

Simple Search Engine

disusun untuk memenuhi tugas
Mata Kuliah penelusuran informasi oleh:

Aqil Fiqran Dzi'Ul Haq
1708107010026



**PROGRAM STUDI INFORMATIKA
FAKULTAS MATEMATIKA DAN PENGETAHUAN ALAM
UNIVERSITAS SYIAH KUALA
DARUSSALAM, BANDA ACEH
2019**

Preprocessing Inverted Index

1. Crawl seed url

```
aqil@DESKTOP-G09DLIV:/mnt/c/xampp/htdocs/search-engine/core$ python Crawl.py -p 200 ../data/crawl
Getting url from : http://indeks.kompas.com/terpopuler/?site=all&date=2020-01-02
Url found : 25
```

Berikut adalah pemberitahuan indeks yang diambil, ia akan berhenti saat mencapai jumlah page yang diinginkan. Jika dalam kasus diatas, ia akan berhenti di page ke 200. Jika telah selesai di crawl. Maka akan tampil data seperti dibawah ini.

Name	Date modified	Type	Size
data1.txt	1/2/2020 9:39 PM	Text Document	2 KB
data2.txt	1/2/2020 9:39 PM	Text Document	2 KB
data3.txt	1/2/2020 9:39 PM	Text Document	2 KB
data4.txt	1/2/2020 9:39 PM	Text Document	2 KB
data5.txt	1/2/2020 9:39 PM	Text Document	3 KB
data6.txt	1/2/2020 9:39 PM	Text Document	3 KB
data7.txt	1/2/2020 9:39 PM	Text Document	3 KB
data8.txt	1/2/2020 9:39 PM	Text Document	2 KB
data9.txt	1/2/2020 9:39 PM	Text Document	2 KB
data10.txt	1/2/2020 9:39 PM	Text Document	2 KB
data11.txt	1/2/2020 9:39 PM	Text Document	2 KB
data12.txt	1/2/2020 9:39 PM	Text Document	2 KB
data13.txt	1/2/2020 9:39 PM	Text Document	3 KB
data14.txt	1/2/2020 9:39 PM	Text Document	2 KB
data15.txt	1/2/2020 9:39 PM	Text Document	3 KB
data16.txt	1/2/2020 9:39 PM	Text Document	3 KB
data17.txt	1/2/2020 9:39 PM	Text Document	2 KB
data18.txt	1/2/2020 9:39 PM	Text Document	0 KB
data19.txt	1/2/2020 8:47 PM	Text Document	2 KB
data20.txt	1/2/2020 8:47 PM	Text Document	2 KB
data21.txt	1/2/2020 8:47 PM	Text Document	1 KB
data22.txt	1/2/2020 8:47 PM	Text Document	2 KB
data23.txt	1/2/2020 8:47 PM	Text Document	2 KB
data24.txt	1/2/2020 8:47 PM	Text Document	2 KB
data25.txt	1/2/2020 8:47 PM	Text Document	2 KB
data26.txt	1/2/2020 8:47 PM	Text Document	3 KB
data27.txt	1/2/2020 8:47 PM	Text Document	3 KB
data28.txt	1/2/2020 8:47 PM	Text Document	2 KB
data29.txt	1/2/2020 8:47 PM	Text Document	2 KB
data30.txt	1/2/2020 8:47 PM	Text Document	2 KB

2. Setelah data berhasil di crawl. Maka bersihkan data tersebut dari symbol dan tanda baca. Di tahap ini juga memisahkan antara judul, kalimat utama dengan kalimat biasa.

```
aqil@DESKTOP-G09DLIV:/mnt/c/xampp/htdocs/search-engine/core$ python Clean.py ../data/crawl ../data/clean
Directory : ../data/crawl
200it [00:00, 310.60it/s]

Done
```

