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

Kohesi: Jurnal Multidisiplin Saintek
Volume 10 No 8 Tahun 2026

EVALUASI SISTEM TEMU BALIK INFORMASI ARTIKEL BERITA KECELAKAAN JALAN TOL MENGGUNAKAN TF-IDF DAN COSINE SIMILARITY

Meliyani[1], Nur Hafizah[2], Raudho Khoirani[3], Della Yulia Citra[4], Dwi Winda Salsabila[5]
[1,2,3,4,5] Sistem Informasi, STMIK Kaputama Binjai
E-mail: meliyani.business@gmail.com[1], fizahafizah763@gmail.com[2], raudhokhairani@gmail.com[3], dellayuliacitra8@gmail.com[4], dwiwindasalsabila26@gmail.com[5]

ABSTRACT

The rapid growth of digital information demands an information retrieval system capable of presenting relevant documents efficiently. In the context of online news, especially related to toll road accidents, users often have difficulty finding articles that match their search needs due to the diversity of text structures and the large number of terms that have similar contexts. This study proposes the implementation of the Term Frequency-Inverse Document Frequency (TF-IDF) method which is used to form a vector representation of each document based on the level of word importance, as well as the Cosine Similarity algorithm to calculate the closeness between user queries and news articles contained in the database. The research process begins with news data collection, followed by preprocessing such as text cleaning, tokenization, stopword removal, and stemming. TF-IDF weights are then calculated to build a search model that is more sensitive to word meaning than simple lexical matching. Test results show that the combination of TF-IDF and Cosine Similarity is able to produce more accurate document rankings, indicated by the increased relevance of search results to user queries. The system also provides a fast response even though the number of documents increases. Through this approach, the research is expected to contribute to the development of more effective text-based news retrieval methods, particularly in the domain of information regarding toll road accidents.

Keywords: TF-IDF, Cosine Similarity, Information Retrieval System, News Search, Highway Accidents, Text Processing, Vector Representation

ABSTRAK

Pertumbuhan informasi digital yang semakin cepat menuntut adanya sistem temu balik informasi yang mampu menyajikan dokumen relevan secara efisien. Dalam konteks berita daring, terutama terkait peristiwa kecelakaan di jalan tol, pengguna sering mengalami kesulitan menemukan artikel yang sesuai dengan kebutuhan pencarian karena keragaman struktur teks dan banyaknya istilah yang memiliki kemiripan konteks. Penelitian ini mengusulkan implementasi metode Term Frequency-Inverse Document Frequency (TF-IDF) yang digunakan untuk membentuk representasi vektor setiap dokumen berdasarkan tingkat kepentingan kata, serta algoritma Cosine Similarity untuk

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

menghitung kedekatan antara kueri pengguna dan artikel berita yang terdapat dalam basis data. Proses penelitian dimulai dari tahap pengumpulan data berita, dilanjutkan dengan praproses seperti pembersihan teks, tokenisasi, penghapusan stopword, dan stemming. Bobot TF-IDF kemudian dihitung untuk membangun model pencarian yang lebih sensitif terhadap makna kata dibandingkan pencocokan leksikal sederhana. Hasil pengujian menunjukkan bahwa kombinasi TF-IDF dan Cosine Similarity mampu menghasilkan peringkat dokumen yang lebih akurat, ditunjukkan oleh meningkatnya relevansi hasil pencarian terhadap kueri pengguna. Sistem yang dibangun juga memberikan respons yang cepat meskipun jumlah dokumen bertambah. Melalui pendekatan ini, penelitian diharapkan dapat memberikan kontribusi terhadap pengembangan metode pencarian berita berbasis teks yang lebih efektif, khususnya pada domain informasi mengenai kecelakaan di jalan tol.
Kata Kunci: TF-IDF, Cosine Similarity, Sistem Temu Balik Informasi, Pencarian Berita, Kecelakaan Jalan Tol, Pemrosesan Teks, Representasi Vektor

1. PENDAHULUAN

Perkembangan media digital telah mendorong lonjakan besar dalam jumlah artikel berita yang tersedia secara online. Hampir setiap menit, berita baru dirilis, terutama terkait dengan peristiwa-peristiwa terkini seperti kecelakaan di jalan raya. Tantangan utama justru terletak pada pencarian artikel relevan, walau informasi yang dibawa amat penting bagi masyarakat. Hal ini disebabkan oleh berbagai portal berita yang memiliki gaya penulisan, struktur paragraf, dan pilihan kata yang bervariasi. Sebagai hasilnya, pencarian yang hanya mengandalkan kata kunci sering kali tidak menghasilkan informasi yang tepat.

Untuk menangani masalah ini, diperlukan sistem yang dikenal sebagai Sistem Temu Balik Informasi (STBI) yang bisa mengolah serta mengatur kembali dokumen berdasarkan seberapa relevannya dengan kebutuhan pengguna. Salah satu metode yang sering dipakai adalah model ruang vektor (Vector Space Model), di mana dokumen dan permintaan diubah menjadi bentuk representasi matematis. Dalam konteks ini, pendekatan pembobotan Term Frequency–Inverse Document Frequency (TF-IDF) berfungsi untuk mengevaluasi sejauh mana sebuah istilah penting dalam dokumen dengan mempertimbangkan seberapa sering istilah itu muncul dan seberapa jarang ia terlihat dalam kumpulan dokumen lainnya. Pendekatan ini dianggap berhasil karena dapat menyoroti istilah yang benar-benar signifikan dalam konteks dokumen.

