

## **RANCANG BANGUN INFORMATION RETRIEVAL SYSTEM (IRS) KAMUS BAHASA-SUNDA.COM DENGAN METODE VECTOR SPACE MODEL (VSM)**

**Sujiliani Heristian<sup>1</sup>; Hanggoro Aji Al Kautsar<sup>2</sup>, Asep Sayfullloh<sup>3</sup>**

Teknik Informatika<sup>1</sup>  
STMIK Nusa Mandiri Jakarta<sup>1</sup>  
<http://nusamandiri.ac.id><sup>1</sup>  
[sujilianiheristian30@gmail.com](mailto:sujilianiheristian30@gmail.com)<sup>1</sup>

Sistem Informasi Akutansi<sup>2</sup>  
Universitas Bina Sarana Informatika<sup>2</sup>  
<http://bsi.ac.id><sup>2</sup>  
[hanggoro.hgr@bsi.ac.id](mailto:hanggoro.hgr@bsi.ac.id)<sup>2</sup>

Sistem Informasi Akutansi<sup>3</sup>  
Universitas Bina Sarana Informatika<sup>3</sup>  
<http://bsi.ac.id><sup>3</sup>  
[asep.alo@bsi.ac.id](mailto:asep.alo@bsi.ac.id)<sup>3</sup>

**Abstract**—Sundanese language is no longer the mother tongue / first language in its own area, but it has become the second language after the Indonesian language. The need for preservation of the Sundanese language in the form of online that can be accessed for its users so it will facilitate the searching of text documents especially sunda documents IRS Software is designed to provide the optimal document search results using VSM method, so users will get fast and accurate search results. The VSM method will weight every document in the database so that the documents have different weights to determine which documents are most similar to queries, the highest-weighted documents are ranked in the search results. The evaluation of IRS search results is done by recall and precision tests. Cosine calculation results known that Document 1 (D1) has the highest level of similarity and then followed by D3 and D2 documents is a point or vector in this space. It proves each query that is given has a different level of closeness to the existing documents, it can be Draw the conclusion dkumen which has a high level of similitas influenced by the given query .

**Keywords:** Sundanese Language, Information Retrieval, Vector Space Model

**Intisari**—Bahasa Sunda bukan lagi bahasa ibu / bahasa pertama di wilayahnya sendiri, tetapi sudah menjadi bahasa kedua setelah bahasa Indonesia. Perlunya pelestarian bahasa Sunda dalam bentuk online yang dapat diakses oleh penggunaanya sehingga akan memudahkan pencarian dokumen teks terutama dokumen sunda. Perangkat Lunak

IRS dirancang untuk memberikan hasil pencarian dokumen yang optimal menggunakan metode VSM, sehingga pengguna akan dapatkan hasil pencarian yang cepat dan akurat. Metode VSM akan memberi bobot pada setiap dokumen dalam basis data sehingga dokumen memiliki bobot yang berbeda untuk menentukan dokumen mana yang paling mirip dengan kueri, dokumen tertimbang tertinggi diberi peringkat dalam hasil pencarian. Evaluasi hasil pencarian IRS dilakukan dengan tes recall dan precision. Hasil perhitungan cosine diketahui bahwa Dokumen 1 (D1) memiliki tingkat kesamaan tertinggi dan kemudian diikuti oleh dokumen D3 dan D2 adalah titik atau vektor dalam ruang ini. Ini membuktikan setiap query yang diberikan memiliki tingkat kedekatan yang berbeda dengan dokumen yang ada, dapat ditarik kesimpulan dkumen yang memiliki tingkat similitas yang tinggi dipengaruhi oleh query yang diberikan

**Kata Kunci:** Bahasa Sunda, Information Retrieval, Vector Space Model

### **PENDAHULUAN**

Perkembangan ilmu pengetahuan yang pesat dewasa ini telah mendorong permintaan akan kebutuhan informasi ilmu pengetahuan itu sendiri. Cara pemenuhan kebutuhan akan informasi ini dapat dilakukan dengan beraneka ragam. Mulai dari sekedar membaca koran, majalah, atau jurnal-jurnal tertulis, hingga menggunakan teknologi digital yang terus berkembang. Akan tetapi semakin luas dan berkembangnya informasi yang beredar, membuat

masyarakat mengalami kesulitan untuk mendapatkan informasi yang dibutuhkannya dari media cetak. Lambat laun masyarakat mulai menggunakan teknologi digital untuk memudahkan mereka dalam mencari informasi yang dibutuhkan.