- Setelah berhasil di bersihkan. Beri skor tiap kata unik dan bobot tiap dokumen untuk kata tersebut. Di file python ini tempat terjadinya proses penghapusan stopwords, stemming dan lain-lain.

```
aqil@DESKTOP-G09DLIV:/mnt/c/xampp/htdocs/search-engine/core$ python Score.py ../data/clean ../data/index/index.txt
Directory : ../data/clean
200it [00:02, 68.36it/s]
unique words : 6107

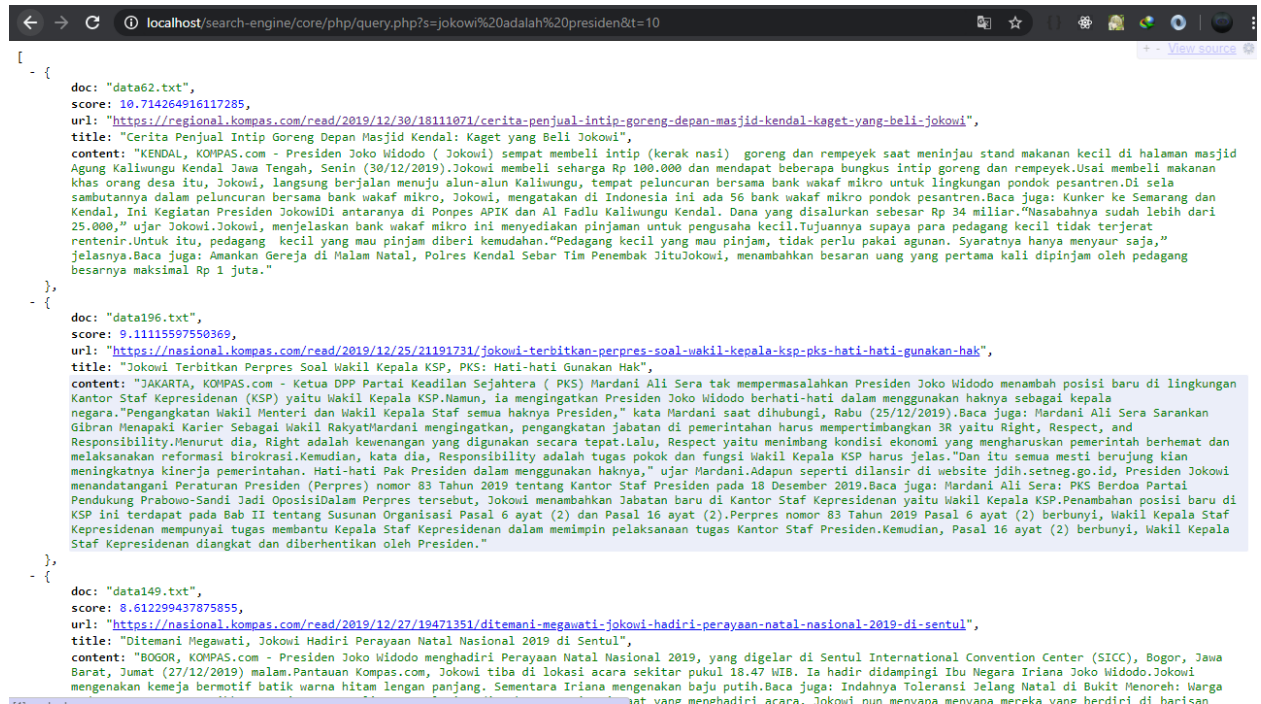
100% | 6107/6107 [00:00<00:00, 6853.15it/s]
done
```

Search Engine

1. Buat file Query.py untuk melakukan proses pencarian dokumen mana yang tampil pertama jika suatu kata di cari. Tampilan dari file ini berupa json yang menampilkan nama dokumen, skor, serta url dari document tersebut.

