuracy of thesis title searches. Thus, the search system developed can facilitate users in finding the titles they are looking for, even when the keywords entered are incomplete or inaccurate. *Keywords: spelling correction; semantic web; Jaro-Winkler Distance algorithm; Jaro-Distance; Jaro-Winkler.*

I. PENDAHULUAN

Ruang Baca Fakultas Teknik merupakan salah satu fasilitas yang Fakultas Teknik Universitas Muhammadiyah Makassar berikan kepada mahasiswa dan civitas akademik berupa layanan peminjaman, pengembalian, berbagai properti perpustakaan seperti skripsi, laporan KKP, buku dan referensi lainnya yang menunjang proses belajar mengajar fakultas teknik.

Berdasarkan penelitian yang dilakukan oleh (Haris, 2022) yang berjudul "Penerapan Web Semantik untuk pencarian judul skripsi pada ruang baca fakultas teknik Universitas Muhammadiyah Makassar", telah melakukan

mengimplementasikan pencarian judul skripsi dengan penerapan semantik pada ruang baca sebagai basis pengetahuannya yang dapat membantu mahasiswa untuk menemukan judul skripsi yang diinginkan yang terdapat pada Ruang Baca Fakultas Teknik khususnya.

Pada penelitian sebelumnya saat memasukkan kata yang akan dicari, terkadang terjadi *Typographical error*. *Typographical error* adalah kesalahan yang dibuat dalam proses pengetikan seperti ejaan atau meninggalkan suatu kata. Salah satu kesalahan yang paling umum adalah transposisi huruf, di mana semua huruf yang diperlukan untuk mengeja kata yang hadir, tetapi dalam urutan yang salah (Yoga

Pratama S, 2017). Oleh karena itu diperlukan pendekatan yang dapat memberikan alternatif kata dari informasi yang dimasukkan oleh pengguna, yaitu *spelling checker*. *Spelling checker* adalah proses memeriksa kata yang salah eja dan memberikan kandidat yang benar (Soleh & Purwarianti, 2011).

Web semantik menurut (Berners-Lee et al., 2001) “*the semantic web is not a separate web but a n extension of the current one in which information is given well-defined meaning, better enabling computers and people to work in*”. Sedangkan algoritma *jaro-winkler distance* adalah varian dari *jaro distance* metrik yaitu sebuah algoritma untuk mengukur kesamaan antara dua *string*, biasanya algoritma ini digunakan di dalam pendeteksi dupikat. Semakin tinggi *jaro-winkler distance* untuk dua *string*, semakin mirip dengan *string* tersebut (Kurniawati et al., 2010).

Berdasarkan hasil penelitian di atas ini menjadi sebuah referensi dan alasan peneliti melakukan penelitian “implementasi perbaikan ejaan pada web semantik pencarian judul skripsi ruang baca fakultas teknik dengan menggunakan algoritma *jaro-winkler distance*”. Penelitian ini bertujuan untuk melakukan penerapan algoritma *jaro-winkler distance* pada web semantik pencarian judul skripsi ruang baca fakultas teknik sehingga dapat mempermudah dalam melakukan pencarian.

II. METODOLOGI PENELITIAN

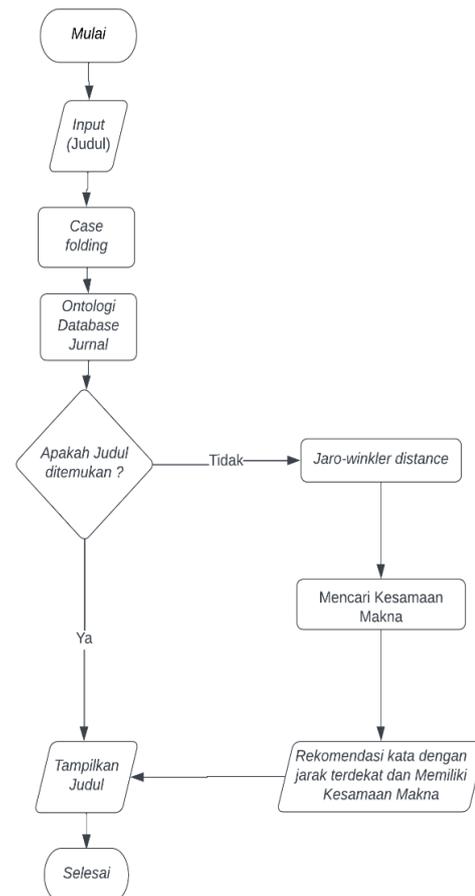
1. Tempat dan waktu penelitian

Lokasi pengambilan data pada penelitian ini dilakukan di Ruang Baca Fakultas Teknik Universitas Muhammadiyah Makassar, Jl. Sultan Alauddin No.259, Kelurahan Gunung Sari, Kecamatan Rappocini, Kota Makassar, Provinsi Sulawesi Selatan 90221.