Setelah dokumen diubah menjadi vektor, kita dapat menghitung sejauh mana kemiripannya dengan kueri menggunakan Cosine Similarity, yang mengukur sudut antara dua

vektor. Metode ini telah terbukti efektif dalam meningkatkan ketepatan pencarian dokumen dalam berbagai penelitian, terutama dalam bidang seperti pencarian berita, sistem rekomendasi, dan pengelompokan teks. Namun, penggunaan kedua metode ini secara spesifik untuk mencari artikel berita tentang kecelakaan di jalan tol masih jarang dibahas, meskipun topik ini memiliki ciri bahasa yang cukup unik, seperti istilah kecelakaan, urutan peristiwa, lokasi tol, tipe kendaraan, dan penjelasan mengenai korban.

Penelitian ini bertujuan untuk menerapkan metode TF-IDF dan Cosine Similarity guna meningkatkan efisiensi dalam mencari artikel berita yang berkaitan dengan kecelakaan di jalan tol. Proses penelitian dilakukan melalui berbagai langkah mulai dari pengumpulan informasi, praproses teks, perhitungan bobot, hingga penghubungan antara kueri dan dokumen. Dengan pendekatan ini, sistem diharapkan dapat memberikan hasil pencarian yang lebih sesuai, cepat, dan tepat dibandingkan dengan metode pencarian konvensional yang hanya mengandalkan pencocokan kata secara langsung.

2. METODE PENELITIAN

Sistem temu balik informasi (STBI) merupakan sistem yang dirancang untuk membantu pengguna dalam menemukan kembali dokumen yang relevan berdasarkan kebutuhan informasi tertentu. Dalam pencarian artikel berita kecelakaan jalan tol, permasalahan yang sering dihadapi adalah banyaknya dokumen dengan struktur teks dan penggunaan istilah yang beragam, sehingga pencarian berbasis kata kunci sederhana belum mampu menghasilkan informasi yang relevan secara optimal. Oleh karena itu, diperlukan metode yang mampu merepresentasikan dokumen teks ke dalam bentuk numerik agar tingkat kemiripan antara dokumen dan kueri dapat dihitung secara komputasional.

Penelitian ini menerapkan model ruang vektor (Vector Space Model) dengan metode Term Frequency–Inverse Document Frequency (TF-IDF) sebagai teknik pembobotan kata serta Cosine Similarity sebagai metode pengukuran tingkat kemiripan antara kueri dan dokumen. Metode TF-IDF digunakan untuk menentukan tingkat kepentingan suatu kata dalam dokumen berdasarkan frekuensi kemunculannya dan tingkat keunikannya dalam keseluruhan dokumen. Sementara itu, Cosine Similarity digunakan untuk mengukur tingkat kemiripan antara dokumen dan kueri berdasarkan sudut kosinus antar vektor.

Metode penelitian yang digunakan adalah metode eksperimental dengan pendekatan kuantitatif. Tahapan penelitian dimulai dari studi literatur hingga evaluasi hasil pencarian. Studi literatur dilakukan untuk memperoleh landasan teori yang berkaitan dengan sistem temu balik informasi, text preprocessing, TF-IDF, dan Cosine Similarity. Selanjutnya dilakukan

pengumpulan data berupa artikel berita kecelakaan jalan tol yang diperoleh dari media berita daring dan dijadikan sebagai korpus data penelitian.

Dokumen berita yang telah dikumpulkan kemudian melalui tahap text preprocessing yang meliputi case folding untuk mengubah seluruh huruf menjadi huruf kecil, tokenization untuk memecah teks menjadi kata-kata, stopword removal untuk menghapus kata-kata umum yang tidak memiliki makna penting, serta stemming untuk mengubah kata berimbuhan menjadi kata dasar. Tahap ini bertujuan untuk mengurangi noise dan menyamakan bentuk kata sehingga proses pembobotan menjadi lebih optimal.

Setelah tahap praproses, setiap dokumen direpresentasikan dalam bentuk vektor numerik menggunakan metode TF-IDF. Perhitungan dimulai dengan menghitung nilai Term Frequency (TF), yaitu jumlah kemunculan suatu kata dalam dokumen, yang dirumuskan sebagai berikut:

$$TF(t,d) = \frac{ft,d}{\sum t' \in dft',d}$$

Selanjutnya dihitung nilai Inverse Document Frequency (IDF) untuk mengetahui tingkat keunikan suatu kata dalam keseluruhan dokumen, dengan rumus:

$$IDF\ (\mathrm{t}) = log\frac{N}{nt}$$

Bobot akhir setiap kata dalam dokumen diperoleh dari hasil perkalian antara nilai TF dan IDF, yang dirumuskan sebagai berikut:

$$TF - IDF\ (t,d) = TF\ (t,d) \times IDF\ (t)$$

Bobot TF-IDF yang dihasilkan menunjukkan tingkat kepentingan suatu kata dalam dokumen. Dokumen dan kueri yang telah direpresentasikan dalam bentuk vektor TF-IDF selanjutnya dihitung tingkat kemiripannya menggunakan metode Cosine Similarity. Sebelum menghitung nilai kemiripan, terlebih dahulu dihitung panjang vektor dokumen dan vektor kueri dengan rumus:

$$|D| = \sqrt{\sum_{i=1}^{n} w_i^2}$$

Kemudian dilakukan perhitungan dot product antara vektor kueri dan vektor dokumen sebagai berikut:

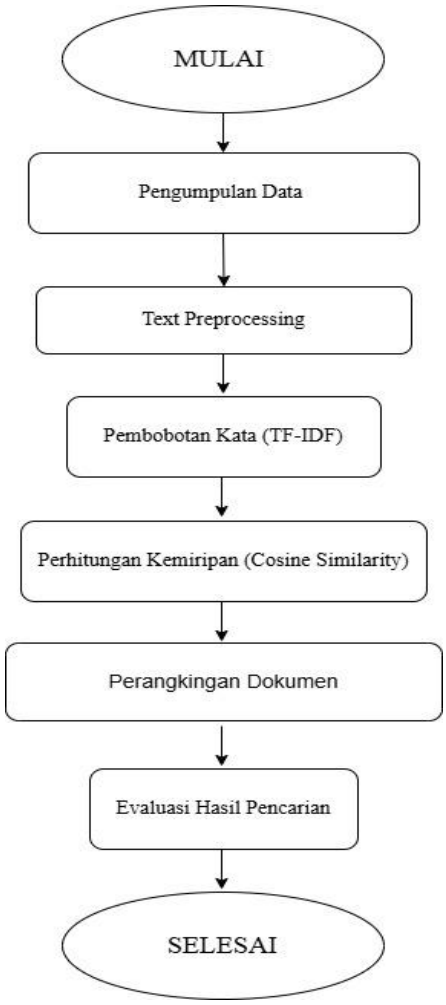
$$Q \cdot D = \sum (w_q \times w_d)$$

Nilai Cosine Similarity antara kueri dan dokumen dihitung dengan rumus:

$$\text{Cosine Similarity}(Q, D) = \frac{Q \cdot D}{|Q| \times |D|}$$

Nilai Cosine Similarity berada pada rentang 0 hingga 1, di mana nilai yang mendekati 1 menunjukkan tingkat kemiripan yang semakin tinggi antara kueri dan dokumen. Nilai kemiripan ini digunakan sebagai dasar dalam proses perangkingan dokumen, sehingga dokumen dengan nilai tertinggi dianggap paling relevan dan ditampilkan pada urutan teratas hasil pencarian.

Untuk memperjelas tahapan penelitian yang dilakukan, digunakan diagram alur penelitian yang menggambarkan proses sistem temu balik informasi secara keseluruhan, mulai dari tahap awal hingga tahap akhir penelitian, sebagaimana ditunjukkan pada Gambar 1.



Gambar 1. Diagram Alir Proses Penelitian

Berdasarkan Gambar 1, proses penelitian dimulai dari studi literatur, dilanjutkan dengan pengumpulan data artikel berita kecelakaan jalan tol. Data kemudian diproses melalui tahap praproses teks sebelum dilakukan pembobotan menggunakan TF-IDF. Selanjutnya

dilakukan perhitungan tingkat kemiripan menggunakan Cosine Similarity untuk memperoleh urutan dokumen yang paling relevan, dan tahap akhir dilakukan evaluasi hasil pencarian.

3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan sejumlah dokumen berita kecelakaan jalan tol sebagai data uji. Dokumen tersebut dipilih untuk merepresentasikan informasi tekstual yang akan diproses menggunakan metode TF-IDF dan Cosine Similarity.

1) Pada penelitian ini digunakan 15 sampel dokumen berita kecelakaan jalan tol yaitu:
   D1 =“Terjadi kecelakaan beruntun di Jalan Tol Medan-Binjai yang melibatkan tiga kendaraan dan menyebabkan kemacetan panjang.”
   D2 =“Sebuah truk terguling di Jalan Tol Medan-Tanjung Pura akibat rem blong sehingga menutup satu lajur jalan.”
   D3 =“Kecelakaan tunggal terjadi di Jalan Tol Medan-Binjai setelah pengemudi hilang kendali akibat jalan licin.”
   D4 =“Kepadatan kendaraan terjadi di beberapa titik Jalan Tol Medan-Tanjung Pura akibat lalu lintas sore hari.”
   D5 =“Di Tol Medan-Binjai KM 8, truk bermuatan berat mengalami rem blong dan menabrak pembatas, sehingga satu jalur ditutup untuk evakuasi.”
   D6 = “Mobil Avanza mogok di lajur cepat Tol Medan-Binjai arah Binjai saat jam pulang kerja, menyebabkan kemacetan panjang.”
   D7 =“Kecelakaan beruntun dua mobil dan satu truk di Tol Medan-Tanjung Pura KM 23 terjadi akibat hujan deras dan jarak pandang terbatas.”
   D8 =“Sebuah minibus tergelincir di Tol Medan-Binjai dekat pintu keluar Binjai karena jalan licin, sehingga lalu lintas dialihkan.”
   D9 =“Truk peti kemas menumpahkan muatan di Tol Medan-Tanjung Pura akibat pintu kontainer tidak terkunci, membuat dua lajur ditutup.”
   D10 =“Sebuah bus pariwisata mengalami pecah ban di Tol Medan-Binjai KM 5 sehingga berhenti di lajur kiri dan membuat arus kendaraan melambat.”
   D11 =“Mobil pick-up menabrak bagian belakang truk di Tol Medan-Binjai KM 7 akibat pengemudi mengantuk, sehingga satu lajur ditutup sementara.”
   D12 =“Sebuah truk pengangkut bahan bangunan mogok di Tol Medan-Tanjung Pura KM 30 dan menyebabkan antrean kendaraan mengular beberapa ratus meter.”
   D13 =“Kecelakaan tunggal terjadi di Tol Medan-Binjai arah Medan ketika sedan

