# LAPORAN TUGAS BESAR

# ALJABAR LINIER DAN GEOMETRI (IF2123)

"Aplikasi Dot Product pada Sistem Temu-balik Informasi"



### **Disusun Oleh**

Haning Nanda Hapsari 13519042

Alifah Rahmatika Basyasya 13519053

Nabilah Erfariani 13519181

# INSTITUT TEKNOLOGI BANDUNG

**KOTA BANDUNG** 

**TAHUN 2020** 

#### **BABI**

#### **DESKRIPSI MASALAH**

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 vektor 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.





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 = (w1, w2,..., wn) di dalam Rn, dimana nilai wi 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:

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

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

# BAB II TEORI SINGKAT

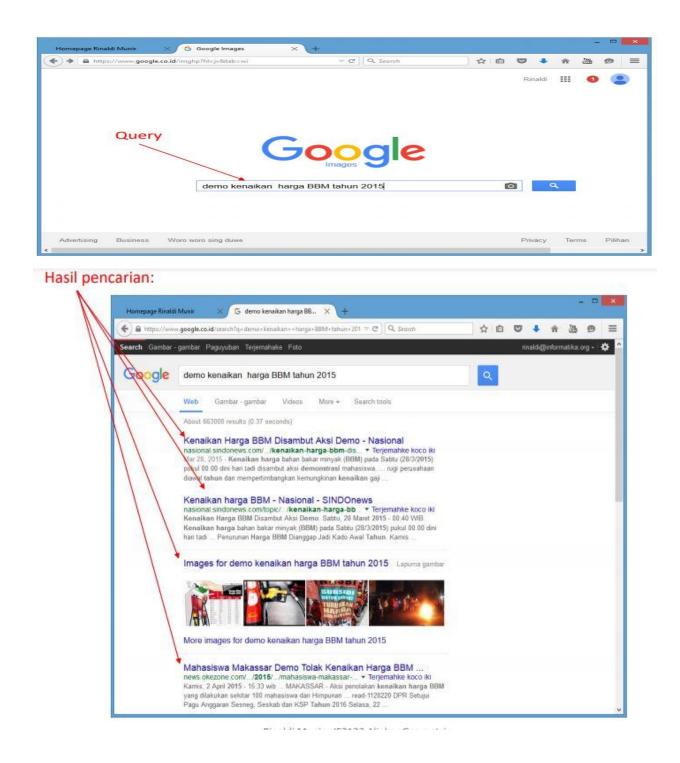
Temu-balik informasi (information retrieval) adalah menemukan kembali (retrieval) informasi yang relevan terhadap kebutuhan pengguna dari suatu kumpulan informasi secara otomatis



Sumber gambar:https://sites.google.com/site/berbagiinformasidanekspresi/arsip/pengantar-temu-kembali-informasi-information-retrieval

Information Retrieval(IR) tidak sama dengan pencarian di dalam basis data (database). IR umumnya digunakan pada pencarian informasi yang isinya tidak terstruktur. Contoh informasi tidak terstruktur tersebut yaitu dokumen dan laman web.

Aplikasi Information Retrieval yaitu pada mesin pencari(*search engine*). Salah satu *search engine* yang paling terkenal yaitu Google.



# Information Retrieval dengan Model Ruang Vektor

Salah satu model IR adalah model ruang vektor. Model ini menggunakan teori di dalam aljabar vektor. Misalkan terdapat n kata berbeda sebagai kamus kata (vocabulary) atau indeks kata

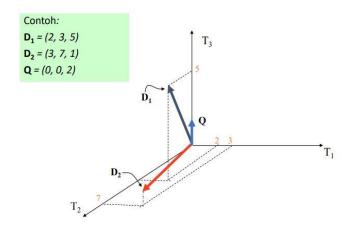
(term index). Kata-kata tersebut membentuk ruang vektor berdimensi n. Setiap dokumen maupun query dinyatakan sebagai :

vektor w = (w1, w2, ..., wn) di dalam R n.

wi = bobot setiap kata i di dalam query atau dokumen

Nilai wi dapat menyatakan jumlah kemunculan kata tersebut dalam dokumen (term frequency).

## Representasi vektor yaitu:



Vektor-vektor tersebut akan digunakan untuk menghitung *cosine similarity* dengan rumus sebagaimana tercantum dalam deskripsi masalah.

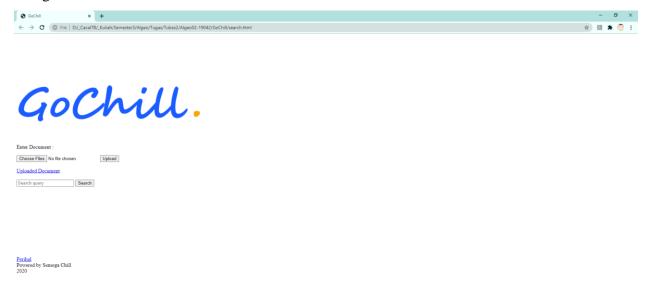
#### **BAB III**

### IMPLEMENTASI PROGRAM

Program dibuat menggunakan HTML, CSS, dan Javascript sebagai *front-end*, PHP sebagai *back-end* dan Xampp sebagai server atau *localhost*. Pada program yang kami buat terdapat beberapa file sebagai berikut.

