

LAPORAN TUGAS BESAR II
IF2123 ALJABAR LINIER DAN GEOMETRI



Laporan ini dibuat untuk memenuhi tugas
Mata Kuliah IF 2123 Aljabar Linier dan Geometri

Disusun Oleh :

Kelompok 17

Irvin Andryan Pratomo (13519162)

Muhammad Dehan Al Kautsar (13519200)

Muhammad Rifat Abiwardani (13519205)

SEKOLAH TEKNIK ELEKTRO DAN INFORMATIKA
INSTITUT TEKNOLOGI BANDUNG
SEMESTER I TAHUN 2020/2021

BAB I

DESKRIPSI MASALAH

Buatlah program mesin pencarian 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.
 - 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

A. Temu Balik Informasi

Temu balik informasi merupakan cabang dari ilmu komputer mengenai proses pencarian suatu informasi di dalam dokumen yang relevan terhadap kumpulan informasi yang diinginkan oleh pengguna. Tujuan dari temu balik informasi adalah untuk menjembatani kebutuhan informasi pengguna dengan sumber informasi yang tersedia. Menurut Belkin (1980) temu balik informasi bertujuan untuk mempertemukan ide yang dikemukakan oleh pengguna dalam bentuk query dengan dokumen yang berisi informasi yang sesuai.

Temu balik informasi memiliki fungsi utama untuk mengidentifikasi informasi yang relevan dengan keinginan pengguna, mengidentifikasi isi dokumen yang berisi informasi sesuai keinginan pengguna, merepresentasikan kebutuhan informasi pengguna dalam bentuk query, mempertemukan query dari pengguna dengan informasi di dalam dokumen menggunakan algoritma tertentu, dan mengembalikan informasi yang diperoleh kepada pengguna.

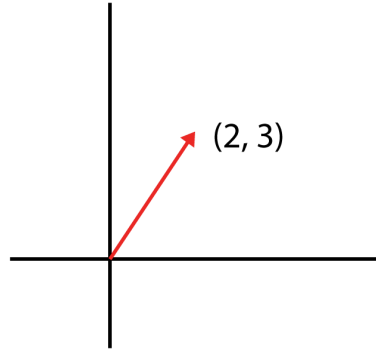
Komponen utama dalam proses temu balik informasi adalah kumpulan dokumen yang berisi informasi-informasi, query yang berisi informasi yang diinginkan oleh pengguna, dan proses pencarian informasi dengan algoritma tertentu untuk memperoleh dokumen yang berisi informasi yang sesuai dengan query dari pengguna.

Dalam penerapannya, temu balik informasi yang berbasis tulisan digunakan dalam mesin pencarian / *search engine* seperti Google, Bing, Yahoo, dll; pencarian judul / isi buku di suatu perpustakaan, dan pencarian judul film dalam web *streaming* seperti Youtube.

B. Vektor

I. Definisi Vektor

Vektor merupakan representasi kuantitas fisik yang memiliki besar dan arah. Secara geometri, vektor dapat dilambangkan sebagai garis berarah. Sebagai contoh, vektor $\mathbf{v} = (2, 3)$ pada ruang dimensi dua memiliki gambaran seperti berikut:



Semua vektor yang ditulis sebagai $\mathbf{v} = (v_1, v_2, \dots, v_n)$ berasal dari titik asal $O(0, 0, \dots, 0)$.

II. Norm

Norm suatu vektor adalah jarak vektor tersebut dari titik asal O pada ruang Euclidean.

Jarak ini adalah panjang vektor tersebut. Panjang vektor $\mathbf{v} = (v_1, v_2, \dots, v_n)$ dapat dihitung menggunakan teorema Pythagoras sebagai berikut:

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

III. Perkalian Titik (Dot Product)

Jika \mathbf{u} dan \mathbf{v} adalah vektor tidak nol di \mathbb{R}^2 atau \mathbb{R}^3 , maka perkalian titik (*dot product*) \mathbf{u} dan \mathbf{v} adalah

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos(\theta)$$

dimana θ adalah sudut yang dibentuk oleh vektor \mathbf{u} dan \mathbf{v} .

Secara umum, jika $\mathbf{u} = (u_1, u_2, \dots, u_n)$ dan $\mathbf{v} = (v_1, v_2, \dots, v_n)$ adalah dua vektor di \mathbb{R}^n , maka perkalian titik \mathbf{u} dan \mathbf{v} adalah

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \dots + u_n v_n$$

C. Cosine Similarity

Cosine Similarity adalah ukuran kemiripan antara dua vektor bukan nol di dalam ruang vektor. Hal ini didefinisikan dari kosinus sudut di antara mereka, yang dihasilkan dari hasil *dot product* kedua vektor yang dibagi oleh kedua vektor tersebut yang telah ternormalisasi. Secara garis besar, kedua data/dokumen dapat dinyatakan *similar* apabila *cosine similarity* kedua vektor dokumen tersebut adalah 1, yang berarti bahwa sudut antara kedua vektor adalah 0 derajat. Sebaliknya kedua dokumen disebut sama sekali tidak *similar* (dalam hal ini *query* dinyatakan oleh dokumen juga) apabila nilai dari *cosine similarity* kedua vektor dokumen tersebut adalah 0, yang mencerminkan bahwa sudut antara kedua vektor adalah 90 derajat (tegak lurus atau ortogonal).

Kesamaan antara dua vektor $\mathbf{Q} = (q_1, q_2, q_3, \dots, q_n)$ dan vektor $\mathbf{D} = (d_1, d_2, d_3, \dots, d_n)$ diukur dengan rumus *cosinus similarity* yang merupakan bagian dari rumus perkalian titik (*dot product*) dua buah vektor dengan rumus berikut:

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

Diketahui bahwa $\mathbf{Q} \cdot \mathbf{D}$ adalah perkalian titik yang didefinisikan sebagai $q_1d_1 + q_2d_2 + q_3d_3 + \dots + q_nd_n$.

