



Deteksi Tingkat Kemiripan Judul Menggunakan Algoritma Oliver Pada Sistem Informasi Pengajuan Skripsi

Nopi Fitrianingsih¹, Marsani Asfi^{1*}, Dwi Prasetyo², Ricky Perdana Kusuma³, Muhammad Afif Sulhan³

¹Fakultas Teknologi Informasi, Program Studi Sistem Informasi, Universitas Catur Insan Cendekia, Cirebon, Indonesia

²Fakultas Sains dan Teknik, Program Studi Ilmu Komputer, Universitas Nusa Cendana, Kupang, Indonesia

³Fakultas Teknologi Informasi, Program Studi Manajemen Informatika, Univ. Catur Insan Cendekia, Cirebon, Indonesia

Email: ¹nopifitrianingsih118@gmail.com, ²marsani.asfi@cic.ac.id, ³dpras.ilkom.undana@gmail.com,

⁴rickyperdanak@cic.ac.id, ⁵muhammad.afif.sulhan@cic.ac.id

Email Penulis Korespondensi: marsani.asfi@cic.ac.id

Abstrak—Skripsi merupakan persyaratan akhir pendidikan akademis. Sistem prosedural skripsi terkadang masih membutuhkan waktu lama karena proses pengecekan judul masih dilakukan berdasarkan daya ingat pimpinan program studi sehingga memungkinkan terjadinya kemiripan judul skripsi. Oleh sebab itu, diperlukan sistem deteksi tingkat kemiripan judul secara otomatis agar dapat mencegah terjadinya *redudansi* judul skripsi. Penelitian ini menerapkan algoritma Oliver yang merupakan sebuah fungsi dalam bahasa pemrograman PHP yang menghasilkan nilai kemiripan dari dua string. Algoritma Oliver belum banyak diterapkan dalam teknik deteksi plagiarisme. Pengujian algoritma oliver menggunakan data uji sebanyak 10 data judul uji dan 217 judul untuk data training. Hasil pengujian menunjukkan algoritma Oliver menghasilkan persentase tingkat kemiripan judul sesuai dengan ketentuan diterima yaitu 60%. Diperoleh 2 dari 10 judul skripsi ditolak dengan persentase kemiripan sebesar 70.06% dan 62,89% dan 8 judul skripsi diterima dengan persentase kemiripan sebesar 31.94%, 30.48%, 26.11%, 31.86%, 29.47%, 31.13%, 35.01% dan 37.08%.

Kata Kunci: Kemiripan; Judul; Skripsi; Algoritma Oliver; Deteksi

Abstract—Thesis is the final requirement of academic education. The procedural system for the title checking process is based on the memory of the huma, thus allowing for the similarity of the thesis title. Therefore, a system can detect the degree of similarity of titles automatically in order to prevent redundancy of thesis titles. This study applies the Oliver algorithm which is a function in the PHP programming language that produces a similarity value of two strings. Oliver's algorithm has not been widely applied in plagiarism detection techniques. Oliver algorithm testing using test data as much as 10 test title data and 217 titles for training data. The test results show that Oliver's algorithm produces a percentage of the title similarity level in accordance with the accepted provisions, namely 60%. It was found that 2 out of 10 thesis titles were rejected with a similarity percentage of 70.06% and 62.89% and 8 thesis titles were accepted with a similarity percentage of 31.94%, 30.48%, 26.11%, 31.86%, 29.47%, 31.13%, 35.01% and 37.08%.

Keywords: Similarity; Title; Thesis; Oliver Algorithm; Detection

1. PENDAHULUAN

Skripsi merupakan salah satu persyaratan akhir pendidikan akademis mahasiswa yang ditulis dalam bentuk karya tulis ilmiah [1] serta sesuai bidang ilmunya [2]. Tahapan pertama dalam prosedur skripsi adalah proses pengajuan judul skripsi. Mahasiswa wajib mengajukan judul skripsi dengan tingkat kemiripan yang tidak sama dengan judul skripsi sebelumnya agar tidak terjadi *redudansi* judul skripsi. Namun, pada tahap pertama dalam prosedur skripsi ini tidak sedikit mahasiswa yang masih ragu akan tingkat kemiripan judul skripsi yang akan diajukan. Sistem yang ada sebelumnya belum dapat mengecek tingkat kemiripan judul yang diajukan dengan judul-judul skripsi sebelumnya. Hal tersebut juga berdampak dalam proses validasi judul skripsi sehingga membutuhkan waktu lama. Proses pengecekan judul masih dilakukan berdasarkan daya ingat pimpinan program studi. Sehingga, hasil pengecekan tersebut tidak efektif karena memungkinkan terjadinya *redudansi*, terlupa karena kemampuan daya ingat manusia. Oleh sebab itu, diperlukan adanya sistem berbasis komputer dengan menerapkan metode yang tepat yang dapat mendeteksi tingkat kemiripan judul skripsi secara otomatis agar dapat mencegah terjadinya *redudansi* judul skripsi sehingga pelaksanaan skripsi dapat lebih optimal dan tidak memakan waktu lama.

