

Sistem Pencarian Similaritas Judul Tugas Akhir Menggunakan Metode TF-IDF

Arif Amrulloh^{1*}, Ipam Fuaddina Adam²

¹Rekayasa Perangkat Lunak, Institut Teknologi Telkom Purwokerto

²Teknik Informatika, Institut Teknologi Telkom Purwokerto

Jl. DI Panjaitan No 128 Purwokerto Selatan, 53147

amrulloh@ittelkom-pwt.ac.id¹, ipam@ittelkom-pwt.ac.id²

Abstract. The final project is a compulsory subject that must be completed by students to get a degree according to the level of higher education taken. Often students or final project supervisors find it difficult to find out whether the title of the proposed final project has been researched before or not. To avoid the title of the final project is the same as the research that has been done previously, a solution is needed, a search system for the similarity of the title of the final project can be used as a solution to the problem. Many methods can be used for the problem of finding similarity or similarity of data. One method that is often used is TF-IDF which works by calculating the weights. In this research, a system for finding the similarity of the title of the final project will be built by applying the TF-IDF method. The results of the TF-IDF trial with 384 data for the title of the final project, found that the title of the final project that has similarities based on keywords is 99 data. The use of TF-IDF method can be used as a solution for lecturers and students in determining the title of the final project.

Keywords: Final Project, Similarity, TF-IDF

Abstrak. Tugas akhir merupakan mata kuliah wajib yang harus diselesaikan oleh mahasiswa untuk mendapatkan gelar sesuai jenjang pendidikan tinggi yg ditempuh. Seringkali mahasiswa atau dosen pembimbing tugas akhir kesulitan untuk mengetahui apakah judul tugas akhir yang diajukan sudah pernah diteliti sebelumnya atau belum. Untuk menghindari judul tugas akhir yang sama dengan penelitian yg telah dilakukan sebelumnya maka diperlukan sebuah solusi, sistem pencarian kemiripan judul tugas akhir bisa dijadikan sebagai solusi dari permasalahan tersebut. Banyak metode yang dapat digunakan untuk permasalahan pencarian similaritas atau kemiripan data. Salah satu metode yang sering digunakan adalah TF-IDF yang bekerja dengan cara menghitung pembobotan. Pada penelitian ini akan dibangun sebuah sistem pencarian similaritas judul tugas akhir dengan menerapkan metode TF-IDF. Hasil ujicoba TF-IDF dengan judul tugas akhir sebanyak 384 data ditemukan judul tugas akhir yang memiliki kesamaan berdasarkan kata kunci sebanyak 99 data. Penggunaan metode TF-IDF bisa dijadikan sebagai solusi bagi dosen dan mahasiswa dalam menentukan judul tugas akhir.

Keywords: *Similarity*, Tugas Akhir, TF-IDF

Received September 2021 / **Revised** November 2021 / **Accepted** December 2021

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



PENDAHULUAN

Tugas Akhir (TA) atau Skripsi merupakan salah satu mata kuliah wajib yang harus diselesaikan oleh mahasiswa tingkat sarjana pada Perguruan Tinggi Negeri maupun Swasta di Indonesia [1]. Banyak mahasiswa yang menyelesaikan TA dengan cara yang tidak kompeten atau tidak layak, yaitu dengan cara *copy paste* dari artike-artikel yang sudah ada sebelumnya atau disebut plagiat [2]. Dalam menulis karya ilmiah tentu ada tantangan agar tulisan yang dibuat orisinal dan tidak plagiat terhadap karya orang lain [3]. Tindakan plagiarisme sering terjadi pada jurnal maupun artikel ilmiah [4]. Unsur plagiarisme dalam sebuah penelitian bisa berasal dari judul maupun isi laporan. [5]. Bila tidak dideteksi, TA yang terus bertambah

^{1*} Corresponding author.

