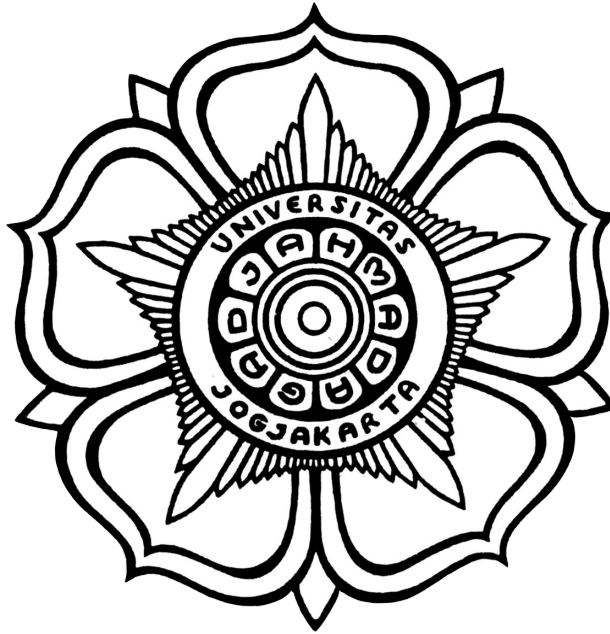


***Sistem Temu Kembali untuk Pasal-Pasal KUHP pada Hukum  
Indonesia***

***Dokumen Metode dan Hasil Transformasi Dataset***

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**

## 1. Pendahuluan

Transformasi data adalah sekumpulan proses berulang (imperatif) yang mengubah sekumpulan data mentah hasil *scrapping* ke dalam suatu format atau struktur data tertentu yang bertujuan untuk memudahkan dilakukan proses lebih lanjut padanya, memudahkan pembuatan model, atau supaya mendapatkan hasil analisis yang optimal. Transformasi data masuk ke dalam *feature engineering*. Pada Sistem Temu kembali informasi, transformasi data biasanya dilakukan untuk memudahkan proses selanjutnya, yaitu pembuatan indeks data sehingga sangat berkaitan dengan pembuatan struktur data yang sesuai.

## 2. Konsep Transformasi

Transformasi pada Sistem Temu Kembali untuk Pasal-Pasal KUHP pada Hukum Indonesia berfokus pada pengategorian dan penyelarasan format-format data kuantitatif, dalam hal ini adalah tanggal dan pemasukan kategori tiap data hasil *scrapping*. Kategori berada pada array dengan isi sebagai berikut: {'bisnis', 'hak-asasi-manusia', 'ilmu-hukum', 'kekayaan-intelektual', 'keluarga', 'ketenagakerjaan', 'perdata', 'perlindungan-kosumen', 'pertanahan-properti', 'pidana', 'profesi-hukum', 'start-up-umkm', 'teknologi'}. Fungsi kategori ini adalah untuk memungkinkan adanya pengelompokkan hasil *retrieval* nanti. Proses penentuan dan penyimpanan kategori ini dapat dilakukan ketika *scrapping* ulang di mana kategori tiap dokumen hasil *scrapping* terdapat pada url-nya dalam bentuk

Selain itu, terdapat penyelarasan format penanggalan artikel pada tanggal hasil *scrapping* di mana sekarang hanya tahun yang disimpan. Hal ini merujuk pada asumsi bahwa makin terbaru suatu dokumen, dapat menjadi rujukan yang lebih baik karena memiliki dasar hukum yang lebih kuat dan lebih aktual daripada yang lebih lama.