Terkait hal tersebut, terdapat beberapa algoritma yang dapat diterapkan untuk mendeteksi kemiripan teks atau *plagiarisme*. Penelitian tentang menguji kemiripan judul pada tugas akhir menggunakan metode *Cosine Similarity* serta penilaian bobot dengan TF-IDF. Penelitian ini diterapkan di Institut Kesehatan Deli Husada Delitua. Penelitian tersebut menunjukkan bahwa 43% dari judul yang diajukan tidak layak diajukan karena memiliki kemiripan dengan judul tugas akhir sebelumnya serta 53% layak untuk dapat diajukan. Hasil penelitian juga menunjukkan rata-rata waktu 0.12117 menit [3]. Penelitian lainnya adalah menggunakan algoritma Levenshtein Distance untuk mendeteksi kesamaan judul skripsi[4]. Penelitian ini menggunakan pencarian judul skripsi dengan multi target untuk 6 kategori serta penentuan threshold ≥ 4 dan bobot similarity $>25\%$ [4]. Penelitian lainnya dilakukan menggunakan algoritma string matching yaitu Boyer-Moore. Dengan algoritma Boyer-Moore pencocokan terhadap string dimulai dari karakter paling kanan selanjutnya ke karakter dibagian kiri. Pergeseran ini didasarkan pada nilai dari *occurence heuristic* (HO) dan *match heuristic* (MH). Hasil penelitian menunjukkan sistem yang dibuat mampu dapat digunakan untuk menentukan kemiripan usulan judul Skripsi berdasarkan data repositori judul skripsi yang telah ada. Hasil pengujian menunjukan bahwa ketika banyak data judul maka rata-rata peningkatan waktu hanya bertambah 34% [5]. Penelitian terdahulu



lainnya juga membahas tentang pengecekan kemiripan judul untuk skripsi dengan menggunakan algoritma *Winnowing* [6][7][8]. Penelitian tersebut menerapkan nilai parameter *k*-gram yang memiliki nilai *similarity* yang tinggi. Nilai *k* semakin besar akan mengurangi nilai kemiripan pada judul [9]. Hasil dari penelitian [9] diharapkan koordinator tugas akhir atau Ketua Program Studi dapat menentukan persentase kemiripan judul sebagai syarat untuk *review* awal atas pengajuan skripsi. Pengujian dilakukan untuk 117 judul skripsi dan diperoleh 11 judul yang dengan tingkat kemiripan lebih besar dari 20 persen [6]. Penelitian lainnya adalah penggunaan algoritma Rabin Karp [10], dalam penelitian ini dilakukan tahapan *preprocessing* untuk membersihkan data judul skripsi. Selanjutnya dilakukan proses *hashing*. Tahapan *string matching* dilakukan dengan Rabin Karp pada setiap kata dan diukur tingkat kemiripannya menggunakan *dice's Coefficient*, sehingga menghasilkan nilai *similarity* dari judul [10].

Mengingat banyaknya algoritma untuk mendeteksi kemiripan teks, maka pada penelitian ini digunakan algoritma *oliver* (PHP *similar_text*) untuk mendeteksi tingkat kemiripan judul skripsi. Algoritma Oliver merupakan sebuah fungsi dalam bahasa pemrograman PHP yang membandingkan 2 *string* dengan hasil nilai kemiripan dari kedua *string* [11] dan diperkenalkan oleh [12]. Penelitian penggunaan algoritma *oliver* sendiri pernah dilakukan oleh [13]. Dalam penelitian tersebut penggunaan algoritma Oliver digabungkan dengan algoritma lainnya yang digunakan dalam perbaikan dan pemberian saran dalam penulisan. Algoritma Oliver belum banyak diterapkan dalam teknik deteksi *plagiarisme*. Dengan menggunakan fungsi *similar text* dapat dilakukan penentuan nilai kesamaan teks dalam bentuk persentase. *Similar_text* adalah fungsi bawaan dalam bahasa pemrograman PHP untuk menghitung kesamaan antara dua *string* [14]. Penelitian yang dilakukan adalah untuk mengembangkan sistem deteksi tingkat kemiripan judul skripsi. Algoritma Oliver digunakan dalam mendeteksi kemiripan judul skripsi. Adapun format penulisan fungsi *similar_text* yaitu :

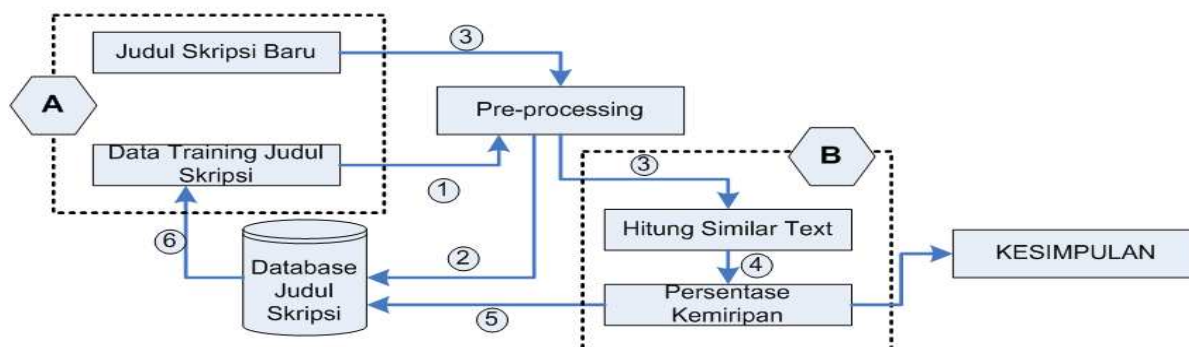
`int similar_text (string $first , string $second[, float &$percent]) [11].`

