

APLIKASI *DOT PRODUCT* PADA SISTEM TEMU BALIK INFORMASI

LAPORAN TUGAS BESAR 2

**Diajukan sebagai salah satu tugas besar
mata kuliah IF2123 Aljabar Linier dan Geometri
Pada Semester 1 Tahun Akademik 2020-2021**



Oleh

Yudi Alfayat (13519051)

Mgs. Tabrani (13519122)

Muhammad Zubair (13519172)

**PROGRAM STUDI TEKNIK INFORMATIKA
SEKOLAH TEKNIK ELEKTRO DAN INFORMATIKA
INSTITUT TEKNOLOGI BANDUNG
BANDUNG**

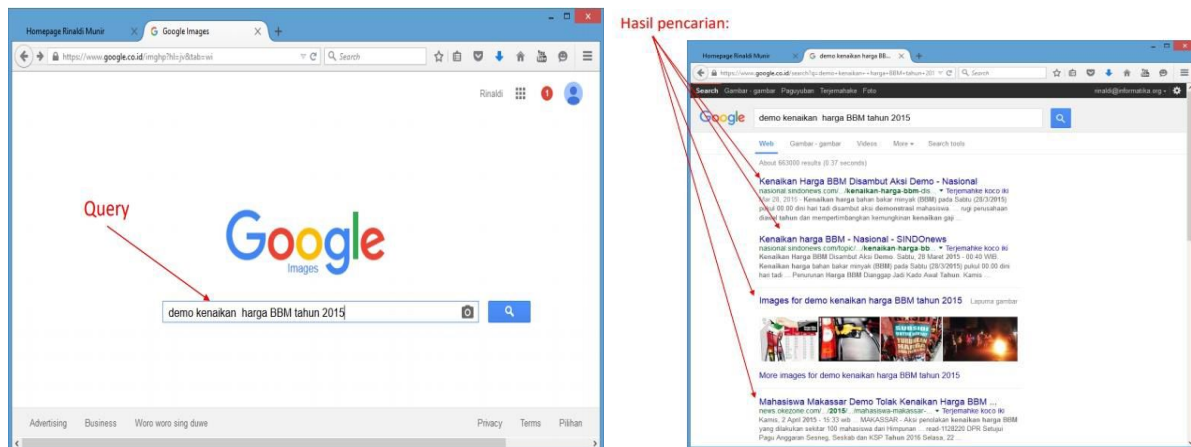
BAB 1

DESKRIPSI MASALAH

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 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 $w = (w_1, w_2, \dots, w_n)$ 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:

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

BAB 2

TEORI SINGKAT

1. Information retrieval

Information retrieval atau yang sering disebut dengan temu kembali informasi merupakan ilmu yang bertujuan mempelajari tahap-tahapan dan metode untuk mencari kembali informasi yang tersimpan dari berbagai sumber yang akurat atau koleksi sumber informasi yang dicari atau dibutuhkan. Dengan adanya tindakan indexing, panggilan , serta pemanggilan data kembali.

Information retrieval merupakan bagian dari computer science yang berkaitan dengan pengambilan informasi dari dokumen-dokumen yang didasarkan pada isi dan konteks dari dokumen itu sendiri. Tujuan dari information retrieval yaitu untuk menemukan informasi yang diinginkan pengguna dengan cara meretrieve dokumen yang akurat atau mengurangi dokumen pencarian yang tidak relevan atau akurat.

Pengertian information retrieval menurut beberapa ahli adalah sebagai berikut :

1. Kowalski

Information retrieval merupakan konsep yang sederhana dalam melakukan pencarian yang dilakukan seseorang. Sebagai contoh ketika pengguna akan mencari informasi yang diinginkan maka sistem menerjemahkan kepada bentuk statement yang kemudian dieksekusi oleh sistem pencari tersebut.

2. William Hersh

IR merupakan bentuk ilmu sistem informasi dan ilmu komputerisasi yang merujuk ke pengurutan dan pengambilan bentuk informasi yang heterogen dan sebagian besar-tekstil.

Information Retrieval memiliki beberapa metode dalam mengambil data dan informasi antara lain inverted index, Boolean retrieval, tokenization, stemming and lemmatization, dictionaries, wildcard queries, dan vector space model.

2. 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 bisa juga ditulis \vec{AB} .

