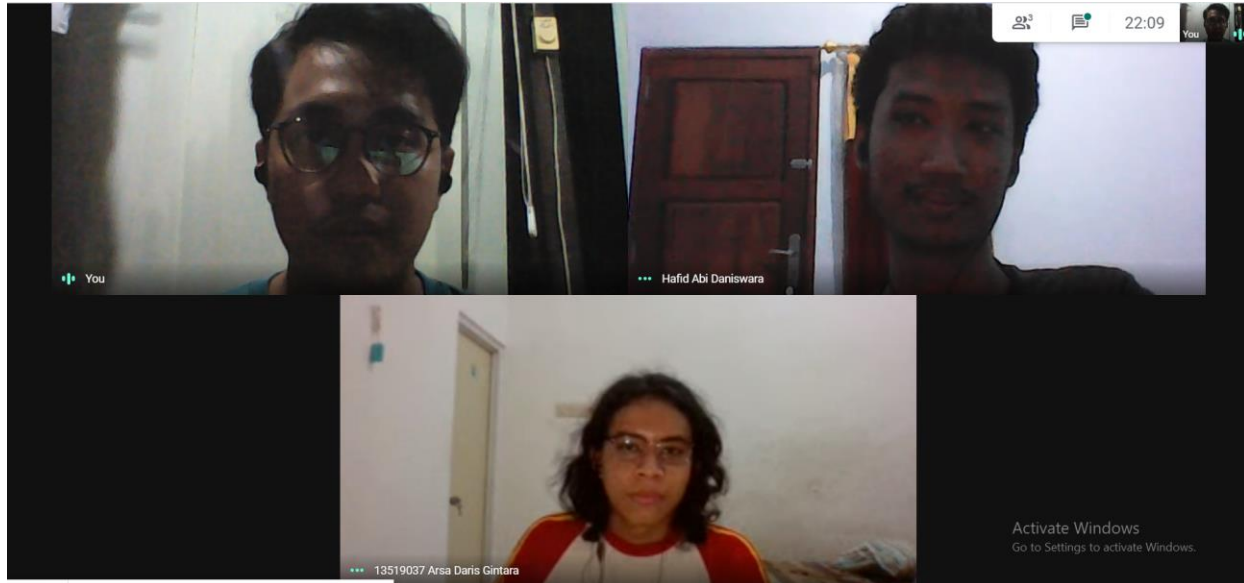


LAPORAN TUGAS BESAR 2 ALJABAR LINIER DAN GEOMETRI IF2123

Aplikasi Dot Product pada Sistem Temu Balik Informasi

Semester 1 Tahun 2020/2021



Disusun oleh :

Hafid Abi Daniswara (13519028)

Arsa Daris Gintara (13519037)

Syamil Cholid Abdurrasyid (13519052)

INSTITUT TEKNOLOGI BANDUNG

2020

BAB 1

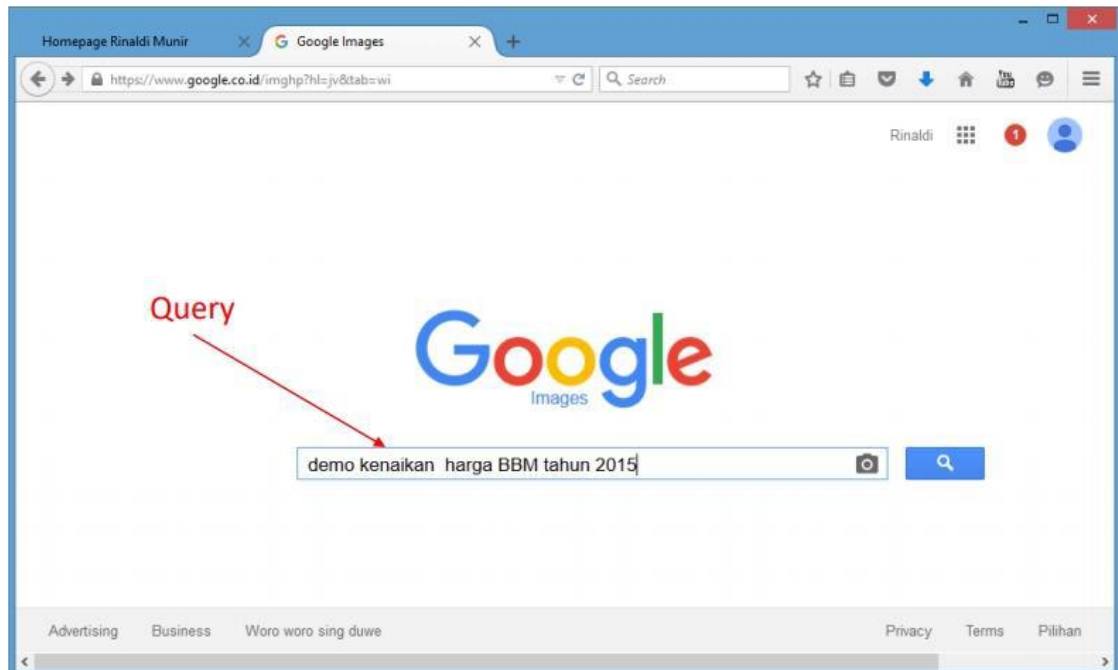
DESKRIPSI MASALAH

A. Deskripsi Singkat

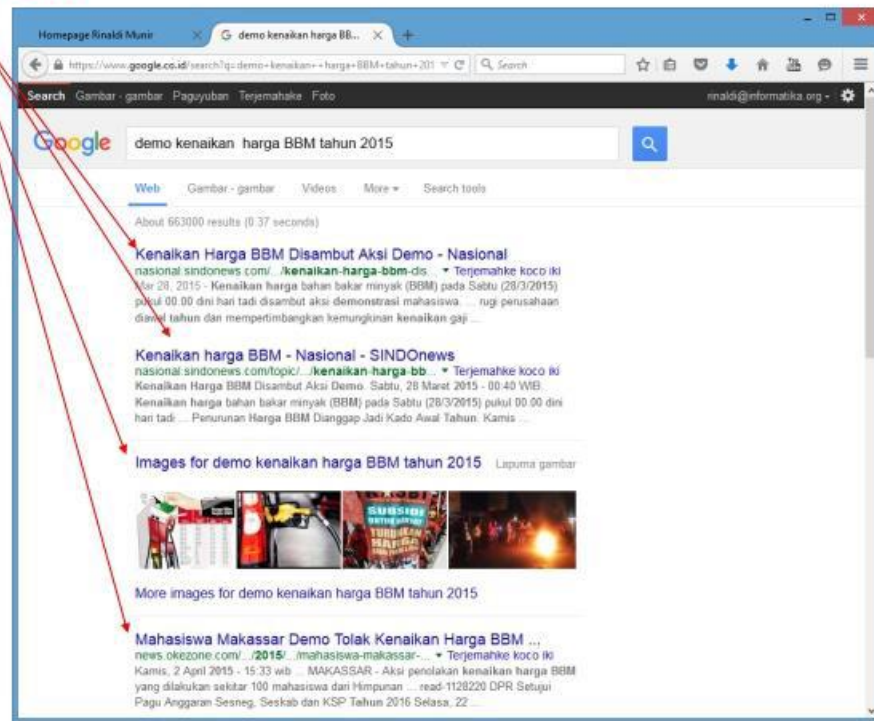
1. Abstraksi

Hampir semua dari kita pernah menggunakan *search engine*, seperti *google*, *bing* dan *yahoo! search*. Setiap hari, bahkan untuk sesuatu yang sederhana kita menggunakan mesin pencarian Tapi, pernahkah kalian membayangkan bagaimana cara *search engine* tersebut mendapatkan semua dokumen kita berdasarkan apa yang ingin kita cari?