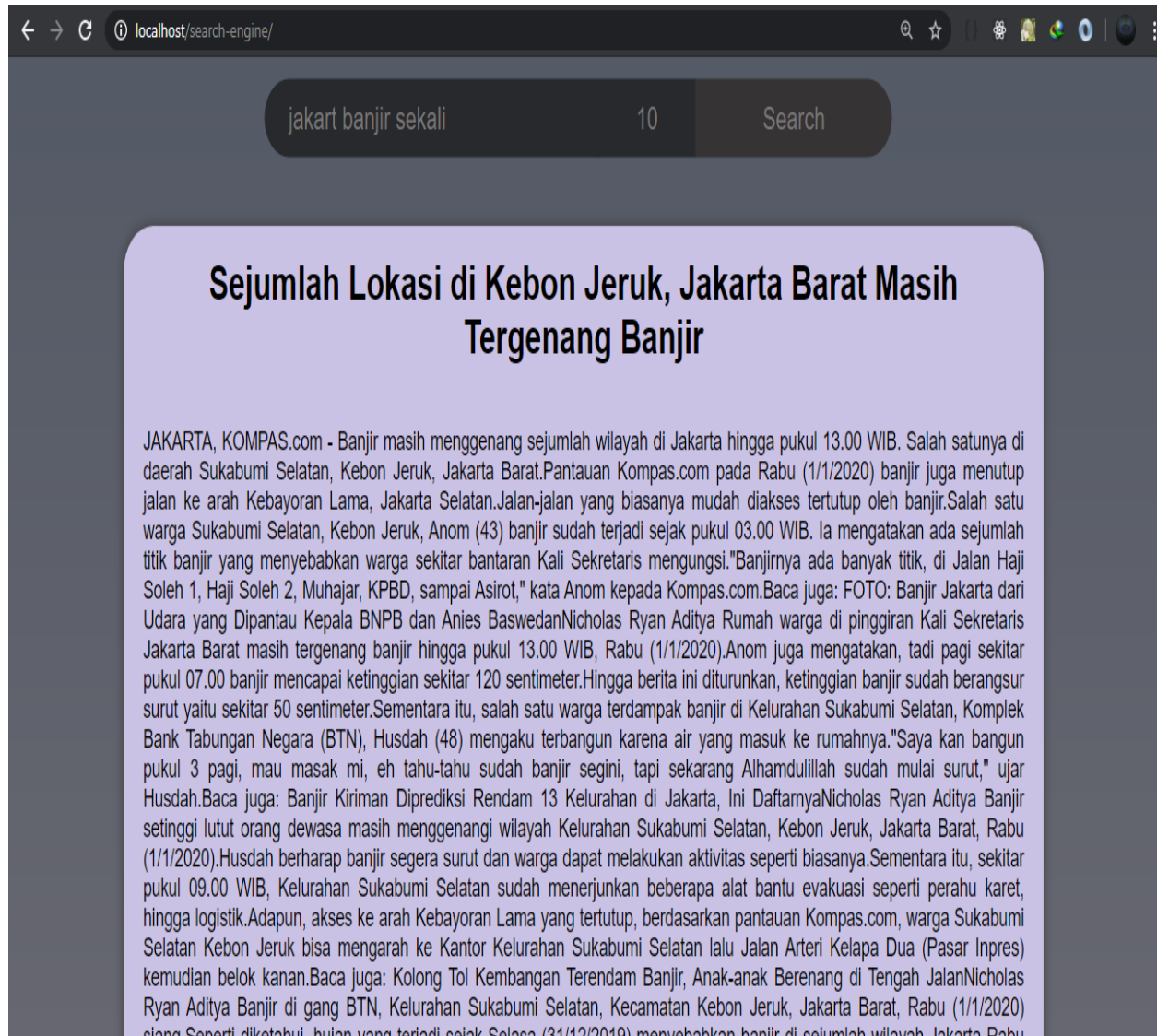
```
aqil@DESKTOP-G09DLIV:/mnt/c/xampp/htdocs/search-engine/core/php$ python3 ../Query.py ../data/index/index.txt ../data/link/link.txt -t 2 presiden jokowi
[{"doc": "data62.txt", "score": 10.714264916117285, "url": "https://regional.kompas.com/read/2019/12/30/18111071/cerita-penjual-intip-goreng-depan-masjid-kendal-kaget-yang-beli-jokowi"}, {"doc": "data196.txt", "score": 9.11115597550369, "url": "https://nasional.kompas.com/read/2019/12/25/21191731/jokowi-terbitkan-perpres-soal-wakil-kepala-ksp-pks-hati-hati-gunakan-hak"}]
aqil@DESKTOP-G09DLIV:/mnt/c/xampp/htdocs/search-engine/core/php$
```