Dimana *\$first* adalah paramater judul skripsi asal, *\$second* adalah paramater judul skripsi kedua, sedangkan *\$percent* adalah nilai bertipe float yang dihasilkan dari hasil perbandingan dua judul *\$first* dan *\$second*

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Metode eksperimen digunakan dalam penelitian ini. Adapun tahapan penelitiannya adalah sebagai berikut :



Gambar 1. Tahapan Penelitian

2.2. Pengumpulan Data

Data yang digunakan adalah dokumen teks berupa daftar judul skripsi mahasiswa sebanyak 217 Judul dari 5 Program Studi yang ada di Fakultas Teknologi Informasi Universitas Catur Insan Cendekia.pada periode 2019/2020. Data tersebut selanjutnya dijadikan data *training* sebanyak 217 judul sedangkan data uji yang digunakan adalah data judul skripsi untuk periode 2020/2021 sebanyak 10 judul. Gambar 1, pada Bagian A merupakan tahapan pengumpulan data. Judul skripsi baru menjadi data uji, sedangkan Data judul-judul skripsi sebelumnya dijadikan data *training*.

Tabel 1. Data Training

N o	Judul Skripsi
1	Sistem Informasi Pendaftaran Seminar pada STMIK CIC Cirebon Berbasis Web
2	Perancangan Media Pembelajaran Tata Surya Menggunakan Teknologi Augmented Reality untuk Siswa Sekolah Dasar
3	Implementasi Mikrotik Routerboard Hap Lite GB941-2ND-TC Untuk Manajemen Jaringan Laboratorium Komputer di SMK Negeri 3 Kuningan



N o	Judul Skripsi
4	Sistem Monitoring dan Kontrol Otomatis Suhu Kandang Peternakan Ayam Menggunakan Wemos dengan Telegram
5	Perancangan Desain User Interface & User Experience Aplikasi Sebagai Media Pengelolaan Sampah di Kota Cirebon
6	Pemanfaatan Mikrotik RB941 untuk Manajemen User Hotspot dengan Metode Queue Free (Studi Kasus : SMA NU Kaplongan Indramayu)
7	Prototype Home Automation pada Keamanan Pintu Rumah Berbasis Arduino UNO
8	Rancang Bangun Akses Kontrol Pintu Rumah dan Keamanan Rumah Berbasis Wemos dengan Konsep Internet of Things (IOT)
9	Sistem Informasi Perhitungan dan Pencatatan Akuntansi Pengupahan Karyawan Menggunakan Metode Accrual Basis Berbasis Web (Studi Kasus : PT. Herona Express Cabang Cirebon)
1	Rancang Bangun Aplikasi Alat Bantu Pembelajaran Pengenalan Hardware Komputer dengan Teknologi
0	Augmented Reality (Studi Kasus : Sekolah Dasar Cirebon Islamic School (CIS))
.	... dan seterusnya.

Sedangkan Tabel 2, adalah data uji untuk digunakan sebagai pengujian pengukuran tingkat kemiripan judul. Data uji yang digunakan sebanyak 10 contoh judul skripsi.

Tabel 2. Data Testing Judul Skripsi[15]

N o	Judul Skripsi
1	Analisis Umur Piutang Atas Biaya Kuliah Dengan Metode Pencatatan <i>Ledgerless Bookkeeping</i> Studi Kasus : Universitas CIC Cirebon.
2	Analisis dan Implementasi <i>Security Network Fail2ban</i> terhadap Serangan Dos Pada Web Server (Studi Kasus : Universitas CIC).
3	Penerapan Metode <i>Customer Satisfaction Index</i> (CSI) untuk Mengukur Tingkat Kepuasan Layanan Manajemen Universitas Catur Insan Cendekia.
4	Sistem Informasi Pengelolaan Alumni berdasarkan IAPS 4.0.
5	Rancangan Bangun Aplikasi LINE <i>Chatbot</i> Informasi dan Edukasi Kesehatan Mental.
6	Implementasi <i>Load Balancing</i> Menggunakan NGINX dengan Metode <i>Round Robin</i> Pada <i>Learning Management System Moodle</i> .
7	Sistem Pendukung Keputusan Pemilihan Jurusan dengan Metode <i>Profiel Matching</i> .
8	Perancangan Aplikasi Monitoring Ruang Menggunakan IP Camera Berbasis Android pada Universitas Catur Insan Cendekia.
9	Analisis Kualitas Pelayanan terhadap Kepuasan Konsumen dengan menggunakan Dimensi TERRA pada Yogya Grand Cirebon
1	Sistem Informasi Penjualan Tunai Obat Menggunakan Metode <i>Cash Basis</i> Pada Apotek Jasa Prima
0	<i>Medical Centre</i> Cirebon.

2.3. Pengolahan Dengan Algoritma Oliver

Algoritma Oliver merupakan fungsi PHP untuk *similarity text*. *Similarity text* digunakan untuk menghitung kesamaan diantara 2 string [11]. Hasil dari *similar text* berupa nilai persentase dari kesamaan 2 string [11]. String yang dimaksud berupa angka, huruf, karakter, ataupun kata yang telah dikelompokkan secara bersama-sama ataupun terpisah sebagai suatu *string* dengan *variable* panjang [14].