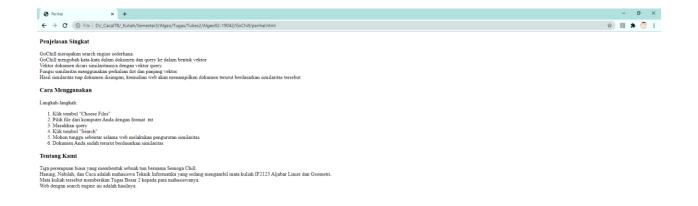
#### 1. index.html

index.html digunakan untuk menampilkan tampilan utama search engine. Terdapat beberapa link, form, dan tulisan yang menyusun halaman ini. Tampilan laman adalah sebagai berikut.



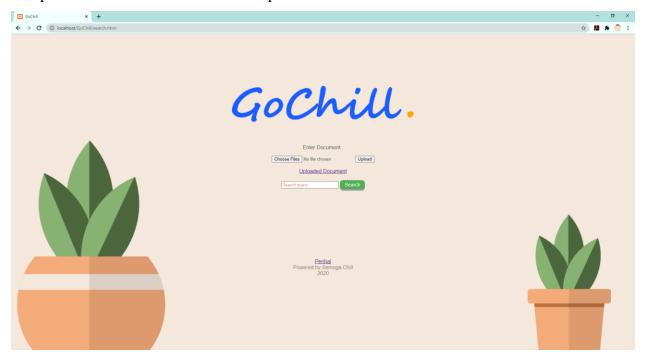
#### 2. perihal.html

perihal.html digunakan untuk menampilkan halaman berisi penjelasan singkat, langkah penggunaan, dan sedikit tentang penyusun.



#### 3. home.css

home.css digunakan untuk pewarnaan atau stilasi pada halaman utama dan perihal. Tampilan laman setelah stilasi adalah seperti berikut ini.





#### 4. hasil.css

hasil.css digunakan untuk pewarnaan atau stilasi pada halaman hasil pencarian. Tampilan laman setelah stilasi adalah seperti berikut ini.



# 5. doc. php

Program doc.php berisi sebuah program yang berfungsi untuk menampilkan dokumen dokumen yang sudah di upload.

### 6. upload.php

Program upload.php berisi program untuk mengupload dokumen dan memasukkan dokumen yang diupload ke dalam folder yang ditentukan. Dalam hal ini folder yang dipilih yaitu folder 'dokumen' sebagai folder penyimpanan dokumen terupload.

#### 7. result.php

Program fungsi.php berisi program untuk memproses dokumen dan query. Dalam fungsi.php dilakukan *stemming*, penghapusan *stopword*, pembuatan vektor, perhitungan *cosine similarity*, dan pemeringkatan dokumen. Pada program ini juga terdapat beberapa fungsi yang dibuat yaitu sim, bubble\_sort, jmlkata, dan tabel. Fungsi sim digunakan untuk melakukan pembentukan vektor dan perhitungan *cosine similarity*, fungsi bubble\_sort digunakan untuk melakukan pemeringkatan dokumen, fungsi jmlkata digunakan untuk menghitung jumlah kata pada masing-masing dokumen, dan fungsi tabel digunakan untuk membentuk tabel term.

## 8. query.txt

File ini digunakan untuk menyimpan sementara inputan query yang dimasukkan oleh pengguna.

Selain itu terdapat dua folder yang berfungsi sebagai penyimpanan dokumen yang telah diupload dan untuk menyimpan library yang digunakan yaitu library sastrawi.

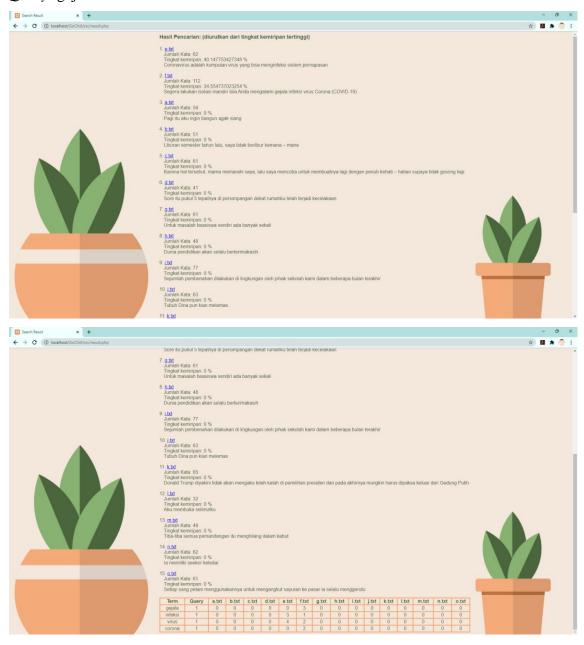
### **BAB IV**

#### **EKSPERIMEN**

Telah dilakukan 3 eksperimen menggunakan 15 file dengan format .txt dan query yang divariasi.