Sebagai contoh, jika vektor \bar{v} merupakan vektor yang berawal dari titik A(x_1, y_1) menuju titik B(x_2, y_2) maka komponen x dan y dari vektor \bar{v} adalah $v_1 = x_2 - x_1$ dan $v_2 = y_2 - y_1$ atau bisa ditulis,

$$\vec{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} x_2 - x_1 \\ y_2 - y_1 \end{pmatrix}$$

Jenis-jenis Vektor

- Vektor Posisi

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

- Vektor Nol

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

- Vektor satuan

Suatu vektor yang panjangnya satu satuan. Vektor satuan dari \bar{v} adalah

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

- 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

$$\vec{I} = (1, 0, 0), \vec{J} = (0, 1, 0), \text{ dan } \vec{K} = (0, 0, 1).$$

Operasi-operasi pada vektor

Operasi vektor di ruang 3 secara umum, memiliki konsep yang sama dengan operasi vektor di ruang 2, hanya saja di ruang 3 terdapat tiga komponen, sedangkan di ruang 2 terdapat 2 buah komponen.

- Panjang vektor

Panjang sebuah vektor dilambangkan dengan $|\vec{v}|$. Untuk menghitung panjang atau jarak antara dua titik vektor dapat dilakukan dengan pengembangan rumus pythagoras. Pada ruang 3, rumus yang digunakan adalah :

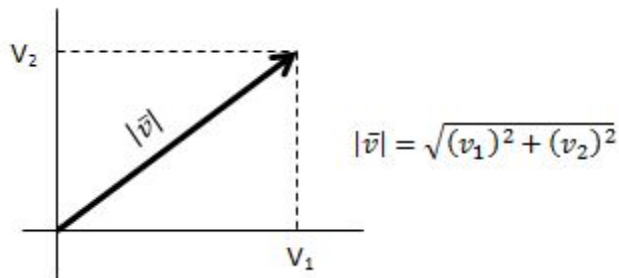
$$AB = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

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

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

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

Sedangkan pada ruang 2, rumus yang digunakan sangat mirip, yaitu sebagai berikut :



- Penjumlahan dan pengurangan

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

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

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

- Perkalian vektor dengan skalar

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

$$k.\vec{v} = \begin{pmatrix} k.v_1 \\ k.v_2 \\ k.v_3 \end{pmatrix}$$

- Perkalian Skalar Dua Vektor (*Dot Product*)

Perkalian skalar dari dua vektor atau disebut juga sebagai hasil kali titik (*dot product*) dua vektor, ditulis sebagai $\vec{a}.\vec{b}$ (a dot b).

Perkalian skalar dua vektor dilakukan dengan mengalikan panjang kedua vektor dengan cosinus θ . Sedangkan θ merupakan sudut yang terbentuk antara 2 vektor tersebut. Perkalian skalar dua vektor atau *dot product* akan menghasilkan suatu skalar. Rumus untuk menghitungnya adalah sebagai berikut,

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

Atau 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 $\vec{a}.\vec{b}$ adalah:

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

3. Cosine Similarity

Cosine similarity digunakan untuk menentukan kemiripan antara dokumen dan query yang diinginkan. Semakin sama suatu vektor dokumen dengan vektor query, semakin relevan dokumen tersebut dengan query. Kesamaan ini (Sim), diukur dengan rumus cosine similarity yang merupakan rumus *dot product* vektor. Misalkan vektor $Q = (q_1, q_2, \dots, q_n)$ dan $D = (d_1, d_2, \dots, d_n)$, dimana setiap komponen vektornya berdasarkan jumlah kemunculan term pada dokumen tersebut. Maka, kesamaannya dapat dihitung dengan menggunakan rumus :

$$\mathbf{Q} \cdot \mathbf{D} = \|\mathbf{Q}\| \|\mathbf{D}\| \cos \theta$$



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

Dengan $\mathbf{Q} \cdot \mathbf{D}$ adalah perkalian titik yang didefinisikan sebagai

$$\mathbf{Q} \cdot \mathbf{D} = q_1 d_1 + q_2 d_2 + \dots + q_n d_n$$

Contoh :

Misalkan terdapat tiga buah kata (T1 , T2 , dan T3), dua buah dokumen (D1 dan D2) serta sebuah query Q. Masing-masing dinyatakan sebagai vektor: D1 = (2, 3, 5), D2 = (3, 7, 1), Q = (0, 0, 2)

D1 = (2, 3, 5) artinya dokumen D1 mengandung 2 buah kata T1 , 3 buah kata T2 , dan 5 buah kata T3 . Contoh: Misalkan T1 = Menteri, T2 = minta, T3 = Korupsi

D1= Menteri olahraga meminta maaf atas perbuatan korupsi. Menteri tersebut terlibat korupsi anggaran. Meminta-minta komisi termasuk korupsi. Korupsi sudah mendarah daging di Indonesia. Korupsi sudah menjadi budaya.

D2 = (3, 7, 1) artinya dokumen D2 mengandung 3 buah kata T1 , 7 buah kata T2 , dan satu buah kata T3 .