Bahasa Sunda di daerah Jawa Barat. Bahasa sunda bukan lagi merupakan bahasa ibu/bahasa pertama di daerah sendiri, tetapi sudah dijadikan bahasa kedua setelah bahasa Indonesia. Pengaruh bahasa Indonesia terhadap kebudayaan di Nusantara sangat besar sehingga di zaman sekarang terutama di kota-kota besar yang tidak lagi mengenal bahasa lokalnya/bahasa ibu khususnya bahasa Sunda untuk daerah Jawa Barat (Pujiatin, Lestari, Susanti, & Informatika, 2018).

Implementasi Vector Space Model dapat dirasakan dan dinikmati pada berbagai bidang keilmuan seperti Computational Linguistics (Erk, Padó, & Padó, 2010), Expert Systems, Medical (Lopes & Ribeiro, 2010), Knowledge-Based Systems, Data and Knowledge Engineering (Mao & Chu, 2007) dan lain sebagainya. Vector space model dapat juga digunakan dalam sistem temu kembali informasi (information retrieval). Sistem temu kembali informasi akan memberikan nilai tambah dalam pencarian informasi jika keinginan user bisa terpenuhi. Penelitian ini diharapkan dapat membuat sistem temu kembali informasi yang bernilai tambah yaitu menghasilkan pencarian informasi dengan cepat dan akurat.

Pencarian informasi saat ini dilakukan dengan menggunakan mesin pencari atau sistem temu kembali informasi, user menuliskan query dan mesin pencari akan menampilkan hasil pencarian. Mesin pencari yang sudah ada dan banyak digunakan saat ini memberikan hasil perolehan pencarian yang banyak (banyak dokumen yang terambil), sehingga diperlukan waktu untuk menentukan hasil pencarian yang relevan. Menentukan hasil yang relevan sesuai dengan keinginan user dengan jumlah hasil pencarian yang banyak akan menyulitkan user. Hal ini terjadi karena dokumen yang terambil oleh sistem jumlahnya banyak, maka sistem berkemungkinan menampilkan hasil pencarian yang tidak relevan. Banyaknya dokumen hasil pencarian ini membuat waktu yang dibutuhkan dalam pencarian menjadi lebih banyak dari yang diharapkan.

Perkembangan penelusuran informasi saat ini menghasilkan recall yang tinggi dan precision yang rendah. Recall yang tinggi diartikan bahwa dokumen yang dihasilkan dalam penelusuran dokumen adalah banyak, sedangkan precision rendah dapat diartikan bahwa dokumen yang diharapkan dapat ditemukan sedikit.

Solusi untuk mengatasi masalah ini adalah dengan membuat software Information Retrieval System (IRS) menggunakan metode Vector Space Model (VSM). Metode VSM dipilih karena cara kerja model ini efisien, mudah dalam representasi dan dapat diimplementasikan pada document-matching.

## BAHAN DAN METODE

### A. Teknik Pengumpulan Data

Untuk mendapatkan data-data yang menunjang teknik yang dilakukan adalah metode observasi, yaitu dengan mengambil data dari Kamus Bahasa-Sunda.Com, selanjutnya metode studi pustaka (library method), yaitu mendapatkan sumber data dari berbagai sumber dengan cara mengumpulkan referensi yang menunjang melalui buku, jurnal dan sumber lainnya yang mendukung penelitian.

### B. Metode Penelitian

Adapun metode yang dilakukan dalam penelitian ini sebagai berikut :

1. Survei Literature dan Studi Pustaka  
Peneliti akan mengumpulkan informasi dari jurnal penelitian sebelumnya yang terkait. Kemudian dari survey literature tersebut, peneliti akan melengkapi dan mendetailkan lebih jauh terkait dengan pemanfaatan sistem yang akan menjadi bagian dari proses pendetailan dari keuntungan bagi masyarakat
2. Identifikasi Masalah  
Melakukan identifikasi terkait permasalahan yang akan dilakukan penelitian mengenai analisa kamus Bahasa sunda
3. Melakukan Hipotesis  
Hipotesis awal yang dirumuskan oleh peneliti Apakah ada pengaruh penggunaan bahasa sunda
4. Observasi, Wawancara dan Kuesioner  
Penelitian akan dilakukan dengan melihat secara langsung proses dan data yang terjadi didalam perusahaan, melakukan wawancara dengan pihak yang berkepentingan serta menggunakan kuesioner untuk mempertajam analisa penelitian yang akan dilakukan
5. Melakukan Analisa  
Peneliti akan menganalisa dan mengolah data yang diperoleh dan memberikan nilai dari masing-masing data yang diperoleh. Setelah dilakukan analisa secara detail maka selanjutnya akan diambil esensi yang utama terkait dengan pemanfaatan yang menjadi

