

Laporan Tugas Besar 2 IF2123 Aljabar Linier dan Geometri

Aplikasi Dot Product pada Sistem Temu-Balik Informasi

Semester I Tahun 2020/2021



Disusun oleh:

Muhammad Fahmi Alamsyah 13519077

Habibina Arif Muzayyan 13519125

Delisha Azza Naadira 13519133

Kelompok 30

PROGRAM STUDI TEKNIK INFORMATIKA

SEKOLAH TEKNIK ELEKTRO DAN INFORMATIKA

INSTITUT TEKNOLOGI BANDUNG

2020

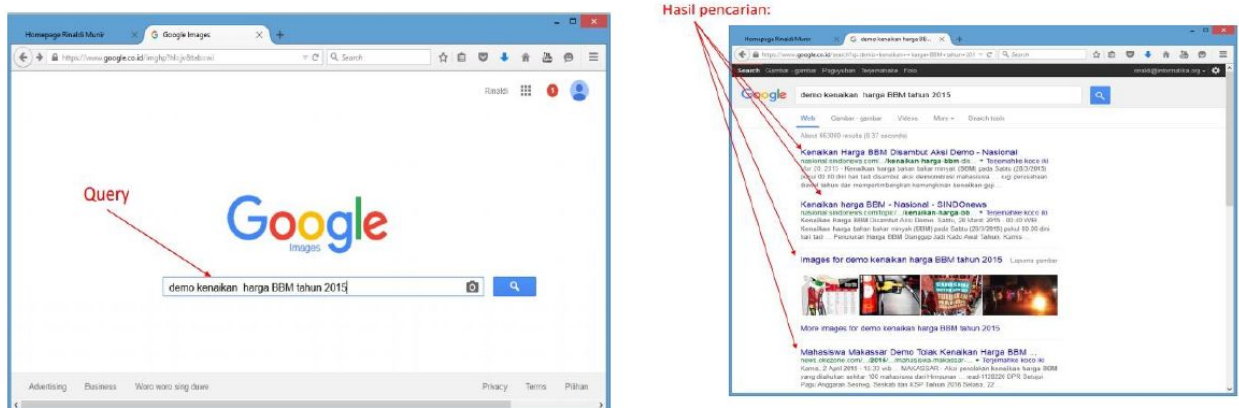
BAB I

DESKRIPSI MASALAH

1.1 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:

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

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

1.2 Spesifikasi Tugas

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

2.1 Sistem Temu Balik Informasi



Gambar ilustrasi Sistem Temu Balik Informasi

Sistem temu balik informasi (*information retrieval system*) adalah sistem yang digunakan untuk menemukan kembali informasi-informasi yang relevan terhadap kebutuhan pengguna dari suatu kumpulan informasi secara otomatis. Salah satu aplikasi umum dari sistem temu kembali informasi adalah *search engine* atau mesin pencarian yang terdapat pada jaringan internet. Pengguna dapat mencari halaman-halaman web yang dibutuhkannya melalui mesin tersebut.

Secara prinsip, penyimpanan informasi dan penemuan kembali informasi adalah hal yang sederhana. Misalkan terdapat tempat penyimpanan dokumen-dokumen dan seseorang (user) merumuskan suatu pertanyaan (*request* atau *query*) yang jawabannya adalah himpunan dokumen yang mengandung informasi yang diperlukan yang diekspresikan melalui pertanyaan user. User bisa saja memperoleh dokumen-dokumen yang diperlukannya dengan membaca semua dokumen dalam tempat penyimpanan, menyimpan dokumen-dokumen yang relevan dan membuang dokumen lainnya.

Tujuan dari sistem *information retrieval* yaitu untuk memenuhi kebutuhan informasi pengguna dengan me-*retrieve* semua dokumen yang mungkin relevan, tetapi pada waktu yang sama me-*retrieve* sesedikit mungkin dokumen yang tidak relevan. Sistem *information retrieval* yang baik memungkinkan pengguna menentukan secara cepat dan akurat apakah isi

dari dokumen yang diterima memenuhi kebutuhannya, karena di dalam *information retrieval* mendapatkan dokumen yang relevan tidaklah cukup.

2.2 Vektor

2.2.1 Definisi

Vector Space Model merupakan model dalam *Information Retrieval System* yang menggambarkan masing-masing *query* dan dokumen sebagai sebuah vektor n-dimensi (Noor, 2011). Tiap dimensi di dalam vektor tersebut merupakan perwakilan kata dokumen atau *query* yang dijadikan acuan.

2.2.2 Model Ruang Vektor dalam *Information Retrieval*

Untuk merepresentasikan proses pencarian informasi dalam *Information Retrieval System*, model ruang vektor dibuat dengan mengambil kata-kata isi yang ada di dalam dokumen kemudian model ini yang akan menentukan tingkat kemiripan (*similarity*) antara dokumen dengan *query* dengan cara menggambarkan dokumen dan *query* masing-masing ke dalam bentuk vektor juga. Tiap kata yang ditemukan di dalam dokumen dan *query* diberi bobot dan disimpan sebagai salah satu unsur vektor (Noor, 2011).

Pada Model Ruang Vektor, setiap dokumen di dalam *database* dan *query* pengguna digambarkan dengan suatu vektor multi-dimensi (Cios, 2007). Dimensi digambarkan sesuai dengan jumlah *term* yang ada di dalam dokumen yang terlibat pada model ini (Noor, 2011):

1. Vocabulary merupakan kumpulan semua term unik yang tersisa dari dokumen setelah preprocessing dan mengandung t term index. Term-term ini membentuk suatu ruang vektor.
2. Setiap term i di dalam dokumen atau query j , diberikan suatu bobot bernilai real W_{ij} .
3. Dokumen dan query digambarkan sebagai vektor t dimensi $d_j = (W_1, W_2, \dots, W_{ij})$ dan terdapat n dokumen di dalam koleksi, yaitu $j = 1, 2, \dots, n$.

Koleksi dokumen digambarkan sebagai matriks *term-document* (atau matriks *term-frequency*) dalam model ruang vektor. Setiap sel dalam matriks diberikan bobot yang

sesuai dengan term yang ada di dalam dokumen yang telah ditentukan. Term tersebut dikatakan tidak berpengaruh dalam dokumen jika bernilai nol.