2.4. Analisa Penentuan Similar Text dengan Algoritma Oliver

Proses analisa kemiripan judul skripsi dengan algoritma Oliver untuk bagian B pada Gambar 1 pada proses 3 dan 4 dapat diilustrasikan dengan *pseudo-code* sebagai berikut :

1. Menentukan 2 string yang akan diuji tingkat *similar text* nya (misal Str_1 dan Str_2).
2. Menghitung panjang string pertama dan kedua, termasuk spasi (misal L_1 dan L_2).
3. Menghitung panjang string total dari string pertama dan kedua (misal $T_{str} = L_1 + L_2$).
4. Melakukan :
 - a. Perhitungan kesamaan karakter yang ada di string kedua dengan karakter di string pertama.
 - b. Simpan jumlah karakter yang sama tersebut (misal T_{kar}).
5. Menghitung nilai rata-rata panjang string, yaitu panjang string total 2 string dibagi dengan 2 (misal $P_{str} = T_{str}/2$).
6. Menghitung persentase similar text dengan cara membagi antara jumlah karakter yang sama (T_{kar}) dengan rata-rata panjang string (P_{str}) dikali dengan 100% (misal $S_{text} = (T_{kar}/P_{str}) * 100\%$).

Misal *similar text* dari string “sisfotenika” dan “sisfo”. Karakter yang sama dari 2 string tersebut adalah karakter s, i, s, f, o dengan panjang 5. Nilai rata-rata panjang stringnya adalah 8 yang diperoleh dari $16 \div 2 = 8$.



Perhitungan *similar textnya* adalah dengan membagi jumlah string yang sama dengan rata-rata panjang string. Sehingga diperoleh nilai persentasenya adalah $(5 \div 8) \times 100\% = 62,5\%$.

2.5. Kesimpulan Hasil Kemiripan

Kesimpulan hasil kemiripan dari judul skripsi yang diuji adalah ditolak atau diterima. Ketentuan yang berlaku pada deteksi kemiripan judul skripsi ini diantaranya adalah :

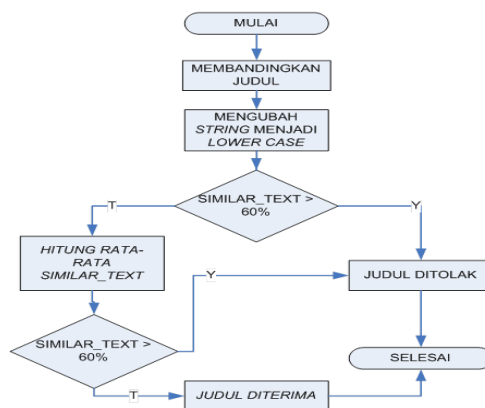
- Apabila terdapat tingkat kemiripan judul lebih besar dari 60%, maka judul usulan ditolak.
- Apabila tingkat kemiripan judul kurang dari atau sama dengan 60%, maka judul usulan diterima.

Plagiat merupakan perbuatan yang sengaja maupun tidak sengaja dilakukan untuk mendapatkan kredit atau nilai pada suatu karya ilmiah. Kegiatan plagiat yang dilakukan yaitu dengan mengutip seluruh ataupun sebagian karya ilmiah pihak lain, lalu diakui sebagai karya ilmiahnya tanpa penyertaan sumber secara tepat dan memadai [16], [17]. Batasan proporsi atau persentase kata atau kalimat mengenai plagiasi adalah sebagai berikut :

- Apabila $< 30\%$, maka tidak plagiasi.
- Apabila $30-60\%$, maka plagiasi sedang.
- Apabila $> 60\%$, maka plagiasi berat atau total.

2.6. Implementasi Algoritma Oliver dalam Sistem

Implementasi sistem dibuat dalam antar muka sistem yang dapat digunakan oleh pengguna untuk menguji tingkat persentase kemiripan judul skripsi yang diajukannya. Gambar 2, menunjukkan tahapan pengecekan tingkat *similarity* dalam sistem yang dikembangkan.



Gambar 2. Implementasi Flowchart Oliver dalam Sistem

Proses pengecekan tingkat kemiripan judul skripsi dengan menggunakan Algoritma Oliver (PHP *Similar_Text*) seperti terlihat pada gambar 2 :

- Mengubah string menjadi *low case*
Dilakukan untuk mengurangi akurasi tingkat kemiripan karena faktor *case sensitive*, semua *string* yang di cek kemiripannya menjadi huruf kecil semua (*low case*).
- Membandingkan judul
Membandingkan judul skripsi yang diajukan dengan judul-judul skripsi yang telah ada.
- Deteksi kemiripan judul (*similar_text*)
Selanjutnya, sistem melakukan pendeteksian kemiripan judul skripsi dengan perhitungan pembagian nilai *similarity text* dengan rata-rata panjang *string*. Hasil yang diperoleh adalah tingkat kemiripan judul dalam bentuk persentase.
- Hasil persentase tingkat kemiripan judul skripsi tersebut kemudian akan di cek dengan ketentuan sebagai berikut :
 - Apabila terdapat tingkat kemiripan judul skripsi lebih besar dari 60%, maka judul yang diajukan ditolak.
 - Apabila tingkat kemiripan judul skripsi kurang dari atau sama dengan 60%, maka judul yang diajukan akan dihitung rata-rata persentasenya. Kemudian, apabila rata-rata persentase kurang dari atau sama dengan 60% maka judul yang diajukan diterima dan sebaliknya apabila rata-rata persentase tersebut lebih besar dari 60% maka judul yang diajukan ditolak.

