

Laporan Tugas Besar 2 IF2123 Aljabar Linier dan Geometri

Aplikasi Dot Product pada Sistem Temu-balik Informasi

13519040 Shafira Naya Aprisadiani

13519104 Nabelanita Utami

13519192 Gayuh Tri Rahutami



PROGRAM STUDI SARJANA TEKNIK INFORMATIKA

SEKOLAH TEKNIK ELEKTRO DAN INFORMATIKA

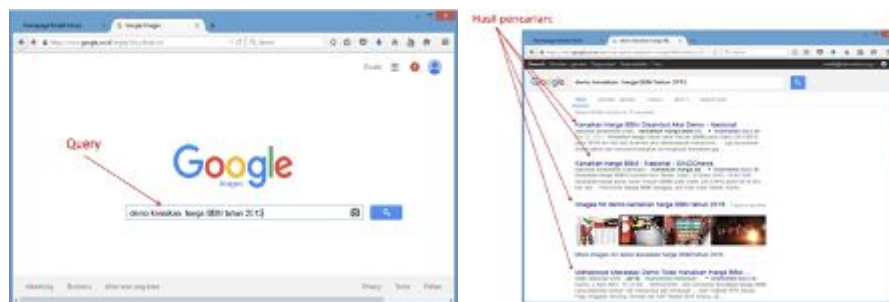
INSTITUT TEKNOLOGI BANDUNG

2020

Bab 1 Deskripsi Masalah

Hampir semua orang pernah menggunakan *search engine*, seperti *google*, *bing* dan *yahoo! search*. Setiap hari, bahkan untuk sesuatu yang sederhana kita menggunakan mesin pencarian Tapi, pernahkah kita 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:

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

Pada kesempatan ini, kami membuat sebuah *search engine* sederhana dengan model ruang vektor dan memanfaatkan cosine similarity.

Bab 2 Teori Singkat

2.1 Information retrieval

Temu-balik informasi (*information retrieval*) adalah menemukan kembali (*retrieval*) informasi yang relevan terhadap kebutuhan pengguna dari suatu kumpulan informasi secara otomatis. *Information retrieval* biasanya digunakan pada pencarian informasi yang tidak terstruktur, misalnya dokumen atau laman web. Aplikasi dari *information retrieval* salah satunya adalah search engine. Salah satu model IR adalah model ruang vektor.

2.2 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 $\mathbf{w} = (w_1, w_2, \dots, w_n)$ di dalam \mathbb{R}^n . Dengan w_i menyatakan jumlah kemunculan kata dalam dokumen (*term frequency*).

2.3 Cosine similarity

Similarity measure adalah penentuan dokumen mana yang relevan dengan query. Kesamaan atau *similarity* antara vektor query $\mathbf{Q} = (q_1, q_2, \dots, q_n)$ dan vektor dokumen $\mathbf{D} = (d_1, d_2, \dots, d_n)$ dapat dihitung dengan rumus *cosine similarity* yang merupakan rumus dot product dari dua vektor.

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

dengan $\mathbf{Q} \cdot \mathbf{D}$ adalah hasil dot product antara vektor \mathbf{Q} dengan vektor \mathbf{D} , yaitu

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

Jika $\cos \theta = 1$ atau $\text{sim}(\mathbf{Q}, \mathbf{D}) = 1$ artinya dokumen \mathbf{D} sama dengan query \mathbf{Q} . Semakin besar nilai cosinus, maka dokumen semakin sesuai dengan query.

Setiap dokumen di dalam koleksi dokumen dihitung nilai similaritasnya terhadap query dengan rumus *cosinus similarity*. Selanjutnya hasil perhitungan diurutkan berdasarkan nilai similaritas, mulai dari yang terbesar (dokumen yang paling relevan dengan query) sampai yang terkecil (dokumen yang paling tidak relevan dengan query).

2.4 Kelebihan dan kelemahan dari model ruang vektor

Model ruang vektor untuk *information retrieval* memiliki beberapa kelebihan dan kelemahan. Kelebihan dari model ruang vektor antara lain:

- Sempel, praktis, dan menggunakan pendekatan berbasis matematika.
- Dapat menyajikan hasil pencarian yang diranking.
- Memungkinkan *partial matching*.
- Dapat menghitung similaritas antara dokumen dengan query, dokumen dengan dokumen, query dengan query, dan lain-lain.