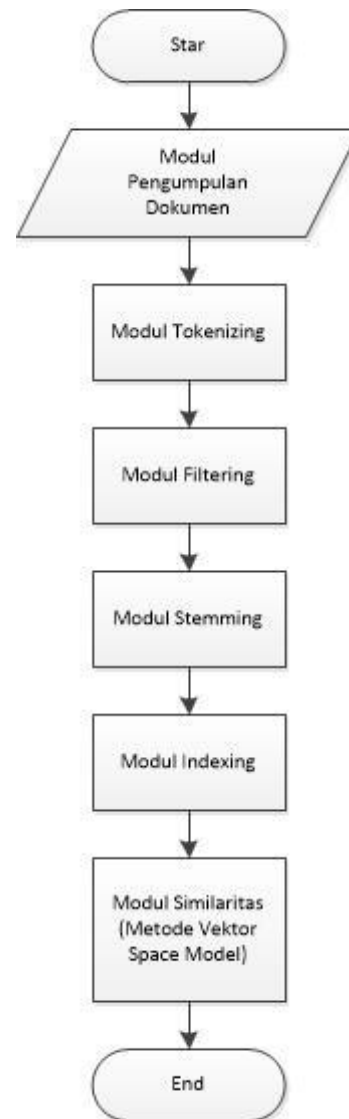
sasaran dalam meningkatkan kamus Bahasa sunda

- Menarik Kesimpulan dan Memberi Saran  
Setelah dilakukan analisa lebih lanjut maka peneliti akan memberikan kesimpulan terkait pengujian hipotesis yang dirumuskan serta memberikan saran yang baik untuk kamus online Bahasa daerah kedepannya menjadi lebih baik.

Model ruang vektor menyediakan pengguna dengan panduan untuk dokumen yang mungkin lebih mirip dan lebih signifikan dengan menghitung jarak atau ukuran sudut antara permintaan dan istilah atau dokumen (Singh, Singh, Vishwavidyalaya, Programmer, & Vishwavidyalaya, 2015). VSM memberikan sebuah kerangka pencocokan parsial Hal ini dicapai dengan menetapkan bobot non-biner untuk istilah indeks dalam query dan dokumen (Kom & Setiawan, 2015) Tidak hanya untuk pencarian teks, pencarian informasi juga dapat query elemen multimedia seperti gambar, suara, (Yulianto, Budiharto, & Kartowisastro, 2017) metode ini melihat tingkat kedekatan atau kesamaan (similarity) term dengan cara pembobotan term. Dokumen dipandang sebagai sebuah vektor yang memiliki magnitude (jarak) dan direction (arah). Pada Vector Space Model, sebuah istilah direpresentasikan dengan sebuah dimensi dari ruang vektor. Relevansi sebuah dokumen ke sebuah query didasarkan pada similaritas diantara vektor dokumen dan vektor query. dokumen dan query direpresentasikan sebagai vektor dan sudut antara keduanya. vektor dihitung menggunakan fungsi cosinus kesamaan. Efektivitas VSM sebagian besar tergantung pada istilah pembobotan yang diterapkan (Information Retrieval System Pada Pencarian File Dokumen Berbasis Teks Dengan Metode Vector Space Model Dan Algoritma Ecs Stemmer 2016). memungkinkan hasil penghitungan menjadi peringkat sesuai dengan ukuran kesamaan (Singh et al., 2015).

Vector Space Model telah memperkenalkan skema berat jangka dikenal sebagai tf-idf weighting. Bobot ini memiliki frekuensi istilah (tf) faktor yang mengukur frekuensi kemunculan istilah dalam dokumen atau teks permintaan dan faktor dokumen inverse frekuensi (idf) mengukur kebalikan dari jumlah dokumen yang berisi istilah kueri atau dokumen (Roshdi & Roohparvar, 2015).

