

APLIKASI DOT PRODUCT PADA SISTEM TEMU BALIK INFORMASI

Laporan Tugas Besar 2 IF 2123 Aljabar Linier dan Geometri

Semester I Tahun 2020/2021



Oleh:

Kelompok 8 - tasdasdadwda

Rexy Gamaliel Rumahorbo	13519010
Dionisius Darryl Hermansyah	13519058
Wilson Tandya	13519209

SEKOLAH TEKNIK ELEKTRO DAN INFORMATIKA

INSTITUT TEKNOLOGI BANDUNG

2020

BAB I

DESKRIPSI MASALAH

1.1. Deskripsi masalah

Pada tugas besar 2 IF2123 Aljabar Linier dan Geometri ini, akan dibuat sebuah sistem temu balik informasi dalam bentuk website dengan memanfaatkan teori *dot product* dari vektor.

1.2. Spesifikasi program

Adapun spesifikasi dari program yang akan dibuat 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 bentuk sebuah website lokal yang sederhana serta 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. Vektor

Vektor merupakan kuantitas fisik yang memiliki besar dan arah. Salah satu contoh vektor adalah vektor di \mathbb{R}^2 dan \mathbb{R}^3 yang masing-masing memiliki 2 dan 3 komponen, serta dapat direpresentasikan ke dalam ruang 2 dimensi dan ruang 3 dimensi.

Sebuah vektor di ruang \mathbb{R}^n umumnya dapat dinyatakan dengan $\mathbf{v} = (v_1, v_2, v_3, \dots, v_n)$. Misalkan terdapat dua buah vektor berdimensi n , $\mathbf{u} = (u_1, u_2, u_3, \dots, u_n)$ dan $\mathbf{v} = (v_1, v_2, v_3, \dots, v_n)$. Hasil penjumlahan keduanya adalah

$$\mathbf{u} + \mathbf{v} = (u_1+v_1, u_1+v_2, u_1+v_3, \dots, u_1+v_n)$$

dan hasil perkalian vektor \mathbf{u} dengan skalar k adalah

$$k\mathbf{u} = k(v_1, v_2, v_3, \dots, v_n) = (kv_1, kv_2, kv_3, \dots, kv_n)$$

Norma/*magnitude* vektor \mathbf{v} adalah

$$\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2 + v_3^2 + \dots + v_n^2}.$$

Terdapat dua operasi perkalian antarvektor: perkalian titik dan perkalian silang. Perkalian titik/*dot product* dari vektor \mathbf{u} dan \mathbf{v} adalah

$$\mathbf{u} \cdot \mathbf{v} = u_1v_1 + u_2v_2 + u_3v_3 + \dots + u_nv_n = \|\mathbf{u}\| \|\mathbf{v}\| \cos a$$

dengan a adalah sudut yang dibentuk \mathbf{u} dan \mathbf{v} . Ketidaksamaan Cauchy-Schwarz menyatakan

$$\mathbf{u} \cdot \mathbf{v} \leq \|\mathbf{u}\| \|\mathbf{v}\|$$

Kesamaan diperoleh saat kedua vektor searah, sementara nilai $\mathbf{u} \cdot \mathbf{v} = 0$ diperoleh saat kedua vektor saling tegak lurus/*orthogonal*, dan nilai $\mathbf{u} \cdot \mathbf{v} = -\|\mathbf{u}\| \|\mathbf{v}\|$ diperoleh saat kedua vektor berlawanan arah.

2.2. Retrieval information

Sistem Temu-Balik Informasi atau *Information Retrieval System* merupakan kegiatan memperoleh informasi yang sesuai dengan kebutuhan dari suatu kumpulan informasi secara

otomatis. Pencarian dapat didasarkan pada teks atau pengindeksan berbasis konten. *Information Retrieval System* digunakan untuk mengurangi atau mengantisipasi *information overload*. Salah satu dari pengaplikasian *Information Retrieval System* ini adalah *search engine*.