2. Perancangan sistem

Perancangan sistem dapat diartikan sebagai tahap setelah analisis dari siklus pengembangan sistem, pendefinisian, dari kebutuhan fungsionalitas dan menggambarkan bagaimana suatu sistem dibentuk (Hidayat, 2009). *Flowchart* adalah suatu jenis diagram yang merepresentasikan algoritma atau langkah-langkah instruksi yang berurutan dalam sistem (Rosaly & Prasetyo, 2019) sedangkan *diagram activity* adalah diagram yang menggambarkan aktivitas pengguna sistem dari keseluruhan menu yang ada pada sistem (Nugroho et al., 2021).

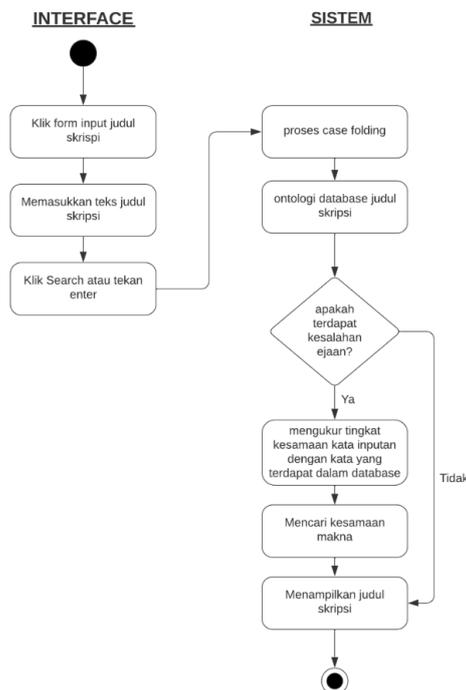
a. Flowchart proses semantik dan *jaro-winkler distance*



Gambar 1. Flowchart proses semantik dan *jaro-winkler distance*

Proses ini dimulai dengan memasukkan teks judul skripsi, kemudian melalui proses *case folding* semua karakter pada teks masukan disamakan menjadi huruf kecil, selanjutnya masuk ke proses ontologi *database* yaitu tahapan mencocokkan teks yang di inputkan dengan ontologi *database*, saat ditemukan maka judul akan ditampilkan dan proses selesai. Akan tetapi, jika teks judul yang dimasukkan tidak ditemukan maka akan masuk ke proses mengoreksi kata menggunakan algoritma *jaro-winkler distance*. Pada tahapan *jaro-winkler distance* sistem akan menghitung nilai kemiripan kata masukan dengan kata pada judul dalam *database*. Setelah mendapatkan kata yang mirip, selanjutnya masuk ke tahapan mencari sinonim dari kata yang mirip. Selanjutnya sistem akan menampilkan rekomendasi judul dengan kata yang mirip dan judul yang memiliki kesamaan makna.

b. Diagram activity proses semantik dan jaro-winkler distance



Gambar 2. Diagram activity proses semantik dan jaro-winkler distance

User mengklik *form input* judul skripsi, user memasukkan teks judul skripsi, user mengklik tombol *search* atau mengklik tombol *enter*, sistem melakukan proses *case folding*, sistem melakukan proses ontologi *database* judul skripsi, jika tidak ada kesalahan ejaan maka sistem akan menampilkan judul skripsi. Jika terdapat kesalahan ejaan maka proses *jaro-winkler distance* akan dijalankan, dengan mengukur tingkat kemiripan kata yang salah eja dengan kata pada judul skripsi di *database*, sistem akan mencari sinonim dari kata yang telah diseleksi pada proses *jaro-winkler distance*, sistem akan menampilkan judul skripsi yang memiliki kemiripan dengan kata salah eja yang dimasukkan.

3. Teknik pengujian sistem

Pada pengujian sistem penelitian ini pengujian sistem yang digunakan adalah metode *white box*. *White box testing* adalah suatu metode pengujian aplikasi yang menggunakan penjelasan struktur kontrol sebagai bagian dari *component-level design* untuk membuat *test cases*. *white box* sendiri memiliki beberapa teknik pengujian seperti *data flow testing*, *control flow testing*, *basic path* dan *loop testing* (Katiyar & Patel, 2019). Pada pengujian yang menggunakan *white box* para penguji harus memahami *source code* yang akan di uji (Khan & Khan, 2012).