Sedangkan, beberapa kelemahannya yaitu:

- Tidak terdapat informasi secara sintaksis (struktur frasa, urutan kata-kata, *proximity information*).
- Tidak terdapat informasi semantik (*word sense*), serta pada model ruang vektor asumsinya adalah *term* saling bebas satu sama lain dan mengabaikan sinonimitas.
- Kurangnya kontrol terhadap Boolean model (mengharuskan sebuah term untuk muncul dalam sebuah dokumen), misalnya diberikan query yang terdiri dari dua term “A B”, model ini bisa saja memilih dokumen 1 yang terdiri dari term A dengan frekuensi kemunculan yang besar, dibanding memilih dokumen 2 yang terdiri dari term A dan B namun dengan frekuensi kemunculan yang lebih kecil.

Bab 3 Implementasi Program

Kami menggunakan bahasa Python dalam pembuatan algoritma *Information Retrieval* serta framework Flask, HTML, CSS, Javascript, dan Bootstrap dalam pembuatan website. Kami juga menggunakan SQLite dalam pembuatan database penyimpanan file yang diupload oleh user.

3.1 searchEngine.py

File searchEngine.py berisi program backend yang berisi fungsi-fungsi. Pada file ini kami mengimpor library nltk, sqlite3, dan math.

3.1.1 Fungsi connect(file)

Fungsi ini akan menghubungkan program dengan database.

3.1.2 Fungsi selectFiles(conn)

Fungsi ini akan mengembalikan file-file pencarian dari database.

3.1.3 Fungsi countWords(file)

Fungsi ini akan mengembalikan jumlah kata dari suatu file teks yang dimasukkan sebagai argumen.

3.1.4 Fungsi searchEngine(query)

Fungsi ini merupakan fungsi yang akan dipanggil ketika user melakukan pencarian. Fungsi ini kemudian akan memanggil fungsi conn untuk menghubungkan program dengan *database*. Fungsi ini lalu akan memanggil fungsi selectFiles untuk mengambil file-file pencarian dari *database* dan menyimpan file-file tersebut dalam bentuk list.

Kemudian, program akan menghapus karakter-karakter non-alphanumeric dan kata-kata *stopwords* dari query dan memisahkan kata-kata dalam query per kata yang kemudian disimpan dalam bentuk tabel yang berisikan kata-kata yang ada di query serta jumlah tiap katanya. Hal yang sama kemudian juga dilakukan terhadap file-file pencarian.

Setelah terbentuk tabel yang berisikan jumlah tiap kata pada query dan file pencarian, program akan menghitung *cosine similarity* dari tiap file pencarian dan mengurutkannya berdasarkan file yang memiliki similaritas terbesar ke terkecil. Fungsi ini akan mengembalikan tabel yang berisi *cosine similarity* tiap file pencarian, tabel yang berisi jumlah tiap kata query pada file, dan tabel yang berisi judul dokumen.

3.2 app.py

File app.py dibangun menggunakan framework Flask yang dibutuhkan agar website dapat berjalan. Pada file ini, terdapat route website, antara lain route ke index.html, upload.html, search.html, results.html, serta halaman artikel.

3.2.1 Route ke index.html (“/”)

Route ini menghubungkan app.py dengan file index.html. File html ini berisi garis besar website serta profil kelompok kami.

3.2.2 Route ke upload.html (“/upload”)

Route ini menghubungkan app.py dengan halaman upload file yaitu upload.html. Pada route ini juga melibatkan database. Ketika user mengupload file, apabila database belum tersedia, yaitu ketika baru pertama kali dilakukan upload, fungsi database() akan membentuk sebuah database baru dengan nama files.db. Apabila sudah ada, maka akan dilakukan penambahan file baru ke dalam database.

Kami menggunakan library flask_wtf untuk membuat Form upload, serta menggunakan bootstrap dan css untuk styling halaman.

3.2.3 Route ke search.html (“/search”)

Route ini menghubungkan app.py dengan halaman input query pencarian yaitu search.html. File search.html akan dirender dengan memanggil fungsi render_template() dengan argumen “search.html” dan variable search yang berisi query dari user. Pada route ini, kami menggunakan library flask_wtf untuk membuat route input text serta menggunakan bootstrap dan css untuk styling. Pada route ini, app.py akan memanggil fungsi results() yang didalamnya akan memanggil fungsi searchEngine() dari file searchEngine.py. Fungsi ini akan mengembalikan hasil searching berdasarkan query yang telah dimasukkan oleh user. Sesaat setelah user menekan tombol Search, hasil akan langsung ditampilkan ke layar.

3.2.4 Route ke results.html (“/results”)

Route ini akan berisi fungsi results() yang akan menampilkan hasil pencarian ke layar. Di dalam fungsi ini, dipanggil fungsi searchEngine() dari file seachEngine.py. Fungsi ini akan mengembalikan tuple berukuran dua yang berisi array hasil pada indeks ke-0. Array ini berisi nama file, tingkat kemiripan file, jumlah kata, serta kalimat pertama dari file tersebut. Isi dari array ini juga sudah terurut menurun berdasarkan tingkat kemiripan. Selain itu, indeks ke-1 berisi array yang merupakan query serta jumlah kata yang ditemukan pada tiap file yang sesuai dengan query. Apabila query berupa stopwords, maka indeks ke-1 berupa array dua dimensi yang kosong.