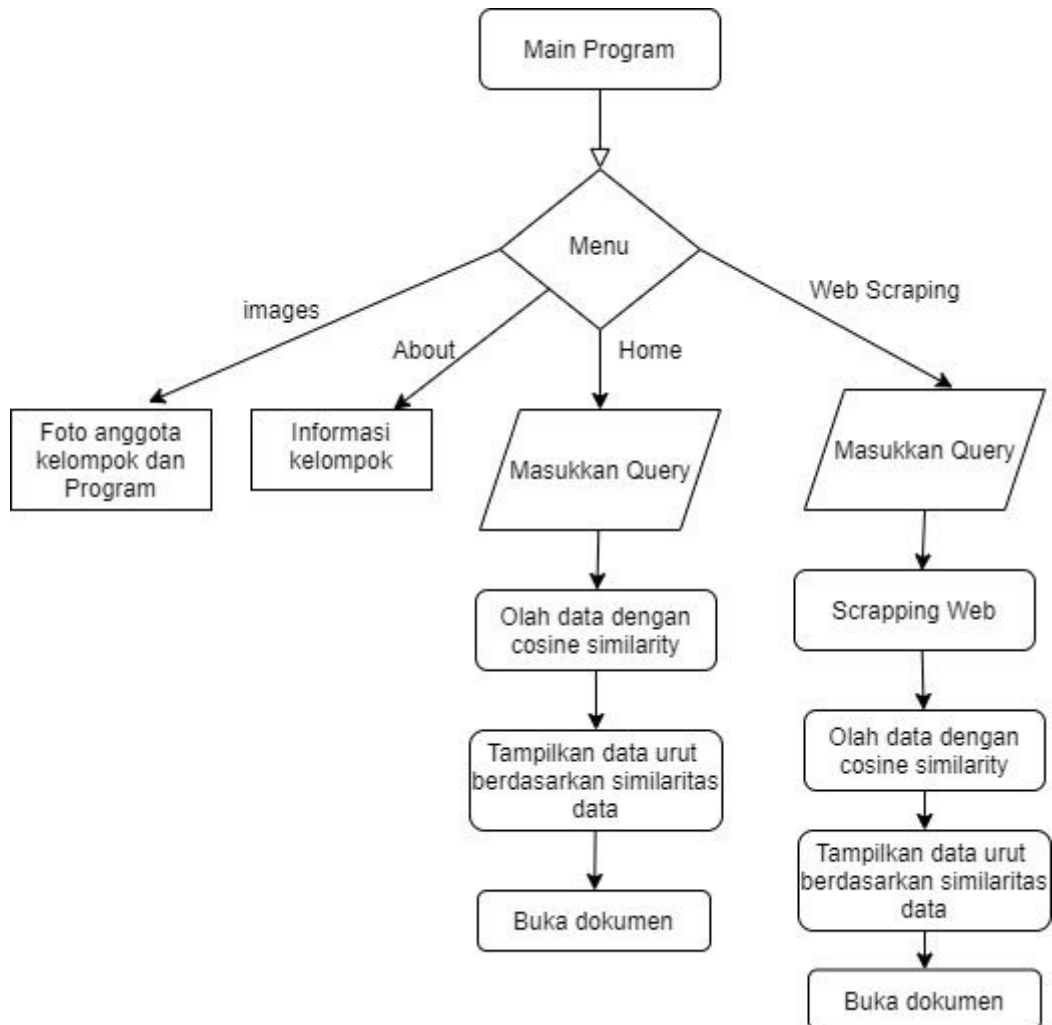
2.3 Cosine Similarity

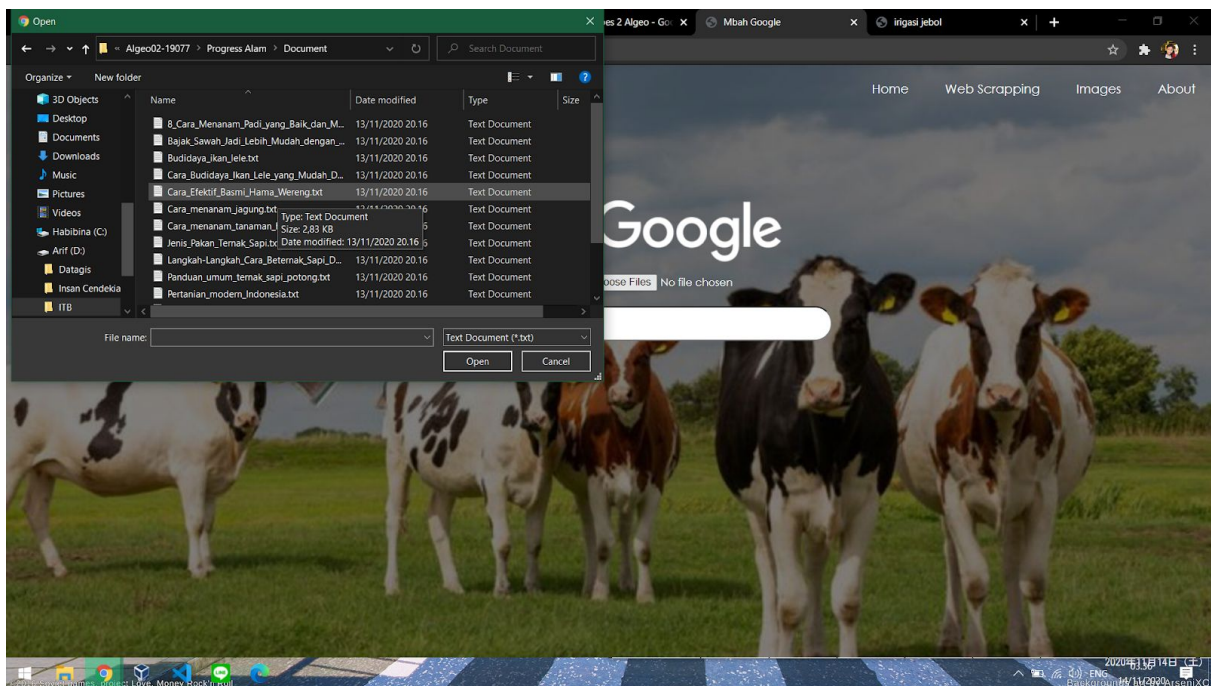
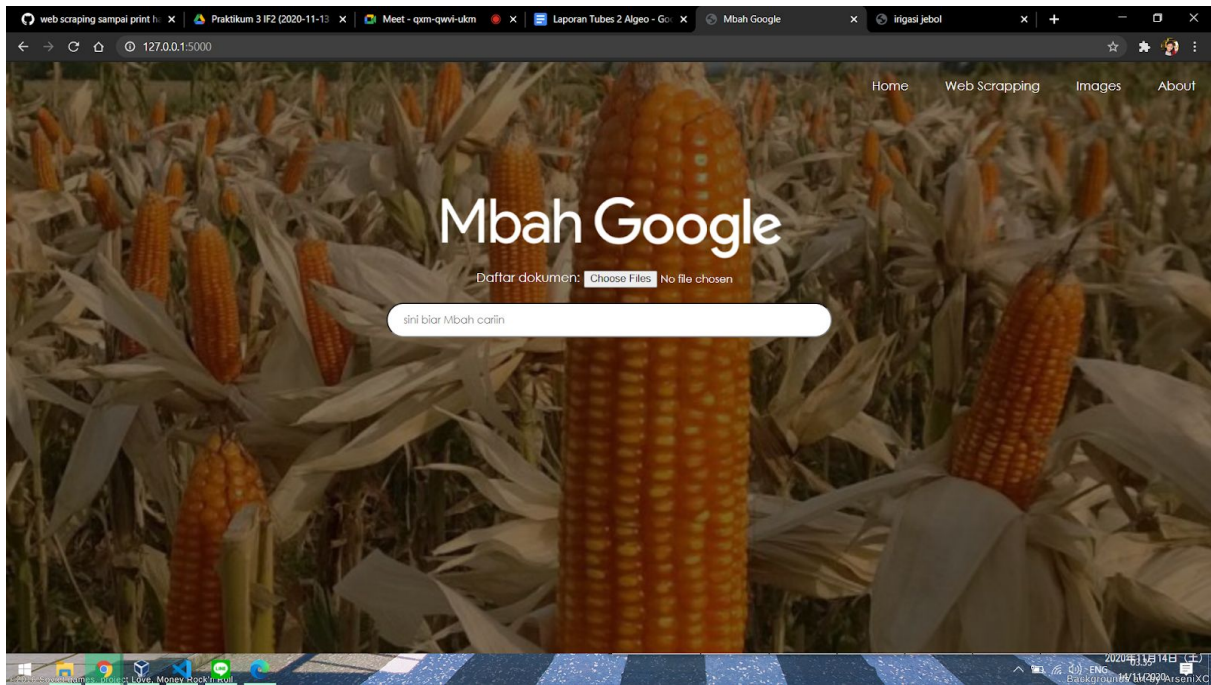
Kemiripan antara dokumen yang satu dengan yang lainnya dihitung menggunakan suatu konsep perkalian dot product 2 vektor yaitu cosine similarity. Fungsi ini yang akan dijadikan perangkingan dokumen sesuai dengan kemiripan yang berhubungan dengan query. Setelah dokumen dirangking, maka bisa dipastikan bahwa dokumen dengan ranking tertinggi memiliki kemiripan dengan query yang dimasukkan oleh pengguna.