Contoh: D2= Gubernur Jabar meminta waktu ketemu Menteri Sosial. Dia meminta Pak Menteri mengunjungi panti. Permintaan yang wajar. Sekretaris Gubernur mengirim surat permintaan kepada Menteri tersebut. Apakah meminta-minta termasuk perbuatan korupsi? Tidak selalu, bukan? Meminta waktu saja.

Q = (0, 0, 2) artinya query Q hanya mengandung 2 buah kata T3 .

Contoh: Q = Korupsi besar atau kecil tetap saja korupsi

Maka,

$$Q \cdot D1 = (2)(0) + (3)(0) + (5)(2) = 10$$

$$Q \cdot D2 = (3)(0) + (7)(0) + (1)(2) = 2$$

$$\|Q\| = \sqrt{4} = 2$$

$$\|D_1\| = \sqrt{38}$$

$$\|D_2\| = \sqrt{59}$$

$$\text{Sim}(Q, D_1) = \cos \theta_1 = \frac{Q \cdot D1}{\|Q\| \|D1\|} = \frac{10}{2\sqrt{38}} = 0.81$$

$$\text{Sim}(Q, D_2) = \cos \theta_2 = \frac{Q \cdot D2}{\|Q\| \|D2\|} = \frac{2}{2\sqrt{59}} = 0.13$$

Karena $0.81 > 0.13$, maka dokumen D_1 lebih sesuai dengan query Q dibandingkan dengan dokumen D_2 .

BAB 3

IMPLEMENTASI PROGRAM

Dalam source code program, terdapat beberapa kelas yang digunakan, yaitu:

1. Class FileUpload pada file models.py berfungsi untuk menyimpan informasi tentang data.
2. Class FileUploadForm pada file forms.py berfungsi untuk membuat form dan memetakan model yang dibuat ke html element.

Ada juga beberapa fungsi yang digunakan, yaitu:

1. Fungsi stemmingDokumen pada file stemming.py menerima array berupa nama-nama dokumen, melakukan pembersihan stopword pada dokumen, kemudian melakukan stemming pada dokumen.
2. Fungsi stemmingQuery pada file stemming.py menerima input berupa query, melakukan pembersihan stopword pada query, kemudian melakukan stemming pada query.
3. Fungsi cosinesimilarity pada file cosinesimilarity.py digunakan untuk menentukan kesamaan semua dokumen dengan query yang dimasukkan. Kesamaan ini dihitung dengan menggunakan rumus cosine similarity.

Parameter dari fungsi ini adalah query yang dimasukkan, dan alamat dari setiap dokumen yang akan di-check kesamaannya.

4. Fungsi conv_txt_to_html pada file uploadfile.py menerima input beberapa file dengan format txt yang sudah diupload, lalu akan mengubah file tersebut menjadi beberapa file dengan format html yang sudah didesain.
5. Fungsi deleteDocuments pada file uploadfile.py menghapus file txt yang sudah diupload dan file html yang sudah diupload.
6. Fungsi home pada file views.py berfungsi untuk menampilkan halaman home
7. Fungsi uploadFiles pada file views.py berfungsi untuk memproses file yang diupload untuk disimpan di dalam server dan juga mengubah file tersebut menjadi file html dengan bantuan fungsi conv_txt_to_html kemudian menampilkan halaman home.
8. Fungsi deleteFiles pada file views.py berfungsi untuk memanggil fungsi deleteDocuments dan menampilkan kembali halaman home.
9. Fungsi about pada file views.py menampilkan halaman about

10. Fungsi search pada file views.py berfungsi untuk mengambil input query lalu memanggil fungsi cosinesimilarity, kemudian menampilkan halaman search.
11. Fungsi dokumen berfungsi untuk menampilkan file yang sudah diupload dan sudah diubah dalam bentuk html pada saat menekan judul dokumen.

BAB 4

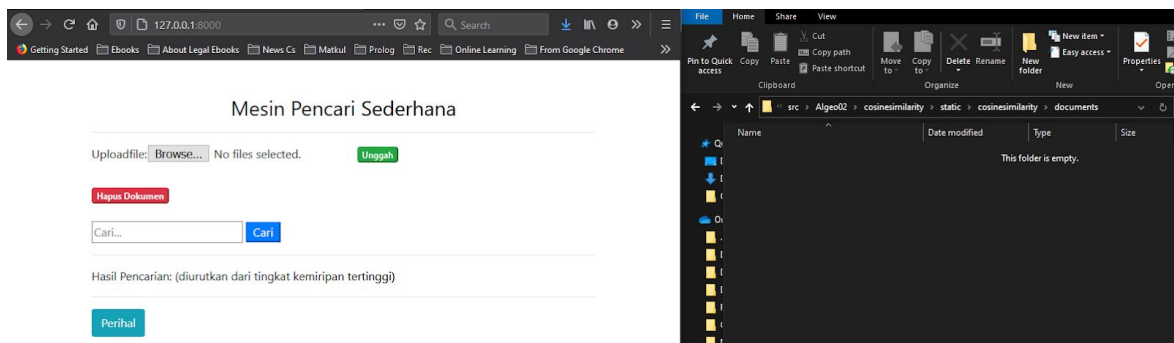
EKSPERIMEN

Berikut uji coba pada program pencarian dokumen berdasarkan hasil similaritas:

