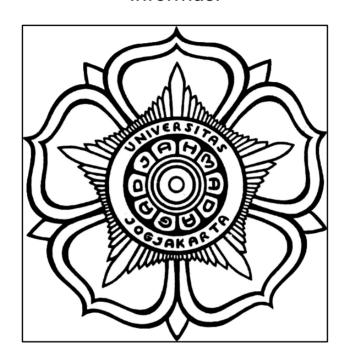
Sistem Temu Kembali untuk Pasal-Pasal KUHP pada Hukum Indonesia

Dokumen Indexing

Ditulis untuk Memenuhi Tugas Mata Kuliah Temu Kembali Informasi



Anggota Kelompok

Ferdiansyah Dwi Nurcahyo (20/459265/PA/19926)

Muny Safitri (22/506376/NUGM/01039)

Raden Mas Garda Zhafir (20/459275/PA/19996)

PROGRAM STUDI ILMU KOMPUTER



FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM UNIVERSITAS GADJAH MADA

Melakukan Crawling

Membuat kolom per kategori yang didapatkan dari website dan juga melakukan perulangan pada tiap dokumen untuk mendapatkan semua data yang diperlukan.

```
| The between types is | form to the import | The import readsts | form to the import | The import readsts | form to the import | The i
```

```
or a in soup.find_all('a', href=True):
    if 'href' in a.attrs:
    temp.append(a.attrs['href']) #menambahkan link ke kolom7
    # print("Found the URL:", a.attrs['href'])
#menambahkan setiap konten ke kolom masing
#menambahkan kolom1,kolom2,kolom4,kolom6 dan nomor
for x in range(0, len(title)):
  count += 1
  kolom1.append(title[x].text.strip())
  kolom2.append(date[x].text.strip())
  kolom4.append(category[x].text.strip())
  #melakukan transformasi pada tanggal
  kolom6.append(date[0].text.strip().split()[2])
  nomor.append(str(k))
  k=k+1
#menambahkan kolom link (kolom7)
for link in range(24,len(temp)-47,4):
kolom7.append('https://www.hukumonline.com'+temp[link])
  # url2 = temp[link]
  # #Membuat Requests
  # r2 = requests.get(url2)
  # #Hasil Response
# request2 = r2.content
  # soup2 = BeautifulSoup(request2, 'html.parser')
#menambahkan kolom3 dan kolom4
for y in range(0,len(mitra_author)-1,2):
  kolom3.append(mitra_author[y].text.strip())
  kolom5.append(mitra_author[y+1].text.strip())
# print("{0}. {1}\n {2} \n{3}".format(y,mitra_author[y].text.strip(),mitra_author[y+1].text.strip(),category[x].text.strip()))
```