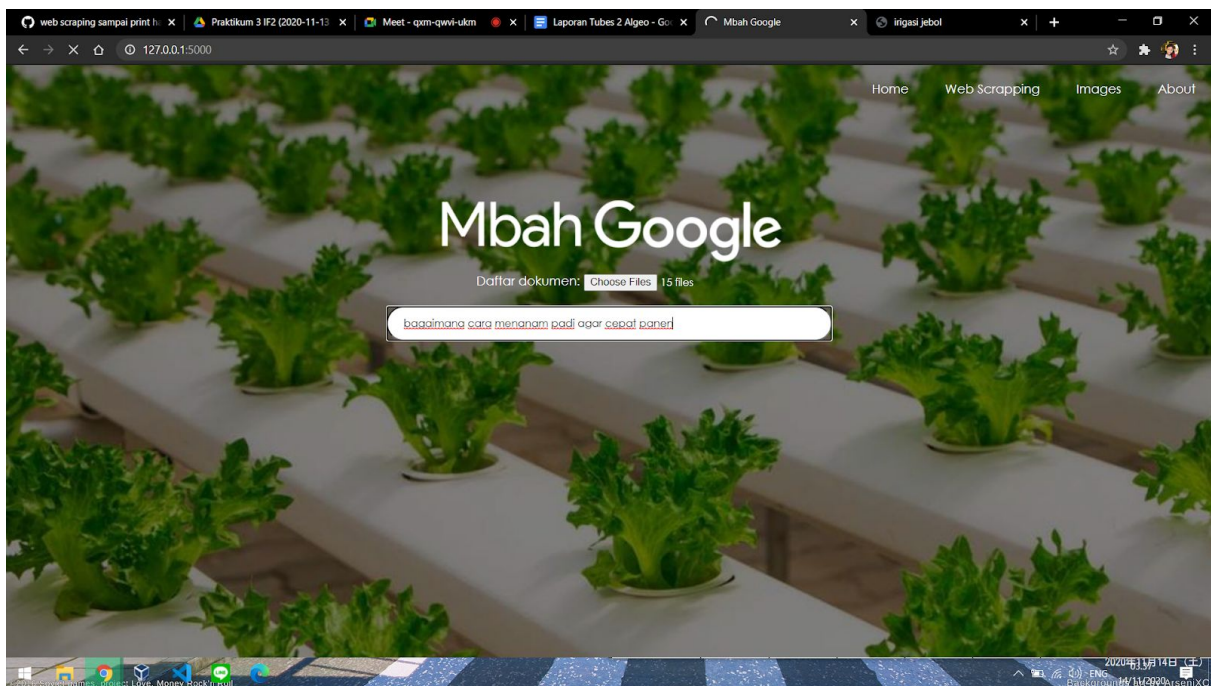
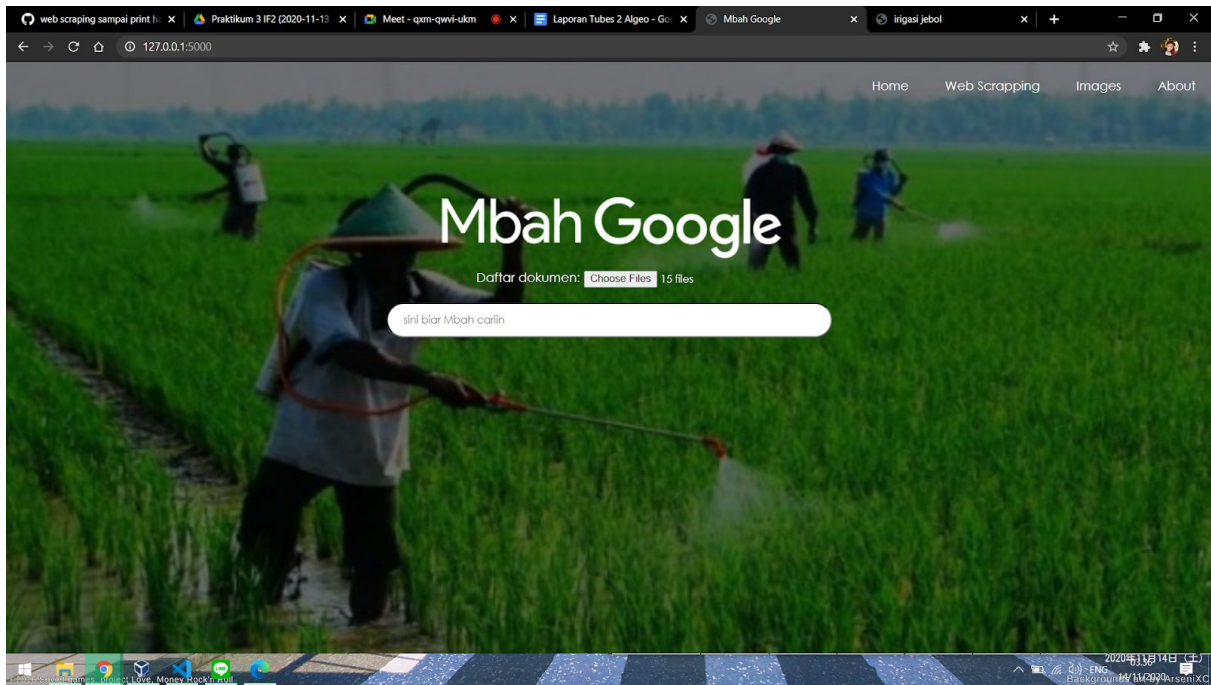
$$sim(Q, D) = \cos \theta = \frac{Q \cdot D}{\|Q\| \|D\|}$$

BAB III

IMPLEMENTASI PROGRAM







web scraping sampai print h x Praktikum 3 IF2 (2020-11-13) x Meet - qm-qm-ukm x Laporan Tubes 2 Algeo - Go x bagaimana cara menanam p x Perihal x + - x

127.0.0.1:5000/result/bagaimana%20cara%20menanam%20padi%20agar%20cepat%20panen#

Mbah Google bagaimana cara menanam padi agar cepat panen Home Web Scapping Images About

8_Cara_Menanam_Padi_yang_Baik_dan_Menguntungkan_Mudah_dan_Praktis.txt
Jumlah kata: 819
Tingkat kemiripan: 52.10%
Lipulan6.com, Jakarta - Cara menanam padi yang baik dan menguntungkan tentunya sangat penting diterapkan.

Cara menanam tanaman hidroponik sederhana.txt
Jumlah kata: 487
Tingkat kemiripan: 34.21%
Salah satu cara untuk mendapatkan sayuran segar tanpa pestisida adalah dengan menanam sendiri sayuran.

Cara menanam jagung.txt
Jumlah kata: 703
Tingkat kemiripan: 28.71%
Cara menanam jagung dengan metode tanpa olah tanah Ada berbagai macam cara menanam jagung salah satunya dengan menerapkan metode tanpa olah tanah (T0T).

Pelani bermigrasi ke pertanian modern.txt
Jumlah kata: 634
Tingkat kemiripan: 28.23%
Tahun 2020 menjadi catatan penting bagi ratusan petani desa Gebang kecamatan Plandaan Kabupaten Jombang Jawa Timur.

Cara Efektif Basmi Hama Wereng.txt
Jumlah kata: 413
Tingkat kemiripan: 26.94%
Hama wereng adalah hama yang paling sering menyerang tanaman padi.

Term	Query	8_Cara_Menanam_Padi_yang_Baik_dan_Menguntungkan_M
bagaimana	1	0
cara	1	13
tanam	1	23
padi	1	30
cepat	1	0
panen	1	11

2020年11月14日 (七) Baskiroun14/11/2020ArseniXC

web scraping sampai print h x Praktikum 3 IF2 (2020-11-13) x Meet - qm-qm-ukm x Laporan Tubes 2 Algeo - Go x Mbah Google x irigasi jebol x + - x