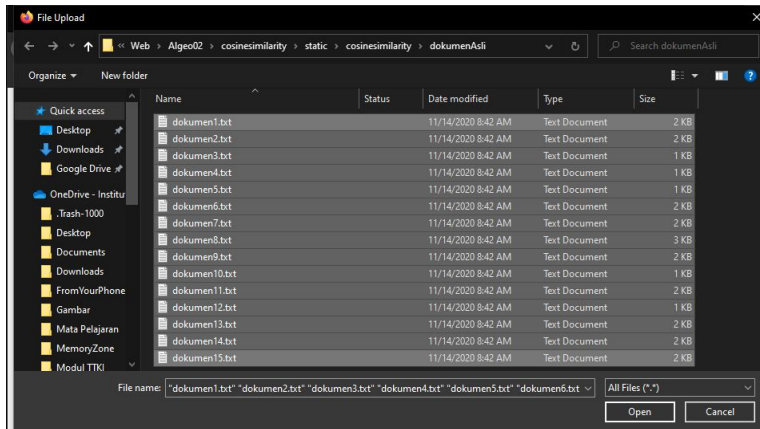
- Uji coba mengupload file dokumen, mengubah file dokumen tersebut ke dalam bentuk html, dan delete file dokumen.
- Uji coba menginput query dan menampilkan hasil pencarian terurut berdasarkan similaritas tertinggi dari hasil teratas hingga terbawah berupa judul dokumen dan kalimat pertama dari dokumen yang sudah diupload.
- Uji coba menampilkan dokumen dengan menekan judul dokumen setelah ditampilkannya hasil pencarian.
- Uji coba menampilkan halaman perihal dengan menekan button perihal dan kembali lagi ke halaman utama setelah menekan button halaman utama pada halaman perihal.

Berikut hasil potongan tangkapan layar uji coba pada program search engine kami.

1. Upload File dan Delete File



Pada gambar di atas, file masih kosong.

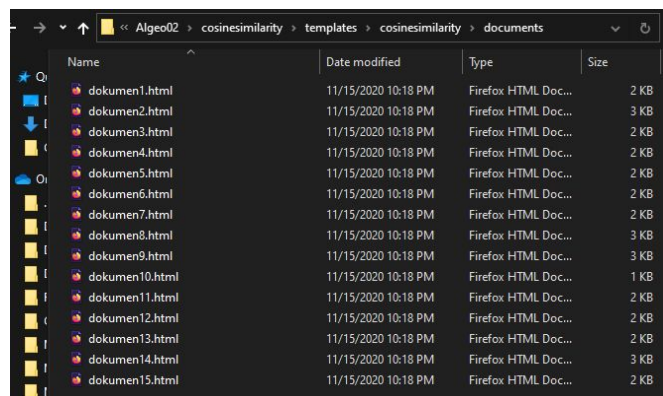
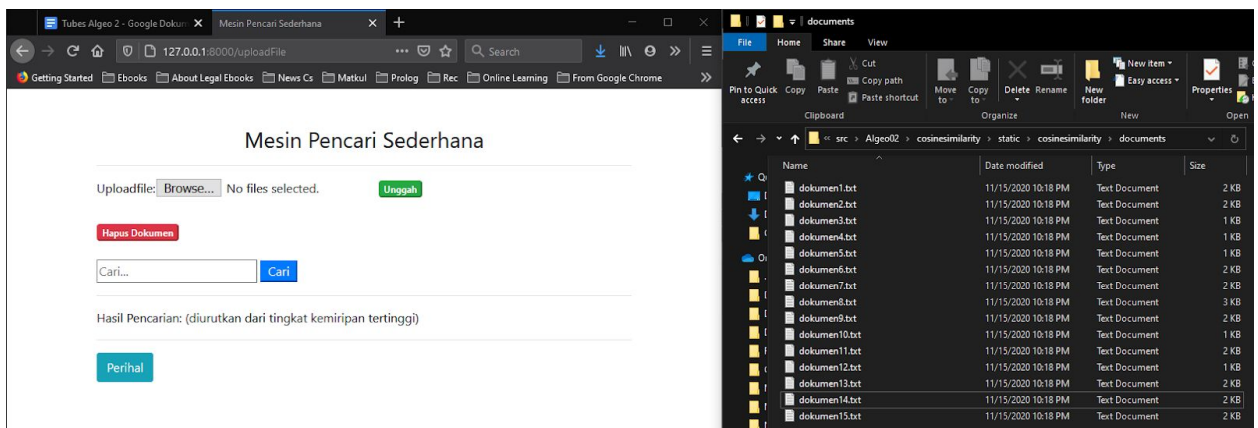


