

ANALISIS PENERAPAN ALGORITMA JACCARD SIMILARITY UNTUK REKOMENDASI BUKU BERDASARKAN KATA KUNCI DAN SINONIMNYA

Alfitra Haskar¹, Muhyddin A M Hayat, S.Kom., M.T², Rizky Yuslian Bakti, S.T., MT³
^{1,2,3}Informatika, Universitas Muhammadiyah Makassar, Makassar, 90221, Indonesia
e-mail koresponden: alfitrahaskar@student.unismuh.ac.id¹, muhyddin@unismuh.ac.id²,
rizkyusliana@unismuh.ac.id³

Received: Februari 11,2025; Accepted: Maret 01, 2025; Published: Maret 30, 2025

Abstrak

Peningkatan jumlah informasi digital, termasuk katalog buku di perpustakaan dan toko buku online, membuat kebutuhan akan sistem rekomendasi semakin penting. Salah satu pendekatan yang dapat digunakan dalam sistem rekomendasi adalah algoritma Jaccard Similarity, yang mengukur kesamaan antara dua himpunan data. Dalam penelitian ini, dilakukan implementasi algoritma Jaccard Similarity untuk rekomendasi buku berdasarkan kata kunci dan sinonimnya. Penggunaan sinonim bertujuan untuk meningkatkan cakupan pencarian dan mengakomodasi variasi istilah yang mungkin digunakan pengguna saat mencari buku. Penelitian ini melibatkan beberapa tahap, mulai dari ekstraksi kata kunci dan sinonim menggunakan sumber daya linguistik seperti kamus sinonim, hingga perhitungan tingkat kesamaan antar kata kunci menggunakan algoritma Jaccard Similarity. Selanjutnya, dilakukan analisis performa algoritma dalam menghasilkan rekomendasi buku yang relevan. Hasil uji coba menunjukkan bahwa penggunaan sinonim dapat meningkatkan akurasi sistem rekomendasi dibandingkan dengan pencarian kata kunci secara eksak. Algoritma ini berhasil mengidentifikasi hubungan antara buku-buku yang relevan, meskipun menggunakan kata kunci yang bervariasi. Kesimpulan dari penelitian ini adalah bahwa algoritma Jaccard Similarity yang diimplementasikan dengan tambahan pemrosesan sinonim efektif dalam meningkatkan relevansi hasil rekomendasi buku. Pengembangan lebih lanjut dapat difokuskan pada integrasi dengan sistem pemrosesan bahasa alami untuk memperkaya kemampuan sistem rekomendasi.

Kata Kunci: Sistem Rekomendasi, Jaccard Similarity, Kata Kunci, Sinonim, Buku

Abstract

The increasing amount of digital information, including book catalogs in libraries and online bookstores, makes the need for recommendation systems increasingly important. One approach that can be used in recommendation systems is the Jaccard Similarity algorithm, which measures the similarity between two data sets. In this research, the Jaccard Similarity algorithm was implemented for book recommendations based on keywords and their synonyms. The use of synonyms aims to increase search coverage and accommodate the variety of terms that users may use when searching for books. This research involves several stages, starting from extracting keywords and synonyms using linguistic resources such as synonym dictionaries, to calculating the level of similarity between keywords using the Jaccard Similarity algorithm. Next, an analysis of the algorithm's performance was carried out in producing relevant book recommendations. The test results show that the use of synonyms can increase the accuracy of the recommendation system compared to exact keyword searches. This algorithm succeeded in identifying relationships between relevant books, even though they used varying keywords. The conclusion of this research is that the Jaccard Similarity algorithm implemented with additional synonym processing is effective in increasing the relevance of book recommendation results. Further development can focus on integration with natural language processing systems to enrich the capabilities of recommendation systems.

Keyword: Recommendation Systems, Jaccard Similarity, Keywords, Synonyms, Books.