127.0.0.1:5000/webscrapping

Mbah Google Web Scapping Home Web Scapping Images About

sini biar Mbah carin

2020年11月14日 (七) Baskiroun14/11/2020ArseniXC


```
Python 3.9.7 Shell
D:\ITB\Semester 3\Algeo\Algeo02-19077\Progress Delisha\python webscrap.py
[https://news.detik.com/berita-jawa-barat/d-5252728/petani-cianjur-sulit-dapat-pupuk-subsidi-mentan-nanti-dikek], [https://news.detik.com/berita-jawa-barat/d-5244718/mengenal-tumbuhan-monokotil-dari-ciri-ciri-dan-contohnya], [https://news.detik.com/berita-jawa-barat/d-5244673/tentang-sumber-daya-alam-yang-dapat-diperbarui-dan-contohnya], [https://news.detik.com/berita-jawa-timur/d-5240193/lapas-banyuwangi-panen-roya-melon-dan-semangka], [https://news.detik.com/berita-jawa-timur/d-5239445/banjir-di-3-kecamatan-ngawi-surut-jalur-u], [https://news.detik.com/berita-jawa-timur/d-5239914/banjir-ngawi-akibat-luapan-sungai-madiun-meluas-ke-3-kecamatan], [https://news.detik.com/berita-jawa-timur/d-5238507/353-hektare-lahan-semangka-dan-blewah-di-jombang-gagal-panen-gegara-hujan], [https://news.detik.com/berita-jawa-timur/d-5237576/kasihan-petani-semangka-dan-blewah-di-jombang-merana-gegara-hujan]
Masukkan query : pertanian di Indonesia
Ini DOCUMENT
[[{"Petani di Cianjur selatan, Jawa Barat kesulitan mendapat pupuk bersubsidi. Akibatnya mereka terpaksa membeli pupuk nonsubsidi dengan harga dua kali lebih mahal dari yang bersubsidi.", "Dedi Rohimat (33), petan asal Kecamatan Cidaun Cianjur mengatakan untuk musim tanam padi kali ini, stok pupuk bersubsidi mengalami kelangkaan. Padahal di awal masa tanam, petani membutuhkan pupuk agar tanamannya tumbuh subur.", "Sudah dua bulan susah pupuk bersubsidi. Tidak tahu kenapa. Tapi bukan hanya di Cidaun di kecamatan lain di selatan juga mengalami hal serupa," kata dia, Kamis (12/11/2020).", "Menurutnya distributor dan toko pupuk hanya memiliki stok pupuk nonsubsidi. Meski harganya di atas pupuk subsidi, petani terpaksa membelinya karena tanaman padi harus segera diberi pupuk.", "Mau tidak mau ya beli, meski harganya Rp 400 ribu per koi atau dua kali harga pupuk subsidi yang paling mahal hanya Rp 200 ribu," kata dia.", "Dia mengungkapkan setiap petani membutuhkan sekitar 2 bintal pupuk untuk lahan seluas setengah hektar. Sulitnya mendapatkan pupuk subsidi dan terpaksa petani membeli pupuk nonsubsidi sebab biaya tanam menjadi bengkok.", "Secara otomatis pendapatan petani dari panen nanti akan minus, karena modal produksinya sudah tinggi.", "tuturnya.", "Kepala Dinas Pertanian Perkebunan Tanaman Pangan dan Hortikultura Cianjur Mammad Nano, mengatakan langkanya pupuk di wilayah selatan disebabkan keterlambatan distribusi.", "Sebelumnya memang ada keterlambatan, karena proses di pusatnya yang terlambat juga. Kalau daerah kan hanya mengusulkan, proses lebih lanjutnya di pusat," kata dia.", "Namun, lanjut Nano, pihaknya sudah mengumpulkan para distributor agar segera mendistribusikan pupuk jika kuota dan stoknya sudah disetujui.", "Kami pastikan dalam waktu dekat sudah aman dan mulai distribusi," kata dia.", "Di sisi lain, Menteri Pertanian Syahrul Yasin Limpo mengaku akan mengecek informasi petani di Cianjur yang sulit mendapatkan pupuk lantaran stok yang belum kunjung ada dan terpaksa membeli pupuk dengan harga mahal.", "Manti saya cek benar tidaknya," ucap Syahrul singkat saat ditemui di Balai Penelitian Tanaman Hias Cipanas Cianjur, Kamis (12/11/2020).", "Menurut biayinya, tumbuhan dikelompokkan menjadi dua yakni tumbuhan monokotil dan tumbuhan dikotil.", "Monokotil adalah tumbuhan biji berkeping satu. Contohnya jagung, padi, kelapa, dan sebagainya.", "Dikotil adalah tumbuhan biji berkeping dua. Contohnya kacang kedelai, kacang tanah, bunga matahari, wortel, dan sebagainya.", "Bentuk akar serabut dan tidak berkambium. Ujung akar dilindungi oleh akar lembaga (coleorhiza).", "Sedangkan dikotil memiliki akar tunggang dan berkambium. Ujung akar tidak memiliki pelindung.", "Letak saluran makanan tidak beraturan. Sedangkan pada dikotil letak saluran makanan beraturan.", "Daun pada umumnya panjang. Bahkan ada yang panjang sekali seperti daun tebu. Selain itu memiliki tulang daun sejajar atau melengkung.", "Sedangkan pada dikotil, panjang daun rata-rata 2-3 kali lebarnya, seperti mangga. Tulang daun utama memiliki banyak cabang-cabang, menyirip atau menjari.", "Kelopak bunga 3 helai atau kelipatannya. Sementara pada tanaman dikotil kelopak bunga 5 helai atau kelipatannya. Beberapa tumbuhan ada yang mempunyai 4 helai kelopak bunga.", "Biji tanaman monokotil ketika berkecambah biji tidak membelah. Sedangkan pada tanaman dikotil, ketika berkecambah biji akan membelah menjadi dua.", "Batang suku Poaceae umumnya berbentuk tabung. Pada ruas batangnya melekat daun.", "Daunnya berbentuk pita dengan pelepah daun yang membungkus batang. Tumbuhan kelompok ini dimanfaatkan sebagai tanaman pangan.", "Contoh tumbuhan kelompok ini yakni jagung, gandum, jagung, alang-alang, serai, tebu, dan bambu betung.", "Zingiberaceae banyak dimanfaatkan sebagai tanaman obat-obatan dan rempah-rempahan.", "Contoh tumbuhan kelompok ini yakni jahe, lempuyang, kunyit, kencur, dan lengkuas.", "Daun kelompok tanaman ini lebar, helainya membentuk lonjong kadang berapis lilin. Contoh tumbuhan kelompok ini yakni pisang batu, pisang manila, dan pisang hawidi.", "Arecaceae banyak dimanfaatkan sebagai sumber protein nabati dan lemak nabati. Contoh tumbuhan kelompok ini antara lain kelapa, sawit, salak, sagu, kelapa, aren, dan kurnia.", "Orchidaceae banyak ditemukan di daerah tropis dan hidup sebagai epifit. Contoh tumbuhan monokotil kelompok ini antara lain anggrek bulan, vanili, dan anggrek merpati.", "Perngertian sumber daya alam yang dapat diperbarui yakni sumber daya alam yang keberadaannya di alam selalu tersedia, dapat didaur ulang, atau dapat dihasilkan kembali.", "Contoh sumber daya alam yang dapat diperbarui di antaranya adalah air, tanah, udara, hewan, dan tumbuhan.", "Sumber daya alam hayati adalah sumber daya alam yang berasal dari makhluk hidup misalnya tumbuhan, hewan, dan hasil pertanian, atau perkebunan.", "Hewan di Indonesia dapat digolongkan menjadi dua yaitu hewan liar dan hewan piaraan.", "Hewan liar adalah hewan yang hidup di alam bebas dan dapat mencari makan sendiri. Hewan liar bergantung pada makanan yang disediakan oleh alam.", "Beragam jenis hewan liar yang terdapat di Indonesia antara lain berbagai jenis burung, ikan, dan serangga.", "Hewan piaraan adalah hewan yang sengaja dipelihara manusia untuk sekedar hobi misalnya burung perkutut, marmut, kucing, dan kakatua.", "Hewan ternak adalah termasuk hewan piaraan yang dikembangkan untuk kemudian dimanfaatkan atau diperjualbelikan hasilnya.", "Jenis hewan yang dternakkan terdiri dari 3 jenis yaitu hewan besar, hewan sedang, dan unggas. Contoh hewan yang digolongkan sebagai hewan besar yaitu sapi, kerbau, dan kuda.", "Beberapa hewan yang digolongkan sebagai hewan sedang yaitu kambing, domba, dan kelinci. Adapun jenis unggas yang banyak dternakkan yakni ayam, bebek, dan burung puyuh.", "Hewan ternak dimanfaatkan daging, kulit, susu, telur, bulu, dan tenaganya. Hewan yang dimanfaatkan tenaganya misalnya kerbau dan kuda.", "Hewan yang dimanfaatkan telurnya misalnya ayam, itik, dan burung puyuh.", "Hewan ternak kebanyakan akan dapat diperbarui. Karena hewan dapat terus dibudidayakan dengan cara dikembangbiakkan.", "Untuk hewan-hewan langka bahkan sudah sering diadakan pembiakan buatan. Hal ini agar hewan tersebut tidak punah dari muka bumi. Sangat disayangkan bila hewan langka yang merupakan kekayaan alam di Indonesia itu punah.", "Selanjutnya halaman halaman sumber daya alam yang dapat diperbarui pengertian sumber daya alam pendidikan contoh sumber daya alam yang dapat diperbarui adalah youthizen.", "Sejumlah petani di Kampung Tanjakan Lengka, Desa Parakanlima, Kecamatan Cikembar, Kabupaten Sukabumi berharap perbaikan saluran irigasi yang jebol akibat longsor.", "Kepala Desa Parakanlima, Mirwanda mengatakan, saluran irigasi jebol, setelah wilayah tersebut diguyur hujan deras.", "Akibatnya, tebing yang lokasinya berada di atas selokan itu, longsor dan menengsal saluran irigasi. Akibatnya, saluran irigasi sepanjang 10 meter dengan lebar sekitar 2,5 meter, jebol dan tidak bisa dilintasi air.", "kata Mirwanda kepada wartawan, Selasa (3/11/2020).", "Mirwanda menjelaskan, saluran di Cibodas sangat dibutuhkan warga untuk mengairi lahan pertanian seluas 50 hektare di dua kedesnusan. Pihak desa juga disebut Mirwanda sudah melaporkan peristiwa longsor ini kepada pemerintah Kecamatan Cikembar, Badan Penanggulangan Bencana Daerah (BPBD) dan Dinas PU Kabupaten Sukabumi.", "Kemarin sudah ada yang meninjau dan survei lokasi dari Petugas Penanggulangan Bencana Kecamatan (P2BK) ke sini. Namun, hasil dari survei belum ada tindak lanjutnya," beber dia.", "Sementara itu, seorang petani Elih (45) asal warga Kampung Tanjakan Lengka, RT 4 RW 4, Desa Parakanlima, Kecamatan Cikembar mengatakan, jebolnya saluran di Cibodas sangat berpengaruh terhadap pertumbuhan tanaman padi. Terlebih lagi, saat ini tengah memasuki bercocok tanam.", "Iya, kalau musim cocok tanam itu, airnya harus maksimal. Kalau tidak, bisa mati tanaman padinya," katanya.", "Ia berharap kepada pemerintah daerah Kabupaten Sukabumi agar dapat memberikan bantuan untuk memperbaiki saluran irigasi yang jebol tersebut. Ini harus segera dilakukan agar para petani dapat maksimal dalam mengelola pertaniannya.", "Kita sudah bekerjasama dengan para petani untuk memperbaiki saluran irigasi yang jebol itu dengan peralatan seadanya. Seperti menggunakan karung dan bambu. Namun, perbaikan ini sifatnya hanya sementara dan tidak akan bertahan lama. Jadi, kalau hujan deras lagi kami khawatir saluran irigasi itu bisa kembali jebol. Apalagi saat ini cuacanya ekstrim," tandas Elih.", "Upaya pembinaan warga binaan Lapas Banyuwangi dilakukan tak hanya furnitur dan perikanan."
Irigasi Jebol, Puluhan Hektar Lahan Pertanian di Sukabumi Terancam
2103
17.21%
Sejumlah petani di Kampung Tanjakan Lengka, Desa Parakanlima, Kecamatan Cikembar, Kabupaten Sukabumi berharap perbaikan saluran irigasi yang jebol akibat longsor.
Kasihan, Petani Semangka dan Blewah di Jombang Merana Gegera Hujan
1940
12.95%
Sejumlah petani di Jombang menghancurkan semangka dan blewah hasil panennya sendiri.
Tentang Sumber Daya Alam yang Dapat Diperbarui dan Contohnya
2159
7.20%
Pengertian sumber daya alam yang dapat diperbarui yakni sumber daya alam yang keberadaannya di alam selalu tersedia, dapat didaur ulang, atau dapat dihasilkan kembali.
Mengenal Tumbuhan Monokotil dari Ciri-ciri dan Contohnya
2183
0.00%
Menurut biayinya, tumbuhan dikelompokkan menjadi dua yakni tumbuhan monokotil dan tumbuhan dikotil.
Banjir di 3 Kecamatan Ngawi Surut, Jalur Utama Kembali Dibuka
1912
0.00%
Banjir akibat luapan Sungai Madiun di Ngawi yang meluas di tiga kecamatan telah surut.
```

```
Python 3.9.7 Shell
Hasil pencarian:
Petani Cianjur Sulit Dapat Pupuk Subsidi, Mentan: Nanti Dikek
2157
24.11%
Petani di Cianjur selatan, Jawa Barat kesulitan mendapat pupuk bersubsidi.
353 Hektare Lahan Semangka dan Blewah di Jombang Gagal Panen Gegera Hujan
2109
23.88%
Dinas Pertanian Jombang mencatat terdapat 353 hektare tanaman semangka dan blewah yang gagal panen karena musim hujan datang lebih awal.
Lapas Banyuwangi Panen Raya Melon dan Semangka
1922
22.25%
Upaya pembinaan warga binaan Lapas Banyuwangi dilakukan tak hanya furnitur dan perikanan.
Irigasi Jebol, Puluhan Hektar Lahan Pertanian di Sukabumi Terancam
2103
17.21%
Sejumlah petani di Kampung Tanjakan Lengka, Desa Parakanlima, Kecamatan Cikembar, Kabupaten Sukabumi berharap perbaikan saluran irigasi yang jebol akibat longsor.
Kasihan, Petani Semangka dan Blewah di Jombang Merana Gegera Hujan
1940
12.95%
Sejumlah petani di Jombang menghancurkan semangka dan blewah hasil panennya sendiri.
Tentang Sumber Daya Alam yang Dapat Diperbarui dan Contohnya
2159
7.20%
Pengertian sumber daya alam yang dapat diperbarui yakni sumber daya alam yang keberadaannya di alam selalu tersedia, dapat didaur ulang, atau dapat dihasilkan kembali.
Mengenal Tumbuhan Monokotil dari Ciri-ciri dan Contohnya
2183
0.00%
Menurut biayinya, tumbuhan dikelompokkan menjadi dua yakni tumbuhan monokotil dan tumbuhan dikotil.
Banjir di 3 Kecamatan Ngawi Surut, Jalur Utama Kembali Dibuka
1912
0.00%
Banjir akibat luapan Sungai Madiun di Ngawi yang meluas di tiga kecamatan telah surut.
```

