

Implementasi Algoritma Vektor Space Model (Vsm) Pada Pencarian Dokumen Teks (Studi Kasus Kantor Sekretariat Biro Hukum Provinsi Gorontalo)

1st HASNIDAR

¹ Fakultas Ilmu Komputer Universitas Ichsan Gorontalo

✉ hasnidardhln@gmail.com

Article Info

Submitted: Juli 01, 2025 *Revised:* ,
Accepted:

Keywords:

Keyword 1; Algoritma 2; Vektor Space Model (VSM), 3; Dokumen Teks 4;

How to cite [Chicago Manual of Style 17th edition (full note)]: 1st Author, 2nd Author and 3rd Author, "Title" Law Social Justice 8, no. 1 (2026): 199–202.

Abstract

Dokumen teks merupakan suatu bentuk penyampaian laporan dalam bentuk tulisan yang terkadang memiliki kemiripan dari sumber teks yang lain. Memiliki banyak dokumen terkadang sangat merepotkan, terutama disaat ingin mencari dokumen yang dibutuhkan secara cepat dan tepat. Salah satu cara untuk mencari dokumen dari tumpukan dokumen adalah dengan me-ranking masing-masing dokumen menggunakan metode Vektor Space Model (VSM). Implementasi Algoritma Vektor Space Model (VSM) merupakan salah satu alternatif yang dapat diimplementasikan untuk memecahkan masalah ini. Dengan metode Vector Space Model dapat dilihat tingkat kedekatan atau kesamaan dengan cara pembobotan term. Ini adalah suatu metode pembobotan data dengan menghitung jarak antar dokumen. Hasil yang diberikan dari pengujian dokume dengan metode Vector Space Model memberikan pembobotan pada judul dokumen, memberikan rekomendasi atau dokumen yang paling diprioritaskan sehingga sesuai dengan kebutuhan.

@2025 – 1st HASNIDAR 2nd, 3rd 4th, 5rd.
Under the license CC BY-SA 4.

Introduction

Sekretariat Biro Hukum Provinsi Gorontalo merupakan salah satu instansi yang bergerak di bidang bantuan hukum. Sekretariat Biro Hukum Provinsi Gorontalo mempunyai tugas menyiapkan bahan perumusan penyusunan produk hukum daerah provinsi; Pembinaan dan pengawasan produk hukum kabupaten/kota, bantuan hukum, dokumentasi dan informasi hukum, penyuluhan hukum dan pemajuan Hak Azasi Manusia (HAM) di daerah serta Tata usaha Biro¹

Perkembangan jaman hingga saat ini sangat berdampak besar terhadap kehidupan sekarang. Salah satu hal yang berubah adalah cara menggunakan data sebagai informasi, pada era big data sekarang ini. Dengan semakin banyaknya jumlah dokumen yang beredar saat ini menimbulkan sebuah masalah untuk melakukan pencarian yang diinginkan dengan cepat dan akurat baik pada media online melalui internet (search engine, e-library) ataupun offline (sistem penyimpanan pada komputer). meskipun saat ini beberapa e-library sudah menerapkan algoritma untuk pencariannya seperti algoritma boolean search namun hal itu belum cukup akurat untuk proses pencarian karena tidak dapat mengenali dokumen-dokumen yang paling relevan atau cukup relevan².

Dalam proses pencarian dokumen teks terdapat beberapa model yang paling sering digunakan salah satunya adalah model Boolean. Model Boolean merupakan model pencarian paling awal dikembangkan dan masih banyak digunakan sampai saat ini. Namun ada beberapa kelemahan pada model ini, yakni menghasilkan pencarian dokumen yang tidak terangkai dan masih menggunakan query pencarian dengan operator matematika. Salah satu metode yang dapat diimplementasikan pada pencarian dokumen teks adalah metode vector space model (VSM). vector space model (VSM) dapat diterapkan untuk membantu menyelesaikan permasalahan retrieval pada pencarian dokumen teks. Metode ini dapat menghasilkan pencarian yang terangkai serta menggunakan query pencarian yang lebih familiar³.