Sebagaimana yang telah diajarkan di dalam kuliah pada materi vector di ruang Euclidean, temu balik informasi (*information retrieval*) merupakan proses menemukan kembali (*retrieval*) informasi yang relevan terhadap kebutuhan pengguna dari suatu kumpulan informasi secara otomatis. Biasanya, sistem temu balik informasi ini digunakan untuk mencari informasi pada informasi yang tidak terstruktur, seperti laman web atau dokumen.



Hasil pencarian:



Gambar 1 : Contoh penerapan Sistem Temu balik pada mesin pencarian

sumber : Aplikasi Dot Product pada Sistem Temu balik Informasi by Rinaldi Munir

Ide utama dari sistem temu balik informasi adalah mengubah *search query* menjadi ruang vektor. Setiap dokumen maupun *query* dinyatakan sebagai vektor $w = (w_1, w_2, \dots, w_n)$ di dalam R_n , dimana nilai w_i dapat menyatakan jumlah kemunculan kata tersebut dalam dokumen (term frequency). Penentuan dokumen mana yang relevan dengan *search query* dipandang sebagai pengukuran kesamaan (*similarity measure*) antara *query* dengan dokumen. Semakin sama suatu vektor dokumen dengan vektor *query*, semakin relevan dokumen tersebut dengan *query*. Kesamaan tersebut dapat diukur dengan *cosine similarity* dengan rumus:

$$\text{sim}(\mathbf{Q}, \mathbf{D}) = \cos \theta = \frac{\mathbf{Q} \cdot \mathbf{D}}{\|\mathbf{Q}\| \|\mathbf{D}\|}$$

Pada kesempatan ini, kalian ditantang untuk membuat sebuah *search engine* sederhana dengan model ruang vector dan memanfaatkan cosine similarity.

2. Penggunaan Program

Berikut ini adalah input yang akan dimasukkan pengguna untuk eksekusi program.

1. **Search query**, berisi kumpulan kata yang akan digunakan untuk melakukan pencarian
2. **Kumpulan dokumen**, dilakukan dengan cara mengunggah multiple file ke dalam web browser.

Tampilan layout dari aplikasi web yang akan dibangun adalah sebagai berikut.

My Simple Search Engine

Daftar Dokumen: <upload multiple files>

Search query

Hasil Pencarian: (diurutkan dari tingkat kemiripan tertinggi)

1. <Judul Dokumen 1>
Jumlah kata:
Tingkat Kemiripan:%
<Kalimat pertama dari Dokumen 1>

2. <Judul Dokumen 2>
Jumlah kata:
Tingkat Kemiripan:%
<Kalimat pertama dari Dokumen 2>

...

<Menampilkan tabel kata dan kemunculan di setiap dokumen>

Perihal

Gambar 2. Tampilan layout dari aplikasi web search engine yang dibangun.

Catatan: Teks yang diberikan warna **biru** merupakan hyperlink yang akan mengalihkan halaman ke halaman yang ingin dilihat. Apabila menekan *hyperlink* <Judul Dokumen 1>, maka akan diarahkan pada sebuah halaman yang berisi *full-text* terkait dokumen 1 tersebut (seperti *Search Engine*).

Anda dapat menambahkan menu lainnya, gambar, logo, dan sebagainya. Tampilan Front End dari website dibuat semenarik mungkin selama mencakup seluruh informasi pada layout yang diberikan di atas.

Data uji berupa dokumen-dokumen yang akan diunggah ke dalam web browser. Format dan extension dokumen dibebaskan selama bisa dibaca oleh web browser (misalnya adalah dokumen dalam bentuk file *txt* atau file *html*). Minimal terdapat 15 dokumen berbeda.

Tabel term dan banyak kemunculan term dalam setiap dokumen akan ditampilkan pada web browser dengan layout sebagai berikut.

Term	Query	D1	D2	...	DN
Term 1					
Term 2					
...					
Term N					

Untuk menyederhanakan pembuatan search engine, terdapat hal-hal yang perlu diperhatikan dalam eksekusi program ini.

1. Silahkan lakukan stemming dan penghapusan *stopwords* pada setiap dokumen
2. Tidak perlu dibedakan antara huruf-huruf besar dan huruf-huruf kecil.
3. *Stemming* dan penghapusan stopword dilakukan saat **penyusunan vektor**, sehingga halaman yang berisi *full-text* terkait dokumen tetap seperti semula
4. Penghapusan karakter-karakter yang tidak perlu untuk ditampilkan (jika menggunakan *web scraping* atau format dokumen berupa html)
5. Bahasa yang digunakan dalam dokumen adalah bahasa Inggris atau bahasa Indonesia (pilih salah satu)

B. Saran Pengerjaan

Anda disarankan untuk membuat program testing pada **backend** terlebih dahulu untuk menguji keberhasilan dari perhitungan cosine similarity tersebut.

C. Spesifikasi Tugas

Buatlah program mesin pencarian dengan sebuah website lokal sederhana. Spesifikasi program adalah sebagai berikut:

1. Program mampu menerima *search query*. *Search query* dapat berupa kata dasar maupun berimbuhan.
2. Dokumen yang akan menjadi kandidat dibebaskan formatnya dan disiapkan secara manual. Minimal terdapat 15 dokumen berbeda sebagai kandidat dokumen. **Bonus:** Gunakan web scraping untuk mengekstraksi dokumen dari website.
3. Hasil pencarian yang terurut berdasarkan similaritas tertinggi dari hasil teratas hingga hasil terbawah berupa judul dokumen dan kalimat pertama dari dokumen tersebut. Sertakan juga nilai similaritas tiap dokumen.
4. Program disarankan untuk melakukan pembersihan dokumen terlebih dahulu sebelum diproses dalam perhitungan cosine similarity. Pembersihan dokumen bisa meliputi hal-hal berikut ini.
 - a. Stemming dan penghapusan stopwords dari isi dokumen
 - b. Penghapusan karakter-karakter yang tidak perlu
5. Program dibuat dalam sebuah website lokal sederhana. Dibebaskan untuk menggunakan *framework* pemrograman website apapun. Salah satu framework website yang bisa dimanfaatkan adalah Flask (Python), ReactJS, dan PHP.
6. Kalian dapat menambahkan fitur fungsional lain yang menunjang program yang anda buat (unsur kreativitas diperbolehkan/dianjurkan).
7. Program harus modular dan mengandung komentar yang jelas.
8. Dilarang menggunakan library cosine similarity yang sudah jadi.

BAB 2

TEORI SINGKAT

A. Sistem Temu balik Informasi

Temu balik informasi adalah aktivitas untuk mendapatkan suatu sumber sistem informasi yang relevan terhadap suatu informasi yang dibutuhkan dari semua ketersediaan informasi. Pencarian bisa berdasarkan suatu kalimat/text tertentu atau berdasarkan suatu konten lain tertentu. Sistem temu balik informasi adalah suatu sains untuk pencarian suatu informasi di suatu dokumen, pencarian dokumen itu sendiri, juga untuk mencari suatu metadata yang menjelaskan suatu data, mencari database text, gambar, ataupun suara.

Sistem temu balik informasi otomatis biasa dipakai untuk mengurangi apa yang disebut information overload atau terlalu banyaknya informasi yang tersedia. Suatu sistem temu balik informasi adalah sebuah perangkat lunak yang memberikan akses kepada kumpulan buku, jurnal, dan dokumen-dokumen lain dan menyimpan serta mengatur dokumen tersebut. Web search engines adalah salah satu aplikasi sistem temu balik informasi yang paling sering ditemukan.