Setiap dokumen di dalam koleksi dokumen dihitung kesamaannya dengan *query* dengan rumus cosinus di atas. Selanjutnya hasil perhitungan di-ranking berdasarkan nilai cosinus dari besar ke kecil sebagai proses pemilihan dokumen yang ‘dekat’ dengan *query*. Nilai cosinus yang besar menyatakan dokumen yang relevan dan nilai cosinus yang kecil menyatakan dokumen yang kurang relevan.

BAB III

IMPLEMENTASI PROGRAM

Pada tugas besar ini, kami mendeklarasikan 2 buah modul dan satu main program, yaitu :

1. `vector_processing.py`

Function	Deskripsi
<code>def dot (u, v)</code>	Mengembalikan nilai perkalian titik vektor u dan vektor v
<code>def norm (v)</code>	Mengembalikan nilai norm vektor v, dihitung sebagai akar dari dot product v dengan v

2. `search_document.py`

`search_document.py` menggunakan `StemmerFactory` dari library `Sastrawi`, `pathlib`, dan `vector_processing.py`.

Function	Deskripsi
<code>def first_sentences (flist, search_size)</code>	Mengembalikan list berisi kalimat pertama dari file berformat .txt dalam flist, sebanyak search_size file
<code>def stemDoc (filename)</code>	Mengembalikan list berisi hasil split dari hasil stemming file berformat .txt dengan path filename
<code>def stemQuery (qString)</code>	Mengembalikan list berisi hasil split dari hasil stemming string qString untuk stemming query
<code>def makeTabTerm (stemDoc)</code>	Mengembalikan list berisi term yang ada dalam stemDoc yang berisi kata-kata hasil stemming dokumen
<code>def makeMasterTabTerm (stemDocs)</code>	Mengembalikan list berisi semua term yang muncul dalam stemDocs, dimana stemDocs adalah list berisi hasil stemming dokumen-dokumen

def extendMaster (masterCount, masterTerm, stemQuery, search_size)	Mengembalikan tuple list (masterCount, masterTerm), dimana term dalam stemQuery yang belum ada dalam masterTerm telah ditambahkan ke masterTerm, dan masterCount telah diperpanjang sebanyak jumlah term baru
def updateMaster (setupData)	Mengembalikan tuple (flist, search_size, titles, openings, docs, masterTerm, lenMaster, masterCount), dimana flist adalah list daftar dokumen-dokumen yang telah di- <i>update</i> , search_size adalah jumlah semua dokumen, titles adalah list judul semua dokumen yang telah di- <i>update</i> , openings adalah list kalimat pertama semua dokumen yang telah di- <i>update</i> , docs adalah list hasil stemming semua dokumen yang telah di- <i>update</i> , masterTerm adalah list semua term yang muncul dalam dokumen-dokumen, lenMaster adalah panjang list masterTerm, masterCount adalah jumlah kemunculan setiap term dalam setiap dokumen. Input setupData juga berbentuk tuple (flist, search_size, titles, openings, docs, masterTerm, lenMaster, masterCount)
def setup ()	Mengembalikan tuple (flist, search_size, titles, openings, docs, masterTerm, lenMaster, masterCount) dimana flist adalah list daftar semua dokumen, search_size adalah jumlah semua dokumen, titles adalah list judul semua dokumen, openings adalah list kalimat pertama semua dokumen, docs adalah list hasil stemming semua dokumen, masterTerm adalah list semua term yang muncul dalam dokumen-dokumen, lenMaster adalah panjang list masterTerm, masterCount adalah jumlah kemunculan setiap term dalam setiap dokumen.
def search (query, setupData)	Mengembalikan tuple (q, workingTerm, sim, indices), dimana q adalah hasil stemming query, workingTerm adalah vektor-vektor kemunculan term-term dalam q dalam setiap dokumen, sim adalah list nilai <i>cosine similarity</i> antara setiap dokumen dan query

	yang telah disortir, dan indices adalah list index dokumen yang berkoresponden dengan sim. Input setupData berbentuk tuple (flist, search_size, titles, openings, docs, masterTerm, lenMaster, masterCount)
--	---

3. webflask.py

webflask.py menggunakan library os, flask, werkzeug.utils, dan search_document.py.

Function	Deskripsi
def mainPage ()	Mengembalikan antarmuka main page website sesuai dengan template dalam file “mainpage.html”
def perihal ()	Mengembalikan antarmuka page perihal pada website sesuai dengan template dalam file “perihal.html”
def uploadPage ()	Mengembalikan antarmuka page untuk upload file sesuai dengan template dalam “file_upload_form.html”
def success ()	Mengembalikan antarmuka page upload setelah berhasil meng-upload suatu file, fungsi ini juga berisi algoritma untuk upload dan save file tersebut
def displayPage (filename)	Mengembalikan antarmuka page yang berisi konten dari dokumen yang dibuka oleh pengguna, dalam fungsi ini juga terdapat algoritma untuk membuka suatu file, membacanya, dan menuliskannya ke dalam format html
def searchPage ()	Mengembalikan antarmuka page hasil pencarian yang berisi judul-judul dokumen yang ada (dengan link menuju page berisi konten dokumen tersebut) terurut berdasarkan tingkat kemiripan dengan query yang dimasukkan oleh pengguna

Kami juga mengimplementasikan web scraping untuk mempersiapkan data. Web scraping menggunakan library wikipedia, dan numpy:

4. wikipedia_scraping.py

Function	Deskripsi
search(query, search_size=1)	Mengambil <i>summary</i> dari wikipedia hasil search query, sejumlah search_size, dan disimpan dalam folder test dengan format <title>.txt