Email addresses: 1st amrulloh@ittelkom-pwt.ac.id (amrulloh), 2nd ipam@ittelkom-pwt.ac.id (adam)

DOI: 10.24014/coreit.v7i2.15338

dari tahun ke tahun dapat menyebabkan penelitian yang sudah ada akhirnya diteliti kembali tanpa mengembangkan penelitian sebelumnya [6].

Untuk menghindari kegiatan plagiat maka perlu diketahui apakah judul TA yang akan dibuat sudah pernah dilakukan sebelumnya atau belum dengan mendeteksi kemiripan atau kesamaan judul TA. Deteksi similaritas dilakukan untuk mencari dan memeriksa kesamaan beberapa dokumen dengan dokumen lain [7]. Untuk menentukan jauh dekatnya hubungan kekerabatan dapat dilakukan dengan menentukan kesamaan (*resemblance* atau *similarity*) [8]. Similaritas merupakan ukuran kedekatan antara satu objek dengan objek lainnya [9], untuk melacak bagaimana hubungan antara objek data yang digunakan [10], menentukan nilai kemiripan antara objek yang sudah ada dengan objek baru yang akan dicari solusinya [11].

Kemudian untuk mencari kemiripan antara objek satu dengan objek lainya diperlukan sebuah metode, metode TF-IDF dapat digunakan untuk mengatasi permasalahan similaritas. Metode TF-IDF merupakan suatu mekanisme pencarian [12] yang akan menghitung nilai *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) pada setiap token (kata) di setiap dokumen [13]. TF-IDF merupakan suatu cara untuk memberikan bobot hubungan suatu kata (*term*) terhadap dokumen [14] [15]. TF-IDF menghasilkan pencarian dokumen sesuai tingkat pembobotan berdasarkan *query* yang dimasukan oleh pengguna [16].

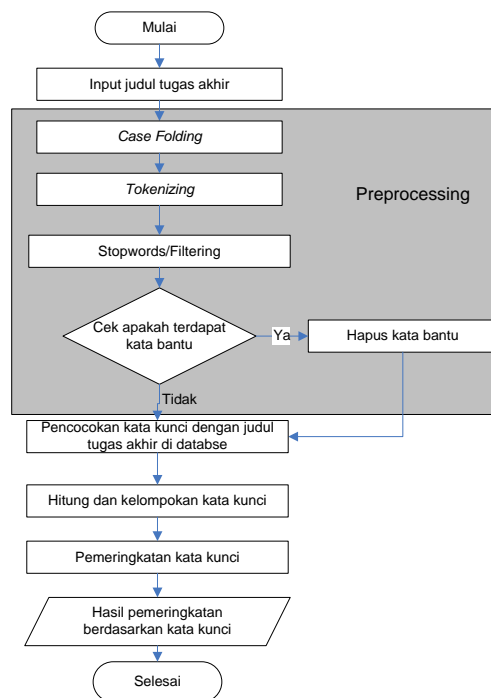
Kajian pembobotan judul tugas akhir menggunakan metode TF-IDF telah dilakukan dalam [17]. TF-IDF digunakan untuk pembobotan dokumen skripsi. Berbeda dengan kajian TF-IDF sebelumnya, Metode TF-IDF pada penelitian ini digunakan untuk pembobotan judul tugas akhir pada text yang sudah tersimpan dalam *database*. Tujuan dari penelitian ini adalah membangun sebuah sistem yang dapat menghitung secara otomatis dengan cara pemeringkatan judul tugas akhir berdasarkan kata kunci menggunakan metode TF-IDF. Diharapkan sistem yang dikembangkan pada penelitian ini dapat membantu mahasiswa dan dosen pembimbing dalam menentukan judul tugas akhir.

METODE

Metode yang digunakan pada penelitian ini adalah metode pengembangan sistem. Adapun langkah-langkahnya adalah sebagai berikut:

1. Pengembangan aplikasi berbasis web menggunakan PHP dengan menerapkan metode TF-IDF.
2. Pengumpulan data judul tugas akhir.
3. Implementasi dan pengujian sistem dengan melakukan pengujian terhadap aplikasi yang telah dibuat dengan menerapkan metode TF-IDF.