3. HASIL DAN PEMBAHASAN

3.1. Implementasi Algoritma

Algoritma Oliver merupakan fungsi yang diimplementasi dalam bahasa pemrograman PHP. Adapun *listing program* dari penerapan algoritma Oliver seperti pada Gambar 3.



```

$total_sim = 0;
$i = 0;
foreach (($skripsi) as $key => $data) {
    $sim = similar_text(strtolower($_POST['judul_usulan']), strtolower($data['judul_skripsi']), $percent);
    $total_sim += $percent;
    $tingkat_sim = $total_sim/$n;
    $arr[$i] = ['percent' => $percent];
    $i++;
    $max = max($arr);
}

```

Gambar 3. Implementasi Algoritma Oliver dalam PHP

3.2. Hasil Preprocessing

Preprocessing dilakukan untuk meningkatkan akurasi kemiripan yang dihasilkan yang disebabkan karena faktor *case sensitive*, maka semua *string* yang di cek kemiripannya menjadi huruf kecil semua (*low case*). Contoh hasil *pre-processing* dari Tabel 1 dan Tabel 2, berikut ini :

Misalkan, judul awal :

Prototype Sistem Keamanan Buka Tutup Atap Jemuran Otomatis Menggunakan Sensor Air dan Light Dependent Resistor (LDR) Berbasis Arduino

Setelah dilakukan *pre-processing* :

prototype sistem keamanan buka tutup atap jemuran otomatis menggunakan sensor air dan light dependent resistor (ldr) berbasis arduino

Untuk proses *preprocessing* dalam penelitian ini dilakukan dengan 217 contoh judul data training, sedangkan untuk pengujian digunakan data uji judul skripsi sebanyak 10 contoh judul skripsi.

3.3. Hasil Perbandingan Similarity

Pada Tabel 1 dan Tabel 2 adalah data setelah dilakukan *preprocessing* selanjutnya dilakukan perbandingan tingkat *similarity* nya.

Judul usulan (data testing ke-1) :

analisis umur piutang atas biaya kuliah dengan metode pencatatan ledgerless bookkeeping studi kasus : universitas cicirebon

Judul skripsi (data training ke-1) :

sistem pendukung keputusan penilaian kinerja karyawan pt. harjamukti jaya mandiri menggunakan metode simple additive weighting

Hasil perbandingan judul usulan (data *testing* ke-1) dengan judul skripsi (data *training* ke-1) didapatkan :

- Panjang *string* judul usulan (data *testing* ke-1) = 125
- Panjang *string* judul skripsi (data *training* ke-1) = 126
- Rata-rata panjang *string* = 125,5
- similarity text* judul usulan (data *testing* ke-1) dengan judul skripsi (data *training* ke-1) = 41
- Maka, % kemiripan judul antara judul usulan (data *testing* ke-1) dengan judul skripsi (data *training* ke-1) = $(41 / 125,5) * 100\% = 32,67\%$.

3.4. Hasil Perhitungan Similarity

Tabel 3 adalah rekapitulasi hasil perhitungan kemiripan teks dilakukan dengan menghitung masing-masing 10 contoh data uji judul skripsi.

Tabel 3. Hasil Rekapitulasi Nilai-Nilai Perhitungan *Similarity* Judul

No	Data Uji Ke-	Similarity Teks	Rata-Rata Panjang	% Kemiripan	MAX
1	Judul-1	37	114,91	32,37%	52,01%
2	Judul-2	35	113,41	30,74%	48,93%
3	Judul-3	31	118,91	26,31%	44,95%
4	Judul -4	25	80,41	31,62%	56,25%
5	Judul-5	26	80,41	31,62%	53,80%
6	Judul-6	34	90,91	29,15%	47,46%
7	Judul-7	30	107,41	31,35%	70,06%
8	Judul-8	39	110,41	35,12%	58,77%



No	Data Uji Ke-	Similarity Teks	Rata-Rata Panjang	% Kemiripan	MAX
9	Judul-9	38	108,41	34,97%	62,89%
10	Judul-10	41	108,91	37,45%	56,50%

3.5. Hasil Keputusan dari Persentase Similarity

Hasil persentase tingkat kemiripan judul skripsi pada Tabel 3 kemudian akan di cek dengan ketentuan sebagai berikut :

- Apabila terdapat tingkat kemiripan judul skripsi lebih besar dari 60%, maka judul yang diajukan ditolak.
- Apabila tingkat kemiripan judul skripsi kurang dari atau sama dengan 60%, maka judul yang diajukan akan dihitung rata-rata persentasenya. Kemudian, apabila rata-rata persentase kurang dari atau sama dengan 60% maka judul yang diajukan diterima dan sebaliknya apabila rata-rata persentase tersebut lebih besar dari 60% maka judul yang diajukan ditolak.

Misalkan hasil perhitungan pada Tabel 3, dari sepuluh data uji terdapat persentase kemiripan judul skripsi yang melebihi batas ketentuan yang telah ditetapkan (lebih besar dari 60%) yaitu 70,06 % (data uji ke-7) dan 62,89% (data uji ke-9). Maka secara langsung kedua judul tersebut ditolak oleh sistem. Sedangkan untuk delapan data testing lainnya kurang dari atau sama dengan 60%, akan dihitung rata-rata persentasenya :

Rata-rata % kemiripan judul skripsi usulan (data testing ke-1), yaitu :

Jumlah judul skripsi (data *training*) = 217

Jumlah % kemiripan judul skripsi usulan (data testing ke-1) = 6929,90%

Rata-rata % kemiripan judul skripsi usulan (data testing ke-1) = 31,94%

Demikian juga akan dilakukan perhitungan rata-rata % kemiripan judul skripsi usulan yang sama untuk data testing lainnya yang nilai maksimum % kemiripan judul skripsinya kurang dari atau sama dengan 60%. Tabel 4 merupakan rangkuman akhir hasil perhitungan similarity dan keputusan akhir yang diberikan.