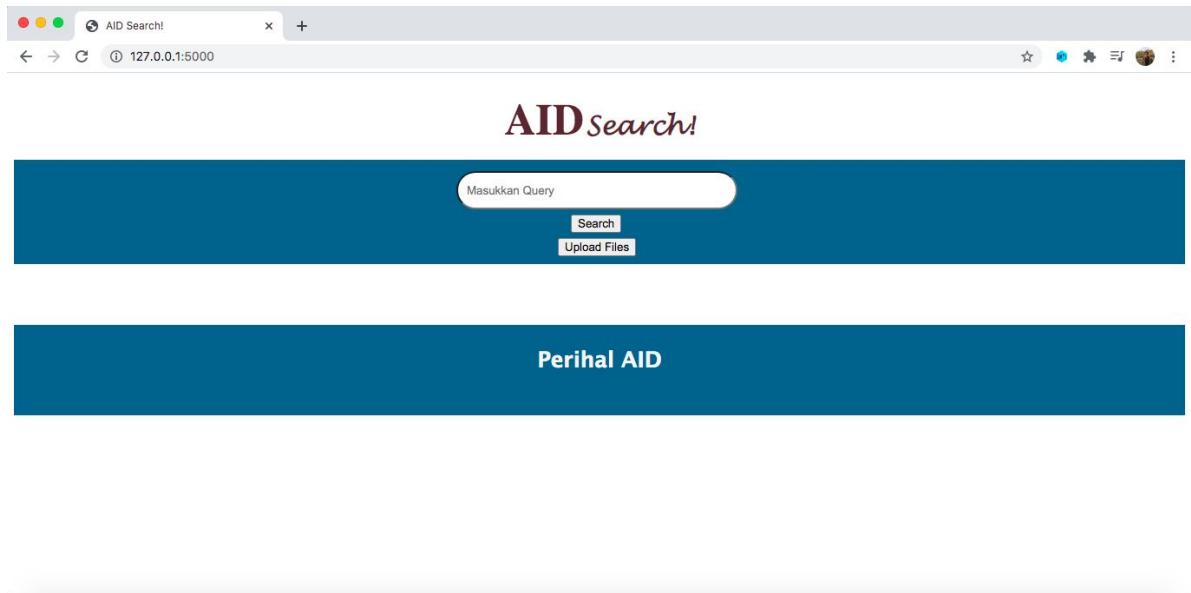
Untuk data uji, digunakan query = “Indonesia” dan search_size = 15.

BAB IV

EKSPERIMEN

1. Main Page

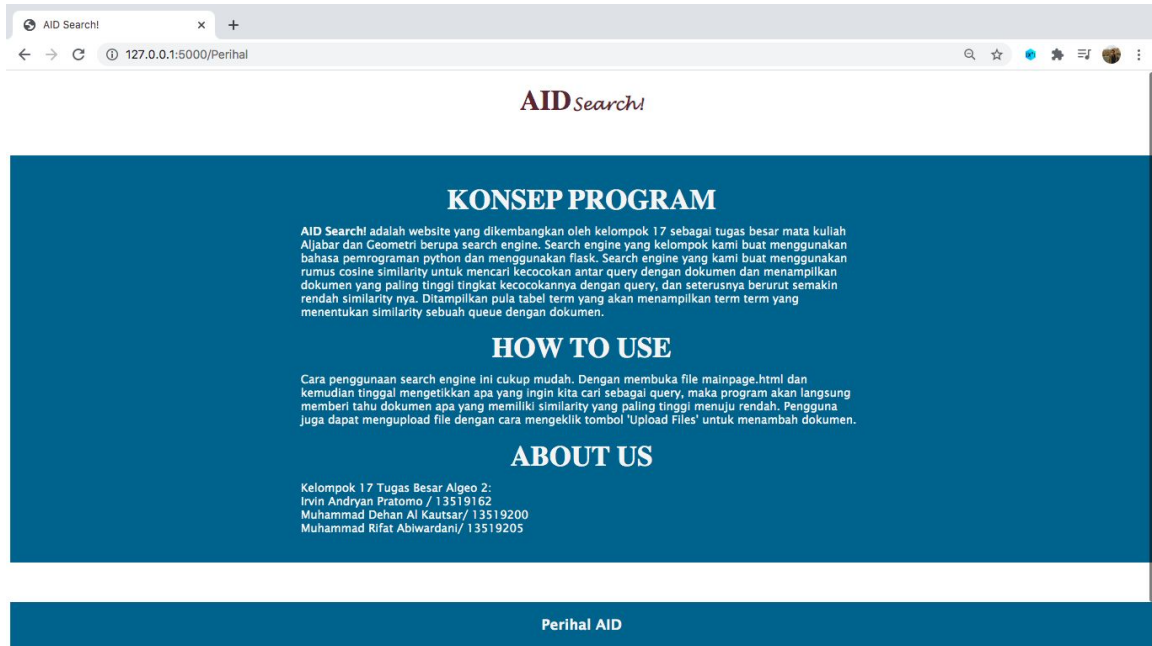
Berikut ini adalah tampilan dari main page *search engine* yang kami buat.