Data yang digunakan pada penelitian ini adalah data judul tugas akhir yang diambil dari basis data di salah satu perguruan tinggi yang berada di Purwokerto, dimana proses perkuliahan untuk mata kuliah tugas akhir sudah tersistem mulai dari pengajuan proposal, bimbingan *online*, pendaftaran sidang sampai proses *input* nilai. Dengan sistem tersebut maka semua judul tugas akhir yang sudah disetujui tersimpan didalam *database*. Proses pencarian judul tugas akhir dapat dilihat pada Gambar 1.



Gambar 1. Flowcart proses pencariia judul Tugas Akhir

Proses pencarian judul tugas akhir dilakukan dalam beberapa tahap yaitu: Pertama menginputkan judul tugas akhir. Kedua adalah tahap *preprocessing* yang terdiri dari [17]:

1. *Case folding* yaitu proses perubahan huruf dari huruf kapital menjadi huruf kecil.
2. *Tokenizing* yaitu proses pemecahan judul tugas akhir berdasarkan karakter spasi yang menghasilkan baris kata dalam bentuk *array*.
3. *Stopwords/Filtering* yaitu proses pengecekan dan penghapusan kata bantu yang tidak penting pada deskripsi melalui pengecekan kata-kata hasil *parsing* deskripsi apakah termasuk di dalam daftar kata tidak penting (*stoplist*). Jika termasuk didalam *stoplist* maka kata-kata tersebut akan dihapus sehingga menyisakan kata-kata yang dianggap sebagai kata-kata penting atau *keywords*.

Kata bantu merupakan salah satu jenis kata dalam penyusunan kalimat, dimana dalam penyusunan kalimat jenis kata dikategorikan menjadi kata konkrit dan kata abstrak. Kata konkrit dibagi menjadi kata benda, kata kerja, kata kerja bantu, kata sifat dan kata bilangan dan kata ganti. Sedangkan kata abstrak terdiri dari kata depan, kata sambung, kata bantu, kata seru dan kata tiruan bunyi [18].

Selain kategori kata dalam [18], kata umum juga dimasukkan kedalam tabel kata bantu. Kata umum adalah kata yang memiliki keluasan cakupan makna, semakin luas cakupan makna sebuah kata maka sifatnya semakin umum [19]. Kata umum merupakan kata yang ruang lingkup maknanya lebih luas [20].

Baris kata yang diperoleh pada proses kedua akan dicek apakah terdapat kata bantu atau tidak. Pengecekan kata bantu dilakukan dengan cara mencocokkan baris kata dengan kata bantu yang sudah disimpan dalam *database*. Kata bantu yang digunakan sebagai parameter pada penelitian ini dapat dilihat pada Tabel 1.

Tabel 1. Kata bantu

NO	Kata bantu
1	Algoritma
2	Dengan
3	Implementasi
4	Menggunakan
5	Metode
6	Pada
7	Penerapan

8	Yang
9	Rancang
10	Bangun
11	Sistem

Jika ditemukan kata bantu maka kata bantu tersebut akan dihapus dari baris kata, sebagai contoh pencarian judul tugas akhir yang terdiri dari sembilan baris kata “sensor inframerah dan mikrokontroler arduino uno dengan *frequency identification*”. Pada judul tersebut terdapat dua kata bantu yang di block dengan warna kuning “dan,dengan”, dua kata bantu tersebut dihapus sehingga menyisakan tujuh baris kata yaitu “sensor inframerah dan mikrokontroler arduino uno dengan *frequency identification*”.

Ketiga adalah pencocokan baris kata dengan data judul tugas akhir yang tersimpan didalam *database*. Setiap ditemukan kecocokan maka akan diberi nilai dengan bobot 1, apabila dalam satu judul ditemukan lebih dari satu kata kunci maka bobot yang diberikan adalah 1 dikali jumlah kata kunci yang muncul.