web scraping sampai print h... Pratikum 3 IF2 (2020-11-13) Meet - qm-qm-ukn Laporan Tubes 2 Algeo - Go petani cianjur sulit dapat pu irigasi jebel

127.0.0.1:5000/webscrapping/petani%20cianjur%20sulit%20dapat%20pupuk#

Mbah Google petani cianjur sulit dapat pupuk Home Web Scrapping Images About

Petani Cianjur Sulit Dapat Pupuk Subsidi, Meranti: Nanti Dicek
 Jumlah kata: 2157
 Tingkat kemiripan: 59.67%
 Petani di Cianjur selatan, Jawa Barat kesulitan mendapat pupuk bersubsidi.

353 Hektare Lahan Semangka dan Blewah di Jombang Gagal Panen
 Jumlah kata: 2109
 Tingkat kemiripan: 16.89%
 Dinas Pertanian Jombang mencatat terdapat 353 hektare tanaman semangka dan blewah yang gagal panen karena musim hujan datang lebih awal.

Lapas Banyuwangi Panen Raya Melon dan Semangka
 Jumlah kata: 1922
 Tingkat kemiripan: 13.99%
 Upaya pembinaan warga binaan Lapas Banyuwangi dilakukan tak hanya furnitur dan perikanan.

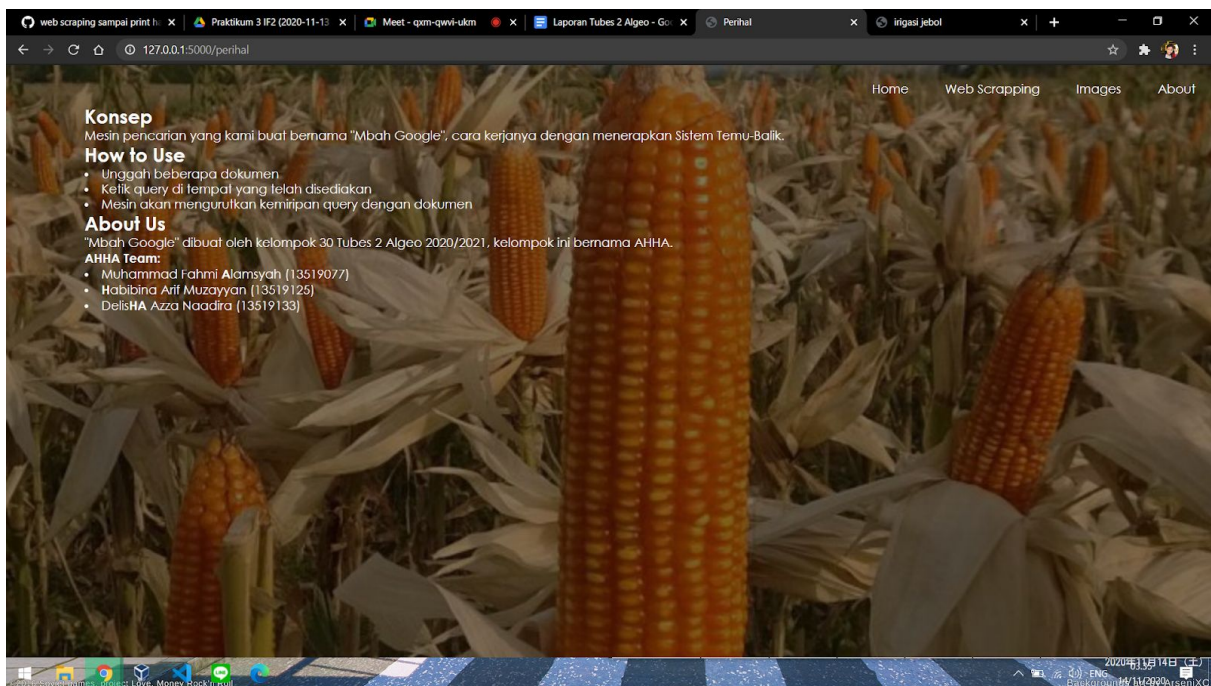
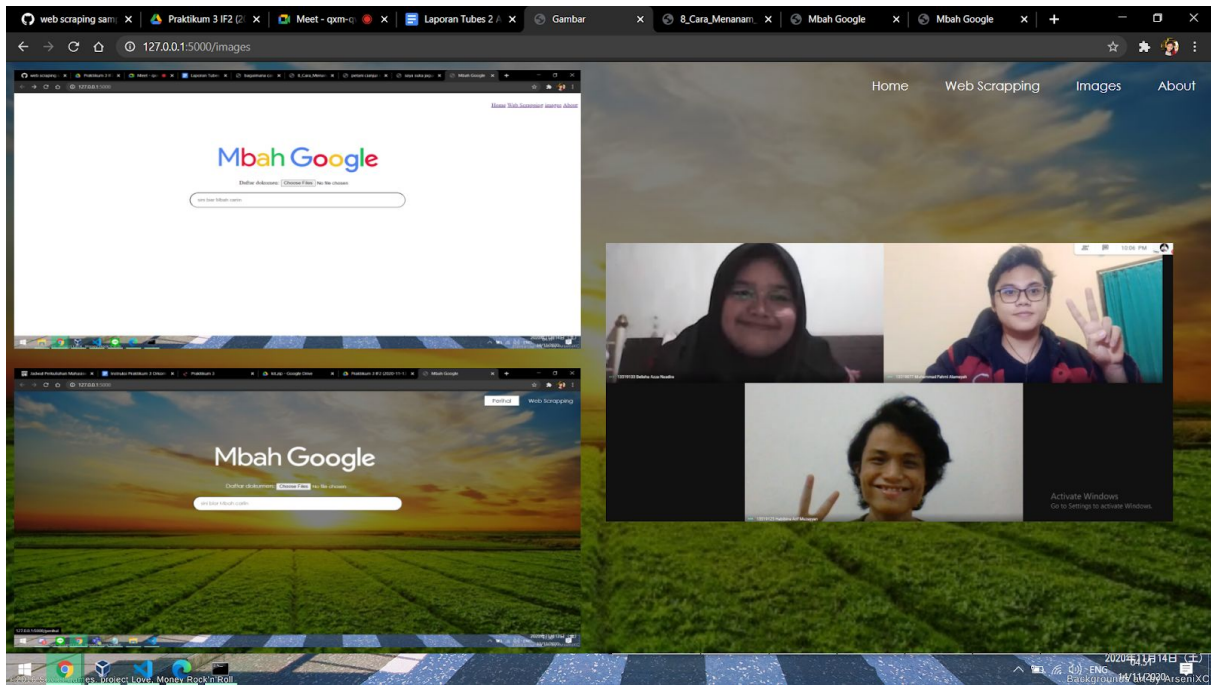
Irigasi Jebol, Puluhan Hektar Lahan Pertanian di Sukabumi Terancam
 Jumlah kata: 2103
 Tingkat kemiripan: 12.17%
 Sejumlah petani di Kampung Tanjakan Lengka, Desa Parakanlima, Kecamatan Cikembar, Kabupaten Sukabumi berharap perbaikan saluran irigasi yang jebol akibat longsor.

