E-ISSN : 2988-1986
https://ejournal.cahayailmubangsa.institute/index.php/kohesi

Kohesi: Jurnal Multidisiplin Saintek
Volume 10 No 8 Tahun 2026

IMPLEMENTASI VECTOR SPACE MODEL PADA SISTEM TEMU BALIK INFORMASI BERITA BENCANA ALAM DENGAN PENGURUTAN INSERTION SORT

Nety Etika Sari [1], Della Ananda [2]
[1,2] Program Studi Sistem Informasi, STMIK Kaputama Binjai
E-mail: etikasarinety@gmail.com

ABSTRACT

The large number of disaster related news articles available on online news portals makes it difficult for users to find information that is relevant to their needs. This study aims to develop an information retrieval system for disaster news using the Vector Space Model (VSM) to improve search relevance. The news data are processed through text preprocessing stages, including case folding, tokenizing, stopword removal, and stemming. Furthermore, term weighting is performed using the TF-IDF method, and the similarity between documents and queries is calculated using cosine similarity. The search results are then sorted in descending order using the insertion sort algorithm. The system is developed using the PHP programming language with a MySQL database on XAMPP. The experimental results show that document D3 achieves the highest cosine similarity value of 0.2171 for the query "the impact of flooding in Aceh on residents," indicating that it is the most relevant document. Therefore, the proposed system is able to present disaster news information more effectively and in a structured manner based on relevance.

Keywords: Information Retrieval System, Vector Space Model, Cosine Similarity, Disaster News

ABSTRAK

Banyaknya berita bencana alam yang tersedia pada portal berita online menyebabkan pengguna kesulitan dalam menemukan informasi yang relevan sesuai kebutuhan. Penelitian ini bertujuan untuk membangun sistem temu balik informasi berita bencana alam menggunakan metode Vector Space Model (VSM) guna meningkatkan relevansi hasil pencarian. Data berita diproses melalui tahapan text preprocessing yang meliputi case folding, tokenizing, stopword removal, dan stemming. Selanjutnya, pembobotan kata dilakukan menggunakan metode TF-IDF dan tingkat kemiripan antara dokumen dan query dihitung menggunakan cosine similarity. Hasil pencarian kemudian diurutkan secara descending menggunakan algoritma insertion sort. Sistem dikembangkan menggunakan bahasa pemrograman PHP dengan database MySQL pada XAMPP. Hasil pengujian menunjukkan bahwa dokumen D3 memiliki nilai cosine similarity tertinggi sebesar 0,2171 terhadap query "dampak banjir aceh terhadap warga", sehingga dinyatakan sebagai dokumen yang paling relevan. Dengan demikian, sistem yang dikembangkan

Article History
Received: Januari 2026
Reviewed:Januari 2026
Published: Januari 2026

Plagiarism Checker No 235
Prefix DOI : 10.8734/Kohesi.v1i2.365
Copyright : Author
Publish by : Kohesi

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

mampu menyajikan informasi berita bencana alam secara lebih efektif dan terstruktur berdasarkan tingkat relevansi.

Kata Kunci: Sistem Temu Balik Informasi, Vector Space Model, Consine Similarity, Berita Bencana Alam

1. PENDAHULUAN

Kemajuan teknologi saat ini memberikan dampak yang sangat signifikan terhadap media informasi. Perkembangan teknologi digital, khususnya internet dan perangkat mobile, telah mengubah cara masyarakat untuk mengakses dan menyebarkan informasi. Informasi yang sebelumnya hanya dapat diperoleh melalui media cetak, kini dapat diakses dengan mudah secara daring dalam hitungan detik. Saat ini, media informasi berbasis online berkembang pesat, salah satunya portal berita online yang menyajikan informasi populer dan khas. Portal berita online dapat diakses melalui perangkat yang terhubung dengan jaringan komputer serta memiliki keunggulan berupa informasi yang unik, selalu diperbarui, bersifat real-time, dan praktis [1]. Detik.com didirikan pada 30 Mei 1998, bertepatan dengan masa awal reformasi. Detik.com merupakan salah satu media online nasional yang secara rutin menyajikan berita bencana alam dalam jumlah yang besar dan terus diperbarui. Banyak informasi yang tersedia, termasuk berita mengenai bencana alam seperti banjir, gempa bumi, tanah longsor dan erupsi gunung berapi. Bencana Alam merupakan kejadian khusus yang merubah keteraturan kehidupan manusia. Dalam Undang-Undang Nomor 24 Tahun 2007 definisi bencana adalah peristiwa atau rangkaian peristiwa yang mengancam dan mengganggu kehidupan dan penghidupan masyarakat yang disebabkan baik oleh faktor alam dan/atau faktor non alam maupun faktor manusia sehingga mengakibatkan timbulnya korban jiwa manusia, kerusakan lingkungan, kerugian harta benda, dan dampak psikologis[2].