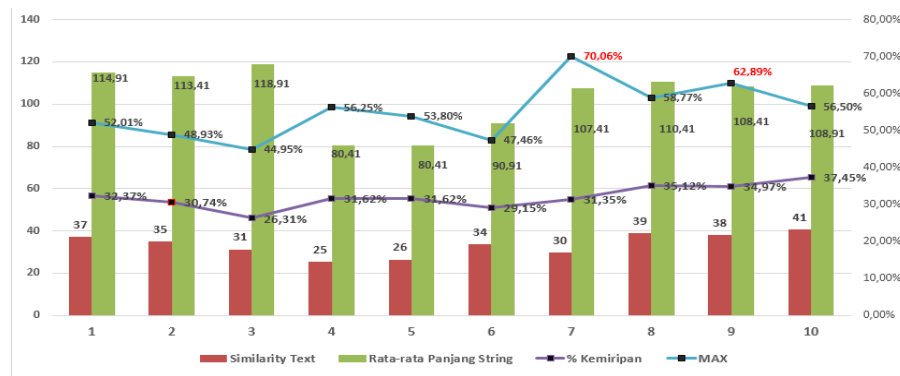
Tabel 4. Contoh Hasil Akhir Keputusan *Similarity* Judul

No	Data Uji Ke-1	% Kemiripan	Keputusan
1	Judul-1	31,94	Diterima
2	Judul-2	30,48	Diterima
3	Judul-3	26,11	Diterima
4	Judul-4	31,86	Diterima
5	Judul-5	29,47	Diterima
6	Judul-6	31,13	Diterima
7	Judul-7	70,06	Ditolak
8	Judul-8	35,01	Diterima
9	Judul-9	62,89	Ditolak
10	Judul-10	37,08	Diterima

3.6. Pembahasan

Gambar 4. menunjukkan beberapa hasil utama dari penggunaan algoritma Oliver ini. Informasi yang diperoleh diantaranya :

- Tingkat *similarity*. Tingkat *similarity* ini menunjukkan nilai kemiripan teks hasil penggunaan fungsi *similar_text* dari algoritma Oliver. Rata-rata tingkat *similarity* dari 10 contoh judul skripsi yang diuji adalah 34.
- Rata-rata panjang String. Nilai rata-rata ini menunjukkan panjang string yang dihasilkan dari proses perbandingan. Rata-rata panjang string dari 10 contoh judul skripsi adalah 103.
- Nilai Maksimum merupakan nilai tertinggi hasil perhitungan nilai prosentase kemiripan semua hasil perbandingan data uji terhadap data training. Dari Gambar 7 diperoleh bahwa contoh judul uji ke-7 diperoleh nilai maksimum 70.06% dan contoh judul uji ke-9 diperoleh nilai maksimum 62.89%. Dari informasi nilai maksimum ini, maka secara otomatis kedua judul tersebut akan ditolak karena memiliki nilai prosentase kemiripan diatas 60%.
- Prosentase Kemiripan. Nilai prosentase kemiripan ini diperoleh dari hasil pembagian dari tingkat similarity dengan rata-rata panjang string. Dari Gambar 7 diperoleh masing-masing prosentase kemiripan untuk 10 contoh judul skripsi yang diujikan, yaitu : 32.07%, 30.74%, 26.31%, 26.31%, 29.15%, 31.35%, 35.12%, 34.97% dan 37.45%. Prosentase kemiripan ini akan menjadi dasar dalam penentuan apakah judul skripsi diterima atau ditolak. Dari hasil ini terlihat bahwa dari 10 contoh judul data uji semuanya akan diterima, tapi jika dilihat dari nilai maksimum yang diperoleh dari hasil perbandingan satu persatu judul diperoleh bahwa ada 2 judul yang ditolak yaitu contoh judul uji ke-7 dengan nilai maksimum 70.06% dan contoh judul uji ke-9 dengan nilai maksimum 62.89%.



Gambar 4. Grafik Perbandingan Tingkat Similarity, Rata-Rata Panjang String, % Kemiripan dan Nilai Max

3.7. Implementasi Sistem

Gambar 5, merupakan implementasi sistem untuk pengecekan judul skripsi. *Interface* ini terdiri dari masukan judul skripsi yang akan diuji tingkat *similarity* nya. *Form* ini digunakan untuk melakukan pengecekan dan pengajuan judul skripsi. Pada *form* ini akan ditampilkan hasil persentase tingkat kemiripan judul usulan terhadap judul skripsi yang sudah ada sebelumnya.

Gambar 5. Implementasi Halaman Pengecekan Kemiripan Judul Skripsi

Gambar 6, merupakan keluaran dari hasil pengecekan judul skripsi. Dari keluaran tersebut akan ditampilkan persentase tingkat kemiripan judul skripsi serta keputusan dapat melanjutkan ke tahapan berikutnya.