Methods

Berdasarkan latar belakang dan kerangka pemikiran yang telah diuraikan, penelitian ini mengambil objek implementasi data mining dalam pencarian dokumen teks di Kantor Biro Hukum Provinsi Gorontalo dengan menggunakan metode Vector Space Model (VSM), di mana konsep dasarnya adalah menghitung jarak antar dokumen dan mengurutkannya berdasarkan tingkat kemiripan dengan query yang diajukan, melalui serangkaian tahapan seperti case folding, tokenizing, filtering, dan stemming untuk membentuk kata-kata (term) yang kemudian diolah menjadi tabel frekuensi dihitung

¹ JDIH, P. (t.thn.). Tugas pokok dan Fungsi Biro hukum.

² Ahmad Ridoh, Y. I. (2021). Perancangan dan Implementasi Sistem Informasi Dokumen Layanan Publik Berbasis Web Untuk Mempermudah Masyarakat Memperoleh Informasi Pada Pemerintah Kabupaten Bungo. *Home > Vol 5, No 5 (2021) > Ridoh, 15*

³ Amin, F. (2012). Sistem Temu Kembali Informasi dengan Metode Vector Space Model. *Jurnal Sistem Informasi Bisnis*, 2 (2), 78-83.

bobotnya menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF), selanjutnya masing-masing dokumen dan query direpresentasikan dalam bentuk vektor sehingga dapat diukur jaraknya dengan rumus cosine similarity untuk menentukan seberapa dekat relevansi antar dokumen dengan query yang dicari, di mana seluruh proses perhitungan mencakup penentuan term frequency, inverse document frequency, bobot akhir, jarak query, jarak dokumen, hingga penghitungan sudut antar vektor, sehingga menghasilkan sistem pencarian dokumen yang lebih cepat, akurat, dan terstruktur sesuai kebutuhan biro hukum tersebut.

Main Heading of the Analysis or Discussion or Results

A. Information Retrieval

Information Retrieval adalah menemukan bahan (biasanya dokumen) dari sesuatu yang tidak terstruktur (biasanya teks) dalam jumlah yang besar untuk menghasilkan informasi yang dibutuhkan (biasanya tersimpan dalam komputer. Prinsip kerja sistem temu kembali informasi jika ada sebuah kumpulan dokumen dan seorang user yang memformulasikan sebuah pertanyaan request atau query, Sistem Temu Kembali Informasi bertujuan untuk menjawab kebutuhan informasi user dengan sumber informasi yang tersedia dalam waktu yang singkat³.

Pengambilan informasi menunjukkan proses pencarian informasi yang diperlukan, Information retrieval umumnya berkaitan dengan pencarian dan pengambilan informasi berbasis pengetahuan sistem information Salah satu penerapan prinsip relevansi yang sejak dahulu digunakan dalam pengembangan sistem Information Retrieval System menemukan informasi yang biasanya dalam bentuk dokumen dari sebuah data yang tidak terstruktur dalam bentuk teks untuk memenuhi kebutuhan informasi dari koleksi data yang sangat besar umumnya tersimpan dalam database⁴.

B. Vector Space Model (VSM)

Model ruang vektor memberikan sebuah kerangka pencocokan parsial Hal ini dicapai dengan menetapkan bobot non-biner untuk istilah indeks dalam query dan dokumen. Tidak hanya untuk pencarian teks, pencarian informasi juga dapat query elemen multimedia seperti gambar, suara. Metode ini melihat tingkat kedekatan atau kesamaan (smilarity) term dengan cara pembobotan term⁴. Vector Space Model (VSM) adalah suatu model yang digunakan untuk mengukur kemiripan antara suatu dokumen dengan suatu query. Query dan dokumen dianggap sebagai vektor-vektor pada ruang n-dimensi, dimana t adalah jumlah dari seluruh term yang ada dalam leksikon. Leksikon adalah daftar semua term yang ada dalam indeks. Selanjutnya akan dihitung nilai cosinus sudut dari dua vektor, yaitu W dari tiap dokumen dan W dari kata kunci⁵

³ Amin, F. (2012). Sistem Temu Kembali Informasi dengan Metode Vector Space Model. *Jurnal Sistem Informasi Bisnis*, 2 (2), 78-83.

⁴ Fauzi, A., & Ginabila, G. (2019). Information Retrieval System Pada File Pencarian Dokumen Tesis Berbasis Text Menggunakan Metode Vector Space Model. *Jurnal Pilar Nusa Mandiri*, 15(1), 41-46.

⁵ Karmayasa, O., & Mahendra, I. B. (2012). Implementasi Vector Space Model dan Beberapa Notasi Metode Term Frequency Inverse Document Frequency (TF-IDF) .