Suatu proses temu balik informasi dimulai dengan menginputkan sesuatu yang disebut 'query' ke dalam sistem. 'Query' adalah suatu statement formal dari informasi yang dibutuhkan. Sebagai contoh, query adalah kalimat yang kita inputkan pada web search engine seperti google untuk mencari tentang hal tersebut. Dalam temu balik informasi, query tidak mengidentifikasi suatu objek unik tunggal melainkan beberapa objek yang masuk atau cocok ke dalam statement query yang kita inputkan dengan berbagai macam derajat relevansi yang berbeda.

Objek adalah suatu entitas yang direpresentasikan oleh informasi dalam suatu koleksi konten atau database. Input query dari pengguna dicocokkan terhadap database informasi. namun, berkebalikan dengan SQL query klasik dari suatu database, dalam temu balik informasi hasil pencarian yang dikembalikan bisa cocok atau bisa juga tidak cocok dengan query tergantung database yang tersedia sehingga hasilnya merupakan suatu peringkat dari yang tercocok atau memiliki kecocokan tertinggi ataupun bisa sebaliknya

(jika dibutuhkan). Pemeringkatan tersebut adalah kunci yang membedakan pencarian temu balik informasi dengan pencarian database biasa.

Berdasarkan pengaplikasian berbagai data objek yang mungkin seperti dokumen text, gambar, audio, mind map, ataupun video, biasanya database dokumen tersebut terpisah atau tidak disimpan secara bersamaan dalam Sistem temu balik informasi melainkan direpresentasikan di dalam sistem dengan perwakilan dokumen atau metadata.

Sebagian besar sistem temu balik informasi mengkomputasi suatu skor numerik berdasarkan seberapa baik relevansi atau kecocokan setiap objek database terhadap query, dan diperingkatkan berdasarkan nilai tersebut. Peringkat-peringkat terbaik kemudian akan ditampilkan kepada pengguna. Proses tersebut berulang seperti itu setiap kali pengguna berkeinginan untuk mengubah query nya.

B. Vektor

Matriks transpose atau transpose dari suatu matriks adalah pertukaran letak tiap elemen antara baris dan kolomnya yang berarti alamat kolom menjadi baris dan baris menjadi kolom. Misal suatu elemen dari matriks berukuran $m \times n$ adalah a_{ij} maka dalam matriks transposenya menjadi a_{ji} dan ukuran matriksnya menjadi $n \times m$. Jika ada suatu matriks A, maka transposenya dilambangkan sebagai A^T .

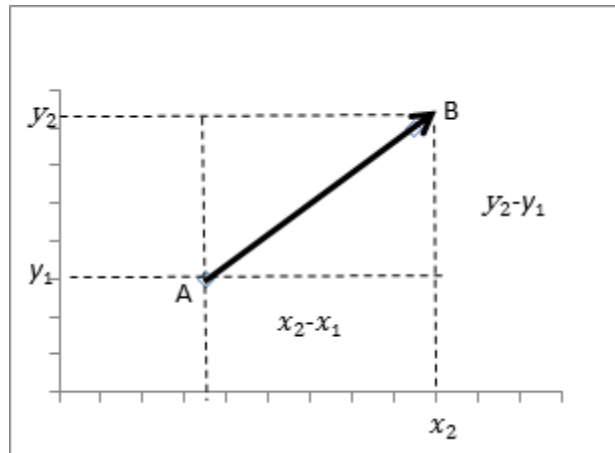
1. Pengertian Vektor

Vektor merupakan sebuah besaran yang memiliki arah. Vektor digambarkan sebagai panah dengan yang menunjukkan arah vektor dan panjang garisnya disebut besar vektor. Dalam penulisannya, jika vektor berawal dari titik A dan berakhir di titik B bisa ditulis dengan sebuah huruf kecil yang di atasnya ada tanda garis/ panah seperti \vec{v} atau \bar{v} atau juga:

$$\vec{AB}$$

Misalkan vektor \vec{v} merupakan vektor yang berawal dari titik $A(x_1, y_1)$ menuju titik $B(x_2, y_2)$ dapat digambarkan koordinat cartesius dibawah. Panjang garis sejajar

sumbu x adalah $v_1 = x_2 - x_1$ dan panjang garis sejajar sumbu y adalah $v_2 = y_2 - y_1$ merupakan komponen-komponen vektor \vec{v} .



Komponen vektor \vec{v} dapat ditulis untuk menyatakan vektor secara aljabar yaitu:

$$\vec{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} x_2 - x_1 \\ y_2 - y_1 \end{pmatrix} \text{ atau } \vec{v} = (v_1, v_2)$$

2. Jenis-Jenis Vektor

Ada beberapa jenis vektor khusus yaitu:

a. Vektor Posisi

Suatu vektor yang posisi titik awalnya di titik 0 (0,0) dan titik ujungnya di A (a_1, a_2)

b. Vektor Nol

Suatu vektor yang panjangnya nol dan dinotasikan $\vec{0}$. Vektor nol tidak memiliki arah vektor yang jelas

c. Vektor satuan

Suatu vektor yang panjangnya satu satuan. Vektor satuan dari $\vec{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$ adalah:

$$\vec{U}_v = \frac{\vec{v}}{|\vec{v}|} = \frac{1}{|\vec{v}|} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

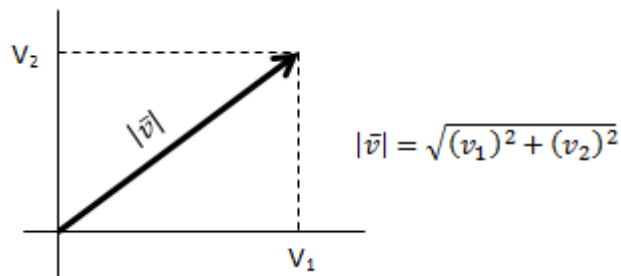
d. Vektor basis

Vektor basis merupakan vektor satuan yang saling tegak lurus. Dalam vektor ruang dua dimensi (R^2) memiliki dua vektor basis yaitu $\vec{i} = (1, 0)$ dan $\vec{j} = (0, 1)$.

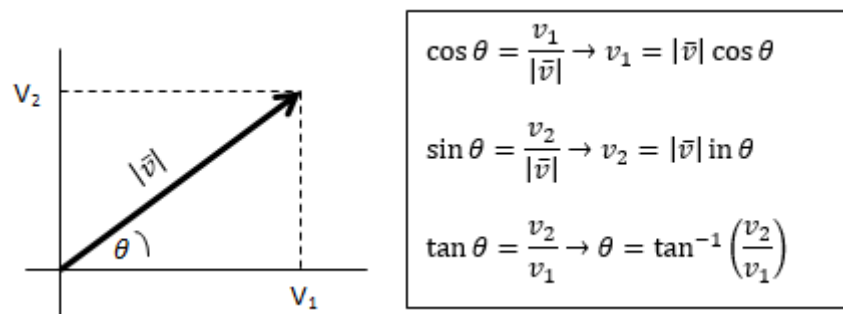
Sedangkan dalam tiga dimensi (R^3) memiliki tiga vektor basis yaitu $\vec{i} = (1, 0, 0)$, $\vec{j} = (0, 1, 0)$, dan $\vec{k} = (0, 0, 1)$

3. Vektor di R^2

Panjang segmen garis yang menyatakan vektor \vec{v} atau dinotasikan sebagai $|\vec{v}|$ Panjang vektor sebagai:



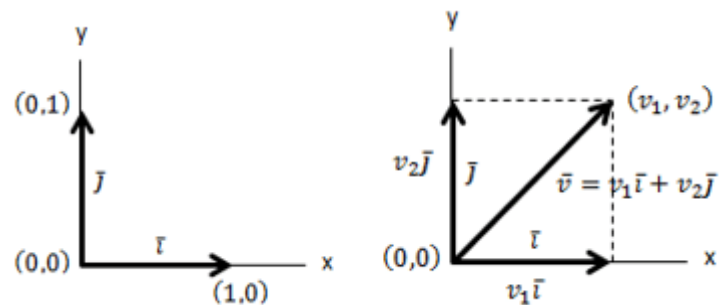
Panjang vektor tersebut dapat dikaitkan dengan sudut θ yang dibentuk oleh vektor dan sumbu x. positif.