Teknik pengujian yang digunakan pada penelitian adalah *basic path*, *basic path* merupakan teknik yang memungkinkan penguji untuk mengukur kompleksitas logika dari rancangan prosedural (Susilo, 2014). *Cyclomatic complexity* adalah besaran perangkat lunak yang menyediakan acuan kuantitatif kompleksitas suatu logika dalam program (Eriana, 2020). Berikut rumus menghitung nilai *cyclomatic complexity* :

$$V(G) = E - N + 2$$

4. Teknik analisis data

(Muhadjir, 1998) mengemukakan pengertian analisis data sebagai “upaya mencari dan menata secara sistematis catatan hasil observasi, wawancara, dan lainnya untuk meningkatkan pemahaman peneliti tentang kasus yang diteliti dan menyajikannya sebagai temuan bagi orang lain. Sedangkan untuk meningkatkan pemahaman tersebut analisis perlu dilanjutkan dengan berupaya mencari makna”. Teknik analisis data yang digunakan dalam penelitian ini adalah analisis data induktif. Analisis data induktif adalah penarikan kesimpulan yang berangkat dari fakta-fakta khusus, untuk kemudian ditarik kesimpulan secara umum.

III. HASIL DAN PEMBAHASAN

1. Proses perancangan

a. Antar muka halaman utama

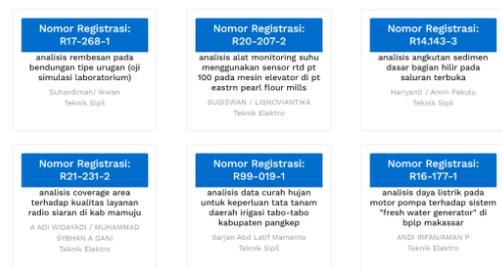


Gambar 3. Halaman pencarian

Halaman tampilan utama adalah halaman yang menampilkan *form input* untuk user memasukkan judul dan melakukan pencarian.

b. Antar muka hasil pencarian

Hasil Pencarian :



Gambar 4. Tampilan hasil pencarian

Pada tampilan pencarian judul skripsi adalah halaman untuk menampilkan hasil pencarian dari kata kunci yang dimasukkan oleh user.

c. Pengujian Algoritma

Dalam pengujian algoritmanya kami menggunakan rumus perhitungan jaro winkler distance dengan keterangan sebagai berikut di bawah :

- m = jumlah karakter yang sama persis
- |s1|= panjang string 1
- |s2|= panjang String 2
- t = jumlah transposisi
- dj = Jaro distance untuk string s1 dan s2
- l = panjang prefiks umum pada awal string dengan nilai maksimal 4 karakter

(panjang karakter yang serupa sebelum ditemukan karakter yang tidak serupa max 4).

p = konstanta scaling factor. Nilai standar untuk konstanta ini menurut Winkler adalah p = 0,1.

Kami menggunakan 20 sampel yang akan dihitung manual sebagai perbandingan validasi kebenaran dari algoritma yang kami gunakan.

Sedangkan untuk pengujian sistem dari aplikasi sendiri kami menggunakan flowgraph.

Pada Tabel 1 disajikan tabel pengujian algoritma dan hasil pengujiannya.

Tabel 1. Pengujian algoritma

No.	Kata yang dimasukkan	Hasil yang Diharapkan	Hasil Pengujian	Valid/tidak valid	Waktu Eksekusi
1	Anlisis	Analisis	Analisis	Valid	1907 ms
2	Pnrapan	Penerapan	Penerapan	Valid	120 ms
3	Pengrh	Pengaruh	Pengaruh	Valid	631 ms
4	Perbandingn	Perbandingan	Perbandingan	Valid	47 ms
5	irigsi	irigasi	Perbandingan	Valid	8769 ms
6	perncangn	perancangan	Perancangan	Valid	11475 ms
7	bndugan	Bendungan	Tidak ada hasil	Tidak valid	10180 ms
8	Rembsan	Rembesan	Rembesan	Valid	27 ms
9	Pasr	Pasir	Pasir	Valid	27 ms
10	Pncanaan	Perencanaan	Perencanaan	Valid	3974 ms
11	gnrtr	generator	Tidak ada hasil	Tidak valid	8117 ms
12	kprlan	Keperluan	Tidak ada hasil	Tidak valid	8663 ms
13	Lirik	Listrik	Listrik	Valid	1638 ms
14	Andorid	Android	Android	Valid	2184 ms
15	Mikrokontrolr	Mikrokontroler	Tidak ada hasil	Tidak valid	12747 ms
16	Arsu	Arus	Arus	Valid	1653 ms
17	Kenangan	Genangan	Genangan	valid	2043 ms
18	Apkilasi	Aplikasi	Aplikasi	Valid	2145 ms
19	Kehalingan	Kehilangan	Kehilangan	Valid	34 ms
20	Pengajian	Pengairan	Pengairan	Valid	67 ms