Setelah itu, file results.html akan dirender dengan memanggil fungsi render_template() dengan argumen berupa “result.html” serta isi dari array yang diperlukan untuk menampilkan hasil ke layar,

antara lain query yang diinputkan user, array terurut hasil pencarian, panjang array pencarian, tabel jumlah kemunculan query pada tiap file, serta panjang baris dan kolom tabel kemunculan query.

3.2.5 Route ke <filename>.html (“</filename>”)

Route ini digunakan untuk menampilkan teks utuh dari setiap dokumen yang ada pada database. Route ini dapat diakses dengan mengklik hyperlink judul dokumen pada halaman hasil pencarian. Terdapat fungsi `get_file()` yang mereturn satu file dari database dengan method `.fetchone()`. File yang dikembalikan berupa array dengan indeks ke-0 berupa nama file dan indeks ke-1 berupa isi teks atau konten. Konten di-decode, kemudian file `text.html` akan dirender dengan fungsi `render_template()` untuk menampilkan teks utuh ke layar pada route `<filename>.html`.

3.3 Templates

Templates merupakan file statik yang diperlukan untuk front-end website ini. File templates yang kami gunakan antara lain `base.html`, `index.html`, `result.html`, `search.html`, `text.html`, dan `upload.html`.

3.3.1 base.html

File ini merupakan *base* untuk pembuatan file templates yang lain. Kami menggunakan file `base.html` ini agar tidak perlu melakukan pengetikan kode HTML yang berulang seperti head, link ke file css, dan sebagainya.

3.3.2 index.html

File ini berisi HTML yang membangun halaman utama dari website. Halaman ini berisi informasi singkat mengenai website

3.3.3 result.html

File ini berisi HTML untuk menampilkan hasil pencarian. Halaman ini berisi hasil pencarian yang telah diranking berupa judul, kalimat pertama dari dokumen, dan nilai similaritas. Pada halaman ini juga terdapat tabel similaritas.

3.3.4 search.html

File ini berisi HTML yang membangun search bar dari website kami. Hasil dari input user akan digunakan sebagai query untuk menampilkan hasil searching. Pada halaman ini juga terdapat tombol untuk beralih ke halaman upload apabila user ingin menambah file baru ke database.

3.3.5 text.html

File ini berisi template HTML untuk menampilkan teks utuh dari setiap dokumen.

3.3.6 upload.html

File ini berisi HTML yang membangun form agar user dapat menambahkan file baru ke database. File yang diupload oleh user akan dijadikan argumen oleh fungsi database() pada file app.py yang akan menuliskan judul dan isi file yang diupload oleh user ke database, atau membuat database baru apabila tidak ada.


3.4 Static

Folder static berisi folder css dan img. Folder css berisi file-file css yang digunakan untuk men-styling html. Kami menggunakan Bootstrap sebagai framework CSS. Folder img berisi logo Sandbox yang ditampilkan pada navbar dan search box.

Bab 4 Eksperimen

Percobaan 1

Pada percobaan pertama, kami memasukkan query berupa “Bake fresh foods”. Ketiga kata ini merupakan kata dasar, sehingga setelah di-*stem* mendapat tiga term yaitu “bake”, “food”, dan “fresh”. Dokumen yang paling relevan adalah dokumen “bakery1.txt” yang memiliki nilai similaritas sebesar 0.35, selanjutnya diikuti oleh resto2.txt, resto4.txt, resto5.txt, bakery2.txt. Dokumen sisanya memiliki nilai similaritas 0.0.

 Home Search Logout

Search results for "Bake fresh foods"

Found 15 results

.bakery1.txt

Word count: 64
Similarity ratio: 0.35
First sentence: "Our founder, Anna and her husband, Roger, have been baking and baking in Belgium since 1992."

.resto2.txt

Word count: 73
Similarity ratio: 0.23
First sentence: "We serve our own sake on draft, and use only the freshest sake available."

.resto4.txt

Word count: 16
Similarity ratio: 0.15
First sentence: "We serve our customers homemade bread and delicious pasta dishes as well as desserts."

.resto5.txt

Word count: 16
Similarity ratio: 0.08
First sentence: "We serve unique wines from around Europe as well our own recipes for you to enjoy during your meal."

