

**Laporan Tugas Besar 1 IF 2123 Aljabar Linier dan Geometri
Sistem Persamaan Linier, Determinan, dan Aplikasinya
Semester I Tahun 2020/2021**



Disusun oleh :

Reinaldo Antolis	(13519015)
Wilbert Fangderson	(13519025)
Rehagana Kevin Christian Sembiring	(13519117)

PROGRAM STUDI TEKNIK INFORMATIKA
SEKOLAH TEKNIK ELEKTRO DAN INFORMATIKA

**INSTITUT TEKNOLOGI BANDUNG
2020**

BAB I

DESKRIPSI MASALAH

Hampir semua mahasiswa pernah menggunakan *search engine* dalam kehidupan sehari-hari untuk memperoleh suatu informasi. Akan tetapi, apakah para mahasiswa mengetahui cara kerja dari *search engine* tersebut untuk memperoleh dokumen berisi informasi yang diinginkan?

Pada tugas besar ini, mahasiswa ditantang untuk membuat sebuah *search engine* sederhana dengan model ruang vektor dan memanfaatkan cosine similarity. 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 II

TEORI SINGKAT

2. 1. Information Retrieval

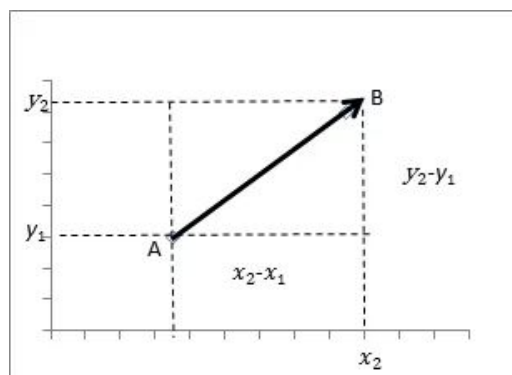
Temu kembali informasi (*information retrieval*) merupakan proses menemukan kembali informasi yang relevan terhadap kebutuhan pengguna dari suatu kumpulan informasi secara otomatis.

Prinsip sistem temu kembali informasi ini pada dasarnya adalah pengguna merumuskan suatu pertanyaan (*query*) yang jawabannya merupakan himpunan dokumen yang mengandung informasi yang diperlukan oleh pengguna tersebut.



2. 2. Vektor

Vektor merupakan sebuah besaran yang memiliki panah. Pada umumnya, vektor digambarkan sebagai panah yang menunjukkan arah vektor dan panjang garisnya disebut besar vektor. Dalam penulisannya, apabila vektor berasal dari titik A dengan koordinat (x_1, y_1) dan berakhir di titik B dengan koordinat (x_2, y_2) , maka vektor dapat ditunjukkan sebagai berikut :



2. 3. Cosine Similarity

Cosine similarity merupakan suatu metode untuk menghitung kesamaan antara dua buah objek yang dinyatakan dalam dua buah vektor dengan menggunakan kata kunci dari dokumen sebagai ukuran. Pada tugas ini, *cosine similarity* akan digunakan untuk melakukan perhitungan kesamaan antara query dengan dokumen. Rumus yang digunakan adalah :

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

BAB III

IMPLEMENTASI PROGRAM

Program app.py :

Nama	Tipe	Parameter	Deskripsi
index	Fungsi		Redirect ke web_scrapping
web_scraping	Fungsi		Melakukan web scraping
search	Fungsi		Melakukan search
printkey	Fungsi		Mencetak tabel tampilan terms
document_details	Fungsi		Menghasilkan isi dari dokumen pada halaman tersendiri
perihal	Fungsi		Pindah ke halaman tentang program dan pembuatnya (Konsep singkat <i>search engine</i> , How to Use, About Us)

Program search_engine.py :

Nama	Tipe	Parameter	Deskripsi
stemming	Fungsi	String	Melakukan stemming pada sentence, yaitu menghilangkan tanda baca, imbuhan, serta stopword yang terdapat pada sentence
length	Fungsi	String	Menghitung panjang dari sebuah sentence

search	Fungsi	List of string, String	Mencari sebuah kata yang terdapat dalam list
make_key	Fungsi	List of String	Membuat vektor(list) kata unik dari input
vectorizer	Fungsi	List of string ,string	Melakukan konversi pada satu nilai menjadi sekumpulan nilai (vektor)
perkalian_dot	Fungsi	List of Integer, List of Integer	Melakukan dot product antar 2 vektor.
panjang_vektor	Fungsi	List of Integer	Melakukan perhitungan mencari panjang dari sebuah vektor
cosine_similarity	Fungsi	List of Integer, List of Integer	Melakukan perhitungan kesamaan antara 2 vektor
hitung_jumlah_kata	Fungsi	String, List of String	Menghitung jumlah kata dari sebuah dokumen
ambil_kalimat_pertama	Fungsi	string, List of string	Mengambil kalimat pertama dari sebuah dokumen
sort	Fungsi	List of string, List of string, List of float	Melakukan sorting dari tingkat similarity tertinggi ke tingkat similarity terendah
search_content	Fungsi	List of string, String	Melakukan pencarian dari dokumen berdasarkan search key

Program web_scraper.py :

Nama	Tipe	Parameter	Deskripsi
web_scraping	Fungsi	List, List	Melakukan web scraping yaitu mengambil dokumen dari website yang diinginkan

BAB IV

HASIL DAN PEMBAHASAN

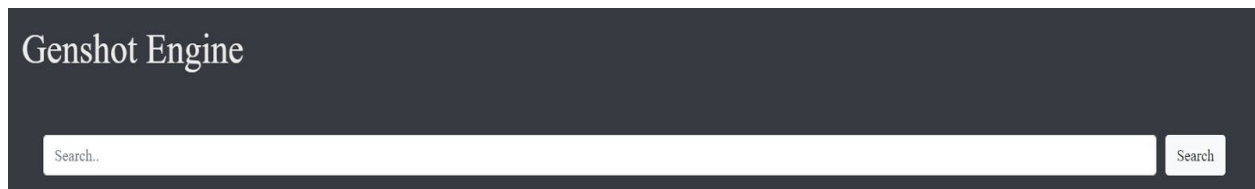
4. 1. Web Scraping

Program melakukan web scraping yaitu mendownload data dari website.