Cara kerja Vector Space Model dimulai dengan case folding, filtering, stemming dan tokenisasi yaitu tahap pemotongan string input berdasarkan tiap kata yang meyusunnya dan memecah dokumen kedalam tabel frekuensi kata. Seluruh kata dalam dokumen dibentuk menjadi satu yang disebut sebagai term. Tiap dokumen ditampilkan sebagai vektor yang akan dibandingkan dengan term yang telah dibentuk. Similarity Analysis untuk mengukur kemiripan dokumen dilakukan dengan menghitung cosinus jarak antara dokumen tersebut⁶.

C. TF-IDF

Metode TF- IDF merupakan suatu cara untuk memberikan bobot hubungan suatu kata (term) terhadap dokumen. Dari hasil pembobotan dengan metode TF- IDF ini nantinya sebuah dokumen dan query bisa direpresentasikan dalam sebuah ruang vektor yang kemudian akan dicari tingkat kedekatannya dengan menggunakan pengukuran cosine silimilarity sehingga mendapatkan dokumen yang paling relevan dengan suatu query. TF-IDF (Term Frequency-Inversed Document Frequency) merupakan salah satu algoritma pembobotan sebuah kata⁵.

D. Cosine Similiarity

Cosine Similarity merupakan salah satu metode untuk mencari suatu nilai similarity pada suatu dokumen dengan menggunakan pemodelan perhitungan vektor⁵.

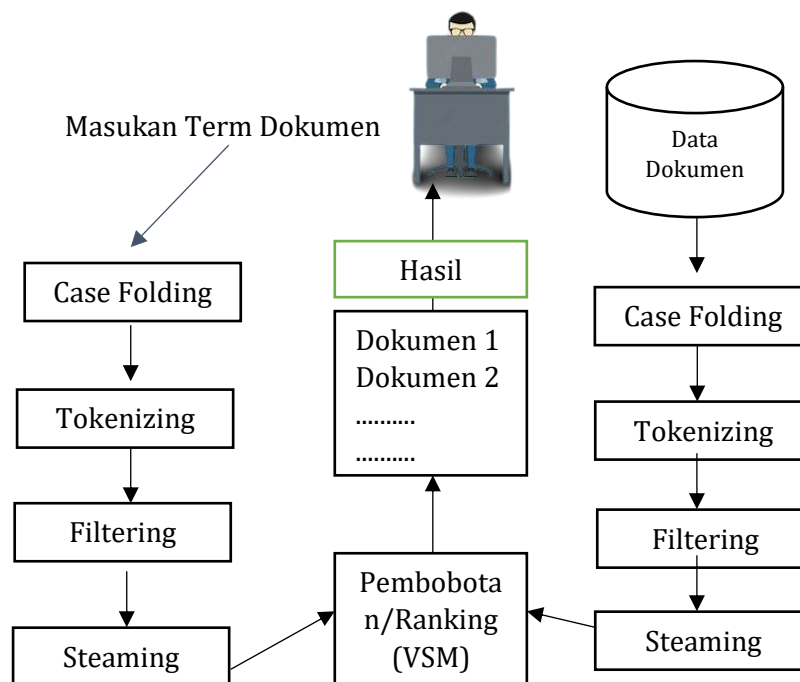
E. Dokumen Teks

Dokumen teks merupakan suatu bentuk penyampaian laporan dalam bentuk tulisan yang terkadang memiliki kemiripan dari sumber teks yang lain. Text mining merupakan salah satu tahapan pengolahan dokumen teks untuk mengetahui similarity dokumen satu dengan yang lainnya dengan beberapa tahapan seperti: stopword, case folding, stemming, tokenizing, dan pembobotan kata serta analyzing dokumen, case folding merupakan tahapan awal text mining mengubah seluruh kata menjadi huruf kecil, stopword merupakan tahapan lanjutan dengan menghapus kata yang termasuk dalam penghubung kata, stemming merupakan tahapan lanjutan untuk mengubah kata menjadi kata dasar tanpa imbuhan algoritma yang digunakan nazief & adriani, tokenizing merupakan tahapan selanjutnya untuk memotong kata perkata, analayzing merupakan tahapan untuk melakukan analisis kata atau melakukan pembobotan kata dalam hal ini pembobotan yang digunakan adalah TF-IDF sedangkan untuk menghitung nilai similarity menggunakan metode cosine similarity

⁴ Fauzi, A., & Ginabila, G. (2019). Information Retrieval System Pada File Pencarian Dokumen Tesis Berbasis Text Menggunakan Metode Vector Space Model. *Jurnal Pilar Nusa Mandiri*, 15(1), 41-46.