menabrak guardrail karena pengemudi kehilangan kendali saat hujan."

D14 ="Tabrakan ringan antara dua mobil terjadi di Tol Medan–Tanjung Pura dekat Gerbang Tol Stabat karena salah satu kendaraan berpindah jalur secara tiba-tiba."

D15 ="Sebuah truk fuso mengalami oli bocor di Tol Medan–Binjai dan meninggalkan tumpahan di jalan, sehingga petugas menutup jalur kiri untuk pembersihan."

Query (Q) = "kecelakaan truk rem blong tol medan binjai"

2) Hitung Bobot

Tabel 1. Proses pencarian nilai IDF (Inverse Document Frequency)

Terms	TF														
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15
Celaka	1	0	1	0	0	0	1	0	0	0	0	0	1	0	0
Truk	0	1	0	0	1	0	1	0	1	0	1	1	0	0	1
Rem	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
Blong	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
Tol	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Medan	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Binjai	1	0	1	0	1	1	0	1	0	1	1	0	1	0	1

3) Dokumen Frequency dan IDF

Jumlah dokumen (N) = 15

$$IDF\ (t)=\log \frac{N}{nt}$$

IDF(celaka) = log(15/4) = 0,574

IDF(truk) = log(15/7) = 0,330

IDF(rem) = log(15/2) = 0,875

IDF(blong) = log(15/2) = 0,875

IDF(tol) = log(15/15) = 0

IDF(medan) = log(15/15) = 0

IDF (binjai) =log(15/9) = 0,221

4) Bobot TF – IDF(W)