1. Pendahuluan

Perpustakaan adalah sumber dan tempat berbagai macam buku dan bahan bacaan lainnya yang umumnya diawasi oleh organisasi untuk digunakan oleh wilayah lokal yang lebih luas untuk data dan informasi. Berbeda dengan perguruan tinggi sendiri, dalam kemajuannya seiring dengan kemajuan teknologi data, perpustakaan juga ikut berkreasi. dengan menyediakan berbagai bahan bacaan dalam struktur terkomputerisasi yang pada umumnya disebut kerangka data perpustakaan.[1][2]

E-Library Pusat Universitas Muhammadiyah Makassar merupakan Unit Pelaksana Teknis (UPT), dimana akan memerankan fungsi vital dalam keberhasilan perkuliahan baik bagi mahasiswa, dosen, juga pengguna lainnya. Perpustakaan Universitas Muhammadiyah Makassar mempunyai enam layanan yaitu Layanan administrasi melayani hal-hal yang berkaitan dengan keanggotaan, layanan sirkulasi melayani peminjaman, pengembalian dan perpanjangan buku yang dipinjam, layanan referensi melayani berbagai macam pertanyaan pengguna E-Library yang mengalami kesulitan dalam menemukan suatu informasi layanan audio visual memberikan layanan dalam mengakses informasi melalui layanan audio visual seperti internet, layanan fotocopy center melayani berbagai macam kebutuhan pengguna E-Library dalam hal fotocopy digital, yang terakhir yaitu layanan pengadaan koleksi melayani berbagai usulan pelayanan pengguna perpustakaan yang berhubungan dengan berbagai macam jenis koleksi perpustakaan. [3][4]

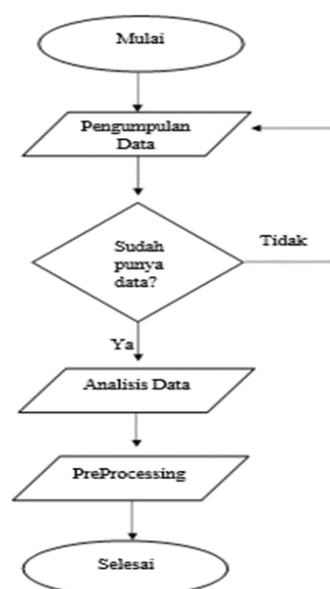
Jaccard Algorithm atau dikenal dengan Jaccard Coefficient dan atau Jaccard Similarity ialah salah satu metode yang dipakai untuk menghitung Similarity antara dua objek Metode ini berguna dalam berbagai aplikasi, termasuk sistem rekomendasi, analisis teks, pengelompokan dokumen, dan lainnya. Algoritma Jaccard Similarity sering digunakan dalam sistem rekomendasi untuk mengidentifikasi kesamaan antara item-item yang ada. Misalnya, dalam konteks rekomendasi buku, algoritma ini dapat digunakan untuk menemukan buku-buku yang memiliki kesamaan kata kunci atau topik, sehingga memungkinkan sistem rekomendasi untuk merekomendasikan buku-buku yang relevan berdasarkan preferensi pengguna[5].

Berdasarkan kondisi yang ada maka peneliti menganggap bahwa perlu adanya sistem yang mampu mendukung dan memudahkan para pengunjung perpustakaan untuk menemukan referensi buku yang dianggap sesuai dengan kata kunci yang diinginkan atau dibutuhkan dengan memanfaatkan perkembangan teknologi yang ada, yaitu menggunakan metode Algoritma Jaccard Similarity. Sebagai metode yang mampu melakukan system rekomendasi.[6]