Vektor dapat disajikan sebagai kombinasi linier dari vektor basis $\vec{i} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ dan $\vec{j} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ sebagai berikut:

$$\vec{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = v_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + v_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$\vec{v} = v_1 \vec{i} + v_2 \vec{j}$$



4. Operasi Vektor di \mathbb{R}^2

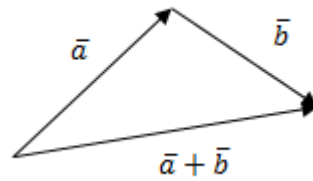
Penjumlahan dan pengurangan vektor di \mathbb{R}^2

Dua vektor atau lebih dapat dijumlahkan dan hasilnya disebut resultan. Penjumlahan vektor secara aljabar dapat dilakukan dengan cara menjumlahkan komponen yang

seletak. Jika $\vec{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ dan $\vec{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$ maka:

$$\vec{a} + \vec{b} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \end{pmatrix}$$

Penjumlahan secara grafis dapat dilihat pada gambar dibawah:



Dalam pengurangan vektor, berlaku sama dengan penjumlahan yaitu:

$$\bar{a} - \bar{b} = \begin{pmatrix} a_1 - b_1 \\ a_2 - b_2 \end{pmatrix}$$

Sifat-sifat dalam penjumlahan vektor sebagai berikut:

- $\bar{a} + \bar{b} = \bar{b} + \bar{a}$
- $\bar{a} + (\bar{b} + \bar{c}) = (\bar{a} + \bar{b}) + \bar{c}$

Perkalian vektor di \mathbb{R}^2 dengan skalar






Suatu vektor dapat dikalikan dengan suatu skalar (bilangan real) dan akan menghasilkan suatu vektor baru. Jika \bar{v} adalah vektor dan k adalah skalar. Maka perkalian vektor:

$$k \cdot \bar{v}$$

Dengan ketentuan:

- Jika $k > 0$, maka vektor $k \cdot \bar{v}$ searah dengan vektor \bar{v}
- Jika $k < 0$, maka vektor $k \cdot \bar{v}$ berlawanan arah dengan vektor \bar{v}
- Jika $k = 0$, maka vektor $k \cdot \bar{v}$ adalah vektor identitas $\bar{0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

Secara grafis perkalian ini dapat merubah panjang vektor dan dapat dilihat pada tabel dibawah:

\vec{v}	$k = 0,5$	$k = -0,5$	$k = 2$	$k = -2$
\vec{v}	$0,5 \cdot \vec{v}$	$-0,5 \cdot \vec{v}$	$2 \cdot \vec{v}$	$-2 \cdot \vec{v}$
				

Secara aljabar perkalian vektor \vec{v} dengan skalar k dapat dirumuskan:

$$k \cdot \vec{v} = \begin{pmatrix} k \cdot v_1 \\ k \cdot v_2 \end{pmatrix}$$

Perkalian Skalar Dua Vektor di \mathbb{R}^2

Perkalian skalar dua vektor disebut juga sebagai hasil kali titik dua vektor dan ditulis sebagai:

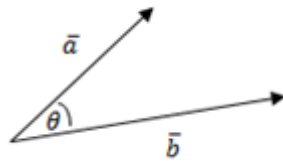
$$\vec{a} \cdot \vec{b} \text{ (dibaca : a dot b)}$$

Perkalian skalar vektor \vec{a} dan \vec{b} dilakukan dengan mengalikan panjang vektor \vec{a} dan panjang vektor \vec{b} dengan cosinus θ . Sudut θ yang merupakan sudut antara vektor \vec{a} dan vektor \vec{b} .

Sehingga:

$$\vec{a} \cdot \vec{b} = |\vec{a}| |\vec{b}| \cos \theta$$

Dimana:



Perhatikan bahwa:

- Hasil kali titik dua vektor menghasilkan suatu skalar
- $\vec{a} \cdot \vec{a} = (\vec{a}^2)$
- $\vec{a} \cdot (\vec{b} + \vec{c}) = (\vec{a} \cdot \vec{b}) + (\vec{a} \cdot \vec{c})$

5. Vektor di \mathbb{R}^3

Vektor yang berada pada ruang tiga dimensi (x, y, z). jarak antara dua titik vektor dalam \mathbb{R}^3 dapat diketahui dengan pengembangan rumus pythagoras. Jika titik $A(x_1, y_1, z_1)$ dan titik $B(x_2, y_2, z_2)$ maka jarak AB adalah:

$$AB = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

Atau jika $\vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$, maka :

$$|\vec{v}| = \sqrt{(v_1)^2 + (v_2)^2 + (v_3)^2}$$

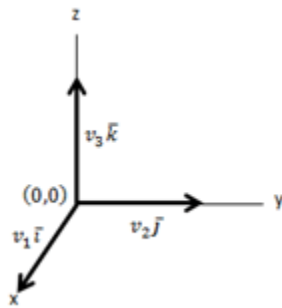
Vektor \vec{AB} dapat dinyatakan dalam dua bentuk, yaitu dalam kolom $\vec{AB} = \begin{pmatrix} b_1 - a_1 \\ b_2 - a_2 \\ b_3 - a_3 \end{pmatrix}$

atau dalam baris $\vec{AB} = (b_1 - a_1, b_2 - a_2, b_3 - a_3)$.

Vektor juga dapat disajikan sebagai kombinasi linier dari vektor basis $\vec{i}(1, 0, 0)$ dan $\vec{j}(0, 1, 0)$ dan $\vec{k}(0, 0, 1)$ sebagai berikut:

$$\vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = v_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + v_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + v_3 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

$$\bar{v} = v_1 \bar{I} + v_2 \bar{J} + v_3 \bar{K}$$



6. Operasi Vektor di R^3

Operasi vektor di R^3 secara umum, memiliki konsep yang sama dengan operasi vektor di R^2 dalam penjumlahan, pengurangan, maupun perkalian.

Penjumlahan dan pengurangan vektor di R^3

Penjumlahan dan pengurangan vektor di R^3 sama dengan vektor di R^2 yaitu:

$$\bar{a} + \bar{b} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \\ a_3 + b_3 \end{pmatrix}$$

dan

$$\bar{a} - \bar{b} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} - \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} a_1 - b_1 \\ a_2 - b_2 \\ a_3 - b_3 \end{pmatrix}$$

Perkalian vektor di R^3 dengan skalar

Jika \bar{v} adalah vektor dan k adalah skalar. Maka perkalian vektor:

$$k \cdot \bar{v} = \begin{pmatrix} k \cdot v_1 \\ k \cdot v_2 \\ k \cdot v_3 \end{pmatrix}$$

Hasil kali skalar dua vektor

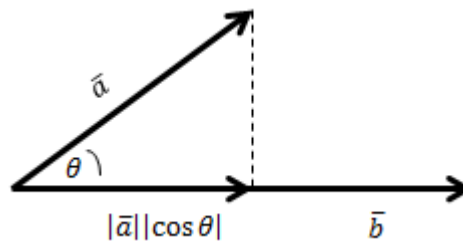
Selain rumus di R^3 , ada rumus lain dalam hasil kali skalar dua vektor.