.bakery2.txt

Word count: 87
Similarity ratio: 0.05
First sentence: "Our unique bakery-coffs, which has been featured in numerous local and national publications, has been featured on TV cooking shows, has been featured on the Today Show, has been featured in the New York Times, has been featured in numerous restaurant and retail press outlets, and have been featured on Martha Stewart Living."

.skincare1.txt

Word count: 64
Similarity ratio: 0.0
First sentence: "The products are all cruelty free, vegan and suitable for everyone."

.skincare2.txt

Word count: 87
Similarity ratio: 0.0
First sentence: "I have a background in professional clinical dermatology since the age 12, I received my doctorate in cosmetic science from Stanford University and trained as an esthetician before applying to work for ten years at Cosmetic Pro Plastics (Singapore)."

.resto1.txt

Word count: 105
Similarity ratio: 0.0
First sentence: "We serve up delicious but affordable delights to our guests."

.hairsalon1.txt

Word count: 103
Similarity ratio: 0.0
First sentence: "We are located on the second floor of the Redstone Inn."

.hairsalon2.txt

Word count: 61
Similarity ratio: 0.0
First sentence: "We're a boutique salon with a focus on quality and the latest haircare techniques, hair services, and styling accessories."

.hairsalon3.txt

Word count: 65
Similarity ratio: 0.0
First sentence: "The beauty salon at The Beauty Inn is the perfect solution for those who are searching for a place to relax and enjoy their time away from the hustle and bustle of the city."

.hairsalon4.txt

Word count: 65
Similarity ratio: 0.0
First sentence: "We are a hair salon with a passion."

.fashion1.txt

Word count: 115
Similarity ratio: 0.0
First sentence: "FACT: FACT is a fashion concept for every occasion."

.fashion2.txt

Word count: 115
Similarity ratio: 0.0
First sentence: "Founded by Amanda Saffron on a beach in Mexico, where life takes you wherever you go, Fotobis is passionate about embracing people of all ages and cultures."

.resto3.txt

Word count: 105
Similarity ratio: 0.0
First sentence: "We serve up delicious but affordable delights to our guests."

Similarity Table

Term	Query	skincare1.txt	skincare2.txt	resto1.txt	bakery1.txt	bakery2.txt	hairsalon1.txt	hairsalon2.txt	hairsalon3.txt	hairsalon4.txt
bake	1	0	0	0	0	0	0	0	0	0
food	1	0	0	0	0	1	0	0	0	0
fresh	1	0	0	0	0	0	0	0	0	0

Percobaan 2

Pada percobaan kedua, kami memasukkan query berupa “serve serving serves”. Setelah di-stem, ketiga kata tersebut menjadi satu kata yang sama yaitu “serv” seperti yang tercantum pada tabel.

SandBox

Home

Search

Upload

Search results for "serve serving serves"

Found 15 results

. resto4.txt

Word count: 116

Similarity rate: 0.35

First sentence: "We serve our customers homemade bread and delicious pasta dishes as well as desserts."

. resto1.txt

Word count: 100

Similarity rate: 0.22

First sentence: "We serve up delicious but affordable delights to our guests."

. resto3.txt

Word count: 100

Similarity rate: 0.22

First sentence: "We serve up delicious but affordable delights to our guests."

. resto5.txt

Word count: 76

Similarity rate: 0.14

First sentence: "We serve unique wines from around Europe as well our own recipes for you to enjoy during your meal."

. resto2.txt

Word count: 73

Similarity rate: 0.1

First sentence: "We serve our own sake on draft, and use only the freshest sake available."

. hairsalon1.txt

Word count: 123

Similarity rate: 0.08

First sentence: "We are located on the second floor of the Redstone Inn."

. skincare1.txt

Word count: 64

Similarity rate: 0.0

First sentence: "The products are all cruelty free, vegan and suitable for everyone."

. skincare2.txt

Word count: 87

Similarity rate: 0.0

First sentence: "I have a background in professional clinical dermatology since the age 12, I received my doctorate in cosmetic science from Stanford University and trained as an esthetician before applying to work for ten years at Cosmetic Pro Plastics (Spartanburg)."

. bakery1.txt

Word count: 84

Similarity rate: 0.0

First sentence: "Our founder, Anna and her husband, Roger, have been baking and baking in Belgium since 1992."

. bakery2.txt

Word count: 87

Similarity rate: 0.0

First sentence: "Our unique bakery-coffee, which has been featured in numerous local and national publications, has been featured on TV cooking shows, has been featured on the Today Show, has been featured in the New York Times, has been featured in numerous restaurant and retail press outlets, and have been featured on Martha Stewart Living."

. hairsalon2.txt

Word count: 81

Similarity rate: 0.0

First sentence: "We're a boutique salon with a focus on quality, and the latest haircare techniques, hair services, and styling accessories."