$$W\ (t,d) = TF\ (t,d) \times IDF\ (t)$$

D1

Celaka = 1 x 0,574 = 0,574

Tol = 1 x 0 =0

Medan = 1 x 0 = 0

Binjai = 1 x 0,221 = 0,221

D2

Truk = 1 x 0,330 = 0,330

Rem = 1 x 0,875 = 0,875

Blong = 1 x 0,875 = 0,875

Tol = 1 x 0 = 0

Medan 1 x 0 = 0

D3

Celaka = 1 x 0,574 = 0,574

Tol = 1 x 0 = 0

Medan = 1 x 0 = 0

Binjai = 1 x 0,221 = 0,221

D4

Tol = 1 x 0 = 0

Medan = 1 x 0 = 0

D5

Truk = 1 x 0,330 = 0,330

Rem = 1 x 0,875 = 0,875

Blong = 1 x 875 = 0,875

Tol = 1 x 0 = 0

Medan = 1 x 0 = 0

Binjai = 1 x 0,221 = 0,221

D6

Tol = 1 x 0 = 0

Medan = 1 x 0 = 0

Binjai = 1 x 0,221 = 0,221

D7

Celaka = 1 x 0,574 = 0,574

Truk = 1 x 0,330 = 0,330

Tol = 1 x 0 = 0

Medan = 1*0 = 0

D8

Tol = 1 x 0 = 0

Medan = 1 x 0 = 0

Binjai = 1 x 0,221 = 0,221

D9

Truk = 1 x 0,330 = 0,330

Tol = 1 x 0 = 0

Medan = 1 x 0 = 0

D10

Tol = 1 x 0 = 0

Medan = 1 x 0 = 0

Binjai = 1 x 0,221 = 0,221

D11

Truk = 1 x 0,330 = 0,330

Tol = 1 x 0 = 0

Medan = 1 x 0 = 0

Binjai = 1 x 0,221 = 0,221

D12

Truk 1 x 0,330 = 0,330

Tol = 1 x 0 = 0

Medan = 1 x 0 = 0

D13

Celaka = 1 x 0,574 = 0,574

Tol = 1 x 0 = 0

Medan = 1 x 0 = 0

Binjai = 1 x 0,221 = 0,221

D14

Tol = 1 x 0 = 0

Medan = 1 x 0 = 0

D15

Truk = 1 x 0,330 = 0,330

Tol = 1 x 0 = 0

Medan = 1 x 0 = 0

Binjai = 1 x 0,221 = 0,221

5) Panjang Vektor

$$| D | = \sqrt{\sum W^2}$$

$| D1 | = \sqrt{(0,574)^2 + (0,221)^2}$

$= \sqrt{0,329 + 0,048}$

$= \sqrt{0,377}$

$= 0,615$

$| D2 | = \sqrt{(0,330)^2 + (0,875)^2 + (0,875)^2}$

$= \sqrt{0,108 + 0,765 + 0,765}$

$= \sqrt{1,638}$

$= 1,281$

$| D3 | = \sqrt{(0,574)^2 + (0,221)^2}$

$= \sqrt{0,329 + 0,048}$

$= \sqrt{0,377}$

$= 0,615$

| D4 | = 0

| D5 | = √(0,330)² + (0,875)² + (0,875)² + (0,221)²
= √0,108 + 0,765+ 0,765 + 0,048
= √1,686
= 1,300

| D6 | = √(0,221)²
= √0,048
= 0,221

| D7 | = √(0,574)² + (0,330)²
= √0,329 + 0,108
= √0,437
= 0,662

| D8 | = √(0,221)²
= √0,048
= 0,221

| D9 | = √(0,330)²
= √0,108
= 0,330

| D10 | = √(0,221)²
= √0,048
= 0,221

| D11 | = √(0,330)² + (0,221)²
= √0,108 + 0,048
= √0,156
= 0,397

| D12 | = √(0,330)²
= √0,108
= 0,330

| D13 | = √(0,574)² + (0,221)²
= √0,329 + 0,048
= √0,377
= 0,615

| D14 | = 0

| D15 | = √(0,330)² + (0,221)²
= √0,108 + 0,048
= √0,156
= 0,397

6) Hitung Dot Product

$$Q.D = \sum(WQ \times WD)$$

Q.D1 = (0,574 x 0,574) + (0,221 x 0,221)
= 0,329 + 0,048
= 0,377

Q.D2 = (0,330 x 0,330) + (0,875 x 0,875) + (0,875 x 0,875)
= 0,108 + 0,765 + 0,765
= 1,638

Q.D3 = (0,574 x 0,574) + (0,221 x 0,221)
= 0,329 + 0,048
= 0,377

Q.D4 = 0

Q.D5 = (0,330 x 0,330) + (0,875 x 0,875) + (0,875 x 0,875) + (0,221 x 0,221)
= 0,108 + 0,765 + 0,765 + 0,048
= 1,686

Q.D6 = (0,221 x 0,221)
= 0,048

Q.D7 = (0,574 x 0,574) + (0,330 x 0,330)
= 0,329 + 0,108
= 0,437

Q.D8 = (0,221 x 0,221)
= 0,048

Q.D9 = (0,330 x 0,330)
= 0,108

Q.D10 = (0,221 x 0,221)
= 0,048

Q.D11 = (0,330 x 0,330) + (0,221 x 0,221)
= 0,108 + 0,048
= 0,156

Q.D12 = (0,330 x 0,330)
= 0,108

Q.D13 = (0,574 x 0,574) + (0,221 x 0,221)
= 0,329 + 0,048
= 0,377

Q.D14 = 0

Q.D15 = (0,330 x 0,330) + (0,221 x 0,221)

= 0,108 + 0,048

= 0,156

7) Cosine Similarity

$$Cos(Q,D) = Q \times D \; / \; | Q | \times | D |$$

$Q = \sqrt{(0{,}574^2 + 0{,}330^2 + 0{,}875^2 + 0{,}875^2 + 0{,}221^2)} = 1{,}421$

Cos(Q,D1) = 0,377 / 1.421 x 0,614

= 0,377 / 0,872

= 0,433

Cos(Q,D2) = 1,638 / 1.421 x 1,279

= 1.638/ 1,817

= 0,901

Cos(Q,D3) = 0,377 / 1.421 x 0,614

= 0,377 / 0,872

= 0,433

Cos(Q,D4) = 0

Cos(Q,D5) = 1,686 / 1.421 x 1,298

= 1,686/ 1,844

= 0,915

Cos(Q,D6) = 0,048 / 1.421 x 0,219

= 0,048 / 0,311

= 0,156

Cos(Q,D7) = 0,437 / 1.421 x 0,661

= 0,437 / 0,939

= 0,466

Cos(Q,D8) = 0,048 / 1.421 x 0,219
= 0,048 / 0,311
= 0,156

Cos(Q,D9) = 0,108 / 1.421 x 0,328
= 0,108 / 0,466
= 0,232

Cos(Q,D10) = 0,048 / 1.421 x 0,219
= 0,048 / 0,311
= 0,156

Cos(Q,D11) = 0,156 / 1.421 x 0,394
= 0,156 / 0,559
= 0,280

Cos(Q,D12) = 0,108 / 1.421 x 0,328
= 0,108 / 0,466
= 0,232

Cos(Q,D13) = 0,377 / 1.421 x 0,614
= 0,377 / 0,872
= 0,433

Cos(Q,D14) = 0

Cos(Q,D15) = 0,156 / 1.421 x 0,394
= 0,156 / 0,559
= 0,280

8) Perangkingan

Proses perangkingan dilakukan berdasarkan nilai Cosine Similarity antara query "kecelakaan truk rem blong tol medan binjai" dengan setiap dokumen (D1–D15) setelah

dilakukan pembobotan menggunakan metode TF-IDF. Dokumen dengan nilai Cosine Similarity tertinggi dianggap paling relevan dan ditempatkan pada urutan teratas hasil pencarian.