Jika $\vec{a} = a_1\vec{i} + a_2\vec{j} + a_3\vec{k}$ dan $\vec{b} = b_1\vec{i} + b_2\vec{j} + b_3\vec{k}$ maka $\vec{a} \cdot \vec{b}$ adalah:

$$\vec{a} \cdot \vec{b} = (a_1b_1) + (a_2b_2) + (a_3b_3)$$

Proyeksi Orthogonal vektor

Jika vektor \vec{a} diproyeksikan ke vektor \vec{b} dan diberi nama \vec{c} seperti gambar dibawah:



Diketahui:

$$\vec{a} \cdot \vec{b} = |\vec{a}| |\vec{b}| \cos \theta \xrightarrow{\text{maka}} \cos \theta = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| |\vec{b}|}$$

Sehingga:

$$|\vec{c}| = |\vec{a}| \cos \theta \quad \text{atau} \quad |\vec{c}| = \left| \frac{\vec{a} \cdot \vec{b}}{|\vec{b}|} \right|$$

Untuk mendapat vektornya:

$$\vec{c} = \left| \frac{\vec{a} \cdot \vec{b}}{|\vec{b}|} \right| \vec{b}$$

C. Cosine Similarity

Cosine Similarity dapat diimplementasikan untuk menghitung nilai kemiripan antar kalimat dan menjadi salah satu teknik untuk mengukur kemiripan teks yang populer. Rumus cosinus dari 2 vektor bukan nol dapat diturunkan dari formula rumus Euclidean dot product :

$$\mathbf{A} \cdot \mathbf{B} = \|\mathbf{A}\| \|\mathbf{B}\| \cos \theta$$

jika diberikan 2 buah vektor A dan B, *cosine similiary*, adalah nilai dari $\cos(\theta)$ yang direpresentasikan dengan dot product dan magnitudo sebagai berikut :

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

Contoh penggunaan *Cosine Similarity* dalam menguji kemiripan dua buah kalimat adalah sebagai berikut:

Diberikan dua buah kalimat yaitu kalimat A dan B, yaitu:

A : Julie loves me more than Linda loves me

B : Jane likes me more than Julie loves me

Dengan:

$$n = |A \cap B|$$

A_i = jumlah kemunculan kata indeks ke - i dari daftar kata pada kalimat A.

B_i = jumlah kemunculan kata indeks ke - i dari daftar kata pada kalimat B.

Tabel Contoh uji kemiripan teks

indeks	Daftar Kata	Jumlah Kemunculan Kata	
		A	B
1	Julie	1	1
2	loves	2	1
3	me	2	2
4	more	1	1
5	than	1	1
6	Linda	1	0
7	Jane	0	1
8	likes	0	1

Berdasarkan rumus tersebut di atas dilakukan penghitungan seperti di bawah ini.

Tingkat kemiripan teks =

$$= \frac{(1 \times 1) + (2 \times 1) + (2 \times 2) + (1 \times 1) + (1 \times 1) + (1 \times 0) + (0 \times 1) + (0 \times 1)}{\sqrt{1^2 + 2^2 + 2^2 + 1^2 + 1^2 + 1^2 + 0^2 + 0^2} \times \sqrt{1^2 + 1^2 + 2^2 + 1^2 + 1^2 + 0^2 + 1^2 + 1^2}}$$

$$= 0.821584$$

BAB 3

IMPLEMENTASI PROGRAM

A. File

Dalam program ini digunakan Django versi 3.x.x dalam membuat program ini. Dalam struktur Django sendiri berbentuk sebagai berikut :

Tubes2Algeo

```
|----- Search_engine
|----- Templates
|----- Tubes2Algeo
|----- Static
|----- Media
|----- Db.sqlite3
|----- Manage.py
```

B. Penjelasan

1. Pada search engine terdapat file-file/class sebagai berikut :

- 1) Admin.py (bawaan Django untuk page admin, namun tidak dipakai karena tidak ada skema login/register/logout).
- 2) Apps.py (untuk melakukan konfigurasi pada app django).
- 3) Models.py (untuk membuat model database), pada models ditaruh 1 models yaitu filestorage sebagai model untuk menyimpan url penyimpanan dokumen asli, penyimpanan url dokumen hasil stemming dan judul dokumen.

```

from django.db import models

# Create your models here.

class filestorage(models.Model):
    url_dokumen = models.CharField(max_length=100)
    url_sastrawee = models.CharField(max_length=100)
    judul = models.CharField(max_length=100)

    class Meta:
        db_table = "fileloc"

```

- 4) objekPendukung.py. Objek pendukung dibuat untuk memudahkan passing query result dari controller ke view dengan membuat objek HasilPencarian sebagai berikut.

```

class HasilPencarian:
    def __init__(self, id, judul, fileloc, sastrawifile):
        self.id = id
        self.title = judul
        self.ori_file = fileloc
        self.stemmed_file = sastrawifile
        self.similaritas = 0
        self.vectorDict = {}
        self.preview = ""
        self.jumlahkata = 0

    def setsimilaritas(self, similaritas):
        self.similaritas = similaritas

    def setpreview(self, previewdocs):
        self.preview = previewdocs

    def setvectorDict(self, vdict):
        self.vectorDict = vdict

    def setJumlahKata(self, number):
        self.jumlahkata = number

```

- 5) supporter.py, berisikan fungsi-fungsi pembantu untuk memudahkan program.

Berikut adalah list fungsi pembantu :

- txtToStr : untuk mengubah dari file txt menjadi string
- ignoreSpace : mengabaikan semua spasi
- stemming : untuk stemming kalimat
- convertDir : untuk mengubah '/' menjadi '\'
- getFileExtension : untuk mendapatkan ekstensi dari suatu file
- read_query : untuk mengubah string jadi list per kata

- `read_doc` : untuk mengubah dokumen txt jadi list perkata
- `listfile` : untuk melist file di suatu direktori
- `getparentdir` : untuk melihat parent dari suatu direktori
- `getfilename` : untuk mendapatkan nama file, bila inputan merupakan detail dalam direktori (ex: asd/fgh/jkl.txt maka akan return jkl.txt)
- `getjudul` : mereturn nama file tanpa .txt (semisal orang berseragam.txt menjadi orang berseragam)
- `dotproduct` : perkalian dot
- `length` : menghitung Panjang dari suatu vector / $|v|$ (magnitudo)
- `sim` : mencari nilai kosinus dari perkalian dot
- `maekQueryVektor` : mengubah suatu query searching menjadi dictionary dengan key adalah term dan value adalah frekuensi kemunculan
- `makeDocsVektor` : mengubah suatu dokumen/file txt menjadi dictionary dengan key adalah term dan value adalah frekuensi kemunculan
- `toVector` : merubah dictionary menjadi vector (list yang berisikan kemunculan term” pada dokumen / query sesuai dengan urutan key pada dictionary)
- `appendKK` : untuk menambahkan kamus kata baru setelah upload dokumen baru
- `makeNolDict` : membuat dictionary dari kamus kata dan memberi semua valuenya dengan 0
- `sortDictionary` : mengurutkan dictionary dengan key adalah judul dokumen dan value adalah cosine similarity dan mengurutkan dari similarity terbesar hingga terkecil
- `getPrevDocs` : mendapatkan preview dari suatu dokumen dengan mengambil 300 huruf pertama dari suatu dokumen dan untuk ditampilkan preview hasil searching nantinya

6) `tests.py` (untuk testing pada Django, tidak dipakai).