Keempat adalah proses penghitungan judul tugas akhir yang memiliki kecocokan dengan kata kunci, setelah dihitung kemudian akan dilakukan pengelompokan nilai berdasarkan kata kunci. Kelima adalah proses pemeringkatan dengan cara menjumlahkan bobot nilai dari masing-masing kata yang sudah dikelompokkan. Proses pembobotan dilakukan dengan menggunakan metode yang sudah dipilih yaitu TF-IDF.

Proses pemeringkatan dilakukan dengan cara menghitung banyaknya jumlah kata yang ditemukan pada judul tugas akhir yang memiliki kesamaan dengan kata kunci yang dimasukan. Kemudian judul tersebut diberi bobot sesuai dengan jumlah kata yang muncul. Pemeringkatan dilakukan dengan cara mengurutkan nilai bobot tertinggi ke bobot terendah. Untuk memudahkan pengguna, informasi mengenai jumlah kata kunci yang muncul dan bobot nilai dari masing-masing judul yang ditemukan ditampilkan dibawah judul seperti pada Tabel 3.

Sistem pencarian judul tugas akhir yang diusulkan menggunakan metode TF-IDF dimana judul tugas akhir akan diberi bobot berdasarkan kata kunci yang di masukan. Metode TF-IDF menggabungkan dua konsep untuk perhitungan bobot, yaitu frekuensi kemunculan sebuah kata dan *inverse* frekuensi data yang mengandung kata tersebut. Frekuensi kemunculan kata di dalam data tersebut menunjukkan seberapa penting kata itu. Frekuensi data yang mengandung kata tersebut menunjukkan seberapa umum kata tersebut. Sehingga bobot hubungan antara sebuah kata dan sebuah data akan tinggi apabila frekuensi kata tersebut tinggi. Rumus yang digunakan untuk TF-IDF adalah [17]:

$$tf = 0,5 + 0,5 \times \frac{tf}{\max(tf)} \quad (1)$$

$$idf = \log \left(\frac{D}{df} \right) \quad (2)$$

$$W_{d,t} = tf_{d,t} \times idf_{d,t} \quad (3)$$

Keterangan:

D = Data ke-d

t = term ke-t dari data

W = bobot data ke-d terhadap term ke-t

tf = banyaknya *term* i pada sebuah data

idf = *Inversed Document Frequency*

df = banyaknya data yang mengandung *term* i

Metode pemeringkatan yang digunakan pada penelitian ini adalah dengan cara menjumlahkan kesamaan similaritas antara judul tugas akhir dengan kata kunci yang diinputkan. Setiap kata kunci yang diinputkan akan dicek apakah terdapat kata bantu, jika terdapat kata bantu maka kata tersebut akan dihilangkan atau tidak dihitung kedalam pemeringkatan. Kata bantu ditentukan terlebih dahulu dengan cara disimpan didalam *database*. Setiap kali dilakukan pencarian judul tugas akhir maka kata kunci yang diinputkan akan

dicek apakah terdapat kata bantu, jika ditemukan kata bantu maka kata bantu tersebut akan dihilangkan dan tidak dihitung.

HASIL DAN PEMBAHASAN

A. Hasil Pengujian

Pada tahap pengujian dilakukan pencarian judul tugas akhir menggunakan kata kunci “sensor inframerah dan mikrokontroler arduino uno dengan *frequency identification*”. Dari hasil pencarian diperoleh hasil pembobotan seperti pada Tabel 2, dan ditemukan similaritas judul tugas akhir sebanyak 99 data seperti pada Tabel 3.

