

***Sistem Temu Kembali untuk Pasal-Pasal KUHP pada Hukum
Indonesia***

Dokumen Crawling

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

(20/459275/PA/19996)

**PROGRAM STUDI ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS GADJAH MADA**

1. Pendahuluan

Web crawling adalah proses mencari tahu kumpulan halaman dari sebuah web untuk dilakukan pengindeksan. Web crawling sangat berguna bagi sebuah website agar mudah ditemukan oleh orang lain. Namun, web crawling bukan satu-satunya cara yang bisa dilakukan. Disini kita melakukan web crawling untuk mencari berita di media online yang berhubungan dengan kasus yang ada pada pasal KUHP.

2. Crawling

Pada crawling yang kami lakukan, kami memilih untuk melakukan crawling dari website <https://www.hukumonline.com/klinik/pidana/>

The screenshot displays the HukumOnline website interface. At the top, there is a dark blue header with the HukumOnline logo, a search bar, and navigation links like 'Pro', 'Solusi', and 'Wawasan Hukum'. Below the header, the breadcrumb trail reads 'Beranda > Klinik > Pidana'. The main heading is 'Arsip Jawaban'. On the left, a sidebar lists various legal categories such as 'Semua Kategori', 'Bisnis', 'Hak Asasi Manusia', 'Ilmu Hukum', 'Kekayaan Intelektual', and 'Keluarga'. The main content area is titled 'Pidana' and lists several articles with their titles, dates, and authors. The articles include: 'Apakah Karyawan Bertanggungjawab Atas Kerugian Perusahaan?' (30 Agt 2022, Haris Satiadi & Partners), 'Perbuatan Tak Penuhi Unsur Delik Pidana, Ini Akibat Hukumnya' (25 Agt 2022, PERSADA UB), 'Pidana Selingkuh Tanpa Bersetubuh bagi Pasangan, Adakah?' (24 Agt 2022, Si Pokrol), and 'Polisi Melakukan Tindak Pidana, Begini Proses Peradilannya' (25 Agt 2022, Si Pokrol). A dark blue footer at the bottom contains the text 'Pantau Kewajiban Hukum Perusahaan Anda Di Sini!' and a small 'x' icon.

Kemudian, media yang kami gunakan pada crawling ini adalah Google Colab yang menggunakan bahasa pemrograman Python.

```
from tkinter import N
from turtle import title
import requests
from bs4 import BeautifulSoup

kolom1 = []
kolom2 = []
kolom3 = []
nomor = []
k=1
for i in range(10):
    url = 'https://www.hukumonline.com/klinik/pidana/page/'+str(i)+'/'
    #Membuat Requests
    r = requests.get(url)
    #Hasil Response
    request = r.content

    soup = BeautifulSoup(request, 'html.parser')
    #Extract Element
    title = soup.findAll('h2', attrs={'class':'header-6 font-weight-bold text-dark'})
    date = soup.findAll('span', attrs={'class':'small text-muted mr-1'})
    author = soup.findAll('a', attrs={'class':'small mx-2 d-none d-md-inline-block'})
    count = 0
    for x in range(0, len(title)):
        count += 1
        # print("{0}. {1}\n {2}".format(count,title[x].text.strip(),date[x].text.strip()))
        kolom1.append(title[x].text.strip())
        kolom2.append(date[x].text.strip())
        kolom3.append(author[x].text.strip())
        nomor.append(str(k))
        k=k+1
```

Setelah dijalankan, kami akan mendapat beberapa data yang kami inginkan, yaitu nomor, judul, tanggal, dan juga author nya.

	nomor		nama	update	author
0	1	Apakah Karyawan Bertanggungjawab Atas Kerugian...	30 Agt 2022		Haris Satiadi & Partners
1	2	Perbuatan Tak Penuhi Unsur Delik Pidana, Ini A...	25 Agt 2022		Haris Satiadi, S.H.
2	3	Pidana Selingkuh Tanpa Bersetubuh bagi Pasanga...	24 Agt 2022		PERSADA UB
3	4	Polisi Melakukan Tindak Pidana, Begini Proses ...	24 Agt 2022	Dr. Aditya Wiguna Sanjaya, S.H., M.H., M.H.Li.	
4	5	Pacar Minta Uang dengan Mengancam Putus, Ini H...	23 Agt 2022		Si Pokrol
...
95	96	Terdakwa Sopan, Bisakah Jadi Alasan Peringan P...	16 Des 2021		Bernadetha Aurelia Oktavira, S.H.
96	97	Jerat Hukum dan Ancaman Pidana Pelaku Sodomi	16 Des 2021		Si Pokrol
97	98	Mengapa Benda Milik Korban Tindak Pidana Malah...	14 Des 2021		Erizka Permatasari, S.H.
98	99	Terjerat Utang dengan Lintah Darat, Ini Hukumnya	07 Des 2021		PBH Peradi
99	100	Pasal untuk Menjerat Oknum Polisi yang Perkosa...	07 Des 2021	Dr. (CN) Maju Posko Simbolon, S.H., M.H.	

100 rows × 4 columns

Setelah mendapat data tersebut, kami langsung mengexport hasil tersebut menjadi dataset yang sudah kami tentukan yaitu menjadi file CSV.

	A	B	C	D	E	F	G	H	I	J
1		nomor	nama	update	author					
2	0	1	Apakah Ka	30 Agt 202	Haris Satiadi & Partners					
3	1	2	Perbuatan	25 Agt 202	Haris Satiadi, S.H.					
4	2	3	Pidana Seli	24 Agt 202	PERSADA UB					
5	3	4	Polisi Mela	24 Agt 202	Dr. Aditya Wiguna Sanjaya, S.H., M.H., M.H.Li.					
6	4	5	Pacar Mint	23 Agt 202	Si Pokrol					
7	5	6	Kasus Pida	19 Agt 202	Bernadetha Aurelia Oktavira, S.H.					
8	6	7	Jika Polisi I	16 Agt 202	Si Pokrol					
9	7	8	Beda Pasa	12 Agt 202	Bernadetha Aurelia Oktavira, S.H.					
10	8	9	Apa Saja A	11 Agt 202	Lembaga Bantuan Hukum Mawar Saron					
11	9	10	Koruptor K	10 Agt 202	Ita Iya Pulina Perangin-angin, S.H.					
12	10	11	Apakah Ka	30 Agt 202	Haris Satiadi & Partners					
13	11	12	Perbuatan	25 Agt 202	Haris Satiadi, S.H.					
14	12	13	Pidana Seli	24 Agt 202	PERSADA UB					
15	13	14	Polisi Mela	24 Agt 202	Dr. Aditya Wiguna Sanjaya, S.H., M.H., M.H.Li.					
16	14	15	Pacar Mint	23 Agt 202	Si Pokrol					
17	15	16	Kasus Pida	19 Agt 202	Bernadetha Aurelia Oktavira, S.H.					
18	16	17	Jika Polisi I	16 Agt 202	Si Pokrol					
19	17	18	Beda Pasa	12 Agt 202	Bernadetha Aurelia Oktavira, S.H.					