Gambar 6. Implementasi Hasil Keluaran dari Pengecekan Judul Skripsi

Sedangkan Gambar 7, merupakan implementasi rekapitulasi hasil perhitungan tingkat kemiripan judul skripsi untuk beberapa data yang digunakan dalam pengujian sistem. Implementasi sistem diterapkan pada aplikasi SIMASITA di Universitas Catur Insan Cendekia [15].



Detail Hasil Algoritma Oliver (Fungsi Similar_Text)					
No.	Judul Usulan	Judul Skripsi	Similarity Text	Rata-rata Panjang String	Persentase Kemiripan
1	analisis umur piutang atas biaya kuliah dengan metode pencatatan ledgerless bookkeeping studi kasus : universitas cic Cirebon	sistem pendukung keputusan penilaian kinerja karyawan pt. harjamukti jaya mandiri menggunakan metode simple additive weighting	41	125.5	32.67 %
2	analisis umur piutang atas biaya kuliah dengan metode pencatatan ledgerless bookkeeping studi kasus : universitas cic Cirebon	sistem pengendalian dan monitoring penerangan rumah menggunakan modul arduino berbasis internet	36	110	32.73 %
3	analisis umur piutang atas biaya kuliah dengan metode pencatatan ledgerless bookkeeping studi kasus : universitas cic Cirebon	pembuatan game rpg "the adventure of sachi" menggunakan engine rpg maker mv	14	100	14 %
4	analisis umur piutang atas biaya kuliah dengan metode pencatatan ledgerless bookkeeping studi kasus : universitas cic Cirebon	sistem pendukung keputusan penilaian kinerja dosen di stmik cic menggunakan metode ahp	32	105.5	30.33 %
5	analisis umur piutang atas biaya kuliah dengan metode pencatatan ledgerless bookkeeping studi kasus : universitas cic Cirebon	prototype sistem keamanan buka tutup atap jemuran otomatis menggunakan sensor air dan light dependent resistor (ldr) berbasis arduino	13	129	10.08 %
6	analisis umur piutang atas biaya kuliah dengan metode pencatatan ledgerless bookkeeping studi kasus : universitas cic Cirebon	perancangan aplikasi augmented reality sebagai alat bantu pemasaran perumahan berbasis android pada perumahan griya kerta kencana weru kabupaten Cirebon	37	138.5	26.71 %

Gambar 7. Implementasi Hasil Pengujian Tingkat Kemiripan

3.8. Hasil Pengujian Tingkat Kemiripan

Teknik pengukuran untuk pengukuran hasil pengujian menggunakan pengukuran persentase tingkat kemiripan judul skripsi. Hasil pengukuran seperti yang terlihat pada Gambar 7, dapat dijelaskan dengan ketentuan berikut :

- Jika tingkat kemiripan judul skripsi lebih besar dari 60%, maka judul skripsi yang diajukan tersebut ditolak.
- Sedangkan jika tingkat kemiripan judul skripsi kurang dari atau sama dengan 60%, maka judul skripsi yang diajukan tersebut akan dihitung rata-rata persentasenya secara total.
- Kemudian, jika rata-rata persentase pengukuran kemiripan judul skripsi kurang dari atau sama dengan 60% maka judul skripsi yang diajukan diterima, sebaliknya jika rata-rata pengukuran persentase judul skripsi secara total lebih besar dari 60% maka judul yang diajukan akan ditolak.

Dari sepuluh data uji diperoleh persentase kemiripan judul skripsi yang melebihi 60% yaitu 70,06 % (data uji ke-7) dan 62,89% (data uji ke-6). Maka secara otomatis langsung kedua judul skripsi tersebut ditolak oleh sistem. Untuk delapan data testing lainnya yang memiliki nilai persentase kurang dari atau sama dengan 60%, akan dihitung rata-rata persentase kemiripannya.

4. KESIMPULAN

Dari hasil pengujian dalam proses pengecekan kemiripan judul skripsi dengan data *testing* sebanyak 10 data judul skripsi yang diusulkan mahasiswa dan data *training* sebanyak 217 judul skripsi menghasilkan persentase tingkat kemiripan judul sesuai dengan ketentuan yang telah ditetapkan, didapatkan 2 judul skripsi ditolak karena persentase tingkat kemiripan judul sebesar **70.06% dan 62,89%** (lebih besar dari 60%), dan 8 judul skripsi lainnya diterima dengan tingkat kemiripan judul diantaranya sebesar **32.07%, 30.74%, 26.31%, 26.31%, 29.15%, 35.12% dan 37.45%** (kurang dari atau sama dengan 60%). Keputusan diterima atau tidak didasarkan prosentase perbandingan satu persatu judul uji dengan judul training. Jika nilai prosentase perbandingan diatas 60% (seperti contoh judul uji ke-7 dan contoh judul uji ke-9) maka judul tersebut ditolak. Sebaliknya jika dibawah 60% proses keputusan diterima data atau tidak didasarkan pada prosentase kemiripan rata-rata yang dihasilkan.