7) `views.py`, untuk controller backend pada Django sebelum direturn ke view.

Mengandung beberapa fungsi sebagai berikut :

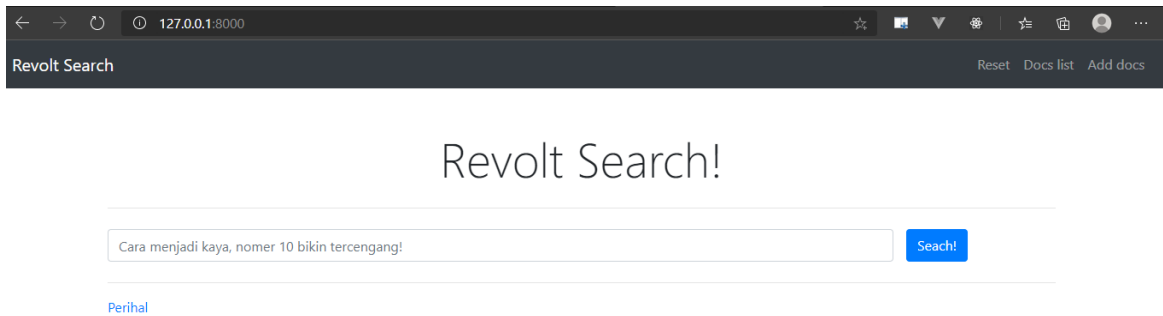
- `cobacek` : untuk menampilkan list dokumen yang telah terinput

- `perihal` : untuk menampilkan perihal dari aplikasi
 - `searchresult` : melakukan searching dan pengelolaan terhadap query lalu mencocokkan similaritasnya dengan semua dokumen terdaftar di aplikasi
 - `reset` : untuk reset semua session, del semua dokumen yang terupload
 - `upfile` : untuk upload file dan stemming file
 - `artikel` : untuk menampilkan artikel
 - `get_item` : fungsi pembantu untuk diakses pada template, untuk memudahkan mendapat value dari dictionary
2. Pada templates terdapat file-file html untuk front end. Dalam hal ini dimanfaatkan html + css + javascript serta framework bootstrap, dalam templates terdapat beberapa file html sebagai berikut :
 - 1) `Basis.html` : berisi meta dan import file, navbar, dan footer
 - 2) `Perihal.html` : berisi template perihal dari program
 - 3) `Qresult.html` : berisi template hasil query searching program
 - 4) `Searchbar.html` : berisi template untuk search bar
 - 5) `Tampilanartikel.html` : berisi template untuk tampilan artikel
 3. Folder `tubes2Algo` terdapat file-file setting dan konfigurasi dasar dari Django, dengan file-file sebagai berikut :
 - 1) `Asgi.py` (konfigurasi ASGI, bawaan dari django)
 - 2) `Setting.py` (untuk settings dan konfigurasi website)
 - 3) `Urls.py` (untuk daftar url dan webpath dalam django)
 - 4) `Wsgi.py` (konfigurasi WSGI, bawaan django)
 4. Pada static terdapat file-file untuk menyimpan gambar yang dibutuhkan dalam web dan file-file css, javascript, dsb dari bootstrap.
 5. Pada folder media berisi file-file txt hasil upload.
 6. `Db.sqlite3` merupakan file database SQL dengan platform sqlite. Berfungsi untuk menyimpan data seperti session dan data” lainnya (merupakan bawaan dari django).
 7. `Manage.py` adalah file yang akan di-*run* untuk menjalankan server atau melakukan migrasi dan hal lainnya.