Tabel 2. Hasil pembobotan

No	Kata Kunci	Jumlah kata ditemukan
1	sensor	24
2	inframerah	5
3	mikrokontroler	17
4	arduino	63
5	uno	30
6	frequency	9
7	identification	4

Tabel 3. Hasil pencarian similaritas judul Tugas Akhir

No	Judul Tugas Akhir
1	rancang bangun sistem smart control room berbasis arduino uno pada ruang tertutup design of smart control room system based on arduino uno in closed room <i>Similaritas : [Arduino : 2 ,Uno : 2]</i> <i>Bobot = 4</i>
2	rancangan sabuk pengemudi mobil menggunakan sensor inframerah dan mikrokontroler berbasis report short message service (sms) <i>Similaritas : [sensor : 1 ,Inframerah : 1 ,Mikrokontroler : 1]</i> <i>Bobot = 3</i>
3	rancang bangun sistem pengurusan dan pengisian air dalam kolam renang menggunakan sensor ultrasonik berbasis arduino uno <i>Similaritas : [Sensor : 1 ,Arduino : 1 ,Uno : 1]</i> <i>Bobot = 3</i>
4	rancang bangun sistem pengamanan jaringan ruangan menggunakan sensor pir (passive infrared receiver) berbasis arduino uno r3 <i>Similaritas : [Sensor : 1 ,Arduino : 1 ,Uno : 1]</i> <i>Bobot = 3</i>
5	rancang bangun penyiraman tanaman otomatis berbasis mikrokontroler arduino uno dengan report sms <i>Similaritas : [Mikrokontroler : 1 ,Arduino : 1 ,Uno : 1]</i> <i>Bobot = 3</i>
6	rancang bangun alat bantu tuna netra menggunakan sensor jarak ultrasonik berbasis arduino uno <i>Similaritas : [Sensor : 1 ,Arduino : 1 ,Uno : 1]</i> <i>Bobot = 3</i>
7	rancang bangun alat bantu parkir mobil menggunakan sensor jarak ultrasonik berbasis arduino uno <i>Similaritas : [sensor : 1 ,Arduino : 1 ,Uno : 1]</i> <i>Bobot = 3</i>
8	thermo-hygrometer digital berbasis arduino uno menggunakan sensor dht 22 <i>Similaritas : [Arduino : 1 ,Sensor : 1]</i> <i>Bobot = 2</i>
9	sistem pengamanan pintu shelter bts otomatis menggunakan password berbasis ponsel cerdas android, modul bluetooth dan arduino uno

	<i>Similaritas : [Arduino : 1 ,Uno : 1] Bobot = 2</i>
10	sistem keamanan pintu ruangan menggunakan rfid dan personal identification number (pin) berbasis arduino <i>Similaritas : [Identification : 1 ,Arduino : 1] Bobot = 2</i>
11	rancang bangun sistem pengunci brankas menggunakan autentifikasi radio frequency identification (rfid) dan short message service (sms) <i>Similaritas : [Frequency : 1 ,Identification : 1] Bobot = 2</i>
12	rancang bangun sistem pemantauan level ketinggian air menggunakan sensor ultrasonik dengan sms sebagai media komunikasi berbasis arduino <i>Similaritas : [sensor : 1 ,Arduino : 1] Bobot = 2</i>
13	rancang bangun alat ukur pendeteksi jarak tempuh kendaraan bermotor dengan sensor efek hall tipe ugn3503 berbasis mikrokontroler atmega8 <i>Similaritas : [sensor : 1 ,mikrokontroler : 1] Bobot = 2</i>
...	
99	aplikasi inframerah untuk pengatur nyala lampu yang dipantau dengan pc <i>Similaritas : [inframerah : 1] Bobot = 1</i>