Mesin Pencari Sederhana

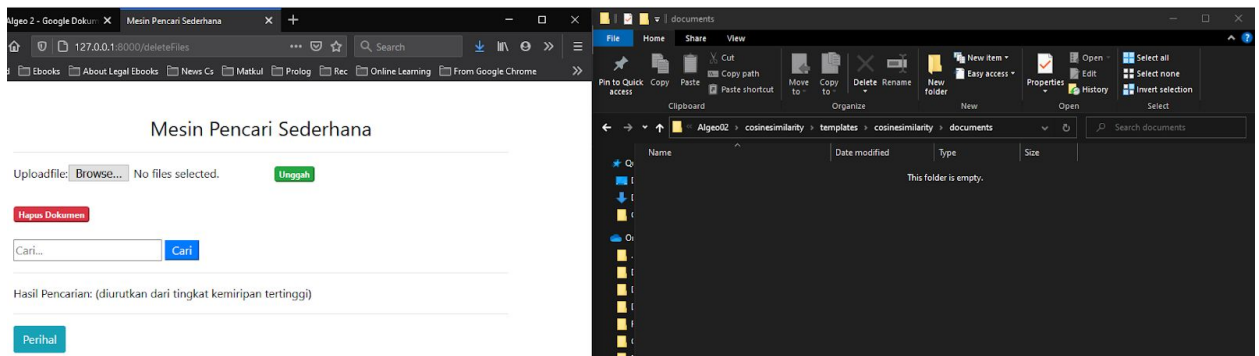
Uploadfile: 15 files selected.

Hasil Pencarian: (diurutkan dari tingkat kemiripan tertinggi)

Pada gambar di atas, file dipilih kemudian di unggah.



Pada gambar di atas, file dokumen berhasil diupload dan diubah ke dalam bentuk html.



Setelah menekan hapus dokumen, dokumen berhasil dihapus.

2. Input query dan Menampilkan Hasil Pencarian Dokumen berdasarkan Similaritas

Mesin Pencari Sederhana

Uploadfile: No files selected.

Hasil Pencarian: (diurutkan dari tingkat kemiripan tertinggi)

Pada gambar di atas, query ditulis.

Querynya : Andi berfikir bahwa STEI ITB berada di Sumatra Selatan. Namun, setelah menonton youtube di komputernya dengan sistem operasi berbasis Linux, Andi memutuskan untuk memilih matkul jurusan fisika dasar.

Alger

127.0.0.1:8000/search

57%

Search

Getting StartedEbooksAbout Legal EbooksNews CsMatkulPrologRecOnline LearningFrom Google ChromeTychonicSemester 3Books

Mesin Pencari Sederhana

Hasil Pencarian :

1. Linux

- Jumlah kata : 122
- Tingkat kemiripan : 25 %

Linux adalah keluarga sistem operasi bebas dan sumber terbuka yang dibangun di atas kernel Linux.

2. macOS

- Jumlah kata : 137
- Tingkat kemiripan : 23 %

macOS atau yang sebelumnya disebut OS X, adalah sistem operasi dengan antarmuka grafis yang dikembangkan dan didistribusikan oleh Apple Inc.

3. Microsoft Windows

- Jumlah kata : 105
- Tingkat kemiripan : 20 %

Microsoft Windows atau yang lebih dikenal dengan sebutan Windows saja adalah keluarga sistem operasi yang dikembangkan oleh Microsoft, dengan menggunakan antarmuka pengguna grafis.

4. Sumatra Selatan

- Jumlah kata : 121
- Tingkat kemiripan : 15 %

Sumatra Selatan (sumunya dingkat merjed Sumed) adalah provinsi di Indonesia yang terletak di bagian selatan Pulau Sumatra.

5. Youtube

- Jumlah kata : 61
- Tingkat kemiripan : 14 %

YouTube adalah salah satu situs terbesar di dunia.

6. Pulau Flores Dasar A

- Jumlah kata : 74
- Tingkat kemiripan : 10 %

Tempat dua pulau bali Flores Dasar A dalam buku Flores Dasar karangan Haliday, Renwick, dan Walker.

7. Sekolah Teknik Elektro dan Informatika

- Jumlah kata : 124
- Tingkat kemiripan : 9 %

Sekolah Teknik Elektro dan Informatika (STE-ITE) yang didirikan pada 1 Januari 2005 merupakan gabungan dua departemen di ITS, yaitu Departemen Teknik Elektro dan Teknik Informatika (25 Rekor No. 013/2007/001/2002).