2. Buat file query.php untuk mengeksekusi apa yang dicari oleh orang saat web akan dijalankan. Output dari file tersebut juga json. Namun telah dilengkapi dengan tampilan konten dan judul.



```
[
  {
    doc: "data62.txt",
    score: 10.714264916117285,
    url: "https://regional.kompas.com/read/2019/12/30/18111071/cerita-penjual-intip-goreng-depan-masjid-kendal-kaget-yang-beli-jokowi",
    title: "Cerita Penjual Intip Goreng Depan Masjid Kendal: Kaget yang Beli Jokowi",
    content: "KENDAL, KOMPAS.com - Presiden Joko Widodo ( Jokowi) sempat membeli intip (kerak nasi) goreng dan rempeyek saat meninjau stand makanan kecil di halaman masjid Agung Kaliwungu Kendal Jawa Tengah, Senin (30/12/2019).Jokowi membeli seharga Rp 100.000 dan mendapat beberapa bungkus intip goreng dan rempeyek.Usai membeli makanan khas orang desa itu, Jokowi, langsung berjalan menuju alun-alun Kaliwungu, tempat peluncuran bersama bank wakaf mikro untuk lingkungan pondok pesantren.Di sela sambutannya dalam peluncuran bersama bank wakaf mikro, Jokowi, mengatakan di Indonesia ini ada 56 bank wakaf mikro pondok pesantren.Baca juga: Kunker ke Semarang dan Kendal, Ini Kegiatan Presiden JokowiDi antaranya di Pongpes APIK dan Al Fadlu Kaliwungu Kendal. Dana yang disalurkan sebesar Rp 34 miliar."Nasabahnya sudah lebih dari 25.000," ujar Jokowi.Jokowi, menjelaskan bank wakaf mikro ini menyediakan pinjaman untuk pengusaha kecil.Tujuannya supaya para pedagang kecil tidak terjerat rentenir.Untuk itu, pedagang kecil yang mau pinjam diberi kemudahan."Pedagang kecil yang mau pinjam, tidak perlu pakai agunan. Syaratnya hanya menyaur saja," jelasnya.Baca juga: Amankan Gereja di Malam Natal, Polres Kendal Sebar Tim Penembak JituJokowi, menambahkan besaran uang yang pertama kali dipinjam oleh pedagang besarnya maksimal Rp 1 juta."
  },
  {
    doc: "data196.txt",
    score: 9.11115597550369,
    url: "https://nasional.kompas.com/read/2019/12/25/21191731/jokowi-terbitkan-perpres-soal-wakil-kepala-ksp-pks-hati-hati-gunakan-hak",
    title: "Jokowi Terbitkan Perpres Soal Wakil Kepala KSP, PKS: Hati-hati Gunakan Hak",
    content: "JAKARTA, KOMPAS.com - Ketua DPP Partai Keadilan Sejahtera ( PKS) Mardani Ali Sera tak memperlakukan Presiden Joko Widodo menambah posisi baru di lingkungan Kantor Staf Kepresidenan (KSP) yaitu Wakil Kepala KSP.Namun, ia mengingatkan Presiden Joko Widodo berhati-hati dalam menggunakan haknya sebagai kepala negara."Pengangkatan Wakil Menteri dan Wakil Kepala Staf semua haknya Presiden," kata Mardani saat dihubungi, Rabu (25/12/2019).Baca juga: Mardani Ali Sera Sarankan Gibran Menapaki Karier Sebagai Wakil RakyatMardani mengingatkan, pengangkatan jabatan di pemerintahan harus mempertimbangkan 3R yaitu Right, Respect, and Responsibility.Menurut dia, Right adalah kewenangan yang digunakan secara tepat.Lalu, Respect yaitu menimbang kondisi ekonomi yang mengharuskan pemerintah berhemat dan melaksanakan reformasi birokrasi.Kemudian, kata dia, Responsibility adalah tugas pokok dan fungsi Wakil Kepala KSP harus jelas."Dan itu semua mesti berujung kian