Aplikasi Pencarian Similaritas Judul Tugas Akhir

<ol style="list-style-type: none"> 1. rancang bangun sistem smart control room berbasis arduino uno pada ruang tertutup design of smart control room system based on arduino uno in closed room <i>Similaritas : [Arduino : 2 ,Uno : 2] Bobot = 4</i> <hr/> 2. rancangan sabuk pengemudi mobil menggunakan sensor inframerah dan mikrokontroler berbasis report short message service (sms) <i>Similaritas : [Sensor : 1 ,Inframerah : 1 ,Mikrokontroler : 1] Bobot = 3</i> <hr/> 3. rancang bangun sistem pengurusan dan pengisian air dalam kolam renang menggunakan sensor ultrasonik berbasis arduino uno <i>Similaritas : [Sensor : 1 ,Arduino : 1 ,Uno : 1] Bobot = 3</i> <hr/> 4. rancang bangun sistem pengamanan jaringan ruangan menggunakan sensor pir (passive infrared receiver) berbasis arduino uno r3 <i>Similaritas : [Sensor : 1 ,Arduino : 1 ,Uno : 1] Bobot = 3</i> <hr/> 5. rancang bangun penyiraman tanaman otomatis berbasis mikrokontroler arduino uno dengan report sms <i>Similaritas : [Mikrokontroler : 1 ,Arduino : 1 ,Uno : 1] Bobot = 3</i> <hr/> 6. rancang bangun alat bantu tuna netra menggunakan sensor jarak ultrasonik berbasis arduino uno <i>Similaritas : [Sensor : 1 ,Arduino : 1 ,Uno : 1] Bobot = 3</i> 	<p>Pencarian berdasarkan kata kunci</p> <table border="0" style="font-size: small;"> <thead> <tr> <th style="text-align: left;">Kata Kunci</th> <th style="text-align: left;">Judul ditemukan</th> </tr> </thead> <tbody> <tr><td>sensor</td><td>: 24</td></tr> <tr><td>inframerah</td><td>: 5</td></tr> <tr><td>mikrokontroler</td><td>: 17</td></tr> <tr><td>arduino</td><td>: 63</td></tr> <tr><td>uno</td><td>: 30</td></tr> <tr><td>frequency</td><td>: 9</td></tr> <tr><td>identification</td><td>: 4</td></tr> </tbody> </table>	Kata Kunci	Judul ditemukan	sensor	: 24	inframerah	: 5	mikrokontroler	: 17	arduino	: 63	uno	: 30	frequency	: 9	identification	: 4
Kata Kunci	Judul ditemukan																
sensor	: 24																
inframerah	: 5																
mikrokontroler	: 17																
arduino	: 63																
uno	: 30																
frequency	: 9																
identification	: 4																

Gambar 2. Tampilan aplikasi pencarian similaritas judul tugas akhir yang dikembangkan

B. Pembahasan

Penelitian ini bertujuan untuk memberikan solusi bagi dosen dan mahasiswa dalam menentukan judul tugas akhir. Pada tahap uji coba dilakukan percobaan pencarian judul tugas akhir menggunakan data sebanyak 384 judul tugas akhir. Kemudian digunakan 11 kata bantu (Tabel 1) untuk menghapus kata kunci

yang tidak dibutuhkan dalam proses pemerigkatan. Penghapusan kata bantu dilakukan karena kata bantu dapat bersifat konkrit, abstrak atau umum yang dapat digunakan oleh judul tugas akhir manapun tanpa perlu dicek similaritasnya.

Kemudian untuk memudahkan dosen dan mahasiswa dalam mengetahui seberapa banyak penelitian yang telah dilakukan oleh peneliti lain sebelumnya, dibuatlah pemeringkatan berdasarkan kata kunci yang diinputkan. Kata kunci yang diinputkan bisa berupa kalimat, tetapi pada proses pencarian, kalimat yang diinputkan akan dipecah per-kata berdasarkan spasi. Kata yang sudah dipecah kemudian akan dicek apakah mengandung kata bantu, jika mengandung kata bantu maka kata bantu tersebut akan dihapus. Kata yang masih tersisa akan digunakan sebagai parameter pencarian, setelah data judul tugas akhir ditemukan kemudian akan dihitung seberapa banyak data yang sama berdasarkan kata kunci yang sudah dipecah dengan cara di kelompokkan seperti pada Tabel 2.