8. Penelitian Kanker di Masa COVID-19

- Jumlah kata : 169
- Tingkat kemiripan : 5 %

Penelitian kanker pada masa COVID-19 ini sangat perlu diperhatikan.

9. Sejarah Bahasa Jawa

- Jumlah kata : 50
- Tingkat kemiripan : 5 %

Bahasa Jawa dibuat oleh James Gasing saat masih bergabung di Sun Microsystems dan dirilis tahun 1995.

10. Sinopsis A Love So Beautiful, Serial Komedi Romantis China

- Jumlah kata : 171
- Tingkat kemiripan : 5 %

A Love So Beautiful merupakan salah satu serial komedi romansa asal China yang hit pada 2017.

11. Ryan Giggs

- Jumlah kata : 68
- Tingkat kemiripan : 4 %

- Jumlah kata : 169
- Tingkat kemiripan : 5 %

Pendidikan Karakter pada masa COVID-19 ini sangat perlu diperhatikan.

- Jumlah kata : 50
- Tingkat kemiripan : 5 %

Bahasa java dibuat oleh James Gosling saat masih bergabung di Sun Microsystems dan dirilis tahun 1995.

- Jumlah kata : 171
- Tingkat kemiripan : 5 %

A Love So Beautiful merupakan salah satu serial komedi romansa asal China yang hit pada 2017.

- Jumlah kata : 88
- Tingkat kemiripan : 4 %

Ryan Joseph Giggs (nama lahir Ryan Joseph Wilson) (lahir di Cardiff, Wales, Inggris, 29 November 1973; umur 46 tahun) adalah mantan pemain sepak bola Wales yang selama karier profesionalnya bermain untuk Manchester United sejak tahun 1990 hingga pensiun pada tahun 2014 dan menjadi Asisten Pelatih Manchester United hingga tahun 2016.

- Jumlah kata : 115
- Tingkat kemiripan : 2 %

Pembaruan perguruan tinggi di Indonesia sangat diperlukan karena lingkungan terus berubah yang ditandai dengan terjadinya demokratisasi ilmu pengetahuan, masyarakat *smartphone*, munculnya generasi terkoneksi, serta terobosan inovasi teknologi dan *engineering*.

- Jumlah kata : 88
- Tingkat kemiripan : 1 %

Bonus demografi yang diprediksi akan dihadapi Indonesia pada tahun 2045 harus benar-benar dimanfaatkan demi mewujudkan Indonesia Emas 2045.

- Jumlah kata : 193
- Tingkat kemiripan : 0 %

Kegiatan ini adalah pembinaan kelima SMART Scholarship Kanwil Bandung.

- Jumlah kata : 70
- Tingkat kemiripan : 0 %

Biodiesel adalah bioenergi atau bahan bakar nabati yang dibuat dari minyak nabati, turunan tumbuh-tumbuhan yang banyak tumbuh di Indonesia seperti kelapa sawit, kelapa, kemiri, jarak pagar, nyamplung kapok, kacang tanah dan masih banyak lagi tumbuh-tumbuhan yang dapat memproduksi bahan minyak nabati dan dalam penelitian ini bahan bakar nabati berasal dari minyak kacang tanah setelah mengalami beberapa proses seperti ekstraksi, transesterifikasi diperoleh metil ester (biodiesel), kemudian biodiesel dicampur dengan bahan bakar solar.

[illegible]

Kegiatan ini adalah pembinaan kelima SMART Scholarship Kanwil Bandung.