Meskipun portal berita online memudahkan akses informasi secara cepat dan real-time, banyaknya berita yang tersedia justru menimbulkan tantangan dalam proses pencarian informasi. Berita bencana alam yang terus diperbarui dengan topik dan konteks yang beragam sering menyebabkan pengguna kesulitan menemukan informasi yang relevan sesuai kebutuhan. Pencarian berbasis query sederhana juga belum mampu menyaring informasi secara optimal, padahal berita bencana alam memiliki dampak penting terhadap kehidupan masyarakat. Oleh karena itu, diperlukan mekanisme pencarian yang mampu mengelola dan menampilkan informasi secara lebih relevan dan terstruktur.

Pencarian informasi saat ini dilakukan dengan menggunakan mesin pencari atau sistem temu kembali informasi, user menuliskan query dan mesin pencari akan menampilkan hasil

pencarian [3]. sistem temu kembali adalah bagian dari ilmu komputer yang berkaitan dengan pengambilan informasi dari dokumen-dokumen berdasarkan isi dari dokumen-dokumen itu sendiri [4]. Dalam penelitian ini, sistem temu balik informasi diterapkan dengan menggunakan Vector Space Model (VSM) untuk merepresentasikan dokumen berita dan query dalam bentuk vektor berdasarkan bobot kata. Tingkat relevansi antara query dan dokumen dihitung menggunakan consine similiarity, sehingga dokumen yang memiliki kesesuaian tertinggi dapat diidentifikasi. Selanjutnya, algoritma insertion sort digunakan untuk mengurutkan hasil pencarian berdasarkan nilai consine similiarity tersebut, sehingga berita bencana alam yang paling relevan dapat ditampilkan terlebih dahulu secara lebih terstruktur dan efisien.

2. METODE PENELITIAN

Metode penelitian ini menjelaskan tahapan sistematis dalam membangun sistem temu balik informasi untuk berita bencana alam. Secara garis besar, proses dimulai dengan pengumpulan data berita yang kemudian diproses melalui beberapa tahapan komputasi untuk menghasilkan informasi yang relevan bagi pengguna. Alur kerja sistem ini meliputi tahap pra-pemrosesan teks (*text preprocessing*), transformasi dokumen ke dalam bentuk vektor menggunakan *Vector Space Model* (VSM), pemberian bobot kata dengan metode TF-IDF, hingga perhitungan tingkat kemiripan dokumen menggunakan *Cosine Similarity*. Sebagai tahap akhir, sistem menerapkan algoritma *Insertion Sort* untuk mengurutkan hasil pencarian agar berita yang paling relevan ditampilkan pada urutan teratas.

2.1 Text Processing

Text preprocessing merupakan tahap awal dalam sistem temu balik informasi yang bertujuan untuk membersihkan dan menormalkan data teks sebelum dilakukan proses pembobotan TF-IDF dan perhitungan cosine similiarity. Pada penelitian ini, text preprocessing dilakukan melalui beberapa tahapan, yaitu case folding, tokenizing, stopword removal, dan stemming. Case folding adalah proses mengubah semua masukan yang berupa huruf besar menjadi huruf kecil (lowercase)[5]. Selanjutnya, Tokenizing Proses penghilangan tanda baca pada kalimat yang ada dalam dokumen sehingga menghasilkan kata-kata yang berdiri masing masing[4]. Pada tahap stopword removal, dilakukan proses penghapusan kata-kata umum yang tidak memiliki makna penting sehingga tersisa kata-kata hasil tokenisasi yang dianggap relevan dan mampu mewakili isi dokumen[5]. Tahap terakhir adalah stemming yaitu Proses mengubah kata menjadi kata dasarnya dengan menghilangkan imbuhan-imbuhan pada kata dalam dokumen.

