**LAPORAN TUGAS BESAR**

**ALJABAR LINIER DAN GEOMETRI**

**IF2123**

**Naufal Yahya Kurnianto – 13519141 – K1**

**Alif Bhadrika Parikesit – 13519186 – K2**

**Rayhan Asadel – 13519196 – K4**

# BAB I

**DESKRIPSI MASALAH**

Search engine merupakan suatu kata yang sangat familiar pada zaman ini. Seluruh orang, mulai dari yang muda hingga yang tua tentu tidak bisa melepas diri dari search engine. Search engine merupakan program yang mencari sesuatu hal dari database dengan memasukkan query. Program akan mencari hal yang berkorelasi dengan query yang dimasukkan kemudian ditampilkan sehingga kita sebagai manusia dapat mendapatkan info atau ilmu baru. Search engine ini dibuat dengan memanfaatkan berbagai ilmu seperti temu-balik informasi yang ada pada materi vektor. Dengan mengubah search query (input pengguna) menjadi ruang vektor. Program dapat menghitung berapa banyak kemunculan kata pada query yang kemudian akan dibandingkan dengan ruang vektor dari dokumen yang ada. Jika terjadi kesamaan, maka dapat ditarik kesimpulan bahwa dokumen tersebut memiliki kesamaan dengan search query dan dapat ditampilkan ke pengguna.

Buatlah program mesin pencarian (search engine) 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.
   1. Stemming dan Penghapusan stopwords dari isi dokumen.
   2. 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 II

**TEORI SINGKAT**

## 2.1 Information Retrieval (Temu-Balik Informasi)

Temu-balik informasi digunakan untuk menemukan kembali informasi yang relevan terhadap kebutuhan pengguna dari suatu kumpulan informasi secara otomatis. Sistem ini umumnya digunakan pada pencarian informasi yang isinya tidak terstruktur. Oleh karena itu, suatu aplikasi umum temu-balik informasi adalah mesin pencarian (search engine) di internet.

Salah satu pemodelan untuk menggunakan temu-balik informasi adalah dengan menggunakan vektor. Misalkan terdapat n kata berbeda sebagai kamus kata atau indeks kata dari query maupun dokumen. Kata-kata berbeda tersebut dapat membentuk ruang vektor berdimensi n. Hasil vektor yang dapat menyatakan dokumen maupun query contohnya adalah sebagai berikut:

w = (w[1],w[2],w[3],...,w[n]); w[i] menyatakan frekuensi kemunculan kata

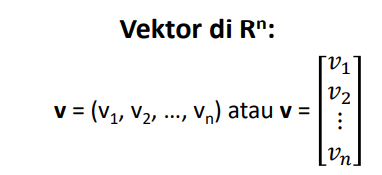
Contohnya adalah, misalkan terdapat empat buah kata berbeda, dua buah dokumen dan sebuah query. Dapat dibentuk:

D1 = (1,2,3,4), D2 = (2,3,4,5), Q = (1,1,1,1)

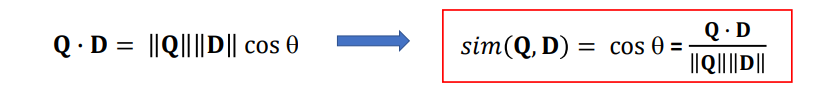
Angka dalam vektor melambangkan frekuensi kemunculan kata yang berbeda. Dapat disimpulkan misalnya bahwa pada dokumen pertama kata 1 muncul sekali; kata 2 muncul dua kali dan seterusnya.

## **2.2 Vektor**

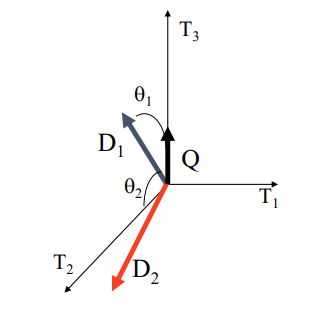
Vektor adalah representasi fisik untuk satuan yang memiliki besar dan arah. Vektor dilambangkan dengan huruf-huruf kecil dan dicetak tebal atau memakai tanda panah jika ditulis tangan. Secara geometri, vektor dinyatakan dengan garis berarah. Setelah itu, ada istilah ruang vektor yang melambangkan ruang tempat vektor didefinisikan.

Vektor dapat mengaplikasikan berbagai aturan matematika seperti penjumlahan yang menggunakan kaidah parallelogram atau kaidah segitiga, pengurangan, dan perkalian. Perkalian juga dibagi dengan perkalian vektor dengan skalar dan sesama vektor. Vektor juga memiliki panjang atau *magnitude* yang dinamakan norma. Berbagai sifat-sifat aljabar juga berlaku untuk penjumlahan antar vektor.

## 2.2 Cosine Similarity

Untuk menentukan dokumen yang relevan dengan query, dapat dilakukan pengukuran kesamaan (similarity measure) antara query dengan dokumen. Semakin besar ukuran kesamaannya, maka semakin relevan dokumen tersebut dengan query. Pengukuran kesamaan ini dapat dilakukan dengan menggunakan vektor-vektor yang didapat dari query dan dokumen. Contohnya, untuk mengukur kesamaan antara dua vektor **Q** dan **D**, gunakan rumus cosine similarity. Rumusnya sebagai berikut:

Rumus demikian diperoleh karena konsep yang digunakan adalah menghitung cosinus, semakin besar cosinus hingga mendekati 1. Maka akan semakin kecil sudut yang memisahkan antara dua vektor tersebut. Dapat dilihat pada visualisasi berikut:

Jika sudut semakin kecil, maka nilai cosinus akan semakin besar. Karena sudut semakin kecil, vektor dokumen akan berjarak lebih dekat dengan vektor query sehingga dokumen akan lebih relevan dengan query. Setelah mendapatkan nilai cosinus antara dokumen pertama dengan query, dilanjutkan dengan menghitung tiap-tiap cosinus antara dokumen lain dengan query. Kemudian, dapat dilakukan pemeringkatan dari yang bernilai paling besar hingga paling kecil untuk menentukan dokumen mana yang paling atau lebih relevan dengan query.

# BAB III

**IMPLEMENTASI PROGRAM**

# BAB IV

**EKSPERIMEN**

# BAB V

**KESIMPULAN, SARAN, DAN REFLEKSI**