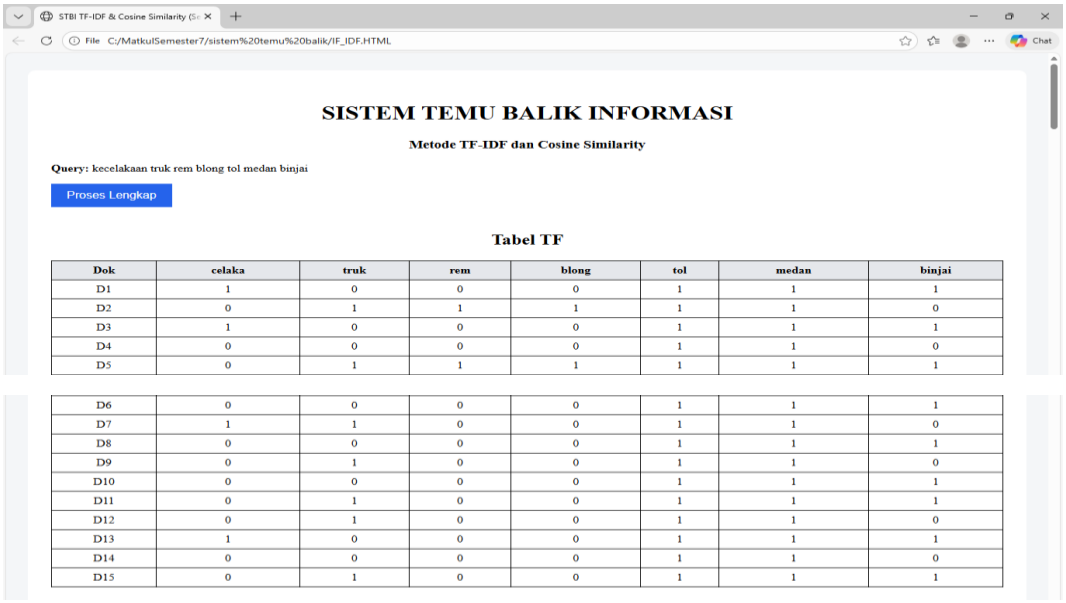
Peringkat 1: Dokumen D5 dengan nilai Cosine Similarity 0,915
Peringkat 2: Dokumen D2 dengan nilai Cosine Similarity 0,901
Peringkat 3: Dokumen D7 dengan nilai Cosine Similarity 0,466
Peringkat 4: Dokumen D1 dengan nilai Cosine Similarity 0,433
Peringkat 5: Dokumen D3 dengan nilai Cosine Similarity 0,433
Peringkat 6: Dokumen D13 dengan nilai Cosine Similarity 0,433
Peringkat 7: Dokumen D11 dengan nilai Cosine Similarity 0,280
Peringkat 8: Dokumen D15 dengan nilai Cosine Similarity 0,232
Peringkat 9: Dokumen D9 dengan nilai Cosine Similarity 0,232
Peringkat 10: Dokumen D12 dengan nilai Cosine Similarity 0,232
Peringkat 11: Dokumen D6 dengan nilai Cosine Similarity 0,156
Peringkat 12: Dokumen D8 dengan nilai Cosine Similarity 0,156
Peringkat 13: Dokumen D10 dengan nilai Cosine Similarity 0,156
Peringkat 14: Dokumen D4 dengan nilai Cosine Similarity 0
Peringkat 15: Dokumen D14 dengan nilai Cosine Similarity 0

Hasil perangkingan menunjukkan bahwa dokumen dengan nilai Cosine Similarity lebih tinggi memiliki tingkat relevansi yang lebih besar terhadap query pencarian, sedangkan dokumen dengan nilai nol dianggap tidak relevan.

9) Hasil Program



Gambar 2. Tampilan Awal



STBI TF-IDF & Cosine Similarity

File C:/MatkulSemester7/sistem%20temu%20balik/IF_IDF.HTML

Chat

SISTEM TEMU BALIK INFORMASI

Metode TF-IDF dan Cosine Similarity

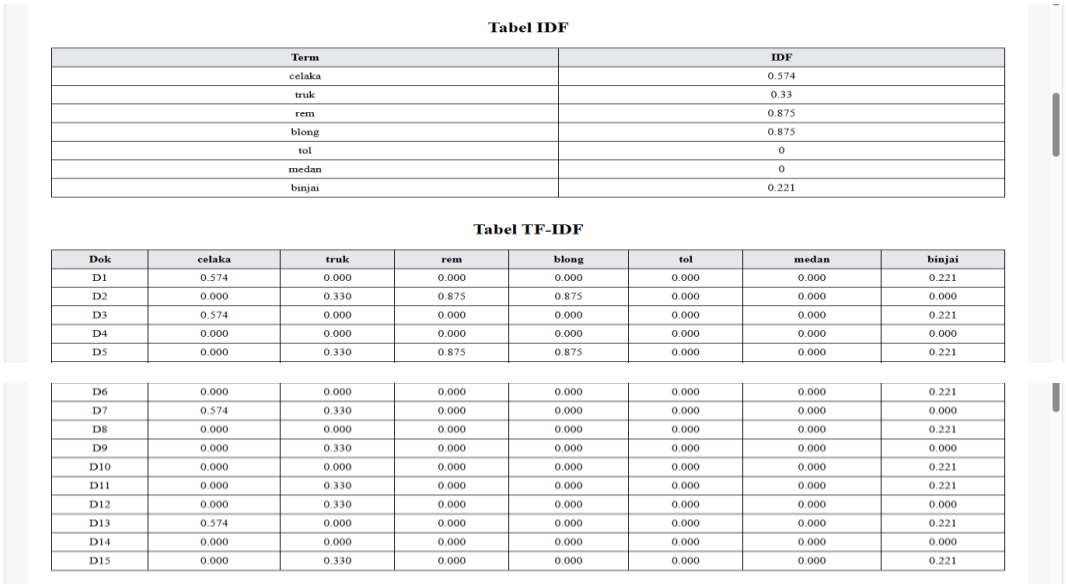
Query: kecelakaan truk rem blong tol medan binjai