. hairsalon3.txt

Word count: 65

Similarity rate: 0.0

First sentence: "The beauty salon at The Beauty Inn is the perfect solution for those who are searching for a place to relax and enjoy their time away from the hustle and bustle of the city."

. hairsalon4.txt

Word count: 95

Similarity rate: 0.0

First sentence: "We are a hair salon with a passion."

. fashion1.txt

Word count: 116

Similarity rate: 0.0

First sentence: "FACT: FACT is a fashion concept for every occasion."

. fashion2.txt

Word count: 118

Similarity rate: 0.0

First sentence: "Founded by Amanda Staffron on a beach in Mexico, where life takes you wherever you go, Fotoba is passionate about embracing people of all ages and cultures."

Similarity Table

Term	Query	skincare1.txt	skincare2.txt	resto1.txt	bakery1.txt	bakery2.txt	hairsalon1.txt	hairsalon2.txt	hairsalon3.txt	hairsalon4.txt
serv	3	0	0	2	0	0	1	0	0	0

10 | IF2123 Aljabar Linier dan Geometri

Percobaan 3

Pada percobaan ketiga, kami memasukkan query berupa “skincare dermatology, skin beautiful salon”. Karena dilakukan stemming pada query, didapatkan term query yaitu “beauti”, “dermatolog”, “salon”, “skin”, “skincar”. Dokumen yang paling relevan dengan query adalah “hairsalon3.txt” dengan nilai similaritas 0.35.

Sanctus

HomeSearchUpload

Search results for "skincare dermatology, skin beautiful salon"

Found 15 results

hairsalon3.txt

Word count: 65

Similarity rate: 0.35

First sentence: "The beauty salon at The Beauty Inn is the perfect solution for those who are searching for a place to relax and enjoy their time away from the hustle and bustle of the city."

skincare1.txt

Word count: 84

Similarity rate: 0.21

First sentence: "The products are all cruelty free, vegan and suitable for everyone."

hairsalon2.txt

Word count: 81

Similarity rate: 0.17

First sentence: "We're a boutique salon with a focus on quality, and the latest haircare techniques, hair services, and styling accessories."

skincare2.txt

Word count: 87

Similarity rate: 0.12

First sentence: "I have a background in professional clinical dermatology since the age 12. I received my doctorate in cosmetic science from Stanford University and trained as an esthetician before applying to work for ten years at Cosmetic Pro Plastics (Spokane)."

hairsalon4.txt

Word count: 96

Similarity rate: 0.12

First sentence: "We are a hair salon with a passion."

hairsalon1.txt

Word count: 123

Similarity rate: 0.1

First sentence: "We are located on the second floor of the Redstone Inn."

resto1.txt

Word count: 100

Similarity rate: 0.0

First sentence: "We serve up delicious but affordable delights to our guests."

bakery1.txt

Word count: 94

Similarity rate: 0.0

First sentence: "Our founder, Anna and her husband, Roger, have been baking and baking in Belgium since 1992."

bakery2.txt

Word count: 87

Similarity rate: 0.0

First sentence: "Our unique bakery- cafe, which has been featured in numerous local and national publications, has been featured on TV cooking shows, has been featured on the Today Show, has been featured in the New York Times, has been featured in numerous restaurants and retail press outlets, and have been featured on Martha Stewart Living"

fashion1.txt

Word count: 116

Similarity rate: 0.0

First sentence: "FACT1-FACT1 is a fashion concept for every occasion."

fashion2.txt

Word count: 116

Similarity rate: 0.0

First sentence: "Founded by Amanda Saffron on a beach in Mexico, where life takes you wherever you go, Fotobae is passionate about embracing people of all ages and cultures."

resto2.txt

Word count: 73

Similarity rate: 0.0

First sentence: "We serve our own sake on draft, and use only the freshest sake available."

resto3.txt

Word count: 100

Similarity rate: 0.0

First sentence: "We serve up delicious but affordable delights to our guests."

resto4.txt

Word count: 116

Similarity rate: 0.0

First sentence: "We serve our customers homemade bread and delicious pasta dishes as well as desserts."

resto5.txt

Word count: 78

Similarity rate: 0.0

First sentence: "We serve unique wines from around Europe as well our own recipes for you to enjoy during your meal."