2.2 Vector Space Model (VSM)

Vector Space Model (VSM) merupakan salah satu metode yang digunakan dalam sistem temu balik informasi untuk mengukur tingkat kemiripan antara kata kunci (keyword) yang dimasukkan oleh pengguna dengan sekumpulan term yang telah dikumpulkan dari dokumen[6] Dalam model ini, dokumen dan query direpresentasikan dalam bentuk vektor numerik pada ruang berdimensi banyak, di mana setiap dimensi merepresentasikan sebuah term. VSM bekerja dengan memanfaatkan proses pembobotan term serta pengukuran kemiripan antar vektor untuk menentukan tingkat relevansi dokumen terhadap query pengguna.

2.3 TF-IDF

Term Frequency–Inverse Document Frequency (TF-IDF) merupakan metode pembobotan term yang digunakan untuk menentukan tingkat kepentingan suatu kata dalam dokumen. TF-IDF terdiri dari dua komponen utama, yaitu Term Frequency (TF) dan Inverse Document Frequency (IDF).

TF (Term Frequency) adalah frekuensi kemunculan suatu istilah *fi* di dalam sebuah dokumen *dj* dibandingkan dengan frekuensi istilah *fi* yang sering muncul pada dokumen itu. Jika dimasukkan dalam rumus matematika didapatkan:

$$\mathrm{TF_{i,j}} = \frac{f_{i,j}}{\max(f_i)}$$

Keterangan :

$TF_{i,j}$ = nilai Term Frequency istilah f_i pada dokumen d_j

$F_{i,j}$ = frekuensi kemunculan istilah f_i pada dokumen d_j

$max(f_j)$ = frenkuensi kemunculan istilah yang paling sering muncul pada dokumen d_j

IDF (Inverse Document Frequency) adalah frekuensi kemunculan suatu istilah *fi* di dalam seluruh dokumen. Penggunaan faktor IDF didasarkan pada istilah yang muncul pada setiap dokumen tidak memberikan suatu ciri khusus untuk menentukan dokumenyang relevan dari yang tidak relevan. Jika jumlah seluruh dokumen di dalam sistem dinyatakan dengan nilai N dan jumlah dokumen yang memiliki istilah *fi* tersebut dinyatakan dengan ni, maka nilai IDFi-nya dapat dinyatakan dengan[7]:

$$\mathrm{IDF_i} = \log\left(\frac{N}{n_i}\right)$$

Keterangan:

IDF_i = nilai Inverse Document Frecuency istilah $\mathbf{f_i}$

N = jumlah seluruh dokumen

n_i = jumlah dokumen yang mengandung istilah f_i

Bobot TF-IDF diperoleh dari hasil perkalian nilai TF dan IDF, yang digunakan untuk merepresentasikan tingkat kepentingan suatu istilah dalam dokumen.

$$W_{i,j} = TF_{i,j} \text{ x } IDF_i$$

Keterangan:

$W_{i,j}$ = bobot TF-IDF istilah f_i pada dokomen d_j

2.4 Cosine Similiarity

Cosine Similarity sendiri adalah salah satu metode yang bisa dimanfaatkan sebagai metode pencarian data di dalam data mining dan sering digunakan untuk mendeteksi dokumen-dokumen yang mirip, cosine similarity akan menghitung tingkat kesamaan antar dua buah atau lebih dari objek yang dinyatakan dalam vektor jumlahnya ada dua vector dengan menggunakan kata kunci (cosine). Jadi cosine similarity dapat digunakanuntuk menemukan kemiripan dokumen dalam data set dengan jumlah yang besar dan dapat lebih cepat dan sesuai saat menemukan dokumen yang dicari [8].