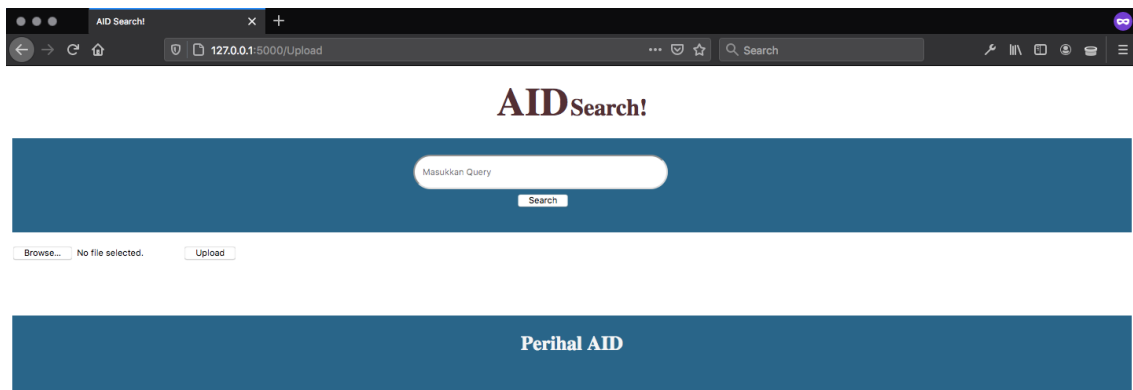
Pada main page terdapat fitur untuk melakukan pencarian dengan memasukkan query yang diinginkan di dalam *search box* lalu menekan enter atau *klik* tombol “search”. Selain itu terdapat fitur untuk upload files yang akan membawa pengguna ke laman khusus untuk melakukan upload file, dan fitur “Perihal AID” yang membawa pengguna ke laman untuk melihat perihal web.

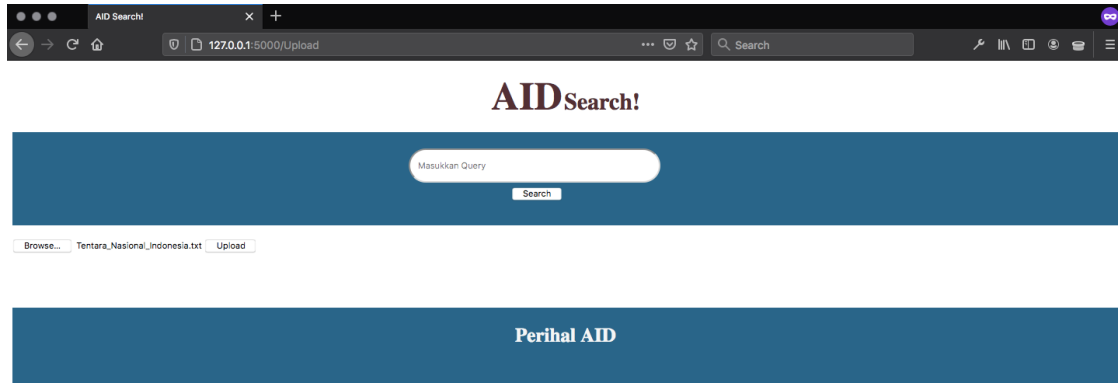
2. Perihal

Laman “Perihal” berisi informasi mengenai konsep program secara garis besar, cara penggunaan program, dan nomor kelompok, nama, serta NIM pembuat program. Berikut ini adalah tampilannya.



3. Upload File





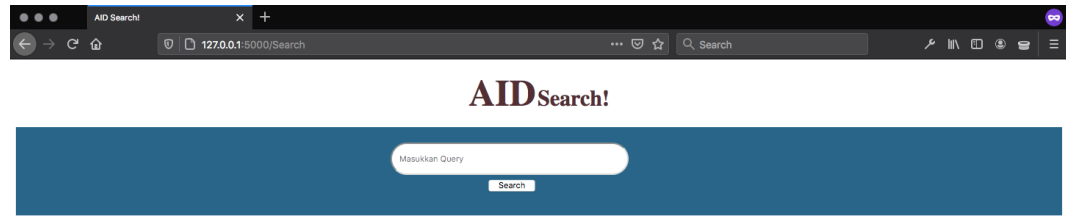
Fitur upload bekerja sesuai spesifikasi, dengan file “Tentara_Nasional_Indonesia.txt” disimpan dalam folder test.

4. Search

Laman “Search” menampilkan daftar dokumen yang dilakukan pencarian, hasil pencarian yang terurut, dimana setiap blok berisi judul dokumen, jumlah kata dokumen, tingkat kemiripan, dan kalimat pertama.

4.1. Search Query 1

Berikut hasil pencarian terhadap query “daerah pada bagian timur pulau”:



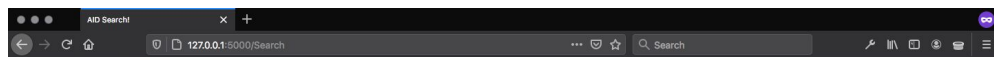
Daftar Dokumen

1. Bahasa_Indonesia.txt
2. Bali.txt
3. Daftar_kabupaten_dan_kota_di_Indonesia.txt
4. Gerakan_Pramuka_Indonesia.txt
5. Indonesia.txt
6. Kalimantan.txt
7. Kecamatan.txt
8. Kepolisian_Negara_Republik_Indonesia.txt
9. Kota_Bandung.txt
10. Kota_Surabaya.txt
11. Kota_Surakarta.txt
12. Orde_Baru.txt
13. Papua.txt
14. Provinsi_di_Indonesia.txt
15. Tentara_Nasional_Indonesia.txt

Hasil Pencarian:

(diurutkan dari tingkat kemiripan tertinggi)