Similarity Table

Term	Query	skincare1.txt	skincare2.txt	resto1.txt	bakery1.txt	bakery2.txt	hairsalon1.txt	hairsalon2.txt	hairsalon3.txt	hairsalon4.txt
beauti	1	2	0	0	0	0	0	0	3	0
dermatolog	1	0	1	0	0	0	0	0	0	0
salon	1	0	0	0	0	0	3	3	2	2
skin	1	2	1	0	0	0	0	0	0	0
skincar	1	0	0	0	0	0	0	0	0	0

11 | IF2123 Aljabar Linier dan Geometri

Percobaan 4

Pada kasus ketiga, kami memasukkan query yang semuanya berupa stopwords, yaitu “and then, I so!” Pada hasil pencarian ditemukan 15 dokumen dengan similarity rate 0. Tabel tidak ditampilkan karena query berupa stopwords semuanya, jadi dianggap tidak ada term query.

 Home Search Upload

Search results for "and then, I so!"

Found 15 results

. skincare1.txt

Word count: 84

Similarity rate: 0

First sentence: "The products are all cruelty free, vegan and suitable for everyone."

. skincare2.txt

Word count: 87

Similarity rate: 0

First sentence: "I have a background in professional clinical dermatology since the age 12, I received my doctorate in cosmetic science from Stanford University and trained as an esthetician before applying to work for ten years at Cosmetic Pro Plastics (Spartans)."

. resto1.txt

Word count: 100

Similarity rate: 0

First sentence: "We serve up delicious but affordable delights to our guests."

. bakery1.txt

Word count: 84

Similarity rate: 0

First sentence: "Our founder, Anna and her husband, Roger, have been baking and baking in Belgium since 1992."

. bakery2.txt

Word count: 87

Similarity rate: 0

First sentence: "Our unique bakery-café, which has been featured in numerous local and national publications, has been featured on TV cooking shows, has been featured on the Today Show, has been featured in the New York Times, has been featured in numerous restaurant and retail press outlets, and have been featured on Martha Stewart Living."

. hairsalon1.txt

Word count: 123

Similarity rate: 0

First sentence: "We are located on the second floor of the Redstone Inn."

. hairsalon2.txt

Word count: 81

Similarity rate: 0

First sentence: "We're a boutique salon with a focus on quality, and the latest haircare techniques, hair services, and styling accessories."

. hairsalon3.txt

Word count: 65

Similarity rate: 0

First sentence: "The beauty salon at The Beauty Inn is the perfect solution for those who are searching for a place to relax and enjoy their time away from the hustle and bustle of the city."

. hairsalon4.txt

Word count: 95

Similarity rate: 0

First sentence: "We are a hair salon with a passion."

. fashion1.txt

Word count: 115

Similarity rate: 0

First sentence: "FACT: FACT is a fashion concept for every occasion."

. fashion2.txt

Word count: 115

Similarity rate: 0

First sentence: "Founded by Amanda Saffron on a beach in Mexico, where life takes you wherever you go, Fotaba is passionate about embracing people of all ages and cultures."

. resto2.txt

Word count: 73

Similarity rate: 0

First sentence: "We serve our own sake on draft, and use only the freshest sake available."

. resto3.txt

Word count: 100

Similarity rate: 0

First sentence: "We serve up delicious but affordable delights to our guests."

. resto4.txt

Word count: 116

Similarity rate: 0

First sentence: "We serve our customers homemade bread and delicious pasta dishes as well as desserts."


. resto5.txt

Word count: 78

Similarity rate: 0

First sentence: "We serve unique wines from around Europe as well our own recipes for you to enjoy during your meal."

Percobaan 5

 Home Search Upload

Search results for "I have a background in professional clinical dermatology since the age 12, I received my doctorate in cosmetic science from Stanford University and trained as an esthetician before applying to work for ten years at Cosmetic Pro Plastics (Sparlona). It was there that this passion grew which led me one amazing journey where i earned my license to practice Dermatology/Racial Dermal Exam, Skin Care & Pregnancy then went on another awesome path by owning Dermatologie LLC back home in Southern California!"

Found 15 results

• skincare2.txt

Word count: 87

Similarity rate: 0.84

First sentence: "I have a background in professional clinical dermatology since the age 12, I received my doctorate in cosmetic science from Stanford University and trained as an esthetician before applying to work for ten years at Cosmetic Pro Plastics (Sparlona)."

• fashion2.txt

Word count: 115

Similarity rate: 0.1

First sentence: "Founded by Amanda baffon on a beach in Mexico, where life takes you wherever you go, Fotobio is passionate about embracing people of all ages and cultures."

• skincare1.txt

Word count: 84

Similarity rate: 0.08

First sentence: "The products are oil cruelty free, vegan and suitable for everyone."

• bakery1.txt

Word count: 84

Similarity rate: 0.07

First sentence: "Our founder, Anna and her husband, Roger, have been baking and baking in Belgium since 1992."

• hairsalon4.txt

Word count: 95

Similarity rate: 0.07

First sentence: "We are a hair salon with a passion."