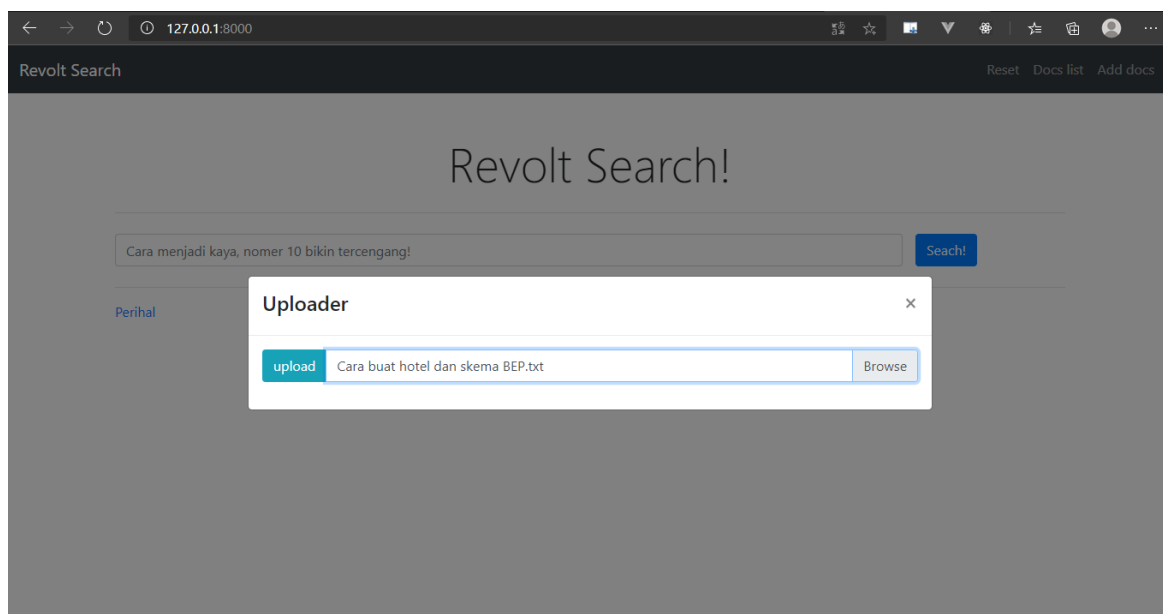
BAB 4

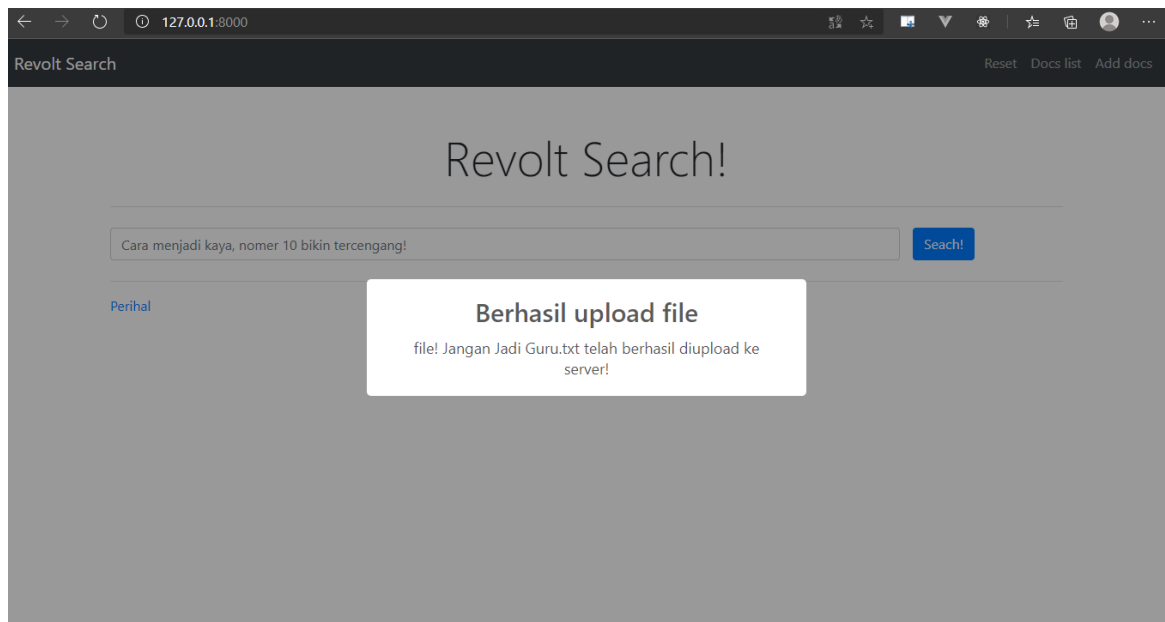
EKSPERIMEN

1. Tampilan Utama

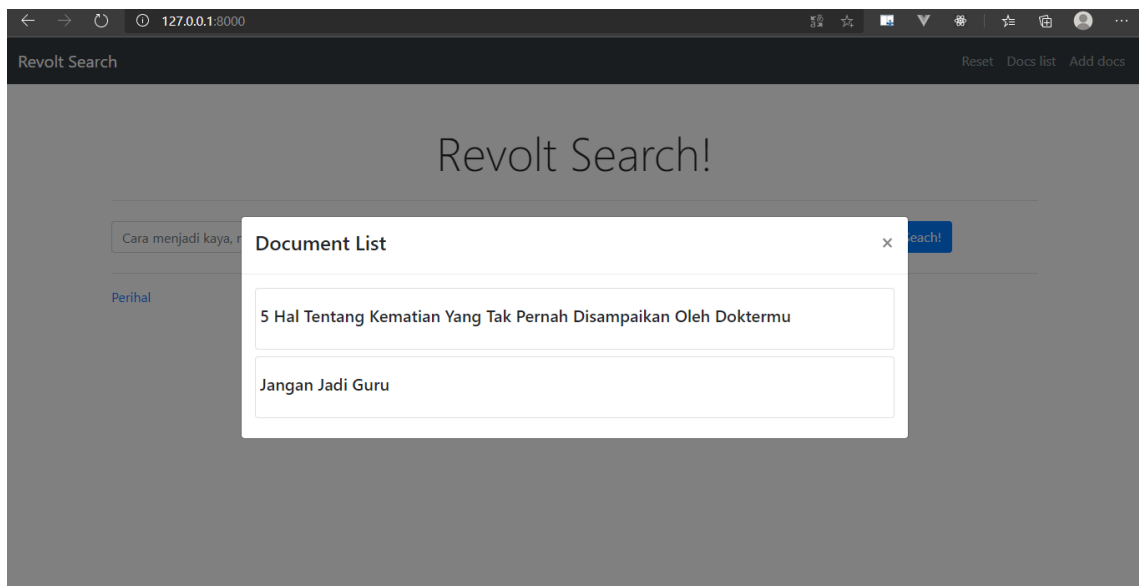


2. Upload Dokumen





3. List Dokumen



4. Hasil Search

The screenshot shows a web browser window with the address bar displaying '127.0.0.1:8000/q/?pencarian=kematian+cara'. The search bar contains the text 'Cara menjadi kaya, nomer 10 bikin tercengang!' and a blue 'Search!' button. Below the search bar, the 'Search Result' section displays two results. The first result is titled '5 Hal tentang Kematian yang Tak Pernah Disampaikan oleh Doktermu' with a 'Details' button. The second result is titled 'Jangan Jadi Guru' with a 'Details' button. Below the search results, the 'Kamus Data' section shows a table with four columns: Term, Query, D1, and D2.

Term	Query	D1	D2
mati	1	18	0
cara	1	5	2

5. Detail Tiap Dokumen

The screenshot shows the 'Cosine similarity details' modal window. It displays the cosine similarity value: 0.3294466331893823. Below this, it states 'Jumlah kata: 505'. A table shows the results of the cosine similarity calculation for the query 'Cara menjadi kaya, nomer 10 bikin tercengang!'. The table has four columns: #, Term, Kemunculan di dokumen, and Kemunculan di query pencarian.

#	Term	Kemunculan di dokumen	Kemunculan di query pencarian
1	mati	18	1
2	seringkali	2	0
3	jadi	16	0
4	topik	1	0
5	yang	9	0
6	hindar	1	0
7	untuk	4	0
8	diskusi	1	0

6. Isi Dokumen

← → ↻ 127.0.0.1:8000/artikel/?konten=5%20Hal%20tentang%20Kematian%20yang%20Tak%20Pernah%20Disamp... ☆ 📄 🔍 🗑️

[Kembali](#)

5 Hal tentang Kematian yang Tak Pernah Disampaikan oleh Doktermu

Kematian seringkali menjadi topik yang dihindari untuk didiskusikan. Tapi diakui atau tidak, hal ini secara pasti akan terjadi dan menghampiri semua makhluk hidup, termasuk manusia. Percakapan mengenai hal yang satu ini pun tak pernah mudah, apalagi ketika seorang dokter harus memberitahu perihal kematian pasiennya kepada keluarga atau bahkan memberi perkiraan umur langsung kepada pasien yang bersangkutan. Tapi, bagaimanapun bukanlah sesuai etikanya seorang dokter harus mengatakan kebenaran? Ketahui 5 fakta mengenai kematian di bawah ini yang mungkin saja tak pernah dikatakan secara jujur oleh doktermu.

Walau waktu persis kematian tidak bisa ditentukan, tapi seperti ibu hamil yang merasakan tanda-tanda umum jelang persalinan, sejatinya ada beberapa kondisi umum yang bisa diukur jadi tanda

Siapa pun pasti akan diliputi rasa khawatir ketika merawat orang tua atau kerabat yang sudah lanjut usia ataupun sedang sakit parah. Agar kamu lebih siap mental dan waspada, maka ada baiknya mengenali waktu-waktu hidup terakhir seseorang. Umumnya, mereka akan mengalami perubahan secara fisik dalam beberapa hari atau jam jelang kematiannya. Lebih sering lelah dan mengantuk misalnya. Sebab, perubahan metabolisme akan membuat pasien jadi tidak bertenaga, lelah dan mengantuk. Mereka akan menghabiskan lebih banyak waktu untuk tidur dan bisa jadi malah tak sadarkan diri dalam tidurnya. Mereka pun akan cenderung menolak makanan dan atau minuman karena merasa kesulitan dalam mengonsumsinya melalui mulut. Hal itu terjadi karena tubuh memang tak lagi sanggup memproses makanan dengan baik. Mengolesi bibir pasien dengan cairan atau lip balm bisa sedikit membantu.

Ketika perubahan napas sudah terjadi, itu berarti kematian telah dekat. Entah napas jadi melambat atau bahkan jadi sangat cepat

Beberapa penelitian menyebutkan, napas seseorang akan berubah jadi lebih cepat, lebih dalam, dan lebih tidak teratur jelang kematiannya. Bisa jadi pula terdapat jeda beberapa waktu di sela tarikan napas. Selain itu, tubuh secara alami juga akan memproduksi dahak dalam sistem pernapasan. Dahak ini pun secara alami akan terbuang melalui batuk. Namun, kalau tubuh pasien sudah tidak mampu bergerak banyak dan mendekati kematian, dahak akan menumpuk dan menimbulkan bunyi pada tarikan napas.

Untuk membantu pasien, kamu bisa meletakkan bantal di bawah kepalanya dan memiringkan kepala ke salah satu sisi. Bisa juga menggunakan oksigen atau penyejuk udara, atau bantuan medis yang bisa membuat pasien bernapas dengan lega.