⁵ Karmayasa, O., & Mahendra, I. B. (2012). Implementasi Vector Space Model dan Beberapa Notasi Metode Term Frequency Inverse Document Frequency (TF-IDF).

⁶ Hendini, A. (2018). KARAOKE, P. M. IMPLEMENTASI VECTOR SPACE MODEL PADA SISTEM.



Gambar 1.1 Ilustrasi skema pencarian dokumen teks

Untuk kepentingan analisis data di dalam melakukan penelitian ini dibutuhkan beberapa sampel data yang diambil dari lima dokumen yaitu :

Q : Penetapan Pejabat Provinsi Gorontalo

- D1 :** PENETAPAN KOORDINATOR PEMUNGUTAN PAJAK KENDARAAN BERMOTOR DAN BEA BALIK NAMA KENDARAAN BERMOTOR PADA UPTD BADAN KEUANGAN PROVINSI GORONTALO
- D2 :** PERHITUNGAN BAGI HASIL PAJAK KENDARAAN BERMOTOR BEA BALIK NAMA KENDARAAN BERMOTOR PAJAK BAHAN BAKAR KENDARAAN BERMOTOR PAJAK AIR PERMUKAAN DAN PAJAK ROKOK UNTUK PEMERINTAH KABUPATEN/KOTA SE PROVINSI GORONTALO
- D3 :** PENETAPAN PEJABAT PELAKSANA TEKNIS KEGIATAN (PPTK) PELAKSANAAN ANGGARAN PENDAPATAN DAN BELANJA DAERAH (APBD) PROVINSI GORONTALO POS ANGGARAN BIRO UMUM, GUBERNUR, WAKIL GUBERNUR GORONTALO DAN PIMPINAN SEKRETARIAT DAERAH PROVINSI GORONTALO TAHUN ANGGARAN 2019
- D4 :** PENETAPAN PEJABAT PENGADAAN BARANG/JASA DI LINGKUNGAN SEKRETARIAT DAERAH PROVINSI GORONTALO TAHUN ANGGARAN 2019

D5 : PENETAPAN PEJABAT PELAKSANA TEKNIS KEGIATAN (PPTK)
PELASANA APBD PROVINSI GORONTALO POS ANGGARAN BIRO HUKUM
DAN ORGANISASI SETDA PROVINSI GORONTALO TAHUN ANGGARAN
2019

Untuk menjawab pertanyaan di atas perlu dilakukan suatu proses perhitungan pembobotan atau perankingan dari dokumen tersebut sehingga dari hasil pembobotan atau perankingan tersebut akan terlihat dokumen mana yang relevan atau yang paling diprioritaskan. Di dalam penerapan metode Vector Space Model ada beberapa tahapan proses pengolahan data terlebih dahulu antara lain.

A. Perhitungan TF-IDF

Berikut adalah langkah terakhir dari pembobotan TF-IDF, yaitu perhitungan hasil TF-IDF yang diperoleh dari perkalian antara nilai TF yang telah dinormalisasi dan nilai IDF. Berikut adalah hasil perhitungan manual TF-IDF menggunakan 5 dokumen.

Table 1 Pembobotan TF-IDF

Nilai TF-IDF					
Q	D1	D2	D3	D4	D5
0.0969	0.0969	0	0.0969	0.0969	0.0969
0	0.6990	0	0	0	0
0	0.6990	0	0	0	0
0	0.0969	0.2907	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0.3979	0.3979	0	0	0
0	0.3979	0.3979	0	0	0
0	0.3979	0.3979	0	0	0
0	0.6990	0	0	0	0