Dalam sistem temu balik informasi, cosine similarity digunakan untuk mengukur tingkat kemiripan antara vektor dokumen dan vektor query yang telah direpresentasikan menggunakan pembobotan TF-IDF.

$$\text{Cosine Similarity}(d_{j,q}) = \frac{\sum_{i=1}^{n} w_{i,j} \; x \; w_{i,q}}{\sqrt{\sum_{i=1}^{n} (w_{i,j})(w_{i,j})^2} \times \sqrt{\sum_{i=1}^{n} (w_{i,q})^2}}$$

Keterangan:

d_j = dokumen ke-j

q = query

$w_{i,j}$ = bobot TF-IDF istilah ke-I pada dokumen d_j

$w_{i,q}$ = bobot TF-IDF istilah ke-I pada query

n = jumlah seluruh istilah (term)

2.5 Insertion Sort

Insertion Sort adalah salah satu algoritma pengurutan yang digunakan untuk menata elemen data secara berurutan dengan cara menyisipkan setiap elemen ke posisi yang sesuai dalam daftar yang sudah terurut, sehingga menghasilkan daftar yang terstruktur dari nilai rendah ke tinggi atau sebaliknya[9].

Dalam kontekstualisasi sistem temu balik informasi, insertion sort digunakan untuk mengurutkan daftar dokumen berita berdasarkan nilai cosine similarity secara descending, sehingga dokumen dengan nilai relevansi tertinggi ditampilkan pada urutan teratas[10]. Meskipun insertion sort memiliki kompleksitas waktu yang relatif sederhana, algoritma ini cukup efektif digunakan pada jumlah data yang tidak terlalu besar dan mudah diimplementasikan. Oleh karena itu, insertion sort dipilih sebagai metode pengurutan hasil pencarian dalam sistem temu balik informasi pada penelitian ini.

3. HASIL DAN PEMBAHASAN

Sistem temu balik informasi berita bencana alam dikembangkan menggunakan bahasa pemrograman PHP dengan dukungan database MySQL pada XAMPP sebagai media penyimpanan data berita. Sistem memungkinkan pengguna untuk mengunggah dataset berita dalam format CSV serta melakukan pencarian berdasarkan kata kunci. Data yang tersimpan kemudian diproses melalui tahapan preprocessing, pembobotan TF-IDF, dan perhitungan cosine similarity untuk menentukan tingkat relevansi dokumen. Hasil pencarian ditampilkan dalam bentuk daftar berita yang telah diurutkan berdasarkan nilai cosine similarity tertinggi, sehingga memudahkan pengguna dalam memperoleh informasi yang relevan. Proses kerja sistem tersebut dijelaskan lebih lanjut pada bagian implementasi sistem.