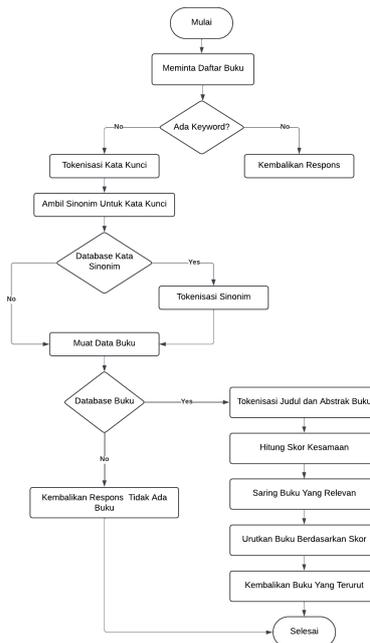
2. Metode Penelitian

2.1 Perancangan Sistem

Untuk mempermudah melakukan analisis data, peneliti merancang *Flowchart*, sehingga analisis data dapat dilakukan secara terstruktur.



Gambar 1. Flowchart Penelitian



Gambar 2. Flowchart Sistem

2.2 Teknik Pengujian Sistem

Untuk menguji sistem pencarian buku yang telah Anda buat, pengujian dapat dilakukan dengan fokus pada empat aspek utama. Pertama, koreksi ejaan dapat diuji dengan memasukkan kata kunci yang salah eja dan memverifikasi apakah sistem mengoreksi kata tersebut dengan benar menggunakan fungsi `check_spelling`. Selanjutnya, proses tokenisasi harus diuji dengan memastikan bahwa kata kunci yang diberikan diubah menjadi token yang sesuai, dan tokenisasi dari judul serta abstrak buku juga berjalan sebagaimana mestinya menggunakan fungsi `tokenize`. Kemudian, pengujian pada perhitungan kesamaan Jaccard perlu dilakukan untuk memastikan bahwa skor kesamaan yang dihasilkan sesuai dengan ekspektasi berdasarkan token yang dibandingkan, di mana fungsi `jaccard_similarity` berperan penting dalam hal ini. Terakhir, kombinasi dari skor kesamaan judul dan abstrak perlu diuji untuk memastikan bahwa skor gabungan dihitung secara akurat dan relevan, sehingga buku-buku yang paling sesuai dengan kata kunci dapat ditampilkan dengan urutan yang benar berdasarkan kesamaan tertinggi. Pengujian ini akan memastikan bahwa sistem pencarian berjalan dengan akurat dan efisien sesuai dengan yang diharapkan.[7]

3. Hasil dan Pembahasan

3.1. Implementasi Aplikasi



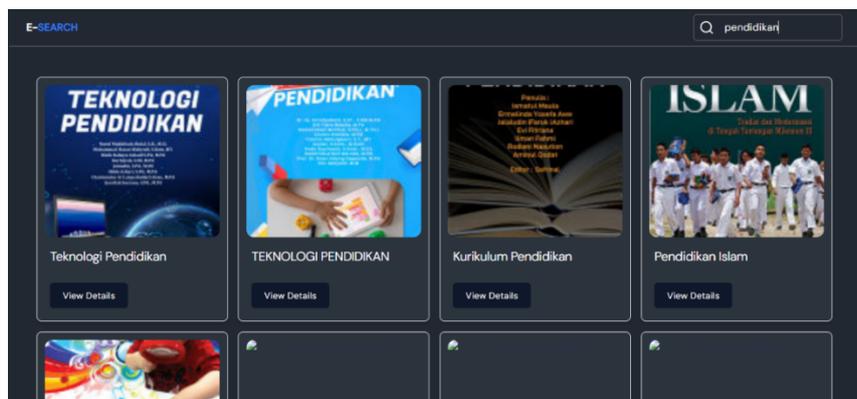
Gambar 3. Tampilan Awal Pencarian

Aplikasi ini menyediakan fitur pencarian yang memungkinkan pengguna untuk mencari buku atau referensi berdasarkan kata kunci tertentu, dalam hal ini "teknologi". Hasil pencarian ditampilkan dalam bentuk kartu atau grid, dengan setiap kartu berisi gambar sampul buku. [8]