Kasihah, Petani Semangka dan Blewah di Jombang Merana Gegara Hujan
 Jumlah kata: 1940
 Tingkat kemiripan: 8.87%
 Sejumlah petani di Jombang menghancurkan semangka dan blewah hasil panennya sendiri.

Term	Query	ngi ya an ta	Banjir di 3 Kecamatan Ngawi Surut, Jalur Utama Kembali Dibuka	Banjir Ngawi Akibat Luapan Sungai Madiun Meluas Hingga 3 Kecamatan	353 Hektare Lahan Semangka dan Blewah di Jombang Gagal Panen Gegara Hujan	Kasihah, Petani Semangka dan Blewah di Jombang Merana Gegara Hujan
tani	1	0	0	0	11	5
cianjur	1	0	0	0	0	0
sulit	1	0	0	0	0	0
pupuk	1	0	0	0	0	0

4

2020年11月14日 ENG Baskiroun18116099ArseniXC



BAB IV

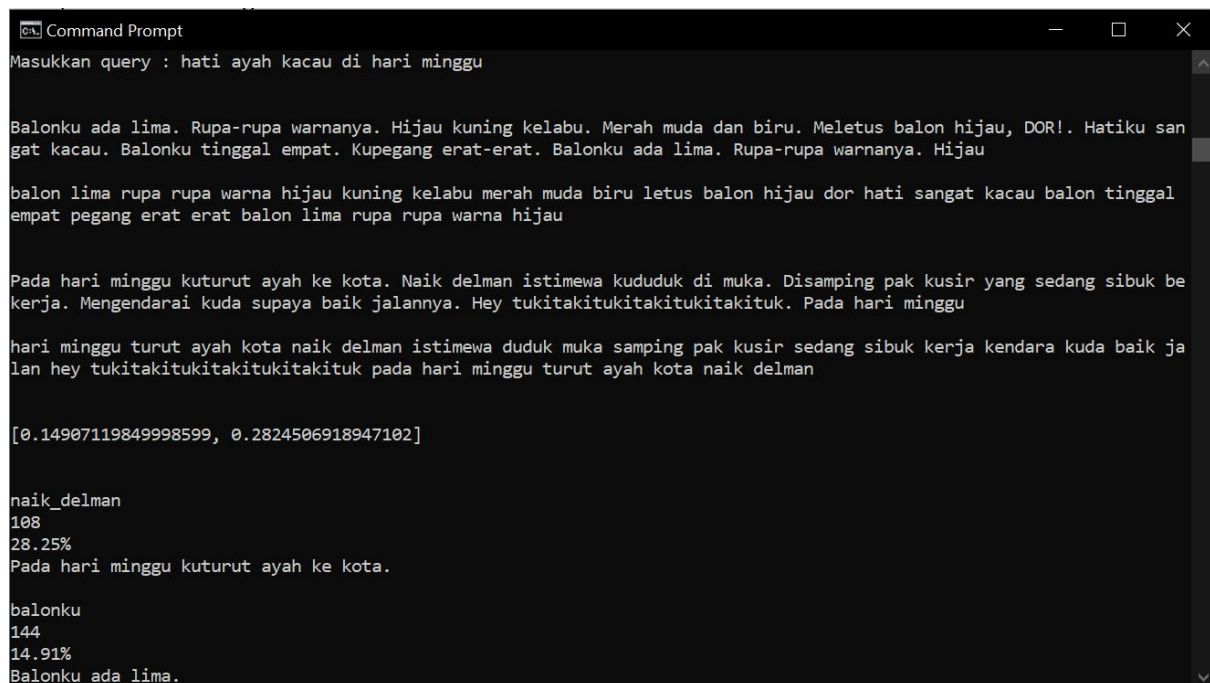
EKSPERIMEN

4.1 Search Program

Pada percobaan pertama kami menggunakan 2 document simple sebagai bahan untuk pencarian query. Analisis yang dilakukan yaitu sebagai berikut :

1. Menghilangkan tanda baca
2. Melakukan stopwords
3. Melakukan stemming
4. Tokenizing kata

Analisis tersebut dilakukan pada query yang dimasukkan oleh user dan document database. Selanjutnya memanggil fungsi cosine similarity untuk mendapatkan perbandingan kesamaan query dan document.



```
Command Prompt
Masukkan query : hati ayah kacau di hari minggu

Balonku ada lima. Rupa-rupa warnanya. Hijau kuning kelabu. Merah muda dan biru. Meletus balon hijau, DOR!. Hatiku san
gat kacau. Balonku tinggal empat. Kupegang erat-erat. Balonku ada lima. Rupa-rupa warnanya. Hijau

balon lima rupa rupa warna hijau kuning kelabu merah muda biru letus balon hijau dor hati sangat kacau balon tinggal
empat pegang erat erat balon lima rupa rupa warna hijau

Pada hari minggu kuturut ayah ke kota. Naik delman istimewa kududuk di muka. Disamping pak kusir yang sedang sibuk be
kerja. Mengendarai kuda supaya baik jalannya. Hey tukitakitukitakitukitak. Pada hari minggu

hari minggu turut ayah kota naik delman istimewa duduk muka samping pak kusir sedang sibuk kerja kendaraan kuda baik ja
lan hey tukitakitukitakitukitak pada hari minggu turut ayah kota naik delman

[0.14907119849998599, 0.2824506918947102]

naik_delman
108
28.25%
Pada hari minggu kuturut ayah ke kota.

balonku
144
14.91%
Balonku ada lima.
```