Dalam dunia kedokteran, ada dua istilah kematian. Yaitu, mati klinis dan mati biologis.

Seorang pasien dinyatakan mati klinis kalau saat dilakukan pemeriksaan tidak ditemukan adanya pernapasan dan denyut nadi, itu artinya sistem

7. Contoh Kasus

a. Query : ‘Cara Mahasiswa Ambis’

← → ↻ 127.0.0.1:8000/q/?pencarian=cara+mahasiswa+ambis 🔍 ☆ 📄 🔍 🗑️

Search Result

[Bagaimana cara menjadi mahasiswa yang ambis](#)

[Detail](#)

1. Belajar dengan Cerdas seorang mahasiswa harus tahu cara belajar dengan cerdas, bukan hanya belajar dengan keras satu malam sebelum ujian. Belajar dengan cerdas dimulai dari dalam kelas. Sebisanya mungkin cobalah duduk di barisan paling depan, sehingga meminimalisir gangguan yang akan diterima. Duduk ... (click title to see details)

[5 Hal tentang Kematian yang Tak Pernah Disampaikan oleh Doktermu](#)

[Detail](#)

Kematian seringkali menjadi topik yang dihindari untuk didiskusikan. Tapi diakui atau tidak, hal ini secara pasti akan terjadi dan menghampiri semua makhluk hidup, termasuk manusia. Percakapan mengenai hal yang satu ini pun tak pernah mudah, apalagi ketika seorang dokter harus memberitahu perihal kematian ... (click title to see details)

[Beban Menjadi Seorang Pria](#)

[Detail](#)

Hari Pria Internasional yang diperingati tiap 19 November, memperingatkan kita bahwa pria tak hanya punya selera, tapi juga punya beban yang tidak sederhana. Terlepas ada tidaknya perbedaan secara deskriptif administratif, pria tentu tak sama dengan lelaki muda. Apalagi anak lelaki remaja. Dibalik sebagai ... (click title to see details)

Kamus Data

Term	Query	D1	D2	D3
cara	1	2	5	3
mahasiswa	1	26	0	0
ambis	1	0	0	0

b. Query : ‘Cara Menghindari Kematian’

Search Result

5 Hal tentang Kematian yang Tak Pernah Disampaikan oleh Doktermu

[Detail](#)

Kematian seringkali menjadi topik yang dihindari untuk didiskusikan. Tapi diakui atau tidak, hal ini secara pasti akan terjadi dan menghampiri semua makhluk hidup, termasuk manusia. Percakapan mengenai hal yang satu ini pun tak pernah mudah, apalagi ketika seorang dokter harus memberitahu perihal kematian ... (click title to see details)

Beban Menjadi Seorang Pria

[Detail](#)

Hari Pria Internasional yang diperingati tiap 19 November, mengingatkan kita bahwa pria tak hanya punya selera, tapi juga punya beban yang tidak sederhana. Terlepas ada tidaknya perbedaan secara deskriptif administratif, pria tentu tak sama dengan lelaki muda. Apalagi anak lelaki remaja. Dilabeli sebagai ... (click title to see details)

Bagaimana cara menjadi mahasiswa yang ambis

[Detail](#)

1. Belajar dengan Cerdas seorang mahasiswa harus tahu cara belajar dengan cerdas, bukan hanya belajar dengan keras satu malam sebelum ujian. Belajar dengan cerdas dimulai dari dalam kelas. Sebis mungkin cobalah duduk di barisan paling depan, sehingga meminimalisir gangguan yang akan diterima. Duduk ... (click title to see details)

Kamus Data

Term	Query	D1	D2	D3
cara	1	5	3	2
mati	1	18	0	0
hindar	1	1	0	0

c. Query : ‘Cara Menjadi Pria’

Search Result

Beban Menjadi Seorang Pria

[Detail](#)

Hari Pria Internasional yang diperingati tiap 19 November, mengingatkan kita bahwa pria tak hanya punya selera, tapi juga punya beban yang tidak sederhana. Terlepas ada tidaknya perbedaan secara deskriptif administratif, pria tentu tak sama dengan lelaki muda. Apalagi anak lelaki remaja. Dilabeli sebagai ... (click title to see details)

5 Hal tentang Kematian yang Tak Pernah Disampaikan oleh Doktermu

[Detail](#)

Kematian seringkali menjadi topik yang dihindari untuk didiskusikan. Tapi diakui atau tidak, hal ini secara pasti akan terjadi dan menghampiri semua makhluk hidup, termasuk manusia. Percakapan mengenai hal yang satu ini pun tak pernah mudah, apalagi ketika seorang dokter harus memberitahu perihal kematian ... (click title to see details)

Bagaimana cara menjadi mahasiswa yang ambis

[Detail](#)

1. Belajar dengan Cerdas seorang mahasiswa harus tahu cara belajar dengan cerdas, bukan hanya belajar dengan keras satu malam sebelum ujian. Belajar dengan cerdas dimulai dari dalam kelas. Sebis mungkin cobalah duduk di barisan paling depan, sehingga meminimalisir gangguan yang akan diterima. Duduk ... (click title to see details)

Kamus Data

Term	Query	D1	D2	D3
pria	1	35	0	0
cara	1	3	5	2
jadi	1	5	16	11

BAB 5

KESIMPULAN DAN SARAN

A. Kesimpulan

Kesimpulan dari laporan tugas besar ini adalah :

1. Pada tugas besar ini telah dibuat program Sistem Temu-Balik Informasi yang berupa web search sederhana dengan menggunakan penerapan rumus *cosine similarity*.
2. Pada program ini dapat dilakukan komparasi antara query dengan beberapa dokumen dan menampilkan dokumen yang memiliki nilai kesamaan tertinggi.
3. Salah satu framework pembuatan web search engine yang dapat digunakan adalah Django seperti yang digunakan pada program ini.

B. Saran

Program bisa dikembangkan di front end dengan menggunakan framework-framework seperti react.js, angular.js, vue.js, atau yang lainnya agar tampilan web lebih menarik dan lebih interaktif. Selain itu, masih mungkin ditambahkan fitur-fitur lain sesuai kebutuhan, serta pengintegrasian dengan web-scraping untuk menyediakan dokumen dari internet seperti search engine yang ada di internet.

C. Refleksi

Refleksi dari pembuatan tugas besar ini adalah :

1. Hafid Abi Daniswara(13519028)
Tugas ini menambah ilmu baru bagi saya. Saya jadi bisa eksplor hal-hal baru.
2. Arsa Daris Gintara (13519037)
Selain menambah ilmu, saya jadi bisa mengeksplor hal-hal baru tentang back end yang belum saya ketahui sebelumnya.
3. Syamil Cholid Abdurrasyid (13519052)
Tubes ini mengajarkan saya tentang asyiknya apabila bisa mengerjakan tugas bersama secara offline.

DAFTAR PUSTAKA

- [1] https://en.wikipedia.org/wiki/Information_retrieval
- [2] <https://www.studiobelajar.com/vektor/>
- [3] <https://informatika.stei.itb.ac.id/~rinaldi.munir/AljabarGeometri/2020-2021/algeo20-21.htm>
- [4] <https://theonotesblog.wordpress.com/2017/05/03/cosine-similarity-indonesia/>
- [5] https://en.wikipedia.org/wiki/Cosine_similarity
- [6] <https://docs.djangoproject.com/en/3.1/>
- [7] <https://pypi.org/project/Sastrawi/>
- [8] <https://pypi.org/project/sweetify/>