Proses Lengkap

Tabel TF

Dok	celaka	truk	rem	blong	tol	medan	binjai
D1	1	0	0	0	1	1	1
D2	0	1	1	1	1	1	0
D3	1	0	0	0	1	1	1
D4	0	0	0	0	1	1	0
D5	0	1	1	1	1	1	1
D6	0	0	0	0	1	1	1
D7	1	1	0	0	1	1	0
D8	0	0	0	0	1	1	1
D9	0	1	0	0	1	1	0
D10	0	0	0	0	1	1	1
D11	0	1	0	0	1	1	1
D12	0	1	0	0	1	1	0
D13	1	0	0	0	1	1	1
D14	0	0	0	0	1	1	0
D15	0	1	0	0	1	1	1

Gambar 3. Tampilan Awal Setelah Mengklik Proses



Tabel IDF

Term	IDF
celaka	0.574
truk	0.33
rem	0.875
blong	0.875
tol	0
medan	0
binjai	0.221

Tabel TF-IDF

Dok	celaka	truk	rem	blong	tol	medan	binjai
D1	0.574	0.000	0.000	0.000	0.000	0.000	0.221
D2	0.000	0.330	0.875	0.875	0.000	0.000	0.000
D3	0.574	0.000	0.000	0.000	0.000	0.000	0.221
D4	0.000	0.000	0.000	0.000	0.000	0.000	0.000
D5	0.000	0.330	0.875	0.875	0.000	0.000	0.221
D6	0.000	0.000	0.000	0.000	0.000	0.000	0.221
D7	0.574	0.330	0.000	0.000	0.000	0.000	0.000
D8	0.000	0.000	0.000	0.000	0.000	0.000	0.221
D9	0.000	0.330	0.000	0.000	0.000	0.000	0.000
D10	0.000	0.000	0.000	0.000	0.000	0.000	0.221
D11	0.000	0.330	0.000	0.000	0.000	0.000	0.221
D12	0.000	0.330	0.000	0.000	0.000	0.000	0.000
D13	0.574	0.000	0.000	0.000	0.000	0.000	0.221
D14	0.000	0.000	0.000	0.000	0.000	0.000	0.000
D15	0.000	0.330	0.000	0.000	0.000	0.000	0.221

Gambar 4. Hasil IDF dan TF-IDF



Panjang Vektor

Dok	|D|
D1	0.615
D2	1.281
D3	0.615
D4	0.000
D5	1.300
D6	0.221
D7	0.662
D8	0.221
D9	0.330
D10	0.221
D11	0.397
D12	0.330
D13	0.615
D14	0.000
D15	0.397

Gambar 5. Ukuran Vektor

Cosine Similarity

Dok	Cos(Q,D)
D1	0.433
D2	0.901
D3	0.433
D4	0.000
D5	0.915
D6	0.156
D7	0.466
D8	0.156
D9	0.232
D10	0.156
D11	0.280
D12	0.232
D13	0.433
D14	0.000
D15	0.280

Gambar 6. Hasil Cosine Similarity

Perangkingan

Peringkat	Dokumen	Nilai
1	D5	0.915
2	D2	0.901
3	D7	0.466
4	D1	0.433
5	D3	0.433
6	D13	0.433
7	D11	0.280
8	D15	0.280
9	D9	0.232
10	D12	0.232
11	D6	0.156
12	D8	0.156
13	D10	0.156
14	D4	0.000
15	D14	0.000

Gambar 7. Hasil Perangkingan

KESIMPULAN

Berdasarkan proses perancangan, implementasi, dan pengujian yang telah dilakukan, dapat disimpulkan bahwa sistem temu balik informasi untuk artikel berita kecelakaan jalan tol yang dibangun mampu memberikan hasil pencarian yang relevan terhadap kebutuhan pengguna. Penggunaan metode TF-IDF berperan dalam memberikan bobot yang proporsional pada setiap kata sehingga istilah yang lebih penting dalam dokumen dapat teridentifikasi dengan baik. Selanjutnya, Cosine Similarity digunakan untuk menghitung tingkat kedekatan antara query dan dokumen sehingga sistem dapat menyusun urutan dokumen berdasarkan tingkat relevansinya. Hasil pengujian menunjukkan bahwa mekanisme perangkingan dokumen berjalan dengan baik, di mana dokumen yang memiliki keterkaitan paling tinggi dengan kata kunci pencarian berada pada posisi teratas. Nilai precision dan recall yang diperoleh menunjukkan bahwa sistem cukup efektif dalam menampilkan dokumen yang sesuai dengan topik pencarian. Dengan demikian, penerapan TF-IDF dan Cosine Similarity dapat dijadikan sebagai solusi yang layak untuk mendukung sistem temu balik informasi berbasis teks, khususnya pada pencarian berita kecelakaan jalan tol. Meskipun demikian, sistem masih

mengandalkan kecocokan kata secara leksikal, sehingga pada penelitian selanjutnya dapat dikembangkan dengan menambahkan pendekatan berbasis semantik agar kualitas hasil pencarian menjadi lebih optimal.