meningkatnya kinerja pemerintahan. Hati-hati Pak Presiden dalam menggunakan haknya," ujar Mardani.Adapun seperti dilansir di website jdih.setneg.go.id, Presiden Jokowi menandatangani Peraturan Presiden (Perpres) nomor 83 Tahun 2019 tentang Kantor Staf Presiden pada 18 Desember 2019.Baca juga: Mardani Ali Sera: PKS Berdoa Partai Pendukung Prabowo-Sandi Jadi OposisiDalam Perpres tersebut, Jokowi menambahkan jabatan baru di Kantor Staf Kepresidenan yaitu Wakil Kepala KSP.Penambahan posisi baru di KSP ini terdapat pada Bab II tentang Susunan Organisasi Pasal 6 ayat (2) dan Pasal 16 ayat (2).Perpres nomor 83 Tahun 2019 Pasal 6 ayat (2) berbunyi, Wakil Kepala Staf Kepresidenan mempunyai tugas membantu Kepala Staf Kepresidenan dalam memimpin pelaksanaan tugas Kantor Staf Presiden.Kemudian, Pasal 16 ayat (2) berbunyi, Wakil Kepala Staf Kepresidenan diangkat dan diberhentikan oleh Presiden."
  },
  {
    doc: "data149.txt",
    score: 8.612299437875855,
    url: "https://nasional.kompas.com/read/2019/12/27/19471351/ditemani-megawati-jokowi-hadiri-perayaan-natal-nasional-2019-di-sentul",
    title: "Ditemani Megawati, Jokowi Hadiri Perayaan Natal Nasional 2019 di Sentul",
    content: "BOGOR, KOMPAS.com - Presiden Joko Widodo menghadiri Perayaan Natal Nasional 2019, yang digelar di Sentul International Convention Center (SICC), Bogor, Jawa Barat, Jumat (27/12/2019) malam.Pantauan Kompas.com, Jokowi tiba di lokasi acara sekitar pukul 18.47 WIB. Ia hadir didampingi Ibu Negara Iriana Joko Widodo.Jokowi mengenakan kemeja bermotif batik warna hitam lengan panjang. Sementara Iriana mengenakan baju putih.Baca juga: Indahnya Toleransi Jelang Natal di Bukit Menoreh: Warga ...at yang menhadiri acara. Jokowi nun menuna menuna mereka yang berdiri di barisan
```

3. Buat file index.html untuk tampilan utama dari web ini. Disini akan memunculkan inputan berupa text untuk pencariannya, dan integer untuk memilih berapa dokumen teratas yang ingin ditampilkan.



4. Buat file app.js untuk menghubungkan antara query.php dengan index.html atau bisa dibilang front-end dengan back-end.

```
assets > js > app.js
1  $(document).ready(() => {
2      $("#search").on('click', () => {
3          let query = $(".text").val()
4          let top = $(".number").val()
5
6          let search = "http://localhost/search-engine/core/php/query.php?t=" + top + "&s=" + query;
7          $.ajax({
8              url: search,
9              success: function (result) {
10                  JSON.parse(result).forEach(el => {
11                      $("#result").append(makeResult(el));
12                  });
13              },
14              error: function (err) {
15                  console.log('err :', err);
16              }
17          });
18      });
19  });
20
21  makeResult = data => {
22      card = '<div class="content">';
23      card += '<div>';
24      card += '<h3>' + data.title + '</h3>';
25      card += '<p>' + data.content + '</p>';
26      card += '<a href="' + data.url + '">Read More</a>';
27      card += '</div></div>';
28      return card;
29  }
```