3.1 Implementasi Sistem

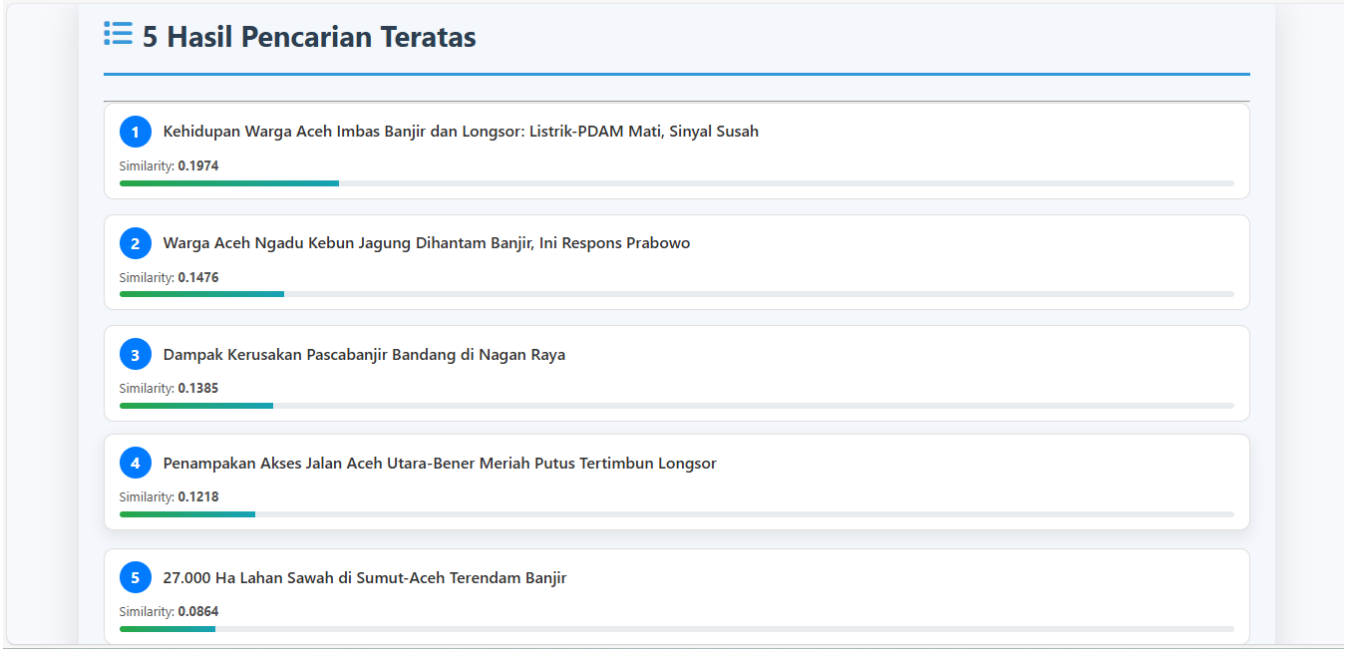
Gambar 1 : Tampilan Halaman Utama



Tampilan antarmuka sistem temu balik berita bencana dirancang sederhana dan mudah digunakan. Pada halaman utama, sistem menampilkan judul aplikasi serta deskripsi singkat yang menjelaskan fungsi sistem sebagai media pencarian informasi bencana berdasarkan dataset berita. Antarmuka terbagi menjadi dua bagian utama, yaitu fitur unggah dataset dan fitur pencarian berita.

Fitur unggah dataset digunakan untuk memasukkan data berita dalam format CSV ke dalam sistem, dengan ketentuan kolom data berupa judul dan isi berita. Data yang diunggah akan disimpan ke dalam database dan selanjutnya digunakan sebagai sumber pencarian. Sementara itu, fitur pencarian berita memungkinkan pengguna memasukkan kata kunci untuk memperoleh informasi yang relevan. Kata kunci tersebut kemudian diproses oleh sistem untuk menghasilkan daftar berita sesuai tingkat relevansinya.

Gambar 2 : Hasil pengurutan insertion sort (descending)



Setelah pengguna memasukkan kata kunci pencarian, sistem menampilkan hasil pencarian dalam bentuk daftar berita teratas. Setiap berita ditampilkan bersama nilai cosine similarity yang menunjukkan tingkat kemiripan antara query dan dokumen berita. Daftar hasil pencarian diurutkan secara menurun (descending) menggunakan algoritma insertion sort, sehingga berita dengan nilai cosine similarity tertinggi ditampilkan pada urutan pertama. Dengan mekanisme ini, sistem mampu menyajikan informasi berita bencana yang paling relevan sesuai dengan kebutuhan pengguna.

3.2 Hasil Pengujian Ssitem

Pengujian sistem dilakukan untuk mengetahui kinerja sistem temu balik informasi dalam menampilkan berita bencana yang relevan berdasarkan kata kunci pencarian. Pada pengujian ini digunakan query “dampak banjir aceh terhadap warga” yang diproses terhadap beberapa dokumen berita.

Tahap awal pengujian adalah data preprocessing, di mana dokumen dan query melalui proses case folding, tokenizing, stopword removal, dan stemming. Hasil preprocessing menghasilkan sekumpulan term penting yang merepresentasikan isi masing-masing dokumen, sebagaimana ditunjukkan pada tabel preprocessing dokumen.