4. 2. Search Query

Query yang diinput oleh pengguna akan dihitung similaritasnya dengan dokumen-dokumen yang tersedia. Program akan mengurutkan dokumen yang dari similaritas yang tertinggi ke similaritas yang lebih rendah dengan membandingkan similaritas antara query dengan dokumen-dokumen yang tersedia. Program juga akan menampilkan table term yang menghitung banyaknya kemunculan term pada setiap dokumen.

Search Engine yang dihasilkan :



The image shows a dark-themed search engine interface. At the top left, the text "Genshot Engine" is displayed in a light gray font. Below this, there is a long, white search input field with the placeholder text "Search..". To the right of the input field is a small, light gray button with the text "Search" in a darker gray font.

User memasukkan query :



The image shows the same dark-themed search engine interface as before. The text "Genshot Engine" is still at the top left. In the white search input field, the query "mondstadt" has been entered. The "Search" button is still to the right. In the bottom left corner of the interface, the word "Perihal" is written in a small, light blue font.

Hasil searching (menampilkan jumlah kata, tingkat kemiripan, dan kalimat pertama dokumen) :

Genshot Engine

Jean

Jumlah Kata : 42

Tingkat Kemiripan : 37.8 %

As the Acting Grand Master of the Knights, Jean has always been devoted to her duties and maintaining peace in Mondstadt.

Diluc

Jumlah Kata : 45

Tingkat Kemiripan : 37.14 %

As the wealthiest gentleman in Mondstadt, the ever-dapper Diluc always presents himself as the epitome of perfection.

Kaeya

Jumlah Kata : 44

Tingkat Kemiripan : 17.68 %

Table of term :

Term	Query	Document 1	Document 2	Document 3	Document 4	Document 5	Document 6	Document 7	Document 8	Document 9
mondstadt	1	2	2	1	1	1	1	1	0	0

4. 3. Perihal

Berisi informasi mengenai konsep singkat *search engine*, *How to use*, dan *About Us*.

[Back to Search Page](#)

Search Engine

“Search engine adalah sebuah program yang berfungsi untuk membantu pengguna mencari dokumen-dokumen yang berisikan informasi yang diinginkan oleh pengguna. Dokumen didapat melalui hasil webscraping dari sebuah website.”



Input Query

User menginput query mengenai informasi yang diinginkan



Similarity Check

Program akan mengecek similaritas antara query dengan dokumen file yang sudah diupload dengan cosine similarity dan mengurutkan dokumen berdasarkan tingkat kesamaannya.



Choose the Document

User dapat memilih judul dokumen untuk melihat isi dokumen

About Us



Reinaldo Antolis

Anti Wibu Wibu Club

📧 : magicianreinaldo



Wilbert Fangderson

Sedang Mencari Jati Diri

📧 : wilbertf



Rehagana Sembiring

Need Keqing right now

📧 : rehagana16

BAB V

PENUTUP

5. 1. Kesimpulan

Dengan adanya program yang telah dibuat, maka dapat diambil kesimpulan bahwa Kelompok 9 Tugas Besar Aljabar Linear dan Geometri IF 2123 2020/2021, mampu menghasilkan sebuah program search engine dengan memanfaatkan *framework* website Flask (Python), melakukan perbandingan similaritas dengan menggunakan *cosine similarity*, serta menampilkan tabel term dan banyaknya kemunculan term pada setiap dokumen.

5. 2. Saran

Program yang telah dibuat dapat dikembangkan menjadi suatu program yang lebih baik dan lebih luas, serta lebih informatif untuk pengguna. Desain dari program juga dapat dibuat lebih menarik tampilannya.

5. 3. Refleksi

Program yang telah dibuat masih memiliki banyak kekurangan dalam segi keterbatasan waktu dan keterbatasan wawasan. Harapan dari kelompok kami dari tugas besar 2 IF 2123 Aljabar Linear dan Geometri ini adalah semoga tugas ini dapat menjadi ajang pembelajaran untuk pengembangan diri kami menjadi lebih baik lagi.

REFERENSI

“Aplikasi Dot Product pada sistem temu balik aplikasi” by Rinaldi Munir

<https://informatika.stei.itb.ac.id/~rinaldi.munir/AljabarGeometri/2020-2021/Algeo12-Aplikasi-dot-product-pada-IR.pdf>

"Create A Simple Search Engine Using Python" by Irfan Alghani Khalid

<https://link.medium.com/yEtxO932Kab>

“TFIDF | TF IDF Python Example” by Cory Maklin

<https://towardsdatascience.com/natural-language-processing-feature-engineering-using-tf-idf-e8b9d00e7e76>

“Genshin Impact Wiki” by Anonymous

https://genshin-impact.fandom.com/wiki/Genshin_Impact_Wiki

“Web Scraping in Python with BeautifulSoup and Flask” by Tanmay Naik

<https://medium.com/code-gandalf/web-scraping-in-python-with-beautifulsoup-and-flask-641efdb6ad5d>

“Search Engine with Flask” by mgokkaya

<https://github.com/mgokkaya/Udacity-cs101-search-engine-with-Flask>