Untuk menghitung kata nilai *jaro-distance* dan *jaro-distance* dapat menggunakan persamaan berikut :

$$d_j = \frac{1}{3} \left(\frac{m}{|s_1|} + \frac{m}{|s_2|} + \frac{m-t}{m} \right) \dots\dots\dots(1)$$

$$d_w = d_j + (lp (1 - d_j)) \dots\dots\dots(2)$$

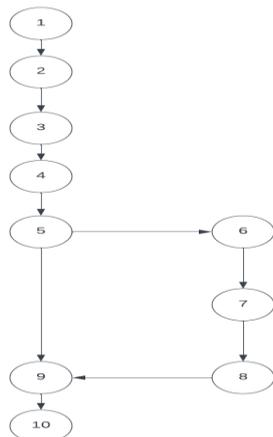
Tabel 2. Hasil Pengujian

No.	String 1	String 2	M	S1	S2	T	Hasil
1	Anlisis	Analisis	7	7	8	0	0.966
2	Pnrapan	Penerapan	7	7	9	0	0.933
3	Pengrh	Pengaruh	6	6	8	0	0.950
4	Perbandingn	Perbandingan	11	11	12	0	1
5	Pncanaan	Perencanaan	9	9	11	0	0.946
6	Irigsi	Irigasi	6	6	7	0	0.971
7	Percangn	Perancangan	9	9	11	0	0.971
8	Bndungan	Bendungna	7	7	9	1.5	0.869
9	Rembsan	Rembesan	7	7	8	0	0.975
10	Pasr	Pasir	4	4	5	0	0.953
11	Gnrtr	Generator	5	5	9	0	0.866
12	Kprlan	Keperluan	6	6	9	0	0.889
13	Lirik	Listrik	5	5	7	0	0.904
14	Andorid	Android	7	7	7	1	0.966
15	Mikrokontrolr	Mikrokontroler	12	12	15	2	0.877
16	Arsu	Arus	4	4	4	1	0.933
17	Kenangan	Genangan	7	8	8	0	0.916
18	Apkilasi	Aplikasi	8	8	8	1	0.966
19	Kehalingan	Kehilangan	10	10	10	1	0.976
20	Pengajian	Pengairan	8	9	9	0	0.963

Dari pengujian di atas didapatkan nilai rata-rata waktu eksekusi data valid 602.625 ms, rata-rata waktu eksekusi data tidak valid 9926.75 ms dan tingkat akurasi pencarian dari pengujian di atas sebanyak 80%.

d. Pengujian sistem

Flowgraph adalah grafik program yang dihasilkan dari pemetaan flowchart program yang ada untuk merepresentasikan aliran kontrol logika program yang ada (Pare, 2013). Flowgraph digunakan pada tahapan pengujian yang berfokus pada penggambaran aliran dari sebuah program (Suprpto et al., 2020).



Gambar 5. Flowgraph proses semantik dan jaro-winkler distance

Menghitung cyclomatic complexity

$$V(G) = (10 - 10) + 2$$

$$V(G) = 0 + 2$$

$$V(G) = 2$$

Hasil independent path pada flowgraph di atas

Path 1 = 1-2-3-4-5-9-10
Path 2 = 1-2-3-4-5-6-7-8-9-10

Menguji path fungsi proses semantik dan jaro-winkler distance

Tabel 3. Pengujian sistem

Path	Jalur	Skenario	Hasil pengujian
1	1-2-3-4-5-9-10	1 Mulai 2 Input judul 3 Case folding 4 Ontologi database judul skripsi 5 Judul ditemukan 9 Menampilkan judul 10 Selesai	Berhasil
2	1-2-3-4-5-6-7-8-9-10	1 Mulai 2 Input judul 3 Case folding 4 Ontologi database judul skripsi 5 Judul tidak ditemukan 6 Proses Jaro-winkler Distance 7 Mencari kesamaan makna 8 Rekomendasi kata dengan jarak terdekat dan memiliki kesamaan makna 9 Menampilkan judul 10 Selesai	Berhasil

Berdasarkan perhitungan dan pengujian di atas didapatkan cyclomatic complexity = 2, region = 2 dan independent path = 2 karena jumlah ketiga parameter sudah sama maka dapat di simpulkan bahwa proses semantik dan jaro-winkler distance sudah terbebas dari kesalahan logika.