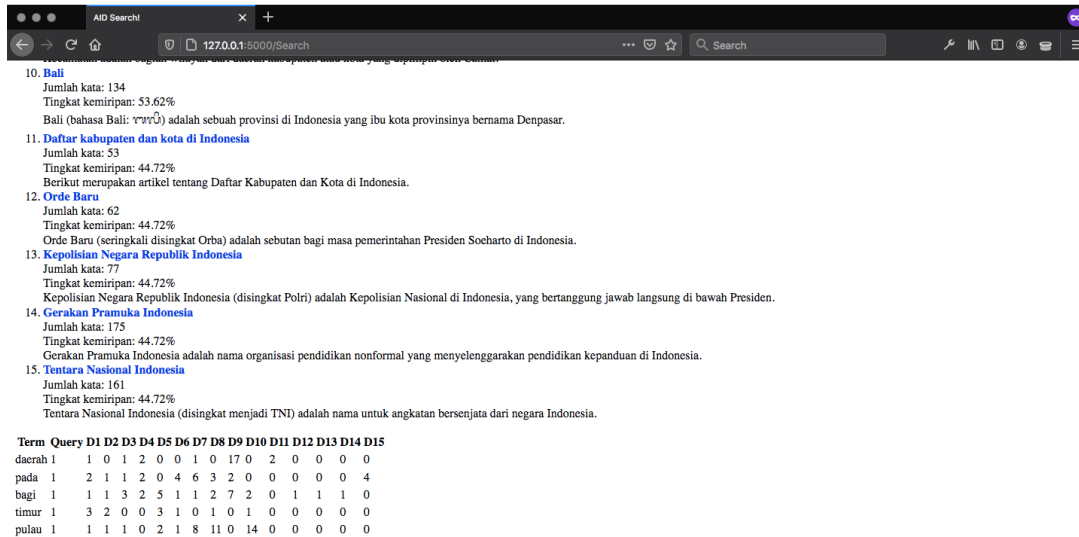
1. [Kota Surabaya](#)
Jumlah kata: 204
Tingkat kemiripan: 89.44%
Kota Surabaya (bahasa Jawa: Hanacaraka: ꦏꦠꦱꦸꦫꦧꦫ, Pegon: سورابايه, Madura: Kottah Sorbhājāh) adalah ibu kota Provinsi Jawa Timur, Indonesia, sekaligus kota metropolitan terbesar di provinsi tersebut.
2. [Kota Surakarta](#)
Jumlah kata: 116
Tingkat kemiripan: 84.51%



Hasil Pencarian:

(diurutkan dari tingkat kemiripan tertinggi)

1. [Kota Surabaya](#)
Jumlah kata: 204
Tingkat kemiripan: 89.44%
Kota Surabaya (bahasa Jawa: Hanacaraka: ꦏꦠꦱꦸꦫꦧꦫ, Pegon: سورابايه, Madura: Kottah Sorbhājāh) adalah ibu kota Provinsi Jawa Timur, Indonesia, sekaligus kota metropolitan terbesar di provinsi tersebut.
2. [Kota Surakarta](#)
Jumlah kata: 116
Tingkat kemiripan: 84.51%
Kota Surakarta (Hanacaraka: ꦱꦸꦫꦏꦂꦠ, Jawa: Kutha Surakarta), juga disebut Solo (ꦱꦺꦴ), adalah wilayah otonom dengan status Kota di bawah Provinsi Jawa Tengah, Indonesia, dengan penduduk 519.
3. [Bahasa Indonesia](#)
Jumlah kata: 293
Tingkat kemiripan: 77.45%
Bahasa Indonesia adalah bahasa resmi Republik Indonesia dan bahasa persatuan bangsa Indonesia.
4. [Provinsi di Indonesia](#)
Jumlah kata: 67
Tingkat kemiripan: 77.45%
Dalam pembagian administratif, Indonesia terdiri atas provinsi, yang dikepalai oleh seorang gubernur.
5. [Papua](#)
Jumlah kata: 75
Tingkat kemiripan: 72.54%
Papua adalah provinsi yang terletak di bagian tengah Pulau Papua atau bagian paling timur wilayah Papua milik Indonesia.
6. [Kota Bandung](#)
Jumlah kata: 268
Tingkat kemiripan: 71.81%
Kota Bandung (Aksara Sunda Baku: ᮊᮧᮒ ᮘᮊᮒ) adalah kota metropolitan terbesar di Provinsi Jawa Barat, sekaligus menjadi ibu kota provinsi tersebut.
7. [Indonesia](#)
Jumlah kata: 498
Tingkat kemiripan: 70.84%
Indonesia disebut juga dengan Republik Indonesia (RI) atau Negara Kesatuan Republik Indonesia (NKRI), adalah negara di Asia Tenggara yang dilintasi garis khatulistiwa dan berada di antara daratan benua Asia dan Australia, serta antara Samudra Pasifik dan Samudra Hindia.
8. [Kalimantan](#)
Jumlah kata: 154
Tingkat kemiripan: 65.43%
Kalimantan (toponim: Kalamantan, Calémantan, Kalemantan, Kelamantan, Kilamantan, Klamantan, Klémantan, K'lemantan, Quallamontan) atau juga disebut Borneo pada zaman kolonial, adalah pulau terbesar ketiga di dunia yang terletak di sebelah utara Pulau Jawa dan di sebelah barat Pulau Sulawesi.
9. [Kecamatan](#)
Jumlah kata: 746



10. [Bali](#)
 Jumlah kata: 134
 Tingkat kemiripan: 53.62%
 Bali (bahasa Bali: ᬩᬲᬩᬶ) adalah sebuah provinsi di Indonesia yang ibu kota provinsinya bernama Denpasar.

11. [Daftar kabupaten dan kota di Indonesia](#)
 Jumlah kata: 53
 Tingkat kemiripan: 44.72%
 Berikut merupakan artikel tentang Daftar Kabupaten dan Kota di Indonesia.

12. [Orde Baru](#)
 Jumlah kata: 62
 Tingkat kemiripan: 44.72%
 Orde Baru (seringkali disingkat Orba) adalah sebutan bagi masa pemerintahan Presiden Soeharto di Indonesia.

13. [Kepolisian Negara Republik Indonesia](#)
 Jumlah kata: 77
 Tingkat kemiripan: 44.72%
 Kepolisian Negara Republik Indonesia (disingkat Polri) adalah Kepolisian Nasional di Indonesia, yang bertanggung jawab langsung di bawah Presiden.

14. [Gerakan Pramuka Indonesia](#)
 Jumlah kata: 175
 Tingkat kemiripan: 44.72%
 Gerakan Pramuka Indonesia adalah nama organisasi pendidikan nonformal yang menyelenggarakan pendidikan kepanduan di Indonesia.