15. Energi Biodiesel

- Jumlah kata : 70
- Tingkat kemiripan : 0 %

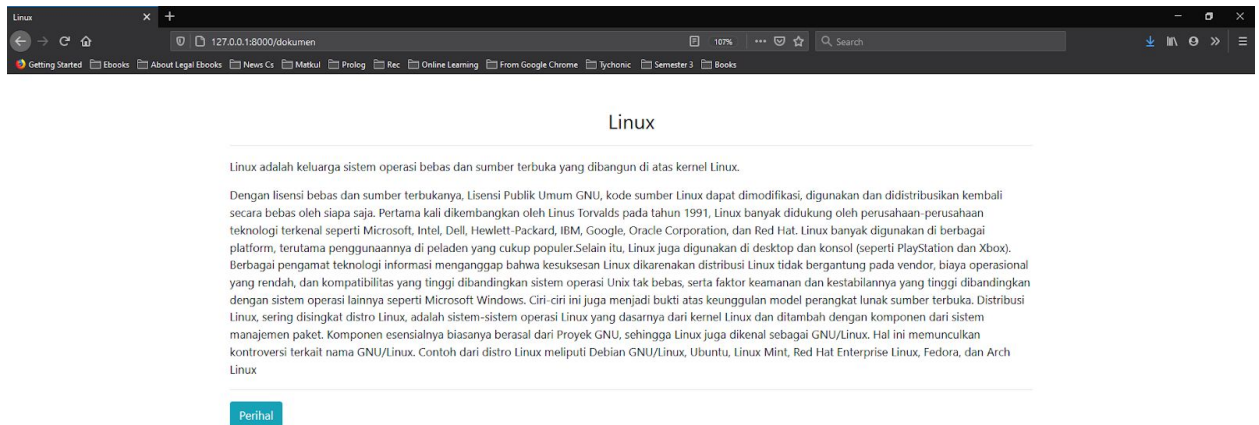
Biodiesel adalah bioenergi atau bahan bakar nabati yang dibuat dari minyak nabati, turunan tumbuh-tumbuhan yang banyak tumbuh di Indonesia seperti kelapa sawit, kelapa, kemiri, jarak pagar, nyamplung kapok, kacang tanah dan masih banyak lagi tumbuh-tumbuhan yang dapat memproduksi bahan minyak nabati dan dalam penelitian ini bahan bakar nabati berasal dari minyak kacang tanah setelah mengalami beberapa proses seperti ekstraksi, transesterifikasi diperoleh metil ester (biodiesel), kemudian biodiesel dicampur dengan bahan bakar solar.

Term	Query	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15
andi	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
berfikiran	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
stei	1	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0
itb	1	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0
sumatra	1	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0
selatan	1	0	0	6	0	0	0	0	1	0	0	0	0	0	0	0
tonton	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
youtube	1	0	0	0	0	0	0	0	1	0	8	0	0	0	0	0
komputer	1	0	1	0	0	1	0	2	0	0	0	0	0	0	0	0
sistem	1	2	1	0	0	8	6	6	0	0	0	0	0	0	0	0
operasi	1	0	1	0	0	8	4	6	0	0	0	0	0	0	0	0
bas	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0
linux	1	0	0	0	0	0	18	1	0	0	0	0	0	0	0	0
andi	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
putus	1	0	0	1	3	1	0	5	1	0	0	0	0	1	0	6
untuk	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
pilih	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
matkul	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
jurus	1	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0
fisika	1	0	0	0	0	0	1	1	0	0	0	3	1	0	0	0
dasar	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	2

Perihal

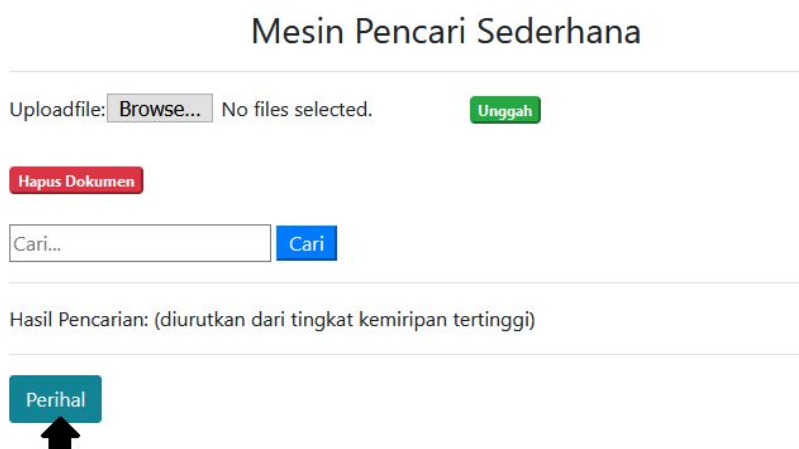
Pada gambar di atas, setelah menulis query, hasil pencarian dokumen dapat ditampilkan berdasarkan similaritas.