2.3. *Cosine similarity*

Cosine similarity adalah besaran yang digunakan untuk mengukur seberapa mirip dokumen berdasarkan kemunculan kata terlepas dari panjang dokumen. Setiap dokumen direpresentasikan oleh vektor frekuensi yang setiap komponennya mewakili banyak kemunculan kata tertentu. Vektor frekuensi biasanya sangat panjang dan jarang (memiliki banyak nilai 0). Secara matematis, *cosine similarity* mengukur nilai kosinus sudut antar dua vektor yang diproyeksikan dalam ruang multidimensi. Sebuah dokumen dapat diwakili oleh ribuan atribut, masing-masing merekam frekuensi kata tertentu (kata kunci). *Cosine similarity* dapat memberikan urutan (*ranking*) berdasarkan kemiripan dokumen terhadap *query* yang diberikan. *Cosine similarity* dapat dihitung dengan persamaan berikut.

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

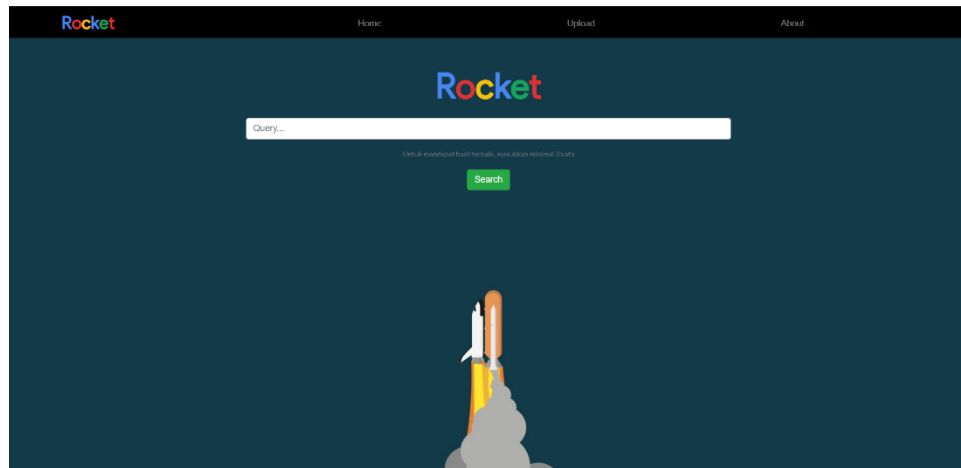
Dengan $\mathbf{Q} = (q_1, q_2, \dots, q_n)$ dan $\mathbf{D} = (d_1, d_2, \dots, d_n)$ dan $\mathbf{Q} \cdot \mathbf{D}$ adalah perkalian titik yang didefinisikan sebagai $\mathbf{Q} \cdot \mathbf{D} = (q_1 d_1, q_2 d_2, \dots, q_n d_n)$. Nilai kosinus 0 menyatakan bahwa kedua vektor saling ortogonal dan tidak memiliki kesamaan. Semakin dekat nilai kosinus ke 1, semakin kecil sudut antar kedua vektor yang menyatakan bahwa nilai kesamaannya semakin tinggi.

BAB III

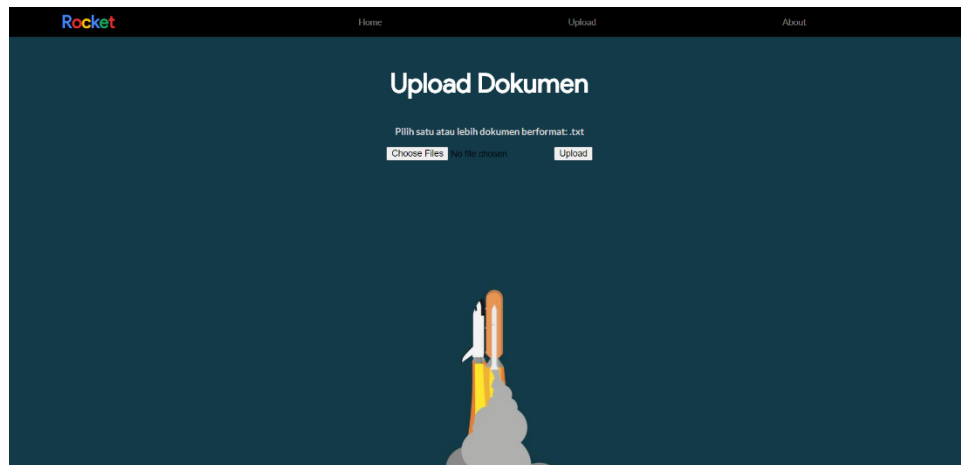
IMPLEMENTASI PROGRAM

Tampilan utama *website* kami ditunjukkan oleh gambar di bawah:

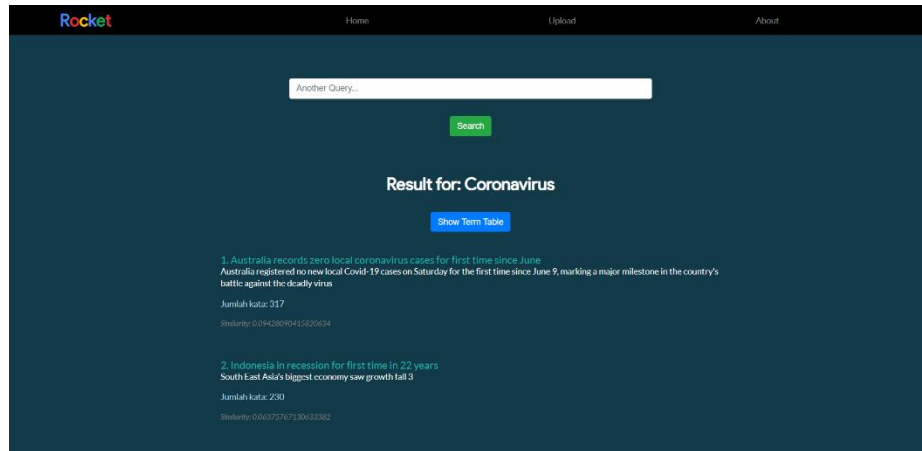
- Tampilan *home page*



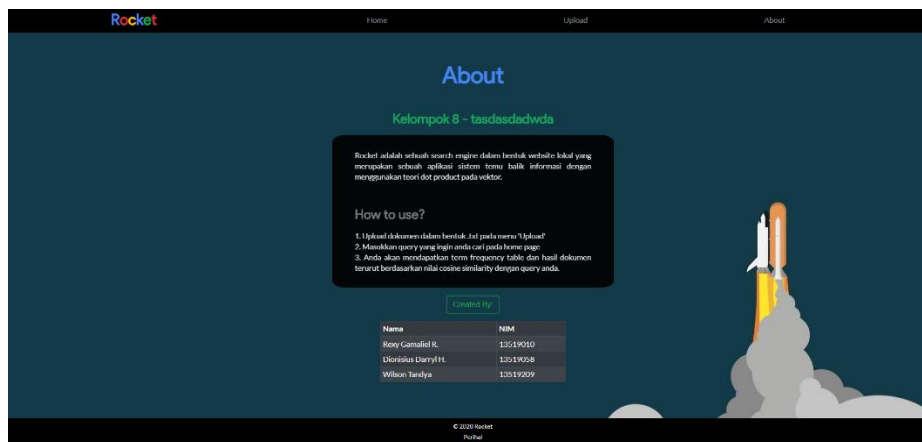
- Tampilan upload dokumen



- Tampilan hasil pencarian query



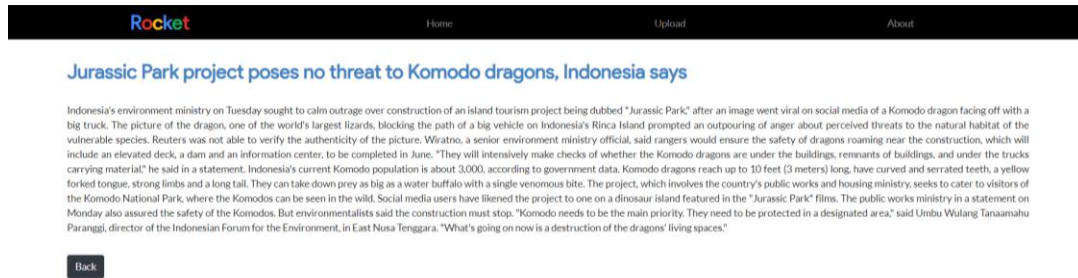
- Tampilan *about* page



- Tampilan *term frequency table*

	D1	D10	D11	D12	D13	D14	D15	D16	D17	D18	D2	D3	D4	D5	D6	D7	D8	D9	Query	Term
1	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	1	joe
13	0	0	0	0	0	0	0	0	0	0	2	1	10	2	0	0	0	0	1	biden
1	0	0	0	0	0	0	0	0	0	0	1	2	0	2	0	0	0	0	1	presidenti
9	0	0	0	0	0	0	0	0	0	0	14	2	3	4	0	0	0	0	1	elect
1	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	1	donald
14	0	0	0	0	0	0	0	0	0	0	3	22	8	4	0	0	0	0	1	trump