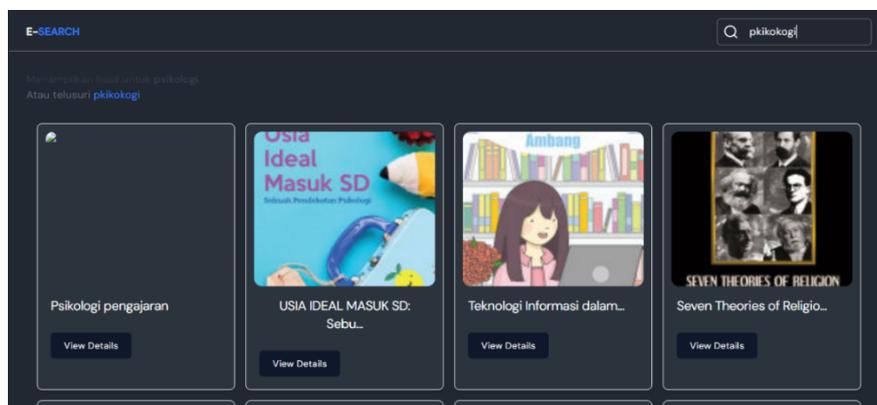
Setiap buku atau referensi yang ditemukan memiliki tombol "View Details" yang kemungkinan besar akan membawa pengguna ke halaman yang lebih rinci mengenai buku tersebut. Halaman detail mungkin mencakup informasi seperti penulis, ringkasan, tahun publikasi, dan tautan untuk mengakses atau mengunduh buku tersebut.[9]

3.2. Spell Checker

Implementasi *Spell Checker* menggunakan pustaka *SpellChecker* dari Python memungkinkan sistem untuk mendeteksi kesalahan ejaan dalam kata kunci yang dimasukkan oleh pengguna. Fitur ini penting untuk memastikan bahwa pencarian buku tidak terhambat oleh kesalahan ketik atau ejaan yang tidak benar. Saat pengguna memasukkan kata kunci yang salah, sistem secara otomatis menyarankan ejaan yang benar, yang kemudian digunakan dalam proses pencarian[10].



Gambar 4. Tampilan Hasil Pencarian



Gambar 5. Tampilan hasil sesudah koreksi ejaan, memperlihatkan perbedaan hasil yang dihasilkan oleh sistem.

```
def check_spelling(word):
    # Connect the spelling
    corrected_word = spell.correction(word)

    # Check if the connection differs from the original word
    if corrected_word != word:
        print(f"Kata '{word}' salah eja. Diduga maksud Anda '{corrected_word}'")
    else:
        print(f"Tidak ada koreksi untuk kata '{word}'.")

    return corrected_word
```

Gambar 6. Kode yang menunjukkan proses pengecekan ejaan dan koreksi otomatis

3.3. Tokenisasi dan Penggunaan Algoritma Jaccard Similarity

Proses *Tokenisasi* memecah teks (judul dan abstrak buku) serta kata kunci yang dimasukkan pengguna menjadi unit-unit kata yang lebih kecil, *Tokenisasi* ini penting untuk memungkinkan algoritma *Jaccard Similarity* bekerja dengan efektif, menghitung kesamaan antara himpunan kata kunci dan teks buku. Algoritma *Jaccard* menghitung rasio antara jumlah kata yang sama (*intersection*) dengan total kata (*union*) dalam dua teks. Hasilnya memberikan skor kesamaan yang digunakan untuk menentukan relevansi buku dengan kata kunci[11].

3.4. Perbedaan Metode Pencarian dengan Integrasi Sinonim

Sinonim dari kata kunci yang dimasukkan oleh pengguna digunakan untuk memperluas cakupan pencarian. Ini dilakukan dengan mengacu pada file `dict.json`, yang berisi daftar kata dan sinonimnya. Dengan menambahkan sinonim dalam proses pencarian, sistem dapat menemukan lebih banyak buku yang relevan, meskipun pengguna tidak menggunakan kata kunci yang tepat.[12]

```
Tidak ada koreksi untuk kata 'perang'.
```

Jenis Pencarian	Jumlah Data Ditemukan
Pencarian Biasa	8
Pencarian Sinonim Jaccard	20

Gambar 7. Contoh hasil pencarian dengan dan tanpa penggunaan sinonim, menunjukkan peningkatan relevansi hasil.