Tabel 1. Dokumen yang sudah di preprocessing

Dokumen	Preprocessing
D_1	warga, aceh, ngadu, kebun, jagung, hantam, banjir, respons, prabowo, presiden, subianto, temu, korban, bandang, posko, ungsi, desa, bambel, tenggara, keluh, rusak, aspirasi, nyata, perintah, buat, baik, rakyat
D_2	gempa, sinabang, aceh, bumi, kuat, magnitudo, dalam, bmkg, pusat, darat, barat, laut, pukul, wib, imbau, warga, waspada, potensi, susulan
D_3	hidup, warga, aceh, imbas, banjir, longsor, listrik, pdam, mati, sinyal, susah, padam, sulit, ponsel, komunikasi, warung, kopi, internet, telepon, ganggu, distribusi, air, langka, bbm, bahan, makan, dampak, butuh, bantu, logistik, bersih
D_4	gunung, semeru, erupsi, pagi, tinggi, awan, panas, kolom, letus, seismograf, status, siaga, masyarakat, kawasan, rawan, potensi, gugur, lava, lahar, jawa, timur

Selanjutnya, sistem melakukan pembobotan kata menggunakan metode TF-IDF untuk menentukan tingkat kepentingan setiap term dalam dokumen. Berdasarkan hasil pengujian, term seperti warga, aceh, dan banjir memiliki bobot yang relatif lebih tinggi pada dokumen yang relevan dengan query sehingga berpengaruh terhadap nilai kemiripan dokumen. Dari hasil preprocessing diperoleh total 92 term, namun pada penelitian ini hanya ditampilkan 10 term sebagai contoh representatif untuk mempermudah penyajian dan analisis, sementara seluruh term tetap digunakan dalam proses perhitungan cosine similarity.

Tabel 2. Hasil perhitungan TF-IDF

No	Term	D1	D2	D3	D4
1	warga	0.1249	0.124938737	0.124938737	0
2	aceh	0.1249	0.124938737	0.124938737	0
3	ngadu	0.6021	0	0	0
4	kebun	0.6021	0	0	0
5	jagung	0.6021	0	0	0
6	hantam	0.6021	0	0	0
7	banjir	0.301	0	0.301029996	0
8	respons	0.6021	0	0	0
9	prabowo	0.6021	0	0	0
10	presiden	0.6021	0	0	0

Tahap berikutnya adalah perhitungan cosine similarity untuk mengukur tingkat kemiripan antara query dan masing-masing dokumen. Hasil perhitungan menunjukkan bahwa dokumen D3 memiliki nilai cosine similarity tertinggi sebesar 0,2171, diikuti oleh D1 dan D2, sedangkan D4 memiliki nilai 0 yang menunjukkan tidak adanya kemiripan dengan query.

Tabel 3. Hasil perhigungan consine similarity

Dokumen	Cosine Similarity
D1	0.0690267
D2	0.02386168
D3	0.21714698
D4	0

Nilai cosine similarity tersebut kemudian diurutkan menggunakan algoritma insertion sort secara descending. Hasil pengurutan menunjukkan bahwa dokumen dengan nilai kemiripan tertinggi ditempatkan pada urutan teratas. Dengan demikian, sistem berhasil menampilkan berita yang paling relevan sesuai dengan kata kunci pencarian yang dimasukkan oleh pengguna.

Tabel 4. Hasil pengurutab insertion sort secara descending

Dokumen	Cosine Similarity
D_1	0.21714698
D_2	0.0690267
D_3	0.02386168
D_4	0

3.4 Pembahasan

Berdasarkan hasil pengujian, sistem temu balik informasi berita bencana yang dikembangkan mampu menentukan tingkat relevansi dokumen terhadap kata kunci pencarian menggunakan metode Vector Space Model (VSM). Proses preprocessing membantu menghasilkan term yang lebih representatif sehingga memengaruhi hasil pembobotan TF-IDF dan perhitungan cosine similarity.