- Tampilan isi dokumen



Arsitektur *website* yang dibuat dibagi ke dalam 2 komponen yaitu, komponen *front-end* dan *back-end*. Komponen frontend dibuat menggunakan React.js dengan bahasa JavaScript dan dihubungkan kepada backend yang diimplementasikan menggunakan framework Flask dengan bahasa Python.

Pada pemrosesan backend ada beberapa kelas dan modul yang dibuat. Modul dan kelas tersebut berjumlah 6, dengan fungsionalitas yang berbeda-beda. Berikut ini adalah penjelasan dari modul dan kelas yang ada tersebut secara lebih rinci:

Nama Fungsi/Prosedur	Keterangan
Main	
Berisi sebuah aplikasi Flask yang digunakan sebagai API untuk melakukan <i>fetch</i> data dari <i>front-end</i> untuk diproses dan ditampilkan. Terdapat berbagai REST API yang didefinisikan menggunakan fungsi yang terdapat pada modul-modul di bawah ini.	
Program	
query_sim	Mencari similarity dari query dari kumpulan dokumen yang ada di folder test. Output hasil query yang telah terurut berdasarkan value dalam bentuk dictionary
term_frequency_table	Menghasilkan term_frequency_table dari query terhadap dokumen yang ada dalam format HTML
sort_dict	Sorting sebuah dictionary berdasarkan value
make_json	Membuat dictionary menjadi berformat json. Mereturn hasil dalam bentuk [{'title': ... , 'sim': ..., 'first_sentence':... } , ...]
Document	

<i>Berisi semua hal yang berkaitan dengan dokumen. Kelas document, preprocessing document, dan menghasilkan token dari sekumpulan dokumen</i>	
<code>__init__</code>	Inisialisasi class document dengan atribut-atributnya
<code>preprocess</code>	Preprocessing text dengan menghapus stopwords dan stemming
<code>get_tokens</code>	Melakukan preprocessing dan menghasilkan token dari sekumpulan dokumen
<code>get_first_sentence</code>	Melakukan preprocessing dan menghasilkan kalimat pertama dari sekumpulan dokumen
<code>get_num_words</code>	Melakukan preprocessing dan menghasilkan jumlah kata dari sekumpulan dokumen
<code>get_table</code>	Menghasilkan tabel perhitungan kata dalam format .json. Hanya menampilkan term global dari query <pre>[{ "Term": ..., "Query": ..., "D1":..., ... }]</pre>
<code>get_table_html</code>	Menghasilkan format HTML dari tabel
Reader	
<i>Berisi semua hal yang berkaitan dengan pembacaan dokumen dari file</i>	
<code>get_files</code>	Mereturn semua file yang ada dalam sebuah folder
<code>read_txt</code>	Membaca file .txt
<code>write</code>	Writing sebuah document
Vector	
<i>Berisi semua hal yang berkaitan dengan vektor. Operasi vektor, menghitung magnitude, dot product, cosine similarity, dan mekonversi token kata menjadi vektor</i>	
<code>vectorize</code>	Membuat vektor token kata dokumen dari sekumpulan token kata dokumen.
<code>term_frequency</code>	Membuat map frekuensi kata dokumen dari sekumpulan token kata dokumen
<code>dot</code>	Melakukan dot product pada vektor a dan b
<code>magnitude</code>	Melakukan dot product pada vektor a dan b
<code>sim</code>	Mencari nilai cosine similarity dari vektor a dan b. Prasyarat: Ukuran a dan b harus sama

BAB IV

EKSPERIMEN

Berdasarkan hasil eksperimen, dapat dianalisis bahwa frekuensi kesamaan kata bukanlah satu-satunya parameter dalam menghitung *cosine similarity*, namun juga jumlah kata unik pada dokumen yang bersangkutan. Semakin banyak kata unik yang ada, maka magnitudo vektor-vektor dokumen pun semakin besar. Hal ini berarti, sesuai dengan rumus:

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

nilai pembagi akan semakin besar, sehingga nilai *similarity* semakin berkurang.