Dibawah ini merupakan flowchart dari metode SMART

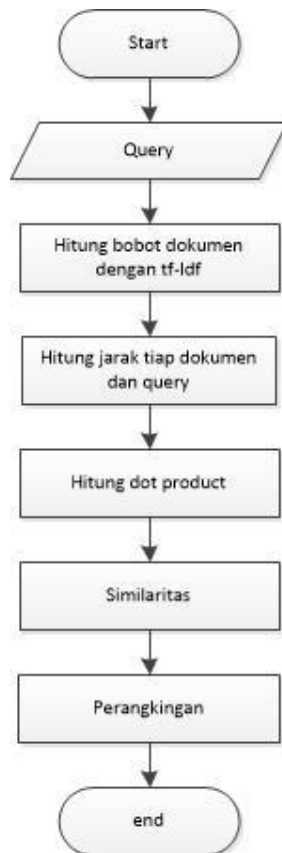


Sumber: (Heristian, Al Kautsar, & Sayfullloh, 2019)  
Gambar 1 Flowchart IRS

Langkah information retrieval system

- Pengumpulan dokumen : yaitu merupakan Proses pengumpulan dokumen-dokumen yang dipilih untuk disimpan dalam korpus
- Proses Tokenizing : merupakan proses pemotongan string input berdasarkan tiap kata yang menyusunnya
- Proses Pembuangan stopword (filtering) : merupakan proses pembuangan term yang tidak memiliki arti atau tidak relevan.
- Proses Pengubahan Kata Dasar (Stemming) : merupakan tahap mencari kata dasar (root) dari tiap kata hasil filtering
- Proses pengindeksan kata : merupakan pengindeksan kata dari koleksi teks yang digunakan untuk mempercepat proses pencarian
- Modul Vector Space Model : Hasil indexing selanjutnya dihitung tingkat kemiripannya

dengan query menggunakan metode vector space model.



Sumber: (Heristian et al., 2019)  
Gambar 2 Flowchart Algoritma VSM

Langkah metode vector space model

1. Menghitung bobot dokumen dengan tf-idf  
 $Idf = \log(D/df)$   
 $Wf(t,d) = tft,d * idft$
2. Menghitung jarak tiap dokumen dan query  
 $Sqrt(Q) = Sqrt(\sum_{j=1}^n Q_{2j})$   
 $Sqrt(D) = Sqrt(\sum_{j=1}^n D_{j,2j})$
3. Menghitung Dot Product  
 $Sum(Q * Di) = \sum_{j=1}^n Q_j Di_{j,j}$
4. Menghitung Similaritas  
 $Cosine \theta = Di / (Q + D) / (|Q| + |Di|)$
5. Membuat Ranking.  
 Setelah menghitung nilai cosinus lalu di buat perangkingan dari dokumen-dokumen tersebut

Indeks adalah bahasa yang digunakan di dalam sebuah buku konvensional untuk mencari informasi berdasarkan kata atau istilah yang mengacu ke dalam suatu halaman (Singh et al., 2015)

Metode TF-IDF (Term Frequency Inverse Document Frequency) merupakan suatu cara untuk memberikan bobot hubungan suatu kata

(term) terhadap dokumen (Roshdi & Roohparvar, 2015).

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 database computer (Manning, 2008).

information retrieval (IRS) merupakan suatu sistem yang menemukan informasi yang sesuai dengan kebutuhan user dari kumpulan informasi secara otomatis. Aplikasi Information Retrieval System sudah digunakan dalam banyak bidang seperti kedokteran, perusahaan dan lain sebagainya. Salah satu aplikasi dari Information Retrieval System adalah mesin pencari yang dapat diterapkan diberbagai bidang. Pada mesin pencari dengan Information Retrieval System user dapat memasukkan query yang bebas dalam arti kata query yang sesuai dengan bahasa manusia dan sistem dapat menemukan dokumen yang sesuai dengan query yang ditulis oleh user.

Prinsip kerja Information Retrieval System jika ada sebuah kumpulan dokumen dan seorang user yang memformulasikan sebuah pertanyaan (request atau query). Jawaban dari pertanyaan tersebut adalah sekumpulan dokumen yang relevan dan membuang dokumen yang tidak relevan. Information Retrieval System akan mengambil salah satu dari kemungkinan tersebut. Information Retrieval System dibagi dalam dua komponen utama yaitu sistem pengindeksan (indexing) menghasilkan basis data sistem dan temu kembali merupakan gabungan dari user interface dan look-up-table. Information Retrieval System didesain untuk menemukan dokumen atau informasi yang diperlukan oleh user.