Hasil pembobotan TF-IDF menunjukkan bahwa term yang relevan dengan query memiliki bobot lebih tinggi dan berkontribusi terhadap nilai cosine similarity. Dokumen dengan nilai cosine similarity tertinggi dianggap paling relevan, sedangkan nilai yang rendah menunjukkan keterkaitan yang minim. Penggunaan algoritma insertion sort berhasil mengurutkan dokumen berdasarkan tingkat relevansi secara descending, sehingga hasil pencarian dapat ditampilkan secara lebih efektif. Dengan demikian, hasil ini menunjukkan bahwa metode Vector Space Model (VSM) efektif untuk pencarian berita bencana berbasis teks.

KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, penerapan metode Vector Space Model (VSM) pada sistem temu balik informasi berita bencana alam mampu menentukan tingkat relevansi dokumen terhadap kata kunci pencarian dengan baik. Proses text preprocessing, pembobotan TF-IDF, dan perhitungan cosine similarity berperan penting dalam merepresentasikan hubungan antara query dan dokumen berita.

Hasil pengujian menunjukkan bahwa dokumen D3 memiliki nilai cosine similarity tertinggi sebesar 0,2171, sehingga dinyatakan sebagai dokumen yang paling relevan terhadap query “dampak banjir aceh terhadap warga”. Dokumen D1 dan D2 memiliki nilai kemiripan yang lebih rendah, sedangkan dokumen D4 memperoleh nilai 0 yang menunjukkan tidak adanya keterkaitan dengan query. Penggunaan algoritma insertion sort secara descending berhasil

mengurutkan dokumen berdasarkan tingkat relevansi tersebut, sehingga sistem mampu menampilkan berita yang paling relevan pada urutan teratas dan membantu pengguna memperoleh informasi secara lebih efektif dan terstruktur.

DAFTAR PUSTAKA

[1] H. Jogja and D. A. N. Tribun, "PADA PORTAL BERITA ONLINE DI YOGYAKARTA," vol. 1, no. 1, pp. 59–69, 2021.

[2] N. Rohmaningtyas, "Wakaf Dan Bencana Alam di Indonesia," vol. 4, pp. 82–91, 2021.

[3] D. A. N. T. Komputer, S. Heristian, H. Aji, A. Kautsar, and A. Sayfulloh, "RANCANG BANGUN INFORMATION RETRIEVAL SYSTEM (IRS) KAMUS BAHASA-SUNDA . COM DENGAN METODE VECTOR SPACE MODEL (VSM)," vol. 5, no. 1, pp. 65–72, 2019.

[4] A. Noviriandini, D. A. Ambarsari, and F. Aprian, "Information Retrieval System Pada Pencarian File Dokumen Berbasis Teks Dengan Metode Vector Space Model," vol. 1, no. 1, pp. 53–57, 2020.

[5] K. Frinta and P. P. Adikara, "Pencarian Berita Berbahasa Indonesia Menggunakan Metode BM25," vol. 3, no. 3, pp. 2589–2595, 2019.

[6] A. Sucipto, "Temu Kembali Informasi Menggunakan Metode Vector Space Model Pada Majalah Suara Muhammadiyah Periode 2010 - 2015," vol. 21, no. 2, 2021.

[7] I. T. Hapsari *et al.*, "APLIKASI INFORMATION RETRIEVAL UNTUK PENCARIAN," pp. 23–28, 2015.

[8] I. P. G. H. Suputra, I. M. W. Wirawan, D. K. P. T. Aji, and A. A. S. Trisnajayanti, "Sistem Temu Kembali Informasi Untuk Mendukung Layanan Frequently Asked Questions (Faq) Pada Platform Sistem Informasi Terintegrasi," *J. Pendidik. Teknol. Dan Kejuru.*, vol. 22, no. 2, pp. 205–215, 2025.

[9] A. Kusumaningrum, "PERBANDINGAN KECEPATAN ANTARA SELECTION SORT, INSERTION SORT, DAN BUBBLE SORT," vol. 3, 2010.

[10] M. Rizky, S. Barokah, T. Saputra, and T. Sutabri, "Perbandingan Kinerja Algoritma Bubble Sort dan Insertion Sort dalam Pengurutan Data Penjualan UMKM," vol. 5, no. 1, pp. 184-

195, 2025.