Untuk membantu dosen dan mahasiswa dalam mengetahui judul apa saja yang memiliki kesamaan dengan kata kunci yang dicari, maka ditampilkan daftar judul tugas akhir yang mengandung kata kunci. Judul yang memiliki kesamaan dengan kata kunci akan diberi tanda berupa *block* warna kuning seperti pada Tabel 3.

Hasil pencarian similaritas judul tugas akhir diharapkan dapat membantu dosen dan mahasiswa dalam menghindari kegiatan yang terindikasi plagiat terhadap judul tugas akhir yang sudah pernah dilakukan sebelumnya.

Dalam penerapan metode TF-IDF pada penelitian yang penulis lakukan masih terdapat kekurangan yaitu:

1. Sistem belum bisa mendeteksi kesalahan penulisan kata kunci, sebagai contoh: penulisan "infra merah" dan "inframerah" dianggap kunci yang berbeda karena salah satunya mengandung karakter spasi.
2. Tidak bisa mendeteksi dua kata yang sebenarnya masih satu kesatuan, sebagai contoh: kata "arduino uno" dianggap dua kata yang memiliki arti sendiri-sendiri padahal kata tersebut masih satu kesatuan.

KESIMPULAN

Sistem pencarian similaritas judul tugas akhir menggunakan metode TF-IDF dapat dijadikan sebagai solusi bagi dosen dan mahasiswa dalam menentukan judul tugas akhir. Metode TF-IDF juga dapat dijadikan sebagai solusi untuk masalah pemeringkatan dan pencarian similaritas untuk kasus yang lain. Dari hasil ujicoba menggunakan kata kunci yang sudah ditentukan ditemukan kecocokan judul tugas akhir sebanyak 99 data dari total 384 judul tugas akhir. Pada sistem yang dikembangkan masih terdapat kekurangan yaitu sistem belum bisa mendeteksi kesalahan penulisan kata kunci dan belum bisa mendeteksi dua kata yang masih satu kesatuan. Penelitian lanjutan dapat dilakukan untuk memperbaiki kekurangan yang ada dengan mengkombinasikan metode TF-IDF dengan metode lain.

REFERENSI

- [1] A. M. Nurdin, "Sistem Pendeteksi Kemiripan Judul Skripsi Menggunakan Algoritma Winnowing," *Jurnal Nasional Informatika dan Teknologi Jaringan (Infotekjar)*, vol. 2, no. 1, pp. 90-97, 2017.
- [2] M. A. D. Lelawati Sibarani, "Analisa Perbandingan Sistem Pendeteksian Kemiripan Judul Skripsi Menggunakan Algoritma Winnowing Dan Algoritma Rabin Karp," *Riset dan E-Jurnal Manajemen Informatika Komputer*, vol. 4, no. 1, pp. 69-82, 2019.
- [3] F. Hermas Yuda Pamungkas, "Deteksi Similaritas Dokumen Ilmiah Menggunakan Algoritma Rabin-Karp," *Jurnal Ilmiah Informatika Komputer*, vol. 24, no. 3, pp. 209-2019, 2019.