DAFTAR PUSTAKA

Amrizal, V. (2018). Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits. Jurnal Teknik Informatika, 11(2), 149–164.

Bening Herwijayanti, Dian Eka Ratnawati, L. M. (n.d.). Tampilan Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity.pdf.

Dhony Syafe'i, H., Sukmawati Nur, E., & Nurdin, B. (2012). Sistem Temu Kembali Informasi pada Dokumen TeksMenggunakan Metode Term Frequency Inverse Document Frequency (TF-IDF). In Sistem Temu Kembali Informasi pada Dokumen Teks Menggunakan Metode Term Frequency Inverse Document Frequency (TF-IDF) (Vol. 20, Issue Information Retrieval, pp. 64–70). https://www.neliti.com/id/publications/132024/sistem-temu-kembali-informasi-pada-dokumen-teks-menggunakan-metode-term-frequenc#cite

Dwi Laxmi, M., & Ali Fauzi, M. (2019). Query Expansion Pada Sistem Temu Kembali Informasi Berbahasa Indonesia Dengan Metode Pembobotan TF-IDF Dan Algoritme Cosine Similarity Berbasis Wordnet. Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer, 3(1), 2548–2964. http://j-ptiik.ub.ac.id

Fahrudin, T. M., Hartanto, M. H., Paramita, A. S., Aulia, A., Maulana, R. A., & Anniswa, I. R. (2022). Temu Kembali Informasi Berita Kegiatan Program Studi Menggunakan Algoritma Pembobotan Tf-Idf Dan Cosine Similarity. Prosiding Seminar Nasional Teknologi Dan Sistem Informasi, 2(1), 270–279. https://doi.org/10.33005/sitasi.v2i1.309

Fitri, M. (2013). Perancangan Sistem Temu Balik Informasi Dengan Metode Pembobotan Kombinasi Tf-Idf Untuk Pencarian Dokumen Berbahasa Indonesia. Jurnal Sistem Dan Teknologi Informasi, Vol. 1(1), 1–6.

Handoko, B. B., Teknologi, I., & Bandung, N. (n.d.). " MAKALAH ALGORITMA TEMU BALIK INFORMASI " ALGORITMA TERM FREQUENCY INVERSE. 1–17.

Hudin, J. M., & Rifai, A. (2017). Perancangan Sistem Temu Kembali Informasi Menggunakan Metode Vector Space Model Pada Pencarian Dokumen Berbasis Teks Berita. Agustus, 2, 128–135.

Irmawati, I. (2017). Sistem Temu Kembali Informasi Pada Dokumen Dengan Metode Vector Space Model. Jurnal Ilmiah FIFO, 9(1), 74. https://doi.org/10.22441/fifo.2017.v9i1.009

Nadirman, F. (2006). Sistem Temu-Kembali Informasi Dengan Metode Vector Space Model Pada Pencarian File Dokumen Berbasis Teks.

Nico, Budiyanto, U., & Fatimah, T. (2022). Implementasi Algoritma Pembobotan TF-IDF dan Cosine Similarity untuk Penetapan Kategori Artikel pada Website Universitas Budi Luhur. Jurnal TICOM: Technology of Information and Communication, 10(3), 218–222.

Rizki Ambarwati, Cindy Meilani, & Hafiz Irsyad. (2025). Relevansi Berita terhadap Kasus Korupsi Dana Iklan Bank BJB Menggunakan TF-IDF dan Cosine Similarity. Journal of Information Technology and Computer Science, 5(2), 146–156. https://doi.org/10.47111/jointecoms.v5i2.21053

Saadah, M. N., Atmagi, R. W., Rahayu, D. S., & Arifin, A. Z. (2013). Sistem Temu Kembali Dokumen Teks Dengan Pembobotan Tf-Idf Dan Lcs. JUTI: Jurnal Ilmiah Teknologi Informasi, 19–22. https://doi.org/10.12962/j24068535.v11i1.a16

Suryadi, U. T. (2018). Sistem Temu Kembali Lokasi Wisata Kabupaten Subang Berbasis Web Menggunakan Metode Term Frequency - Inverse Document Frequency (Tf- Idf). Jurnal Teknologi Informasi Dan Komunikasi, 1–15.

Tanuwijaya, W., Setiawan, C. E., Irsyad, H., & Rahman, A. (2025). Implementasi TF-IDF dan Cosine Similarity untuk Penyaringan Dokumen Berita Program Makan Siang Gratis Pemerintah Indonesia. 6(2), 322–334