Berikut merupakan hasil eksekusi *website* dan analisis terhadap contoh-contoh dokumen yang diberikan:

Query	Hasil
“Coronavirus 2020”	<p style="text-align: center;">Urutan dokumen:</p> <p style="text-align: center;">Result for: Coronavirus 2020</p> <p style="text-align: center;">Show Term Table</p> <ol style="list-style-type: none"> <p>1. Indonesia in recession for first time in 22 years South East Asia's biggest economy saw growth fall 3</p> <p>Jumlah kata: 230</p> <p>Similarity: 0.09016696346674323</p> <p>2. Australia records zero local coronavirus cases for first time since June Australia registered no new local Covid-19 cases on Saturday for the first time since June 9, marking a major milestone in the country's battle against the deadly virus</p> <p>Jumlah kata: 317</p> <p>Similarity: 0.06666666666666665</p> <p>3. Trump's biggest campaign promises Did he deliver As Election Day approaches, President Donald Trump and his Democratic challenger Joe Biden are making their final pitches to American voters, touting what is to come and, especially, boasting about their accomplishments</p> <p>Jumlah kata: 906</p> <p>Similarity: 0.04556976628693896</p> <p>4. Minister ensures Komodo National Park follows conservation principles The Environmental Affairs and Forestry (KLHK) Ministry has made assurance that the development of facilities and infrastructure to support tourism at the Komodo National Park, East Nusa Tenggara (NTT), aligns with the principles of conservation</p>

Term frequency table:

D1	D10	D11	D12	D13	D14	D15	D16	D17	D18	D2	D3	D4	D5	D6	D7	D8	D9	Query	Term
1	0	0	0	0	0	0	0	0	0	2	1	0	1	1	1	0	1	1	coronaviru
1	1	0	0	0	0	0	0	0	0	0	2	1	1	1	1	1	0	1	2020

"Joe Biden Presidential Election and Donald Trump"

Urutan dokumen:

Result for: Joe Biden Presidential Election and Donald Trump

Show Term Table

1. No Winner Yet, But Biden's Lead Grows in US Presidential Election

WASHINGTON - Americans waited Thursday to see if vote count updates from several states would make clear the winner of the presidential election, a day after wins in two battleground states put former Vice President Joe Biden within reach of victory and President Donald Trump's campaign launched a series of legal challenges

Jumlah kata: 829

Similarity: 0.3330779473894949

2. No Foreign Meddling Detected in Presidential Vote, US Officials Say

US election security officials say no matter who emerges as the winner of the country's presidential election, it will be American voters, not foreign adversaries, who determined the outcome

Jumlah kata: 646

Similarity: 0.26302545492234575

3. Trump's biggest campaign promises Did he deliver

As Election Day approaches, President Donald Trump and his Democratic challenger Joe Biden are making their final pitches to American voters, touting what is to come and, especially, boasting about their accomplishments

Jumlah kata: 906

Similarity: 0.2543272627158346

Term frequency table:

	D1	D10	D11	D12	D13	D14	D15	D16	D17	D18	D2	D3	D4	D5	D6	D7	D8	D9	Query	Term
1	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	1	joe
13	0	0	0	0	0	0	0	0	0	0	2	1	10	2	0	0	0	0	1	biden
1	0	0	0	0	0	0	0	0	0	0	1	2	0	2	0	0	0	0	1	presidenti
9	0	0	0	0	0	0	0	0	0	0	14	2	3	4	0	0	0	0	1	elect
1	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	1	donald
14	0	0	0	0	0	0	0	0	0	0	3	22	8	4	0	0	0	0	1	trump

Urutan dokumen:

Result for: Building Komodo Jurassic Park in Indonesia

Show Term Table

1. Jurassic Park project poses no threat to Komodo dragons, Indonesia says

Indonesia's environment ministry on Tuesday sought to calm outrage over construction of an island tourism project being dubbed "Jurassic Park," after an image went viral on social media of a Komodo dragon facing off with a big truck