15. [Tentara Nasional Indonesia](#)
 Jumlah kata: 161
 Tingkat kemiripan: 44.72%
 Tentara Nasional Indonesia (disingkat menjadi TNI) adalah nama untuk angkatan bersenjata dari negara Indonesia.

Term	Query	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15
daerah	1	0	1	2	0	0	1	0	17	0	2	0	0	0	0	
pada	1	2	1	1	2	0	4	6	3	2	0	0	0	0	4	
bagi	1	1	1	3	2	5	1	1	2	7	2	0	1	1	1	0
timur	1	3	2	0	0	3	1	0	1	0	1	0	0	0	0	0
pulau	1	1	1	1	0	2	1	8	11	0	14	0	0	0	0	0

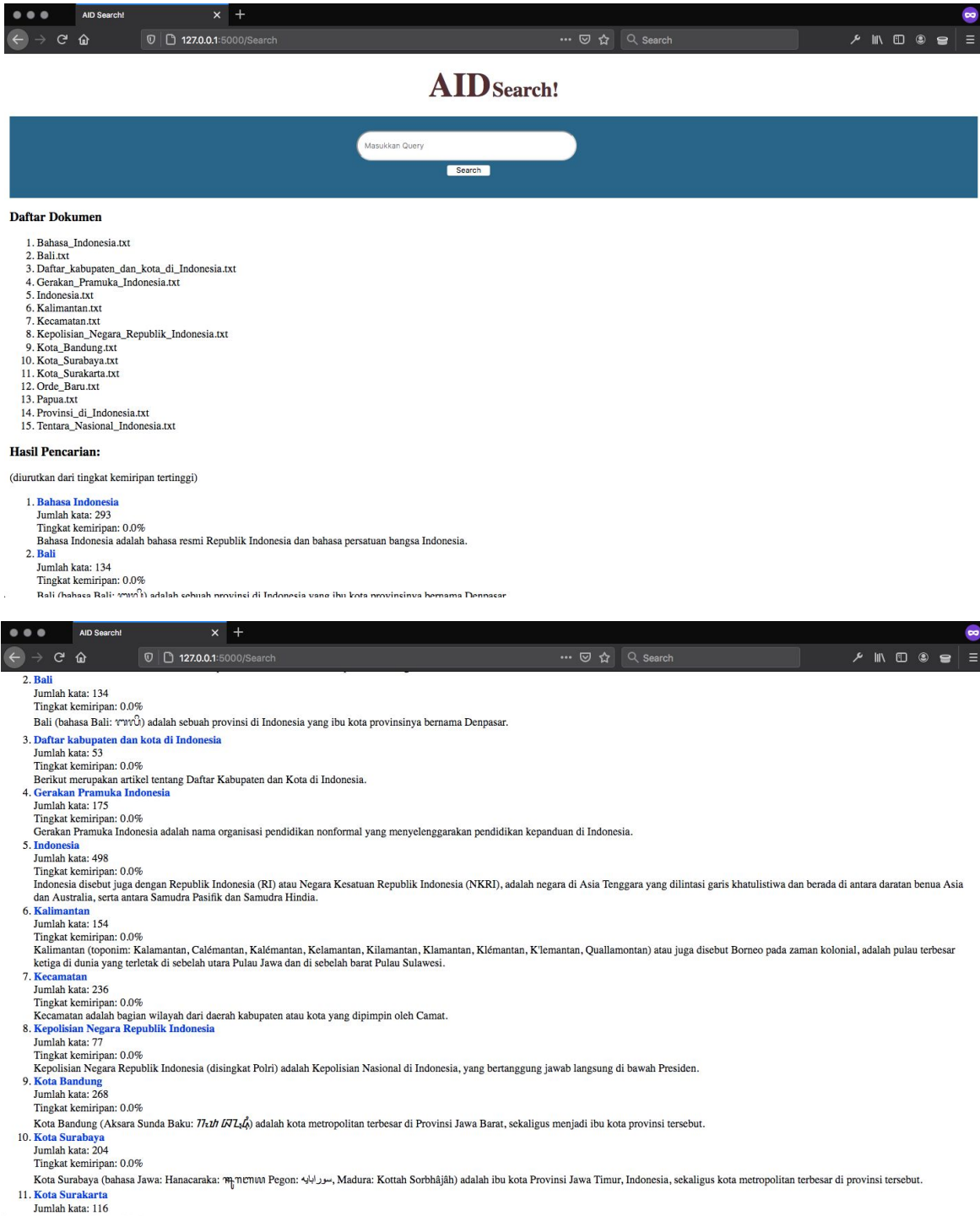
Perihal AID

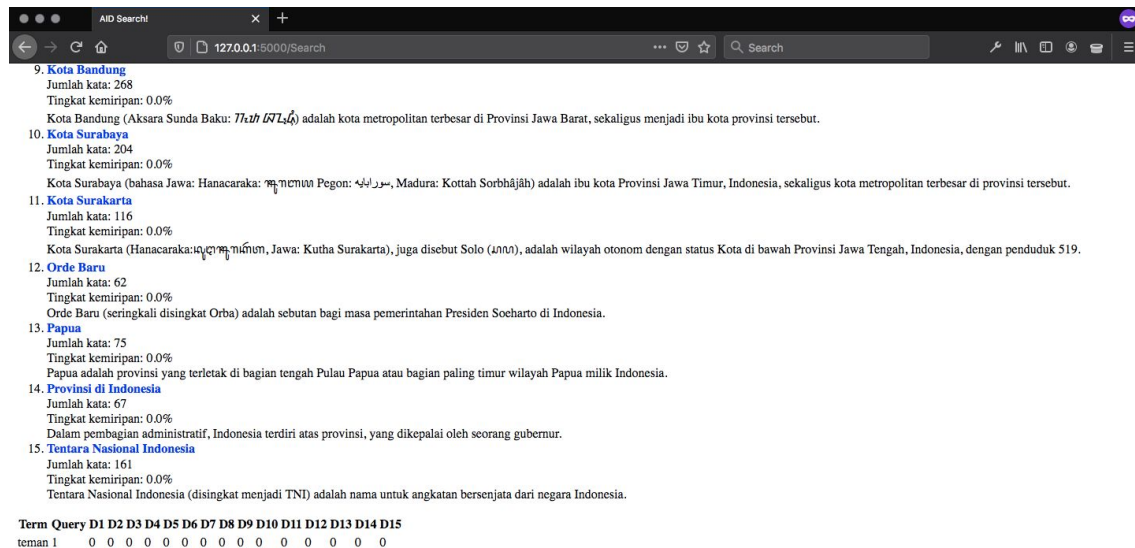
Hasil search adalah sesuai ekspektasi, fungsionalitas *sorting* bekerja dengan baik, dan dapat dikonfirmasi secara visual bahwa hasil *cosine similarity* vektor Query dengan vektor D1 hingga D10 terurut mengecil.