• hairsalon1.txt

Word count: 123

Similarity rate: 0.04

First sentence: "We are located on the second floor of the Radstone Inn."

• resto1.txt

Word count: 105

Similarity rate: 0.03

First sentence: "We serve up delicious but affordable delights to our guests."

• resto3.txt

Word count: 100

Similarity rate: 0.03

First sentence: "We serve up delicious but affordable delights to our guests."

• resto5.txt

Word count: 78

Similarity rate: 0.02

First sentence: "We serve unique wines from around Europe as well our own recipes for you to enjoy during your meal."

• fashion1.txt

Word count: 115

Similarity rate: 0.01

First sentence: "TACT: FACT is a fashion concept for every occasion."

• resto2.txt

Word count: 73

Similarity rate: 0.01

First sentence: "We serve our own sake on draft, and use only the freshest sake available."

• bakery2.txt

Word count: 87

Similarity rate: 0.0

First sentence: "Our unique bakery-coffee, which has been featured in numerous local and national publications, has been featured on TV cooking shows, has been featured on the Today Show, has been featured in the New York Times, has been featured in numerous restaurant and retail press outlets, and have been featured on Martha Stewart Living."

• hairsalon2.txt

Word count: 81

Similarity rate: 0.0

First sentence: "We're a boutique salon with a focus on quality, and the latest haircare techniques, hair services, and styling accessories."

• hairsalon3.txt

Word count: 85

Similarity rate: 0.0

First sentence: "The beauty salon at The Beauty Inn is the perfect solution for those who are searching for a place to relax and enjoy their time away from the hustle and bustle of the city."

• resto4.txt

Word count: 116

Similarity rate: 0.0

First sentence: "We serve our customers homemade bread and delicious pasta dishes as well as desserts."

Similarity Table

Term	Query	skincare1.txt	skincare2.txt	resto1.txt	bakery1.txt	bakery2.txt	hairsalon1.txt	hairsalon2.txt	hairsalon3.txt
id	1	0	1	0	0	0	0	0	0
age	1	0	1	0	0	0	0	0	0
amaz	1	1	1	0	0	0	0	0	0
cnath	1	0	1	0	0	0	0	0	0
appl	1	0	1	0	0	0	0	0	0
awesom	1	0	1	0	0	0	0	0	0
back	1	0	1	0	0	0	0	0	0

13 | IF2123 Aljabar Linier dan Geometri

Pada percobaan ini, kami memasukkan query berupa teks yang sama dengan isi dokumen skincare2.txt, yaitu: “I have a background in professional clinical dermatology since the age 12, I received my doctorate in cosmetic science from Stanford University and trained as an esthetician before applying to work for ten years at Cosmetic Pro Plastics (Sparlona). It was there that this passion grew which led me one amazing journey where i earned my license to practice Dermatology/Racial Dermal Exam, Skin Care & Pregnancy then went on another awesome path by owning Dermatologie LLC back home in Southern California!”

Dokumen yang paling relevan dengan query adalah dokumen “skincare2.txt” dengan nilai similaritas yang sangat tinggi yaitu 0.94.

Percobaan 6

Pada percobaan ini, kami memasukkan query berupa “I love Italian foods and desserts, so I like bakery!”. Dari 10 kata, didapatkan 6 term query karena sisanya merupakan stopwords, yaitu “I”, “so”, dan “and”. Berdasarkan hasil pencarian, “resto4.txt” merupakan dokumen yang paling relevan dengan query, dengan nilai similaritas 0.18.

SanDisk

HomeSearchUpload

Search results for "I love Italian foods and desserts, so I like bakery!"

Found 15 results

• resto4.txt

Word count: 116

Similarity rate: 0.18

First sentence: "We serve our customers homemade bread and delicious pasta dishes as well as desserts."

• resto5.txt

Word count: 78

Similarity rate: 0.06

First sentence: "We serve unique wines from around Europe as well our own recipes for you to enjoy during your meal."

• bakery2.txt

Word count: 87

Similarity rate: 0.03

First sentence: "Our unique bakery-cake, which has been featured in numerous local and national publications, has been featured on TV cooking shows, has been featured on the Today Show, has been featured in the New York Times, has been featured in numerous restaurant and retail press outlets, and have been featured on Martha Stewart Living."

• hairsalon1.txt

Word count: 103

Similarity rate: 0.03

First sentence: "We are located on the second floor of the Redstone Inn."

• skincare1.txt

Word count: 94

Similarity rate: 0.0

First sentence: "The products are all cruelty free, vegan and suitable for everyone."