3.5. Peringkat Buku Berdasarkan Skor Kesamaan

Buku-buku yang ditemukan oleh sistem diberi peringkat berdasarkan skor kesamaan gabungan (judul dan abstrak). Skor ini dihitung dengan menggabungkan skor kesamaan judul dan abstrak yang dihitung menggunakan Algoritma *Jaccard Similarity*. Buku dengan skor tertinggi dianggap paling relevan dengan kata kunci yang dicari dan ditempatkan di urutan teratas dalam hasil pencarian[13].

```
# Tokenize book titles and abstracts
tokenized_judul_buku = [tokenize(buku["judul"]) for buku in data_buku]
tokenized_abstrak_buku = [tokenize(buku["abstrak"]) for buku in data_buku]

# Calculate Jaccard similarity score
skor_kesamaan_judul = [jaccard_similarity(judul, tokenized_keyword) for judul in tokenized_judul_buku]
skor_kesamaan_abstrak = [jaccard_similarity(abstrak, tokenized_keyword) for abstrak in tokenized_abstrak_buku]

# Combine title and abstract similarity scores
skor_kesamaan_gabungan = [skor_judul + skor_abstrak for skor_judul, skor_abstrak in zip(skor_kesamaan_judul, skor_kesamaan_abstrak)]
```

Gambar 8. menunjukkan distribusi skor kesamaan dari buku-buku yang ditemukan

```

"books": [
  {
    "judul": "Perang suci",
    "skor_kesamaan": 0.5
  },
  {
    "judul": "Peranan Islam dalam perang Aceh dan perjuangan kemerdekaan Indonesia",
    "skor_kesamaan": 0.111
  },
  {
    "judul": "Osamu Dazai: Bambu Biru dan Kisah Fantasi Lainnya",
    "skor_kesamaan": 0.01
  },
  {
    "judul": "Sapiens Grafis vol.2: Pilar-pilar Peradaban",
    "skor_kesamaan": 0.01
  },
  {
    "judul": "Islam Itu Ramah Bukan Marah",
    "skor_kesamaan": 0.01
  },
  {
    "judul": "Titik Nol",
    "skor_kesamaan": 0.007
  },
  {
    "judul": "Apa Jadinya Dunia Tanpa Islam?",
    "skor_kesamaan": 0.006
  },
  {
    "judul": "The Art of War Khalid bin Al-walid",
    "skor_kesamaan": 0.005
  }
]
    
```

Gambar 9. Hasil pencarian diatas menunjukkan beberapa judul buku yang relevan berdasarkan jumlah skor tertinggi

3.6. Sumber Data

Data yang digunakan dalam penelitian ini mencakup dua komponen utama: dataset buku dan korpus sinonim. Dataset buku berisi informasi seperti judul, abstrak, dan gambar buku yang digunakan dalam pengujian sistem rekomendasi. Dataset ini diperoleh dari perpustakaan digital Universitas Muhammadiyah Makassar, yang berisi koleksi buku dan jurnal yang tersedia secara elektronik di perpustakaan tersebut. Data ini diambil dalam format JSON dan diintegrasikan ke dalam sistem sebagai sumber utama untuk proses pencarian dan rekomendasi.[14]

Selain itu, korpus sinonim yang digunakan dalam penelitian ini disimpan dalam file dict.json. File ini dibuat dengan mengumpulkan kata-kata dan sinonim yang relevan, yang diambil dari berbagai sumber, termasuk kamus digital, literatur akademik, dan sumber lainnya yang terkait dengan terminologi informatika. Korpus sinonim ini membantu memperluas cakupan pencarian dengan memasukkan variasi sinonim dari kata kunci yang dimasukkan oleh pengguna[15][16].