Informasi dokumen ditampilkan sesuai spesifikasi (judul yang merupakan link kepada file .txt, jumlah kata dokumen, tingkat kemiripan dokumen, dan kalimat pertama).

4.2. Search Query 2

Berikut hasil pencarian terhadap query “teman”:





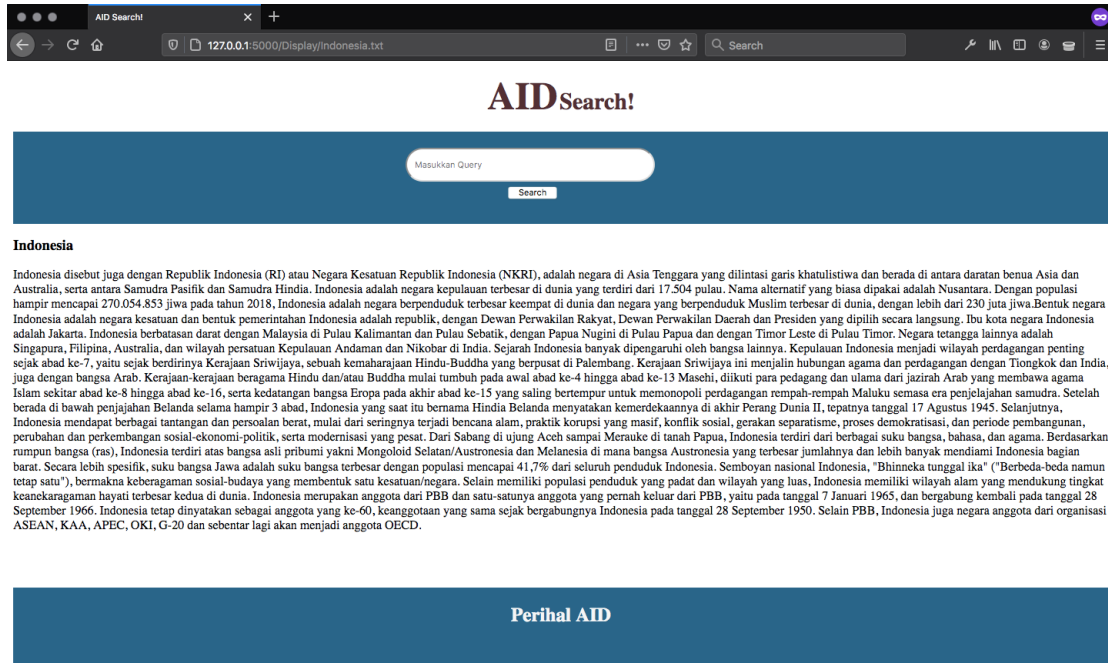
Perihal AID

Hasil search adalah sesuai ekspektasi, tidak ada kata (baik akar maupun padanan) dari “teman” dalam dokumen-dokumen tentang Indonesia. Hasil *cosine similarity* juga sesuai, yaitu 0.

Informasi dokumen ditampilkan sesuai spesifikasi (judul yang merupakan link kepada file .txt, jumlah kata dokumen, tingkat kemiripan dokumen, dan kalimat pertama).

5. Display File

Berikut hasil display file “Indonesia.txt”:



AID Search!