0	0.6990	0	0	0	0
0	0.6990	0	0	0	0
-0.1461	-0.1461	-0.1461	-0.2923	-0.1461	-0.2923
-0.2041	-0.2041	-0.2041	-0.6124	-0.2041	-0.4082
0	0	0.6990	0	0	0
0	0	0.6990	0	0	0
0	0	0.6990	0	0	0
0	0	0.6990	0	0	0
0	0	0.6990	0	0	0
0	0	0.6990	0	0	0
0	0	0.6990	0	0	0
0	0	0.6990	0	0	0
0	0	0.6990	0	0	0
0	0	0.6990	0	0	0
0	0	0.6990	0	0	0
0.2218	0	0	0.2218	0.2218	0.2218
0	0	0	0.2218	0	0.4437
0	0	0	0.3979	0	0.3979
0	0	0	0.3979	0	0.3979
0	0	0	0.3979	0	0.3979
0	0	0	0.1938	0	0.1938
0	0	0	0.6990	0	0

B. Menghitung Panjang Vektor

Tabel 2 Menghitung Panjang Vektor

Dokumen	Panjang Vektor & Query
Q	0,3488
D1	1,7320
D2	2,4488
D3	2,3066
D4	1,5527
D5	1,8517

C. Menghitung Indeks Term

Tabel 3 Menghitung Indeks Term

Dokumen	indeks term
D1	9,5929
D2	0,0596
D3	10,256
D4	12,8176
D5	6,027

D Menghitung Similiarity Dokumen

Dari proses Similiarity di atas dapat di ambil ranking dari masing-masing Dokumen yaitu :

Dokumen 1 = 15,8796 Dengan Rank (2)
Dokumen 2 = 0,0689 Dengan Rank (5)
Dokumen 3 = 12,7482 Dengan Rank (3)
Dokumen 4 = 23,5560 Dengan Rank (1)
Dokumen 5 = 9,3167 Dengan Rank (4)

Jadi Dokumen yang paling relevan dengan kata kunci adalah Dokumen 4. Hasil Pencarian Dokumen dengan kata kunci “Penetapan Pejabat Provinsi Gorontalo” Menghasilkan bobot dokumen yang tertinggi adalah 4 yaitu “23,5560” dengan bobot terendah di peroleh dari dokumen 2 yaitu “0,0689”.

Conclusion

Berdasarkan analisa dan pembahasan, implementasi dan hasil pengujian yang telah dilakukan pada bab-bab sebelumnya maka dapat di ambil kesimpulan bahwa tujuan penelitian ini telah tercapai yaitu:

Metode Vector Space Model (VSM) dapat diterapkan pada algoritma pencarian dokumen teks dinamis. Dengan algoritma Metode Vector Space Model (VSM) ini pencarian Dokumen dapat dilihat nilai bobot dari masing-masing dokumen dan nilai rangking yang paling relevan.

References

- JDIH, P. (t.thn.). Tugas pokok dan Fungsi Biro hukum.
- Ahmad Ridoh, Y. I. (2021). Perancangan dan Implementasi Sistem Informasi Dokumen Layanan Publik Berbasis Web Untuk Mempermudah Masyarakat Memperoleh Informasi Pada Pemerintah Kabupaten Bungo. *Home > Vol 5, No 5 (2021) > Ridoh, 15*.
- Amin, F. (2012). Sistem Temu Kembali Informasi dengan Metode Vector Space Model. *Jurnal Sistem Informasi Bisnis*, 2 (2), 78-83.
- Fauzi, A., & Ginabila, G. (2019). Information Retrieval System Pada File Pencarian Dokumen Tesis Berbasis Text Menggunakan Metode Vector Space Model. *Jurnal Pilar Nusa Mandiri*, 15(1), 41-46.
- Karmayasa, O., & Mahendra, I. B. (2012). Implementasi Vector Space Model dan Beberapa Notasi Metode Term Frequency Inverse Document Frequency (TF-IDF) .
- Hendini, A. (2018). KARAOKE, P. M. IMPLEMENTASI VECTOR SPACE MODEL PADA SISTEM.
- Amin, F. (2012). Sistem Temu Kembali Informasi dengan Metode Vector Space Model. *Jurnal Sistem Informasi Bisnis*, 2 (2), 78-83.
- Sanjaya, F. (2017). Pemanfaatan Sistem Temu Kembali Informasi dalam Pencarian Dokumen Menggunakan Metode Vector Space Model. *J-INTECH (Journal of Information and Technology)*, 5(02), 147-153.
- Ileman, D., & Andesa, K. (2017, November). Implementasi Vector Space Model Untuk Meningkatkan Kualitas Pada Sistem Pencarian Buku Perpustakaan. In *Seminar Nasional Informatika (SNIf)* (Vol. 1, No. 1, pp. 8-15).