1	Teknologi Informasi & Komunikasi 1	Tidak diketahui		http://books.google.com/books/content?id
2	Sistem Peralatan Hidup dan Teknologi	Prof. Dr. Alo Liliweri	Buku digital ini berjudul "Sistem Peralatan Hidup dan Teknologi", merupa	http://books.google.com/books/content?id
3	Media dan Teknologi Pembelajaran	Dr. Muhammad Yaumi,	Dalam buku ini, penulis menyajikan sepuluh pokok bahasan mulai dari Ba	http://books.google.com/books/content?id
4	TEKNOLOGI INFORMASI & MANAJEMEN	Zulkifli, S.E., M.M., Anief Budi	Dalam pengelolaan bisnis dan organisasi, Teknologi Informasi memberika	http://books.google.com/books/content?id
5	KESADARAN RUNTUH TEKNOLOGI BE	Muhammad Ilham, dkk.		http://books.google.com/books/content?id
6	Pengantar Teknologi Pendidikan	Mohamad Mustari, Habib Z	Selamat datang dalam dunia yang penuh inovasi dan transformasi, di ma	http://books.google.com/books/content?id
7	Pengantar Teknologi Informasi	Arif Muhamad Nurdin, Nort	Buku ini hadir sebagai sumber pengetahuan yang komprehensif bagi par	http://books.google.com/books/content?id
8	Antara Teknologi Dan Teologi	Mulyadi	Dr. Mulyadi, M.Hum., lahir di Bojen Pandeglang, 03 Agustus 1977. Alumni	http://books.google.com/books/content?id
9	Antara Teknologi Dan Teologi Theopa	Dr. Mulyadi, M.Hum.	Dr Mulyadi, MHum, lahir di Bojen Pandeglang, 03 Agustus 1977 Alumni P	http://books.google.com/books/content?id
10	Teknologi Dan Dampak Kebudayaan	Tidak diketahui		http://books.google.com/books/content?id
11	Pemidanaan atas Kejahatan yang Berf	Dr. Oksidelfa Yanto, S.H., M.	Lahirnya buku ini diyakini berangkat dari kegelisahan penulis dalam melit	http://books.google.com/books/content?id
12	LITERASI TEKNOLOGI DI ERA DISRUP	Ayu Nawang Wulan, dkk.		http://books.google.com/books/content?id
13	Manajemen Teknologi Pendidikan	Feri Ardiansah, Muhammad	Manajemen Teknologi Pendidikan merupakan mata kuliah kependidikan y	http://books.google.com/books/content?id

Gambar 10. Database Judul Buku

4. Kesimpulan

Hasil dari implementasi *Spell Checker*, *Tokenisasi*, *Algoritma Jaccard Similarity*, dan integrasi sinonim dalam sistem rekomendasi buku menunjukkan peningkatan signifikan dalam akurasi dan relevansi hasil pencarian. Sistem yang dibangun mampu mempersempit ruang pencarian dan meningkatkan pengalaman pengguna dalam menemukan referensi yang tepat.