Information Retrieval System bertujuan untuk menjawab kebutuhan informasi user dengan sumber informasi yang tersedia dalam kondisi seperti sebagai berikut (Salton, 1989);

- a. Mempresentasikan sekumpulan ide dalam sebuah dokumen menggunakan sekumpulan konsep.
- b. Terdapat beberapa pengguna yang memerlukan ide, tapi tidak dapat mengidentifikasi dan menemukannya dengan baik.
- c. Information Retrieval System bertujuan untuk mempertemukan ide yang dikemukakan oleh penulis dalam dokumen dengan kebutuhan informasi pengguna yang dinyatakan dalam bentuk key word query/istilah penelusuran.

Fungsi utama Information Retrieval System

- Mengidentifikasi sumber informasi yang relevan dengan minat masyarakat pengguna yang ditargetkan
- Menganalisis isi sumber informasi (dokumen)
- Merepresentasikan isi sumber informasi dengan cara tertentu yang memungkinkan untuk dipertemukan dengan pertanyaan pengguna
- Merepresentasikan pertanyaan (query) user dengan cara tertentu yang memungkinkan untuk dipertemukan sumber informasi yang terdapat dalam basis data. Mempertemukan pernyataan pencarian dengan data yang tersimpan dalam basis data
- Menemu-kembalikan informasi yang relevan
- Menyempurnakan unjuk kerja sistem berdasarkan umpan balik yang diberikan oleh user.

**HASIL DAN PEMBAHASAN**

Dalam hal ini penulis melakukan penelitian dengan kamus Bahasa sunda dan beberapa contoh kata yang penulis ambil untuk dijadikan sample dalam penghitungannya, dalam penghitungan ini berikut adalah kalimat dan kata yang diambil dan dijadikan penghitungan oleh penulis.

Contoh :

Query (Q) = Nalika Sunda

Dokumen 1 (D1) = Abdi hoyong diajar nalika sunda karena abdi kantun didaerah jawa kulon

Dokumen 2 (D2) = Angklung sarta suling mangrupa pakakas musik tradisional anu asalna ti jawa kulon sedengkeun sasando mangrupa pakakas musik ti wewengkon nusa tenggara wetan

Dokumen 3 (D3) = Nalika sunda mangrupa basa indung anu diomongkeun sahenteuna opat dua juta jalmi, nalika sunda oge nalika penutur terbanyak kadua sanggeus jawa

Tabel 1. Index

Index		
D1	D2	D3
Abdi	Angklung	Nalika
Hoyong	Sarta	Sunda
Ajar	Suling	Mangrupa
Nalika	Mangrupa	Basa
Sunda	Pakakas	Indung
Abdi	Musik	Omong
Kantun	Tradisional	Sahanteun
Daerah	Asal	juta
Jawa	Jawa	Jalmi
kulon	Kulon	Nalika
	Sasando	Sunda
	Mangrupa	Nalika

Pakakas	Penutur
Music	Banyak
Wewengkon	Dua
Nusa	Sanggeus
Tenggara	Jawa
Wetan	Opat
	Dua

Sumber: (Heristian et al., 2019)

Dalam tahap index ini kami memecah kata-kata yang terdapat dalam dokumen yang kami jadikan contoh, dari kalimat yang ada didokumen yang ada dijadikan kata dan dipilih atau yang diproses hanya kata-kata dasar saja.

Tabel 2 Perhitungan tf

Token	Tf			df	
	Q	D1	D2		D3
Abdi	0	2	0	0	1
Ajar	0	1	0	0	1
Angklung	0	0	1	0	1
Asal	0	0	1	0	1
Banyak	0	0	0	1	1
Basa	0	0	0	1	1
Daerah	0	1	0	0	1
Dua	0	0	0	2	1
Hoyong	0	1	0	0	1
Indung	0	0	0	1	1
Jalmi	0	0	0	1	1
Jawa	0	1	1	1	3
Juta	0	0	0	1	1
Kantun	0	1	0	0	1
Kulon	0	1	1	0	2
Mangrupa	0	0	2	1	2
Musik	0	0	2	0	1
Nalika	1	1	0	0	1
Nusa	0	0	1	0	1
Omong	0	0	0	1	1
Opat	0	0	0	1	1
Pakakas	0	0	1	0	1
Penutur	0	0	0	1	1
Sahanteun	0	0	0	1	1
Sanggeus	0	0	0	1	1
Sasando	0	0	1	0	1
Suling	0	0	1	0	1
Sunda	1	0	0	1	1
Tenggara	0	0	1	0	1
Tradisional	0	0	1	0	1
Wetan	0	0	1	0	1
Wewengkon	0	0	1	0	1