Mengambil judul artikel sebagai dokumen dan juga menyiapkan array,

```
[ ] #mengambil judul artikel sebagai dokumen
    array = [] #dipake untuk menghitung frequency nanti
    for i in df.index:
      array.append(df['judul_artikel'][i])
      print("Judul: "+df['judul artikel'][i])
    Judul: Konversi Utang Jadi Setoran Saham, Ini Caranya
    Judul: Tanpa NDA, Bisakah Karyawan Digugat Pelanggaran Rahasia Dagang?
    Judul: Langkah Jika Mantan Karyawan Membocorkan Rahasia Perusahaan
    Judul: Pinjaman Melebihi Modal Dasar PT, Perlukah Persetujuan RUPS?
    Judul: Persamaan Asosiasi dengan Perkumpulan dan Prosedur Pendiriannya
    Judul: 7 Alasan Mengapa Investor Asing Berinvestasi di Indonesia
    Judul: Fungsi Lembaga Pengelola Investasi dan Bedanya dengan PIP atau BPKM
    Judul: Dapatkah Aset Kripto Jadi Alat Pembayaran Utang dalam Kepailitan?
    Judul: Biar Tak Diblokir, Ini Cara Pendaftaran PSE Lingkup Privat
    Judul: PT Tak Punya IUP Eksplorasi? Ini Sanksi Pidananya
    Judul: Perbedaan Layanan Pinjaman Bank Digital dengan Pinjol
    Judul: Anak Berkendara Hingga Tabrak Orang, Bagaimana Proses Hukumnya?
    Judul: Dapatkah Tender Terbatas Dikategorikan Persekongkolan Tender?
    Judul: Izin Usaha dan Kode KBLI Perdagangan Batu Kapur
    Judul: Begini Hak Opsi dalam Sewa Guna Usaha (Leasing)
    Judul: Jatuh Talak Suami, Ini Hak-hak Istri dan Anak
     Judul: Dana Sumbangan Diselewengkan, Dapatkah Yayasan Dibubarkan?
    Judul: 3 Perbedaan Mediasi dan Arbitrase
    Judul: Bingung Cari Nama PT yang Bagus? Simak Aturannya
    Judul: Luas Rumah KPR Berbeda dengan di IMB, Ini Konsekuensinya
    Judul: Dapatkah Menyimpangi Isi Homologasi dengan Novasi?
    Judul: Pembayaran Mandek, Bagaimana Status Uang dalam Perjanjian Akuisisi?
    Judul: Syarat Perusahaan Menjadi Trader Batubara
    Judul: Sengketa yang Tak Dapat Diselesaikan Melalui Penyelesaian Sengketa Alternatif
    Judul: Kenali Bank Digital di Indonesia dan Syarat Pendiriannya
     Judul: Klasifikasi Bahan Baku, Barang Konsumsi, dan Bahan Penolong
     Judul: Pemindahan Kepemilikan Perusahaan (Akuisisi) oleh Pemegang Saham
    Judul: Haruskah Pengumuman Tender Swasta Melalui Media Massa?
    Judul: Skema Pembayaran Utang Debitur Tak Boleh di Luar Homologasi
    Judul: Pembayaran ke Kreditur Pasca Homologasi, Haruskah Izin Pengurus?
```

Preprocessing dan Indexing

Memasukkan stopword dan juga imbuhan untuk menghilangkan hal yang tidak diperlukan. Serta membuat tokenisasi dari hasil tersebut.

Mendapatkan hasil index dari token

```
[ ] #daftar term(token) berserta id document(dictiniories)
     ele = {}
     kolom1=[]
     kolom2=[]
     index = 1
     for i in range(len(token)):
       for a in token[i]:
         ele[a] = index
          kolom1.append(a)
         kolom2.append(str(index))
print(a +',' + str(index))#token dan id dokumen
       index = index+1
     # print(ele['karyawan'])
     keluarga,82
di,82
     lapas,82
     selama,82
pandemi,82
     covid,82
     19,82
     aspek,83
     perlindungan,83
     data,83
     pribadi,83
     dalam,83
     jasa,83
     pengiriman,83
barang,83
hukumnya,84
mewajibkan,84
     karyawan,84
share,84
     live,84
     location,84
     di,84
luar,84
jam,84
     kerja,84
```

Hasil yang didapat sebelum dilakukan sorting

		id doo	 	
1	term	id_doc		
2	konversi	1		
3	utang 	1		
4	jadi	1		
5	setoran	1		
6	saham	1		
7	ini	1		
8	caranya	1		
9	tanpa	2		
10	nda	2		
11	bisakah	2		
12	karyawan	2		
13	digugat	2		
14	pelanggara	2		
15	rahasia	2		
16	dagang	2		
17	langkah	3		
18	jika	3		
19	mantan	3		
20	karyawan	3		
21	memboco	3		
22	rahasia	3		
23	perusahaa	3		
24	pinjaman	4		
25	melebihi	4		
26	modal	4		
27	dasar	4		
28	pt	4		
29	nerlukah	4		

Hasil yang didapat setelah sorting

1 term id doc					
	term	id_doc			
2	0	555			
3	0	545			
4	1	234			
5	10	455			
6	10	445			
7	10	139			
8	11	127			
9	14	282			
10	17	121			
11	19	80			
12	19	88			
13	19	82			
14	19	81			
15	19	519			
16	2	312			
17	2	323			
18	21	129			
19	3	413			
20	3	104			
21	3	18			
22	3	63			
23	3	389			
24	3	94			
25	3	172			
26	4	260			
27	4	314			
20	4	250			