Referensi

- [1] Al kautsar Muhammad Dehan. (2021). *Pemrosesan Bahasa Alami Berbasis Teks (NLP Berbasis Teks)*. <https://medium.com/@dehanalkautsar/pemrosesan-bahasa-alami-berbasis-teks-nlp-berbasis-teks-5949235effc6>
- [2] Alkaff, M., Khatimi, H., & Eriady, A. (2020). *Sistem Rekomendasi Buku Menggunakan Weighted Tree Similarity dan Content Based Filtering*. 20(1), 193–202. <https://doi.org/10.30812/matrik.v20i1.617>
- [3] Aqimuddin1, I. M., Pamungkas2, S. M., Gunawan3, C., Yaqin4, M. A., & Fauzan5, A. C. (2023). Analisis Kemiripan Model Proses Bisnis PMBoK dan Scrum menggunakan Metode Jaccard Coefficient Similarity dan Semantic Similarity. *ILKOMNIKA: Journal of Computer Science and Applied Informatics*, 5(2). <https://doi.org/10.28926/ilkomnika.v5i2.396>
- [4] Binus University. (2020). *sistem rekomendasi*. <https://mti.binus.ac.id/2020/11/17/sistem-rekomendasi-content-based>
- [5] Desena, W., & Solichin, A. (2021). Pencarian Abstrak Tugas Akhir Mahasiswa Berdasarkan Tingkat Kemiripan Menggunakan Algoritma Winnowing dan Jaccard Similarity pada Universitas Budi Luhur. *Informatik : Jurnal Ilmu Komputer*, 17(2), 112. <https://doi.org/10.52958/iftk.v17i2.3628>
- [6] Husnul, A. (2021). *Pengertian Analisis Menurut Para Ahli, Kenali Fungsi, Tujuan, dan Jenisnya*. <https://www.liputan6.com/hot/read/4569178/pengertian-analisis-menurut-para-ahli-kenali-fungsi-tujuan-dan-jenisnya?page=2>
- [7] Mayola, L., Hafizh, M., & Putra, D. M. (2024). *Algoritma Jaccard Similarity untuk Deteksi Kemiripan Judul Disertasi dengan Pendekatan Variasi Stop Word Removal*. 8, 477–487. <https://doi.org/10.30865/mib.v8i1.7109>
- [8] *pengertian kata kunci dan ciri-cirinya*. (n.d.). <https://www.merdeka.com/jatim/pengertian-kata-kunci-dan-ciri-cirinya-dalam-bahasa-indonesia-menarik-dipelajari.html>
- [9] Rachman, F. I., Alam, N., Wahyuni, T., & Anas, L. (2024). *Implementasi Perbaikan Ejaan Pada Web Semantik Ruang Baca Fakultas Teknik Dengan Menggunakan Algoritma Jaro-Winkler Distance*. 12(1).
- [10] *pengertian kata kunci dan ciri-cirinya*. (n.d.). <https://www.merdeka.com/jatim/pengertian-kata-kunci-dan-ciri-cirinya-dalam-bahasa-indonesia-menarik-dipelajari.html>
- [11] *pengertian sinonim*. (2022). <https://www.merdeka.com/jatim/pengertian-kata-kunci-dan-ciri-cirinya-dalam-bahasa-indonesia-menarik-dipelajari.html>
- [12] Rajadoli, J. V. (2022). Pembangunan Aplikasi Sistem Rekomendasi Resep Makanan Berdasarkan Bahan Yang Ada Menggunakan Teknologi Image Recognition Dan Api “Masak Apa.” In *Rajadoli* (Issue 004).
- [13] Yusriani1, Lukman Hakim2, M. K., & 1). (2018). *PELAYANAN AKADEMIK E-LIBRARY; DI UNIVERSITAS MUHAMMADIYAH MAKASSAR*. 4, 235–251. <https://journal.unismuh.ac.id/index.php/kolaborasi/article/view/1459/1262>
- [14] Hilmi Tsaqif. (2021). MENGGUNAKAN METODE WINNOWER ALGORITHM DAN SYNONYM RECOGNITION SKRIPSI Oleh : HILMI TSAQIF. *Universitas Islam Negeri Maulana Malik Ibrahim*.
- [15] Rahma, S. L., & Taufiq, U. (2024). Analisis Tingkat Akurasi Metode Pendeteksian Plagiarisme Ide dengan menggunakan Yake dan Sentence Transformer. *Journal of Internet and Software Engineering*, 5(1), 15–22. <https://doi.org/10.22146/jise.v5i1.9073>
- [16] D. Anggreani, "Grid Search Hyperparameter Analysis in Optimizing The Decision Tree Method for Diabetes Prediction," *Indonesian Journal of Data and Science*, vol. 5, no. 3, pp. 190–197, Dec. 2024.