## 3. Implementasi

Berikut adalah *source code* proses transformasi data yang berbentuk proses scrapping ulang:

```
from tkinter import N

from turtle import title

import requests

from bs4 import BeautifulSoup
```

```

categories = ['bisnis',
'hak-asasi-manusia','ilmu-hukum','kekayaan-intelektual','keluarga',
'ketenagakerjaan','perdata','perlindungan-kosumen','pertanahan-properti','
pidana','profesi-hukum','start-up-umkm','teknologi']

# categories = ['bisnis',
'hak-asasi-manusia','ilmu-hukum']# 'kekayaan-intelektual','keluarga',
'ketenagakerjaan','perdata','perlindungan-kosumen','pertanahan-properti','
pidana','profesi-hukum','start-up-umkm','teknologi']

kolom1 = []

kolom2 = []

kolom3 = []

kolom4 = []

kolom5 = []

kolom6 = []

nomor = []

k=1

for j in range(len(categories)):

    for i in range(10):#pengulangan 10 kali halaman setiap kategory

        url =
'https://www.hukumonline.com/klinik/'+categories[j]+'/page/'+str(i)+'/'

        #Membuat Requests

        r = requests.get(url)

        #Hasil Response

```

```

request = r.content

soup = BeautifulSoup(request, 'html.parser')

#Extract Element

category = soup.findAll('span', attrs={'class':'d-inline-flex flex-row justify-content-center align-items-center mx-2 badge badge-primary badge-pill'})

title = soup.findAll('h2', attrs={'class':'header-6 font-weight-bold text-dark'})

date = soup.findAll('span',attrs={ 'class':'small text-muted mr-1'})

mitra_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

    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

```

```

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()))

    y=y+2

```

Berikut adalah hasilnya dalam bentuk dataframe:

No	judul_artikel	update	mitra	kategori	author	tahun
1	Langkah Jika Mantan Karyawan Membocorkan Rahas...	09 Sep 2022	Si Pokrol	Bisnis	Bernadetha Aurelia Oktavira, S.H.	2022
2	Pinjaman Melebihi Modal Dasar PT, Perlukah Per...	08 Sep 2022	Tjoetiar Tandjung & Associates	Bisnis	Syafrida Tandjung, S.H., LL.M	2022
3	Persamaan Asosiasi dengan Perkumpulan dan Pros...	06 Sep 2022	Si Pokrol	Bisnis	David Christian, S.H.	2022
4	7 Alasan Mengapa Investor Asing Berinvestasi d...	01 Sep 2022	Si Pokrol	Bisnis	David Christian, S.H.	2022
5	Fungsi Lembaga Pengelola Investasi dan Bedanya...	31 Agt 2022	Pusat Kajian Hukum Bisnis FH Unair	Bisnis	Dr. Widhayani Dian Pawestri S.H., M.H.	2022
...	...	...	...	...	...	...
1177	Hukumnya Berbohong dan Menyebar Fitnah di Medsos	26 Feb 2019	Si Pokrol	Teknologi	Tri Jata Ayu Pramesti, S.H.	2019
1178	Hukum Menyebarluaskan Lagu yang Bermuatan Peng...	26 Feb 2019	Si Pokrol	Teknologi	Tri Jata Ayu Pramesti, S.H.	2019
1179	Mengunggah Screenshot SMS Menjengkelkan ke Med...	22 Feb 2019	Si Pokrol	Teknologi	Tri Jata Ayu Pramesti, S.H.	2019
1180	Apakah Mengirim Link Berita tentang Seks Terma...	19 Feb 2019	Si Pokrol	Teknologi	Tri Jata Ayu Pramesti, S.H.	2019
1181	Recovery File Foto Pribadi di HP dan Masalah H...	15 Feb 2019	Indonesia Cyber Law Community (ICLC)	Teknologi	Teguh Arifiyadi, S.H., M.H.	2019

1181 rows x 6 columns

Hasil *scrapping* kemudian dikonversi ke dalam bentuk csv

```
df_hasil.to_csv('hasil-crawling_scrapping-web-hukumOnline-pidana.csv')
```

## 5. Saran dan Pengembangan Lebih Lanjut

Ke depannya, dapat dicek kegunaan bagian lagi dari data tanggal, seperti hari dan bulan untuk mendukung pemberian *retrieval* yang lebih relevan.