Jumlah kata: 319

Similarity: 0.41905817746174695

2. Viral photo sparks concerns about Indonesia's 'Jurassic Park'

A photo of a Komodo dragon facing a truck has raised concerns about a "Jurassic Park" attraction being built on an Indonesian island

Jumlah kata: 529

Similarity: 0.3425395354310701

3. Komodo

The Komodo dragon (Varanus komodoensis), also known as the Komodo monitor, is a species of lizard found in the Indonesian islands of Komodo, Rinca, Flores, and Gili Motang

Jumlah kata: 329

Similarity: 0.254714264556119

4. Ministry highlights rising Komodo dragon population in past few years

Jakarta (ANTARA) - The population of Komodo dragons has increased in the past few years, reaching 3,022 in 2019, according to the

“Building Komodo Jurassic Park in Indonesia”

Term frequency table:

	D1	D10	D11	D12	D13	D14	D15	D16	D17	D18	D2	D3	D4	D5	D6	D7	D8	D9	Query	Term
2	0	0	0	0	0	0	0	0	0	0	0	0	1	2	0	0	0	2	1	build
8	0	0	0	0	0	0	0	0	0	0	12	10	9	6	0	0	0	0	1	komodo
2	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	1	jurass
3	0	0	0	0	0	0	0	0	0	0	5	1	4	1	0	0	0	0	1	park
3	0	0	0	0	0	0	0	0	0	0	3	0	0	0	7	8	3	0	1	indonesia

Urutan dokumen:

Result for: US Election 2020 Affecting Indonesian Stock Market

Show Term Table

1. No Foreign Meddling Detected in Presidential Vote, US Officials Say

US election security officials say no matter who emerges as the winner of the country's presidential election, it will be American voters, not foreign adversaries, who determined the outcome

Jumlah kata: 646

Similarity: 0.16603244517199559

2. IHSG remains in the green tracking rise in global markets

Jakarta (ANTARA) - The Jakarta Composite Index (IHSG) remained in the green on Tuesday as the strengthening of global stock exchanges and positive domestic cues lifted investor sentiments

Jumlah kata: 237

Similarity: 0.14736689024108546

3. Trump's biggest campaign promises Did he deliver

As Election Day approaches, President Donald Trump and his Democratic challenger Joe Biden are making their final pitches to American voters, touting what is to come and, especially, boasting about their accomplishments

Jumlah kata: 906

Similarity: 0.08119354887670824

4. No Winner Yet, But Biden's Lead Grows in US Presidential Election

WASHINGTON - Americans waited Thursday to see if vote count updates from several states would make clear the winner of the presidential election, a day after wins in two battleground states put former Vice President Joe Biden within reach of victory and President Donald Trump's campaign launched a series of legal challenges

Jumlah kata: 829

“US Election 2020 Affecting Indonesian Stock Market”

“Donald Trump or Covid Cause Indonesia's Economic Recession”

Term frequency table:

D1	D10	D11	D12	D13	D14	D15	D16	D17	D18	D2	D3	D4	D5	D6	D7	D8	D9	Query	Term
1	0	0	2	0	0	0	0	0	0	0	6	0	0	0	2	1	0	1	us
14	0	0	0	0	0	0	0	0	0	0	2	9	4	0	3	0	0	1	elect
0	1	1	0	0	1	1	0	0	0	0	2	1	1	0	0	1	0	1	2020
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	affect
0	0	0	0	1	0	0	0	0	0	1	0	0	0	3	0	1	2	1	indonesian
0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	2	0	1	stock
0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	1	market

Urutan dokumen:

Result for: Donald Trump or Covid Cause Indonesia's Economic Recession

Show Term Table

1. Indonesia in recession for first time in 22 years

South East Asia's biggest economy saw growth fall 3

Jumlah kata: 230

Similarity: 0.2409813463559399

2. Trump's biggest campaign promises Did he deliver

As Election Day approaches, President Donald Trump and his Democratic challenger Joe Biden are making their final pitches to American voters, touting what is to come and, especially, boasting about their accomplishments