Percobaan pertama sukses mendapatkan cosine similarity antara 2 document simple. Selanjutnya kami melakukan pengujian terhadap ke-15 document yang telah disiapkan dan sukses menghasilkan perbandingan berdasarkan cosine similarity.

```
Command Prompt
Masukkan query : pertanian di indonesia

Liputan6.com, Jakarta - Cara menanam padi yang baik dan menguntungkan tentunya sangat penting diterapkan. Apalagi, olahan dari padi atau nasi merupakan bahan makanan pokok untuk orang Indonesia. Menanam padi di

liputan6com jakarta cara tanam padi baik untung tentu sangat penting terap apalagi olah padi nasi rupa bahan makan pokok orang indonesia tanam padi iklim tropis indonesia buat tumbuh subur namun

Alat pertanian modern tentu akan membuat aktivitas pertanian Anda menjadi semakin mudah. Peralatan ini sudah dilengkapi dengan teknologi yang canggih sehingga penggunanya tidak harus mengandalkan metode manual lagi. Dengan adanya

alat tani modern buat aktivitas tani jadi makin mudah alat lengkap teknologi canggih guna andal metode manual lagi ada alat alat modern harap hasil panen dapat lebih maksimal tarik mulai

Panduan lengkap budidaya ikan lele Ikan lele merupakan salah satu jenis ikan yang sanggup hidup dalam kepadatan tinggi. Ikan ini memiliki tingkat konversi pakan menjadi bobot tubuh yang baik. Dengan

pandu lengkap budidaya ikan lele ikan lele rupa salah satu jenis ikan sanggup hidup padat tinggi ikan milik tingkat konversi pakan jadi bobot tubuh baik sifat ini budidaya ikan lele

Merdeka.com - Ikan kaya akan gizi seperti omega-3, serat protein, asam amino, zat besi dan yodium. Selain itu ikan juga kaya akan vitamin seperti vitamin A, D, B6 dan B12. Untuk

merdekacom ikan kaya gizi seperti omega 3 serat protein asam amino zat besi yodium ikan kaya vitamin vitamin a d b6 b12 itu konsumsi ikan selalu saran masyarakat indonesia ikan milik

Hama wereng adalah hama yang paling sering menyerang tanaman padi. Hama ini biasanya akan hinggap di batang tanaman padi dan mulai menghisap cairan di dalamnya. Selain itu, hama wereng juga

hama wereng hama paling sering serang tanam padi hama biasa hinggap batang tanam padi mulai menghisap cair dalam itu hama wereng kerap sebab virus sangat rugi tani rusak pengaruh

Cara menanam jagung dengan metode tanpa olah tanah Ada berbagai macam cara menanam jagung salah satunya dengan menerapkan metode tanpa olah tanah (TOT). Pengertian tanpa olah tanah di sini adalah
```

```
Command Prompt
cara tanam jagung metode olah tanah ada berbagai macam cara tanam jagung salah satu terapan metode olah tanah tot erti olah tanah sini cara tanam laku siap lahan balik gembur

Salah satu cara untuk mendapatkan sayuran segar tanpa pestisida adalah dengan menanam sendiri sayuran. Yang jadi masalah adalah tidak semua orang mempunyai lahan untuk menanam sayuran. Oleh karena itu ada

salah satu cara dapat sayur segar pestisida tanam sendiri sayur yang jadi masalah semua orang punya lahan tanam sayur cara tanam sayur manfaat air media ganti tanah sebut hidroponik bagi

Pakan ternak sapi merupakan hal terpenting yang perlu diperhatikan dalam usaha ternak sapi. Karena pakan sangat berpengaruh terhadap pertumbuhan dan perkembangan sapi. Sapi yang biasa di jadikan ternak yaitu jenis

pakan ternak sapi rupa penting perlu perhati usaha ternak sapi pakan sangat pengaruh tumbuh kembang sapi sapi biasa jadi ternak jenis sapi dagang sapi perah dua jenis sapi sebut sangat

Sapi termasuk hewan ternak yang bisa dibilang sulit untuk dirawat. Apalagi untuk dikembangkan sebagai usaha agribisnis, tentu bukanlah menjadi sebuah peluang usaha yang mudah. Namun, siapa sangka jika ternyata ternak

sapi masuk hewan ternak bilang sulit rawat kembang usaha agribisnis bukan jadi buah peluang usaha mudah namun siapa sangka nyata ternak sapi bernilai untung besar meski untung bisnis besar alam

Panduan umum ternak sapi potong Saat ini, usaha ternak sapi potong yang paling menguntungkan adalah penggemukan. Memelihara sapi dari ukuran bakalan hingga ukuran siap untuk dipotong. Kebanyakan ternak sapi di

pandu umum ternak sapi potong saat ini usaha ternak sapi potong paling untung gemuk pelihara sapi ukur bakal hingga ukur siap potong banyak ternak sapi indonesia menitikberatkan usaha gemuk jarang

Pertanian Modern Indonesia Saat Ini Dan Di Masa Depan Sistem pertanian modern saat ini dapat dibagi menjadi tiga hal. Pertama adalah penelitian benih unggul. Kedua, penggunaan teknologi yang canggih. Terakhir yaitu

tani modern indonesia masa depan sistem tani modern bagi jadi tiga hal pertama teliti benih unggul dua guna teknologi canggih akhir guna bahan kimia saat tani banyak tumpu pasok eksternal upa
```



```
Command Prompt
ACEH BARAT DAYA - Meski teknologi pertanian modern sudah berkembang pesat, bukan berarti para petani tidak lagi menggunakan sarana tradisional untuk mengolah lahan mereka. Bukti, hingga saat ini masih ada

aceh barat daya meski teknologi tani modern kembang pesat bukan arti tani guna sarana tradisional olah lahan mereka bukti hingga sebahagian tani kabupaten aceh barat daya abnya memanfaatkan tenaga kerbau

Tahun 2020 menjadi catatan penting bagi ratusan petani desa Gebang kecamatan Plandaan Kabupaten Jombang Jawa Timur. Jika sebelumnya dalam bercocok tanam mereka menggunakan cara lama atau konvensional, maka sejak awal

tahun 2020 jadi catat penting ratus tani desa gebang camat plandaan kabupaten jombang jawa timur belum cocok tanam guna cara lama konvensional sejak awal bulan juni 2020 ini mereka sudah

Jakarta - Produk Peternakan Indonesia mempunyai potensi besar untuk mengisi pasar di beberapa negara tetangga, hal ini disampaikan oleh Direktur Jenderal Peternakan dan Kesehatan Hewan (Dirjen PKH) Kementan, I Ketut

jakarta produk terna indonesia punya potensi besar isi pasar beberapa negara tetangga sampai direktur jenderal terna sehat hewan dirjen pkh tani ketut diarmita jakarta sabtu 31 03 saat beberapa

1.Tanaman Padi Sawah a.Iklim Tanaman padi dapat hidup di daerah yang berhawa panas dan banyak mengandung uap air. Curah hujan rata-rata 200 mm per bulan, dengan distribusi selama 4 bulan,

1tanaman padi sawah aiklim tanam padi hidup daerah hawa panas banyak kandung uap air curah hujan rata rata 200 mm per bulan distribusi lama 4 bulan curah hujan hendak per

[0.03515571357047489, 0.22794676290092006, 0.005288562998090665, 0.013410441545721692, 0.0466590635868752, 0.022605003945421216, 0.0, 0.02353429828549472, 0.0, 0.08305693677313693, 0.6137846542001532, 0.10341753799900381, 0.23091347559276848, 0.21325808235055563, 0.05775006795788362]

Pertanian modern Indonesia
777
61.38%
Pertanian Modern Indonesia Saat Ini Dan Di Masa Depan
Sistem pertanian modern saat ini dapat dibagi menjadi tiga hal.
```

```
Command Prompt
Petani bermigrasi ke pertanian modern
634
23.09%
Tahun 2020 menjadi catatan penting bagi ratusan petani desa Gebang kecamatan Plandaan Kabupaten Jombang Jawa Timur.

Bajak Sawah Jadi Lebih Mudah dengan 5 Alat Pertanian Modern Ini
553
22.79%
Alat pertanian modern tentu akan membuat aktivitas pertanian Anda menjadi semakin mudah.

Produk Peternakan Indonesia Punya Potensi Besar Penuhi Pasar Negara Tetangga
516
21.33%
Jakarta - Produk Peternakan Indonesia mempunyai potensi besar untuk mengisi pasar di beberapa negara tetangga, hal ini disampaikan oleh Direktur Jenderal Peternakan dan Kesehatan Hewan (Dirjen PKH) Kementan, I Ketut Diarmita di Jakarta, Sabtu (31/03).

Petani Abnya Masih Gunakan Kerbau untuk Membajak Sawah
301
10.34%
ACEH BARAT DAYA - Meski teknologi pertanian modern sudah berkembang pesat, bukan berarti para petani tidak lagi menggunakan sarana tradisional untuk mengolah lahan mereka.

Panduan umum ternak sapi potong
910
8.31%
Panduan umum ternak sapi potong
Saat ini, usaha ternak sapi potong yang paling menguntungkan adalah penggemukan.

Syarat tumbuh tanaman padi
1375
5.78%
1.Tanaman Padi Sawah
a.Iklim
Tanaman padi dapat hidup di daerah yang berhawa panas dan banyak mengandung uap air.

Cara Efektif Basmi Hama Wereng
413
4.67%
Hama wereng adalah hama yang paling sering menyerang tanaman padi.
```



```
Command Prompt

8 Cara Menanam Padi yang Baik dan Menguntungkan, Mudah dan Praktis
819
3.52%
Liputan6.com, Jakarta - Cara menanam padi yang baik dan menguntungkan tentunya sangat penting diterapkan.

Jenis Pakan Ternak Sapi
395
2.35%
Pakan ternak sapi merupakan hal terpenting yang perlu diperhatikan dalam usaha ternak sapi.

Cara menanam jagung
703
2.26%
Cara menanam jagung dengan metode tanpa olah tanah
Ada berbagai macam cara menanam jagung salah satunya dengan menerapkan metode tanpa olah tanah (TOT).

Cara Budidaya Ikan Lele yang Mudah Diterapkan, Cocok untuk Pemula
1179
1.34%
Merdeka.com - Ikan kaya akan gizi seperti omega-3, serat protein, asam amino, zat besi dan yodium.

Budidaya ikan lele
1627
0.53%
Panduan lengkap budidaya ikan lele
Ikan lele merupakan salah satu jenis ikan yang sanggup hidup dalam kepadatan tinggi.

Cara menanam tanaman hidroponik sederhana
487
0.00%
Salah satu cara untuk mendapatkan sayuran segar tanpa pestisida adalah dengan menanam sendiri sayuran.

Langkah-Langkah Cara Beternak Sapi Dengan Benar
622
0.00%
Sapi termasuk hewan ternak yang bisa dibilang sulit untuk dirawat.
```

Percobaan terakhir yaitu kami mencoba menerapkan 15 document pada web dengan percobaan sebanyak 5 kali. Didapatkan hasil bahwa 100% document berhasil di ranking berdasarkan cosine similarity terhadap query yang dimasukkan user.