IV. KESIMPULAN

Perbaikan ejaan pada pencarian informasi judul skripsi dengan penerapan algoritma jaro-winkler distance pada web semantik ruang baca sebagai basis pengetahuannya mampu membantu mahasiswa untuk menemukan judul skripsi yang diinginkan yang terdapat pada ruang baca fakultas teknik khususnya.

Pada hasil pengujian white box ditemukan bahwa alur sistem dari aplikasi telah bekerja sesuai dengan apa yang peneliti harapkan.

V. REFERENSI

Haris, N. A. (2022). Penerapan Web Semantik untuk pencarian judul skripsi pada ruang baca fakultas teknik Universitas Muhammadiyah Makassar. Makassar.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American*, 284(5), 34–43.

Eriana, E. S. (2020). Pengujian Sistem Informasi Aplikasi Perpustakaan Berbasis Web Dengan White Box Testing. *ESIT*, 15(2), 28–33.

Hidayat, R. (2009). Penerapan Sistem Informasi Manajemen Sebagai Alat Pelaksanaan Sistem Administrasi Akademik. *Jurnal Ilmiah Pendidikan Tinggi*, 2(2), 56–59. <https://scholar.google.co.id/scholar?oi=bib&s&cluster=11164046142770467310&btnI=1&hl=id>

Katiyar, V. P., & Patel, S. (2019). White-box testing technique for finding defects. *Glob. J. Res. Anal*, 8(7), 83–85.

Khan, M. E., & Khan, F. (2012). A comparative study of white box, black box and grey box testing techniques. *International Journal of Advanced Computer Science and Applications*, 3(6). <https://thesai.org/Publications/ViewPaper?Volume=3&Issue=6&Code=IJACSA&SerialNo=3>

Kurniawati, A., Puspitodjati, S., & Rahman, S. (2010). Implementasi Algoritma Jaro-Winkler Distance untuk Membandingkan Kesamaan Dokumen Berbahasa Indonesia. *Skripsi Program Studi Sistem Informasi*.

Muhadjir, N. (1998). Metodologi Penelitian Kualitatif Pendekatan Positivistik.

- Rasionalistik, Fenomenologik, Dan Realisme Metaphisik Telaah Studi Teks Dan Penelitian Agama.*
- Nugroho, N., Rahmanto, Y., Rusliyawati, R., Alita, D., & Handika, H. (2021). Software development sistem informasi kursus mengemudi (kasus: kursus mengemudi Widi Mandiri). *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, 5(1), 328–336. <https://tunasbangsa.ac.id/ejurnal/index.php/jsakti/article/view/325/304>
- Pare, S. (2013). Desain Dan Implementasi E-Commerce Pada Toko As 88 Celluler Merauke. *Jurnal Ilmiah Mustek Anim Ha*, 2(3). <https://ejournal.unmus.ac.id/index.php/mustek/article/view/201/144>
- Rosaly, R., & Prasetyo, A. (2019). Pengertian Flowchart Beserta Fungsi dan Simbol-simbol Flowchart yang Paling Umum Digunakan. *Program Studi Teknik Informatika Politeknik Purbaya.*
- Soleh, M. Y., & Purwarianti, A. (2011). A non word error spell checker for Indonesian using morphologically analyzer and HMM. *Proceedings of the 2011 International Conference on Electrical Engineering and Informatics*, 1–6.
- Suprpto, D. D. A., Fauziah, F., Fitri, I., & Hayati, N. (2020). Pengembangan Aplikasi Sistem Informasi Smart Register Online Berbasis Android Menggunakan Algoritma BruteForce. *Edumatic: Jurnal Pendidikan Informatika*, 4(1), 47–56. <https://doi.org/10.29408/edumatic.v4i1.2106>
- Susilo, J. (2014). Aplikasi Pengujian White Box IBII Online Jugde. *Jurnal Informatika Dan Bisnis*, 3(2), 56–69. <https://jurnal.kwikkiangie.ac.id/index.php/JIB/article/view/130>
- Yoga Pratama S, N. (2017). *Pendeteksian Kesalahan pengetikan Menggunakan Metode Double Metaphone Pada Teks Berbahasa Indonesia.* Universitas Komputer Indonesia. <https://elib.unikom.ac.id/download.php?id=358540>