Masukkan Query

Search

Indonesia

Indonesia disebut juga dengan Republik Indonesia (RI) atau Negara Kesatuan Republik Indonesia (NKRI), adalah negara di Asia Tenggara yang dilintasi garis khatulistiwa dan berada di antara daratan benua Asia dan Australia, serta antara Samudra Pasifik dan Samudra Hindia. Indonesia adalah negara kepulauan terbesar di dunia yang terdiri dari 17.504 pulau. Nama alternatif yang biasa dipakai adalah Nusantara. Dengan populasi hampir mencapai 270.054.853 jiwa pada tahun 2018, Indonesia adalah negara berpenduduk terbesar keempat di dunia dan negara yang berpenduduk Muslim terbesar di dunia, dengan lebih dari 230 juta jiwa. Bentuk negara Indonesia adalah negara kesatuan dan bentuk pemerintahan Indonesia adalah republik, dengan Dewan Perwakilan Rakyat, Dewan Perwakilan Daerah dan Presiden yang dipilih secara langsung. Ibu kota negara Indonesia adalah Jakarta. Indonesia berbatasan darat dengan Malaysia di Pulau Kalimantan dan Pulau Sebatik, dengan Papua Nugini di Pulau Papua dan dengan Timor Leste di Pulau Timor. Negara tetangga lainnya adalah Singapura, Filipina, Australia, dan wilayah persatuan Kepulauan Andaman dan Nikobar di India. Sejarah Indonesia banyak dipengaruhi oleh bangsa lainnya. Kepulauan Indonesia menjadi wilayah perdagangan penting sejak abad ke-7, yaitu sejak berdirinya Kerajaan Sriwijaya, sebuah kemaharajaan Hindu-Buddha yang berpusat di Palembang. Kerajaan Sriwijaya ini menjalin hubungan agama dan perdagangan dengan Tiongkok dan India, juga dengan bangsa Arab. Kerajaan-kerajaan beragama Hindu dan/atau Buddha mulai tumbuh pada awal abad ke-4 hingga abad ke-13 Masehi, diikuti para pedagang dan ulama dari jazirah Arab yang membawa agama Islam sekitar abad ke-8 hingga abad ke-16, serta kedatangan bangsa Eropa pada akhir abad ke-15 yang saling bertempur untuk memonopoli perdagangan rempah-rempah Maluku semasa era penjelajahan samudra. Setelah berada di bawah penjajahan Belanda selama hampir 3 abad, Indonesia yang saat itu bernama Hindia Belanda menyatakan kemerdekaannya di akhir Perang Dunia II, tepatnya tanggal 17 Agustus 1945. Selanjutnya, Indonesia mendapat berbagai tantangan dan persoalan berat, mulai dari seringnya terjadi bencana alam, praktik korupsi yang masif, konflik sosial, gerakan separatisme, proses demokratisasi, dan periode pembangunan, perubahan dan perkembangan sosial-ekonomi-politik, serta modernisasi yang pesat. Dari Sabang di ujung Aceh sampai Merauke di tanah Papua, Indonesia terdiri dari berbagai suku bangsa, bahasa, dan agama. Berdasarkan rumpun bangsa (ras), Indonesia terdiri atas bangsa asli pribumi yakni Mongoloid Selatan/Austronesia dan Melanesia di mana bangsa Austronesia yang terbesar jumlahnya dan lebih banyak mendiami Indonesia bagian barat. Secara lebih spesifik, suku bangsa Jawa adalah suku bangsa terbesar dengan populasi mencapai 41,7% dari seluruh penduduk Indonesia. Semboyan nasional Indonesia, "Bhinneka tunggal ika" ("Berbeda-beda namun tetap satu"), bermakna keberagaman sosial-budaya yang membentuk satu kesatuan/negara. Selain memiliki populasi penduduk yang padat dan wilayah yang luas, Indonesia memiliki wilayah alam yang mendukung tingkat keanekaragaman hayati terbesar kedua di dunia. Indonesia merupakan anggota dari PBB dan satu-satunya anggota yang pernah keluar dari PBB, yaitu pada tanggal 7 Januari 1965, dan bergabung kembali pada tanggal 28 September 1966. Indonesia tetap dinyatakan sebagai anggota yang ke-60, keanggotaan yang sama sejak bergabungnya Indonesia pada tanggal 28 September 1950. Selain PBB, Indonesia juga negara anggota dari organisasi ASEAN, KAA, APEC, OKI, G-20 dan sebentar lagi akan menjadi anggota OECD.

Perihal AID

BAB V

KESIMPULAN, SARAN, DAN REFLEKSI

Kesimpulan

1. Kami telah membuat sebuah program *search engine* yang mampu mencari dokumen di dalam *database*, yang paling sesuai dengan *query* yang dimasukkan oleh pengguna, dan menampilkan hasilnya terurut berdasarkan tingkat similaritasnya.
2. Kami telah membuat sebuah program yang mampu mencari tingkat similaritas suatu dokumen dengan *query* yang dimasukkan oleh pengguna menggunakan *cosine similarity*.

Saran

Kedepannya *search engine* dapat dibuat dengan lebih baik apabila diberikan waktu pengerjaan yang lebih lama. Lebih baik dalam arti dengan membuat *user interface* yang lebih indah, sehingga *search engine* ini lebih ramah pengguna. Penulis juga menyarankan untuk menggunakan *pandas library* dalam membuat *masterTabTerm* dan *masterTabCount* yang berisi hasil stemming kata-kata dalam file, sehingga proses stemming dapat berjalan lebih cepat dan program dapat lebih cepat dalam melakukan inisialisasi.

Refleksi

Dalam mengerjakan tugas besar ini, terdapat beberapa kendala. Yang pertama, kurangnya pemahaman kami terhadap pemrograman berbasis python yang menggunakan Flask, sehingga dibutuhkan waktu yang lumayan lama untuk kami memahami bagaimana cara kerja dari Flask. Kedua, kami kesulitan dalam melakukan integrasi terhadap frontend dan backend program, dibutuhkan waktu yang cukup lama pula untuk mengintegrasikan hal tersebut.

Kesalahan tersebut murni datang dari kekurangan kami, sehingga kami akan lebih serius dan belajar lebih giat kembali agar hal-hal yang sebelumnya menjadi hambatan kami dapat teratasi dengan baik.

BAB VI

REFERENSI

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

<https://informatika.stei.itb.ac.id/~rinaldi.munir/AljabarGeometri/2020-2021/Algeo-10-Vektor-di-Ruang-Euclidean-Bag1.pdf>

<https://informatika.stei.itb.ac.id/~rinaldi.munir/AljabarGeometri/2020-2021/Algeo-11-Vektor-di-Ruang-Euclidean-Bag2.pdf>

<https://programminghistorian.org/en/lessons/creating-apis-with-python-and-flask>

<https://pypi.org/project/Sastrawi/>

<https://flask.palletsprojects.com/en/1.1.x/quickstart/>

<https://www.javatpoint.com/flask-file-uploading>

<https://stackoverflow.com/questions/60911565/re-rendering-with-render-template-not-working>

<https://medium.com/better-programming/building-your-first-website-with-flask-part-1-903a8b44e806>