## 1. Eksperimen 1

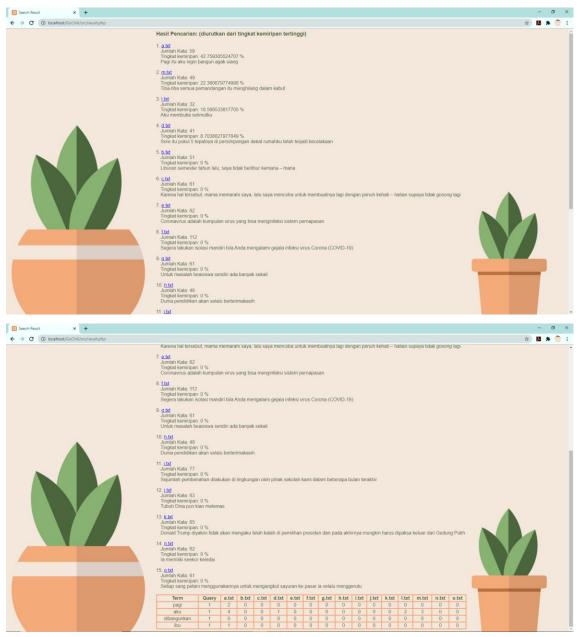
Query: gejala infeksi virus corona



Query yang dimasukkan seluruhnya merupakan kata dasar. Semua kata pada *query* digunakan di dalam vektor. Web berhasil menampilkan pengurutan similaritas dengan baik sesuai dengan *query*.

### 2. Eksperimen 2

Query: pagi itu aku dibangunkan ibu



Dapat dilihat kata 'itu' pada *query* tidak muncul. Hal ini terjadi karena telah dilakukan *stemming* dan penghapusan *stopwords* pada *query* dan dokumen. Hasil pencarian tetap terurut dari yang paling mirip.

# 3. Eksperiman 3

Query: kumpulan virus tidak akan mengaku



Kata 'tidak' dan 'akan' juga dihapus karena kedua kata tersebut temasuk dalam kategori *stopwords*. Hasil pencarian yang muncul tetap terurut dari tingkat similaritas paling tinggi.

# 4. Eksperimen 4

# Query: GEJALA INFEKSI VIRUS CORONA



Eksperimen 4 menggunakan *query* eksperimen 1 yang seluruhnya menggunakan huruf kapital. Dapat dilihat web menampilakan hasil yang sama. Web tidak membedakan huruf kapital dan huruf kecil.

#### **BAB V**

### KESIMPULAN, SARAN, DAN REFLEKSI

### 5.1 Kesimpulan

Search engine ini mampu menerima berbagai dokumen inputan user dengan ekstensi .txt, melakukan pencarian dokumen berdasarkan query dengan metode Cosine Similarity. Tampilan akhir berupa urutan dokumen dengan tingkat kemiripan paling tinggi, kalimat pertama dari tiap dokumen, persen kemiripan dokumen dengan query, serta tabel frekuensi kemunculan term pada query dan dokumen-dokumen.

Pada Tugas Besar Aljabar Linier dan Geometri ini, kelompok Go Chill dapat menyelesaikan tugas dengan baik.

#### 5.2 Saran

Penulis berharap *search engine* ini dapat dikembangkan lagi hingga dapat menerima berbagai inputan *user* seperti gambar, link, dll agar dapat digunakan oleh banyak orang. Selain itu, penulis juga berharap pada *search engine* ini ditambahkan fitur-fitur lainnya sehingga tidak kalah dengan *search engine-search engine* yang sudah terkenal saat ini. Penulis juga terbuka akan setiap saran dan kritik yang diberikan.

#### 5.3 Refleksi

Penulis memperoleh ilmu-ilmu baru dalam bidang programming selama mengerjakan tugas besar ini, seperti pemrograman berbasis web dan bahasa pemrograman PHP. Penulis juga belajar *time management* dan *team working* yang baik sehingga pekerjaan tugas besar ini dapat dilakukan dengan lancar.

### **REFERENSI**

Munir, Rinaldi. 2020. Aplikasi Dot Product pada Sistem Temu-balik Informasi

https://www.w3schools.com/html/default.asp (Terakhir diakses pada Minggu, 15 November 2020)

https://www.w3schools.com/css/default.asp (Terakhir diakses pada Minggu, 15 November 2020)

https://www.w3schools.com/php/default.asp (Terakhir diakses pada Minggu, 15 November 2020)

Brent. 5 Juni 2020. *How to merge multidimensional array in PHP?*. <a href="https://stitcher.io/blog/merging-multidimensional-arrays-in-php">https://stitcher.io/blog/merging-multidimensional-arrays-in-php</a> (Terakhir diakses pada Minggu, 15 November 2020)

https://stackoverflow.com/questions/17777273/in-php-number-of-rows-and-columns-in-a-2-d-array (Terakhir diakses pada Minggu, 15 November 2020)

https://stackoverflow.com/questions/54221447/using-ifisset-post-method-to-see-if-the-user-enter-the-data-so-in-will (Terakhir diakses pada Sabtu, 14 November 2020)