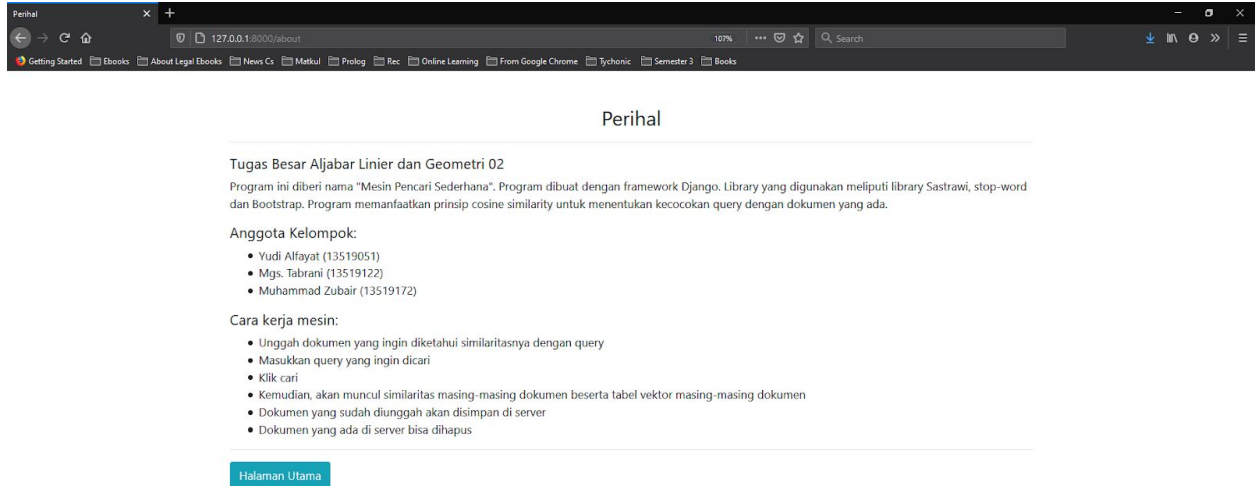
3. Menampilkan dokumen



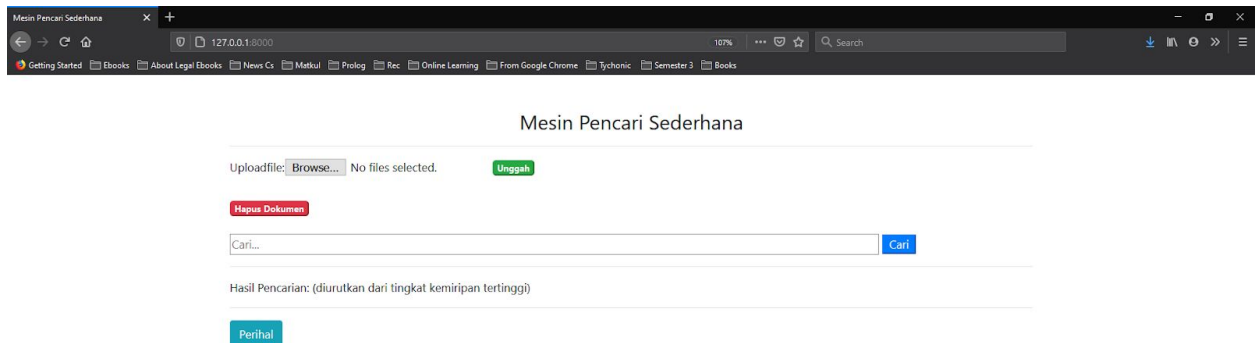
Pada gambar di atas, dokumen berhasil ditampilkan setelah menekan judul dokumen pada tampilan hasil pencarian dokumen.

4. Menampilkan halaman perihal dan dapat kembali ke halaman utama





Pada gambar di atas, setelah menekan button perihal, halaman perihal dapat ditampilkan.



Pada gambar di atas, setelah menekan button halaman utama, halaman utama akan kembali ditampilkan.

BAB 5

KESIMPULAN, SARAN, DAN REFLEKSI

Kesimpulan

Berdasarkan percobaan di atas, dapat disimpulkan bahwa program pencarian dokumen berdasarkan hasil similaritas tertinggi tersebut berhasil menampilkan hasil pencarian dokumen terurut berdasarkan similaritas tertinggi dengan cara melakukan perhitungan cosine similarity.

Saran

Pada proses pembentukan vektor dari setiap dokumen, algoritma untuk pengecekan setiap kata yang ada di dalam dokumen tersebut bisa menggunakan yang lebih optimal dan efisien. Sehingga waktu yang dibutuhkan untuk mencari kesamaan antara isi dokumen dan query dapat lebih sedikit.

Refleksi

Algoritma yang digunakan pada program ini untuk mencari vektor kesamaan dari setiap dokumen dengan query belum optimal dan efisien. Sehingga ketika melakukan pencarian memerlukan waktu yang cukup lama.

REFERENSI

1. <https://garudacyber.co.id/artikel/1436-pengertian-dan-konsep-information-retrieval>
2. <https://onlinelearning.binus.ac.id/computer-science/post/metode-metode-information-retrieval/>
3. <https://www.seputarpengetahuan.co.id/2020/05/vektor.html>
4. <https://pypi.org/project/Sastrawi/>
5. <https://pypi.org/project/stop-words/>
6. <https://www.w3schools.com/html/default.asp>
7. <https://docs.djangoproject.com/en/3.1/ref/templates/builtins/>
8. <http://informatika.stei.itb.ac.id/~rinaldi.munir/AljabarGeometri/2020-2021/Algeo-12-Aplikasi-dot-product-pada-IR.pdf>
9. <https://docs.djangoproject.com/en/3.1/topics/db/models/>
10. <https://docs.djangoproject.com/en/3.1/topics/forms/>