Sumber: (Heristian et al., 2019)

D1, D2, D3 = Dokumen

tf = banyak kata yang dicari pada sebuah dokumen

D = total dokumen



df = Banyak dokumen yang mengandung kata yang dicari

Tabel 3 Perhitungan tf\*idf

idf	Tf*idf			
Log(D/df)	Q	D1	D2	D3
0.47712125	0	0.9542	0	0
0.47712125	0	0.4771	0	0
0.47712125	0	0	0.4771	0
0.47712125	0	0	0.4771	0
0.47712125	0	0	0	0.4771
0.47712125	0	0	0	0.4771
0.47712125	0	0.4771	0	0
0.47712125	0	0	0	0.9542
0.47712125	0	0.4771	0	0
0.47712125	0	0	0	0.4771
0.47712125	0	0	0	0.4771
0.47712125	0	0.4771	0	0
0.47712125	0	0	0	0.9542
0.47712125	0	0.4771	0	0
0.47712125	0	0	0	0.4771
0.47712125	0	0	0	0.4771
0.47712125	0	0	0	0
0.47712125	0	0	0	0.4771
0.47712125	0	0.4771	0	0
0.17609126	0	0.1761	0.1761	0
0.17609126	0	0	0.3522	0.1761
0.47712125	0	0	0.9542	0
0.47712125	0.4771	0.4771	0	0
0.47712125	0	0	0.4771	0
0.47712125	0	0	0	0.4771
0.47712125	0	0	0	0.4771
0.47712125	0	0	0.4771	0
0.47712125	0	0	0	0.4771
0.47712125	0	0	0	0.4771
0.47712125	0	0	0.4771	0
0.47712125	0.4771	0	0	0.4771
0.47712125	0	0	0.4771	0
0.47712125	0	0	0.4771	0
0.47712125	0	0	0.4771	0
0.47712125	0	0	0.4771	0
0.47712125	0	0	0.4771	0

Sumber: (Heristian et al., 2019)

Tabel 4 Perhitungan jarak Q-D

Jarak Q-D			
Q	D1	D2	D3
0	0.910579	0	0
0	0.227645	0	0
0	0	0.227645	0
0	0	0.227645	0
0	0	0	0.227645
0	0	0	0.227645
0	0.227645	0	0

0	0	0	0.910579
0	0.227645	0	0
0	0	0	0.227645
0	0	0	0.227645
0	0	0	0
0	0	0	0.227645
0	0.227645	0	0
0	0.031008	0.031008	0
0	0	0.124033	0.031008
0	0	0.910579	0
0.227645	0.227645	0	0
0	0	0.227645	0
0	0	0	0.227645
0	0	0	0.227645
0	0	0.227645	0
0	0	0	0.227645
0	0	0	0.227645
0	0	0	0.227645
0	0	0.227645	0
0	0	0.227645	0
0.227645	0	0	0.227645
0	0	0.227645	0
0	0	0.227645	0
0	0	0.227645	0
0	0	0.227645	0
0.455289	2.07981	3.342066	3.445679

SQRT SQRT (D)

0.674751 1.442155 1.828132 1.856254

Sumber: (Heristian et al., 2019)

Tabel 5 perhitungan Dot Product

Dot produk		
Q*D1	Q*D2	Q*D3
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0.051822	0	0
0	0	0
0	0	0
0	0	0



0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0,051822
0	0	0
0	0	0
0	0	0
0	0	0
<b>SUM (Q*D)</b>		
<b>0.051822</b>	0	0.051822

Sumber: (Heristian et al., 2019)

Setelah semua penghitungan yang sudah dilakukan dapat dicari untuk rangking dari dokumen yang ada.