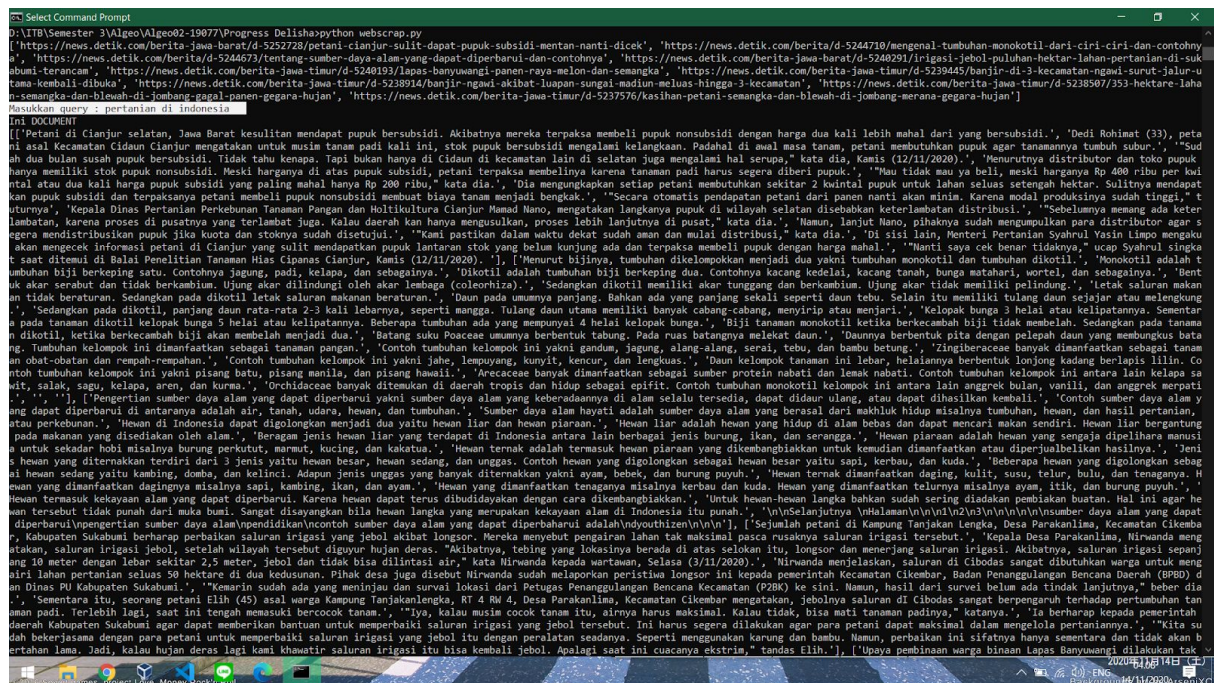
The screenshot shows a web browser window with multiple tabs. The active tab displays search results for the query "bagaimana cara menanam padi agar cepat panen". The results list several documents with their word counts and similarity percentages. A table is overlaid on the right side of the page, showing the frequency of specific terms across the query and the documents.

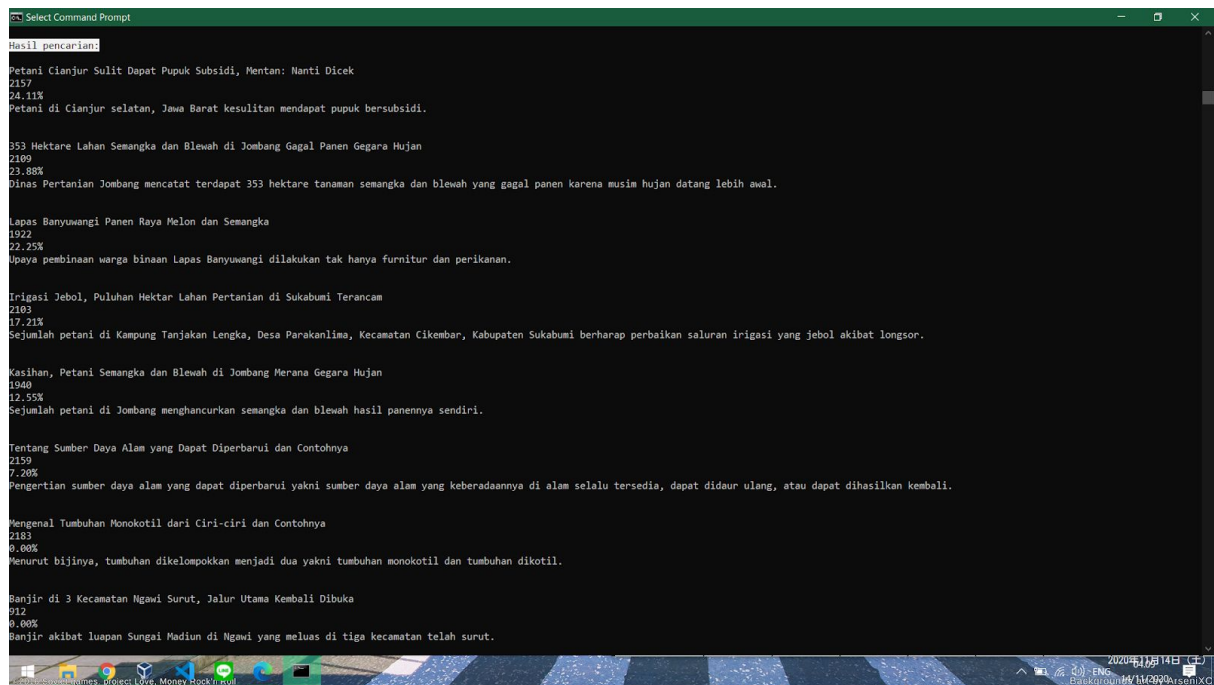
Term	Query	8_Cara_Menanam_Padi_yang_Baik_dan_Menguntungkan_M
bagaimana	1	0
cara	1	13
tanam	1	23
padi	1	30
cepat	1	0
panen	1	11

Pada percobaan pertama kami menggunakan 1 alamat web sebagai bahan untuk pencarian query. Analisis yang dilakukan yaitu sebagai berikut :

1. Mendapatkan seluruh link yang berada pada web
2. Membuka seluruh link yang berada pada web
3. Inspect element
4. Mendapatkan isi dari seluruh document di web
5. Menghilangkan tanda baca
6. Melakukan stopwords
7. Melakukan stemming
8. Tokenizing kata

Analisis tersebut dilakukan pada query yang dimasukkan oleh user dan document database. Selanjutnya memanggil fungsi cosine similarity untuk mendapatkan perbandingan kesamaan query dan document.





Percobaan pertama sukses mendapatkan cosine similarity terhadap seluruh document pada web kompas.com. Selanjutnya yaitu kami mencoba menerapkan web scraping pada web yang telah dibuat dengan flask dengan percobaan sebanyak 5 kali. Didapatkan hasil bahwa 100% document berhasil di ranking berdasarkan cosine similarity terhadap query yang dimasukkan user.

Term	Query	ngi ya an ta	Banjir di 3 Kecamatan Ngawi Surut, Jalur Utama Kembali Dibuka	Banjir Ngawi Akibat Luapan Sungai Madiun Meluas Hingga 3 Kecamatan	353 Hektare Lahan Semangka dan Blewah di Jombang Gagal Panen Gegara Hujan	Kasihani, Petani Semangka dan Blewah di Jombang Merana Gegara Hujan
tani	1	0	0	0	11	5
cianjur	1	0	0	0	0	0
sulit	1	0	0	0	0	0
pupuk	1	0	0	0	0	0

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Information Retrieval System adalah suatu sistem yang secara otomatis melakukan pencarian terhadap informasi yang tidak terstruktur dan menemukan kembali informasi yang relevan terhadap kebutuhan pengguna. Kebutuhan pengguna tersebut disebut dengan *query*, dan informasi yang dimiliki oleh sistem disebut dokumen. Sistem akan mengukur kesamaan dokumen-dokumen yang ada dengan *query* yang diberikan pengguna, kemudian menampilkan dokumen-dokumen sesuai dengan relevansinya dengan *query* tersebut, yaitu semakin mirip dengan *query*, dokumen tersebut akan ditampilkan lebih tinggi. Berdasarkan hasil program yang telah dibuat, dapat disimpulkan bahwa setelah menghitung kesamaan dokumen dengan *query* menggunakan *cosine similarity*, dokumen yang memiliki similaritas yang lebih tinggi akan memiliki nilai cosinus yang lebih tinggi.

Seluruh percobaan yang dilakukan baik search program dan web scraping menghasilkan persentase keberhasilan 100%. Selain itu, Front-end yang diimplementasikan pada web juga sangat interaktif dan menarik dengan adanya background yang selalu berganti setiap menit.

5.2 Saran

Dalam pengembangan web selanjutnya, web di implementasikan pada web online yang sesungguhnya agar orang-orang lain juga dapat menggunakannya. Selain itu, di tambahkan juga fitur-fitur lain seperti dark mode, login account, dan search image agar nantinya dapat menambah pengalaman yang menarik terhadap pengguna web.