• skincare2.txt

Word count: 87

Similarity rate: 0.0

First sentence: "I have a background in professional clinical dermatology since the age 12. I received my doctorate in cosmetic science from Stanford University and trained as an esthetician before applying to work for ten years at Cosmetic Pro Plastics (Singapore)."

• resto1.txt

Word count: 100

Similarity rate: 0.0

First sentence: "We serve up delicious but affordable delights to our guests."

• bakery1.txt

Word count: 84

Similarity rate: 0.0

First sentence: "Our founder, Anna and her husband, Roger, have been baking and baking in Belgium since 1992."

• hairsalon2.txt

Word count: 81

Similarity rate: 0.0

First sentence: "We're a boutique salon with a focus on quality, and the latest haircare techniques, hair services, and styling accessories."

• hairsalon3.txt

Word count: 85

Similarity rate: 0.0

First sentence: "The beauty salon at The Beauty Inn is the perfect solution for those who are searching for a place to relax and enjoy their time away from the hustle and bustle of the city."

• hairsalon4.txt

Word count: 95

Similarity rate: 0.0

First sentence: "We are a hair salon with a passion."

• fashion1.txt

Word count: 115

Similarity rate: 0.0

First sentence: "FACT: FACT is a fashion concept for every occasion."

• fashion2.txt

Word count: 115

Similarity rate: 0.0

First sentence: "Founded by Amanda Salton on a beach in Mexico, where life takes you wherever you go, Fotaba is passionate about embracing people of all ages and cultures."

• resto2.txt

Word count: 73

Similarity rate: 0.0

First sentence: "We serve our own sake on draft, and use only the freshest sake available."

• resto3.txt

Word count: 100

Similarity rate: 0.0

First sentence: "We serve up delicious but affordable delights to our guests."

Similarity Table

Term	Query	skincare1.txt	skincare2.txt	resto1.txt	bakery1.txt	bakery2.txt	hairsalon1.txt	hairsalon2.txt	hairsalon3.txt	hairsalon4.txt
bakery	1	0	0	0	0	0	0	0	0	0
dessert	1	0	0	0	0	0	0	0	0	0
food	1	0	0	0	0	1	0	0	0	0
salon	1	0	0	0	0	0	0	0	0	0
like	1	0	0	0	0	0	0	0	0	0
love	1	0	0	0	0	0	1	0	0	0

15 | IF2123 Aljabar Linier dan Geometri

Bab 5 Kesimpulan, Saran, Refleksi

Kesimpulan

Temu-balik informasi (*information retrieval*) adalah menemukan kembali (*retrieval*) informasi yang relevan terhadap kebutuhan pengguna dari suatu kumpulan informasi secara otomatis yang biasanya digunakan pada pencarian informasi yang tidak terstruktur. Salah satu aplikasi dari *information retrieval* adalah search engine. Salah satu metode yang dapat digunakan untuk membuat suatu sistem *search engine* adalah dengan menghitung *Cosine Similarity*.

Hal ini dilakukan dengan menghitung nilai similaritas tiap dokumen di dalam koleksi dokumen terhadap query dengan rumus *cosine similarity*. Selanjutnya hasil perhitungan diurutkan berdasarkan nilai similaritas, mulai dari yang paling relevan dengan query sampai yang paling tidak relevan dengan query.

Dengan mengimplementasikan dua hal di atas, kami kemudian membuat sebuah *website* pencarian.

Saran

Ada beberapa hal yang dapat ditingkatkan untuk membuat program ini menjadi lebih baik. Pertama, kami bisa menambahkan web scraping untuk mengambil dokumen dari internet sehingga user tidak perlu mengunggah dokumen secara manual. Kedua, desain *website* bisa dibuat lebih baik lagi sehingga user merasa lebih nyaman ketika menggunakan *website* kami.

Refleksi

Dalam pengerjaan tugas besar ini, kami tidak melalui kendala yang berarti. Pada saat awal pembuatan program, kami memang mendapatkan sedikit kesulitan karena tidak ada satupun dari kami yang berpengalaman dalam membuat *back-end website*. Namun, hal tersebut dapat terlewati dengan melakukan pembagian tugas yang jelas sejak awal sehingga anggota yang mendapatkan bagian untuk membuat *back-end* mempunyai waktu yang cukup untuk mempelajari framework yang kami gunakan, yaitu flask.

Daftar Referensi

<http://informatika.stei.itb.ac.id/~rinaldi.munir/AljabarGeometri/2020-2021/Algeo-12-Aplikasi-dot-product-pada-IR.pdf>

<https://sites.cs.ucsb.edu/~tyang/class/293S17/slides/Topic2IRModels.pdf>

<https://www.ccs.neu.edu/home/jaa/CSG339.06F/Lectures/vector.pdf>