Tabel 6 Perhitungan Rangking

<b>RENGKING</b>		
<b>Dokumen</b>	<b>Nilai</b>	<b>Rengking</b>
D1	0.053255	1
D2	0	3
D3	0.041375	2

Sumber: (Heristian et al., 2019)

Setelah dilakukan penghitungan rangking hasil yang didapatkan bahwa dari tiga dokumen yang diujikan bahwa dokumen 1 atau D1 mendapatkan rangking 1 diikuti dengan dokumen 3 lalu yang terakhir adalah dokumen 3.

### KESIMPULAN

Mengambil informasi dari database yang besar cukup memakan waktu terutama jika informasi tersebut tidak terstruktur. Banyak algoritma dan teknik telah dikembangkan di bidang penambangan data dan pengambilan informasi namun mengambil data dari basis data besar terus menjadi masalah. Dalam penelitian ini, penulis menggunakan model ruang vector space model. Pertama, menghitung skor kemiripan menggunakan rata-rata tertimbang dari setiap item. Ukuran kosinus kemudian menghitung ukuran kemiripan dan untuk menentukan sudut antara vektor dokumen dan vektor query karena VSM didasarkan pada geometri di mana setiap istilah memiliki dimensi sendiri dalam ruang multi-dimensi, pertanyaan dan dokumen adalah titik atau vektor dalam ruang ini. Ukuran kosinus sering digunakan. Penelitian ini bekerja sangat signifikan.

Dalam hal ini bertujuan untuk merancang alat yang akan memungkinkan pengguna untuk mengambil informasi secara lebih efisien dan efektif, seperti Sistem Perpustakaan Buku atau Layanan Pengambilan Dokumen Komersial, akan berubah secara konstan saat dokumen-dokumen ditambahkan, diubah, dan dihapus. Ini membatasi jenis struktur data dan algoritma yang dapat digunakan untuk Information Retrieval.kamus. Hasil perhitungan Cosine diketahui bahwa Dokumen 1 (D1) memiliki tingkat similaritas tertinggi kemudian disusul dengan D3 dan D2 dokumen adalah titik atau vektor dalam ruang ini.hal tersebut pembuktian setiap query yang di berikan memiliki tngkat kedekatan yang berbeda dengan dokumen yang ada, maka bisa di Tarik kesimpulan dkumen yang memiliki tingkat similatas yang tinggi di pengaruhi oleh query yang di berikan.

### REFERENSI

- Erk, K., Padó, S., & Padó, U. (2010). A Flexible , Corpus-Driven Model of Regular and Inverse Selectional Preferences, (June).
- Heristian, S., Al Kautsar, H. A., & Sayfullloh, A. (2019). Laporan Akhir Penelitian Mandiri.
- INFORMATION RETRIEVAL SYSTEM PADA PENCARIAN FILE DOKUMEN BERBASIS TEKS DENGAN METODE VECTOR SPACE MODEL DAN ALGORITMA ECS STEMMER. (2016). *Jurnal Insand Comtech*, 1(1).
- Kom, M., & Setiawan, D. (2015). Rancang Bangun Information Retrieval System (IRS) Bahasa Jawa Ngoko pada Palintangan Penjebar Semangad dengan Metode Vector Space Model (VSM), 0624097401.
- Lopes, C. T., & Ribeiro, C. (2010). Using Local Precision to Compare Search Engines in, 835-836.
- Mao, W., & Chu, W. W. (2007). The phrase-based vector space model for automatic retrieval of free-text medical documents q, 61, 76-92. <https://doi.org/10.1016/j.datak.2006.02.008>
- Pujiatin, D., Lestari, U., Susanti, E., & Informatika, T. (2018). 3 1,2,3. *Aplikasi Transliterasi Bahasa Sunda The Application Sundanese Language Transliteration*, 5(2), 1-7.
- Roshdi, A., & Roohparvar, A. (2015). Review :

Information Retrieval Techniques and Applications, 3(9), 373-377.

Singh, V. K., Singh, V. K., Vishwavidyalaya, G., Programmer, S. A., & Vishwavidyalaya, G. G. (2015). VECTOR SPACE MODEL: AN INFORMATION RETRIEVAL, 141-143.

Yulianto, B., Budiharto, W., & Kartowisastro, I. H. (2017). The Performance of Boolean Retrieval and Vector Space Model in Textual Information, 11(1), 33-39.