REFERENCES

- [1] Badan Pengembangan dan Pembinaan Bahasa, "Skripsi." Accessed: Jun. 20, 2020. [Online]. Available: <https://kbbi.kemdikbud.go.id/entri/skripsi>.
- [2] R. Fitriyani, "Perancangan Sistem Prosedur Skripsi Menggunakan Metode BPMN dan RAD Pada Kampus STMIK Mercusuar," *J. TEKNO*, vol. 2, no. 2, pp. 1–13, 2018.
- [3] I. Mawanta, T. S. Gunawan, and Wanayumini, "Uji Kemiripan Kalimat Judul Tugas Akhir dengan Metode Cosine Similarity dan Pembobotan TF-IDF," *J. MEDIA Inform. BUDIDARMA*, vol. 5, no. 2, pp. 726–738, Apr. 2021, doi: 10.30865/mib.v5i2.2935.
- [4] B. Harira Irawan, M. H. Sahat Simarankir, and S. MIC Cikarang, "DETEKSI KEMIRIPAN JUDUL SKRIPSI MENGGUNAKAN ALGORITMA LEVENSHTEIN DISTANCE PADA KAMPUS STMIK MIC CIKARANG," *J. Ilm. Edutic Pendidik. dan Inform.*, vol. 7, no. 2, pp. 143–149, May 2021, doi: 10.21107/EDUTIC.V7I2.10051.
- [5] I. Ahmad, R. I. Borman, G. G. Caksana, and J. Fakhrurozi, "Implementasi String Matching Dengan Algoritma Boyer-Moore Untuk Menentukan Tingkat Kemiripan Pada Pengajuan Judul Skripsi/Ta Mahasiswa (Studi Kasus: Universitas XYZ)," *SINTECH (Science Inf. Technol. J.)*, vol. 4, no. 1, pp. 53–58, Apr. 2021, doi: 10.31598/SINTECHJOURNAL.V4I1.699.
- [6] N. Nurdin and A. Munthoha, "Sistem Pendeteksian Kemiripan Judul Skripsi Menggunakan Algoritma Winnowing,"



- InfoTekJar (Jurnal Nas. Inform. dan Teknol. Jaringan)*, vol. 2, no. 1, pp. 90–97, Sep. 2017, doi: 10.30743/infotekjar.v2i1.165.
- [7] F. R. N. Wulan, A. Kunaefi, and A. Permadi, “Deteksi Plagiasi Dokumen Skripsi Mahasiswa Menggunakan Metode N-GRAMS Dan WINNOWNING,” *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 9, no. 2, pp. 1021–1032, Nov. 2018, doi: 10.24176/simet.v9i2.2535.
- [8] W. Pramudita, B. P. Tomasouw, Z. A. Leleury, and M. E. Rijoly, “Perancangan Sistem Deteksi Plagiarisme Skripsi (Judul Dan Abstrak) Berbasis Matlab Menggunakan Algoritma Winnowing | Tensor: Pure and Applied Mathematics Journal,” 2021. <https://ojs3.unpatti.ac.id/index.php/tensor/article/view/4439> (accessed Aug. 11, 2022).
- [9] W. Hidayat, E. Utami, and A. D. Hartanto, “Pemilihan Parameter Terbaik pada Algoritma Winnowing dalam Mendeteksi Tingkat Kesamaan Dokumen Bahasa Indonesia,” *Creat. Inf. Technol. J.*, vol. 7, no. 2, p. 119, Mar. 2021, doi: 10.24076/citec.2020v7i2.256.
- [10] S. Adi, “Penerapan Algoritma Rabin Karp Untuk Mendeteksi Kemiripan Judul Skripsi,” *Terakreditasi DIKTI*, vol. 22, no. 1, pp. 125–130, Jul. 2018, Accessed: May 21, 2021. [Online]. Available: <http://ejurnal.pelitanusantara.ac.id/index.php/mantik/article/view/472>.
- [11] “similar_text,” *php.net*. <https://www.php.net/manual/en/function.similar-text.php> (accessed Jun. 20, 2020).
- [12] I. Oliver, *Programming classics: implementing the world’s best algorithms*. New York: Prentice Hall, 1994.
- [13] N. Sembilu, F. Samopa, and M. Er, “Perbandingan Algoritma Kemiripan Teks untuk Perbaikan dan Saran Penulisan Frasa Dalam Bahasa ALami,” *Sisfo*, vol. 08, no. 01, Sep. 2018, doi: 10.24089/J.SISFO.2018.09.002.
- [14] M. Wali and S. Safrizal, “Similar text sebagai Pengkodean Aplikasi Plagiarisme,” *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 2, no. 1, p. 11, Jan. 2018, doi: 10.35870/jtik.v2i1.43.
- [15] M. Asfi and N. Fitrianiingsih, “Implementasi Algoritma Naive Bayes Classifier sebagai Sistem Rekomendasi Pembimbing Skripsi,” *InfoTekJar J. Nas. Inform. dan Teknol. Jar.*, vol. 5, no. 1, pp. 44–50, Sep. 2020, doi: 10.30743/infotekjar.v5i1.2536.
- [16] Menteri Pendidikan Nasional, “Permendiknas No 17 Tahun 2010 Tentang Pencegahan Dan Penanggulangan Plagiat di Perguruan Tinggi,” pp. 1–9, 2010, [Online]. Available: http://sumberdaya.ristekdikti.go.id/wp-content/uploads/2016/02/permendiknas-no-17-tahun-2010_pencegahan-plagiat.pdf.
- [17] Rosihan Ari Yuana, “Cara Mengetahui Tingkat Kemiripan Teks (Text Similarity) dengan PHP,” 2012. <https://blog.rosihanari.net/cara-mengetahui-tingkat-kemiripan-teks-text-similarity-dengan-php/> (accessed Nov. 02, 2021).