Jumlah kata: 906

Similarity: 0.19486451730409976

3. Indonesia Slips Into First Recession in More Than Two Decades

Indonesia has fallen into a recession after the \$1-trillion economy shrank for the second consecutive quarter in the July to September period, the statistics agency said on Thursday

Jumlah kata: 385

Similarity: 0.16398401233815754

4. No Winner Yet, But Biden's Lead Grows in US Presidential Election

WASHINGTON - Americans waited Thursday to see if vote count updates from several states would make clear the winner of the presidential election, a day after wins in two battleground states put former Vice President Joe Biden within reach of victory and President Donald Trump's campaign launched a series of legal challenges

Jumlah kata: 829

Term frequency table:

	D1	D10	D11	D12	D13	D14	D15	D16	D17	D18	D2	D3	D4	D5	D6	D7	D8	D9	Query	Term
	0	0	1	0	0	0	0	0	0	0	1	0	1	0	1	1	0	0	1	donald
	0	0	3	0	0	0	0	0	0	0	22	0	14	0	8	4	0	0	1	trump
	0	5	0	0	2	3	0	0	0	0	0	3	0	0	0	0	0	5	1	covid
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	caus
	7	0	0	3	0	0	0	0	0	0	0	3	0	8	0	0	3	0	1	indonesia
	2	0	0	0	0	0	0	0	0	0	1	3	0	0	0	0	0	0	1	econom
	1	0	0	0	0	0	0	0	0	0	0	3	0	3	0	0	0	0	1	recess

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Kesimpulan yang dapat ditarik dari tugas besar ini yaitu sebuah sistem temu balik informasi dapat dirancang menggunakan teori aljabar vektor yaitu ruang vektor dan *dot product* dengan indikator utama *cosine similarity* serta dapat diimplementasikan secara nyata dalam bentuk sebuah *website* lokal menggunakan teknologi Flask dan React.js.

5.2. Saran

Untuk pengembangan lagi kedepannya, penulis memberikan saran sebagai berikut:

1. Mendeploy website ke internet agar dapat diakses khalayak umum.
2. Memperbanyak fitur-fitur *search engine* pada website.

5.3. Refleksi

Melalui pengerjaan tugas besar 2 IF2123 Aljabar Linier dan Geometri ini, penulis memperoleh banyak hal baik dari segi akademik maupun non-akademik. Penulis dapat memahami materi tentang vektor serta aplikasinya secara lebih mendalam dan jelas. Melalui implementasi algoritma-algoritma secara nyata dalam program, hal ini membantu penulis dalam memahami materi yang ada. Selain itu, penulis juga belajar untuk bekerja sama dalam tim terutama dalam mengerjakan sebuah proyek serta dapat melatih skill manajemen waktu. Seluruh hasil yang telah direfleksikan ini, diharapkan dapat membantu penulis untuk berkembang ke arah yang lebih baik lagi.

DAFTAR REFERENSI

- Akeela, K. 2020. *Implementing the TF-IDF Search Engine*. Dilansir dari https://medium.com/@kartheek_akella/implementing-the-tf-idf-search-engine-5e9a42b1d30b.
- Grimberg, M. 2020. *How To Create a React + Flask Project*. Dilansir dari <https://blog.miguelgrinberg.com/post/how-to-create-a-react--flask-project>.
- Khalid, I. A. 2020. *Create A Simple Search Engine Using Python*. Dilansir dari <https://towardsdatascience.com/create-a-simple-search-engine-using-python-412587619ff5>.
- Munir, R. 2020. *Aplikasi Dot Product pada Sistem Temu-balik Informasi*. Bandung: Institut Teknologi Bandung.
- N.N. 2020. *Tugas Besar 2 IF 2123 Aljabar Linier dan Geometri Aplikasi Dot Product pada Sistem Temu-balik Informasi Semester I Tahun 2020/2021*. Bandung: Institut Teknologi Bandung.
- Symth, P. 2018. *Creating Web APIs with Python and Flask*. Dilansir dari <https://programminghistorian.org/en/lessons/creating-apis-with-python-and-flask>.
- Wahyuni, R. T. 2017. *Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF pada Sistem Klasifikasi Dokumen Skripsi*. Semarang: Universitas Negeri Semarang.