- [4] Amardeep, "Deteksi Similaritas Artikel Ilmiah Dengan Teknik Pencocokan String Boyer Moore," *Jurnal Ilmiah Informatika Komputer*, vol. 25, no. 2, pp. 127-136, 2020.
- [5] A. L. A. J. Dewa Ayu Rai Ariantini, "Pengukuran Kemiripan Dokumen Teks bahasa Indonesia Menggunakan Metode Cosine Similarity," *E-Journal Teknik Informatika*, vol. 9, no. 1, pp. 1-8, 2016.
- [6] M. S. H. S. E. Bei Harira Irawan, "Deteksi Kemiripan Judul Skripsi Menggunakan Algoritma Levenshtein Distance Pada Kampus STMIK MIC Cikarang," *Jurnal Ilmiah Edutic*, vol. 7, no. 2, pp. 143-149, 2021.
- [7] K. B. U. Abdul Najib, "Deteksi Similaritas Dokumen Abstrak Tugas Akhir Menggunakan Metode Levenshtein Distance," *JUST TI*, vol. 10, no. 1, pp. 58-62, 2018.
- [8] A. R. F. G. J. S. P. Maisuna Kundariati, "Analisis Hubungan Kekerabatan *Drosophila* sp. (Lalat Buah) Dari Tuban, Kediri, dan Tulungagung Berdasarkan Indeks Similaritas dan Dendrogram," *JBIO: JURNAL BIOSAINS*, vol. 7, no. 1, pp. 10-17, 2021.
- [9] S. W. Anton Setiawan, "Case Based Reasoning Untuk Mendiagnosa Penyakit Dan Hama Pada tanaman Mangga Menggunakan Algoritma Simliaritas Sorgenfrei," *Jurnal DINAMIK*, vol. 23, no. 1, pp. 1-10, 2018.
- [10] H. M. M. R. Fera Anugreni, "Pengembangan Pendekatan Matriks Hubungan Untuk Pengukuran Similaritas," *Jurnal Inotera*, vol. 1, no. 1, pp. 46-50, 2016.
- [11] W. T. T. Y. Saputra, "Case Based Reasoning dalam Menentukan Titik Indikasi Gangguan Jin untuk Ruqyah Menggunakan Similaritas Braun-blanquet," *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*, vol. 6, no. 2, pp. 53-59, 2018.
- [12] ., A. P. W. U. P. Meiga Ayu Ariyanti Nur Fitroh, "Metode term frequency - invers document frequency pada mekanisme pencarian judul skripsi," *TEKNO Jurnal Teknologi, Elektro, dan Kejuruan*, vol. 28, no. 2, pp. 177-190, 2018.
- [13] E. N. Zezen Zenal Abidin, "Sistem Peringkat Teks Otomatis Multi Dokumen Kliping Artikel Berita Gempa Menggunakan Metode TF-IDF," *Jurnal Teknologi Informasi dan Komunikasi*, vol. 13, no. 1, pp. 52-69, 2020.
- [14] F. A. S. R. A. Riki Ruli A Siregar, "Aplikasi Penentuan Dosen Penguji Skripsi Menggunakan Metode TF-IDF dan Vector Space Model," *Computatio: Journal of Computer Science and Information Systems*, vol. 1, no. 2, pp. 171-186, 2017.
- [15] Y. A. N. H. Sari Wahyunita, "Analisa Sentimen Tweet Berbahasa Indonesia dengan Menggunakan Metode Pembobotan Hybrid TF-IDF pada Topik Transportasi Online," *REPOSITOR*, vol. 2, no. 2, pp. 185-192, 2020.
- [16] A. E. A. Habsyi, "Aplikasi Website Dan E-DMS Kejaksaan Negeri Kraksaan Dengan Metode TF-IDF," *Jurnal SimanteC*, vol. 8, no. 1, pp. 10-16, 2019.
- [17] D. P. E. S. Rizki Tri Wahyuni, "Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF pada Sistem Klasifikasi Dokumen Skripsi," *Jurnal Teknik Elektro*, vol. 9, no. 1, pp. 18-23, 2017.
- [18] S. Liana, "Kesalahan Penggunaan Kata Bantu Struktural "DE" Mahasiswa STBA-PA Medan," *stbapia.ac.id*, pp. 1-9, 2017.
- [19] A. Maharani, "Pemakaian Diksi Dalam Penulisan Caption Media Sosial Instagram," *diksi*, vol. 28, no. 2, pp. 179-189, 2010.
- [20] F. H. R. F. S. Mohammad Mastur, "Efektifitas Penggunaan Stoplist Kata Umum Dari Dokumen Hasil Klasifikasi Pretopology," *KONVERGENSI*, vol. 13, no. 1, pp. 1-10, 2017.