▾ Laporan Praktikum M8

Arya Duta Gordra Sumitro Putra

162112133017

▾ Import Library

```
import requests as rq
from bs4 import BeautifulSoup as bs
import pandas as pd
import re
import scrapy as sc
```

▾ Scraping Python Requests

```
response = rq.get('https://www.unair.ac.id/tantangan-bonus-demografi-indonesia-dan-pembang
preq = response.text
```

```
preq
```

Indonesia","PIN","Terbaik","UNAIR","universitas","universitas
Airlangga"],"articleSection":["berita"],"inLanguage":"id-ID"},
{"@type":"WebPage","@id":"https://www.unair.ac.id/tantangan-bonus-demografi-
indonesia-dan-pembangunan-
berkelanjutan/","url":"https://www.unair.ac.id/tantangan-bonus-demografi-
indonesia-dan-pembangunan-berkelanjutan/","name":"Tantangan Bonus Demografi
Indonesia dan Pembangunan Berkelanjutan\xa0 - Universitas Airlangga Official
Website","isPartOf":
{"@id":"https://www.unair.ac.id/#website"},"primaryImageOfPage":
{"@id":"https://www.unair.ac.id/tantangan-bonus-demografi-indonesia-dan-
pembangunan-berkelanjutan/#primaryimage"},"image":
{"@id":"https://www.unair.ac.id/tantangan-bonus-demografi-indonesia-dan-
pembangunan-
berkelanjutan/#primaryimage"},"thumbnailUrl":"https://www.unair.ac.id/wp-
content/uploads/2022/10/adw.png","datePublished":"2022-10-
24T07:47:16+00:00","dateModified":"2022-10-
24T07:47:19+00:00","description":"Sejalan dengan itu, UKM Kependudukan
Universitas Airlangga menggelar Airlangga Demography Festival (ADFEST) 2022.
\xa0","breadcrumb":{"@id":"https://www.unair.ac.id/tantangan-bonus-demografi-
indonesia-dan-pembangunan-berkelanjutan/#breadcrumb"},"inLanguage":"id-
ID","potentialAction":[{"@type":"ReadAction","target":
["https://www.unair.ac.id/tantangan-bonus-demografi-indonesia-dan-pembangunan-
berkelanjutan/"]}]},{"@type":"ImageObject","inLanguage":"id-
ID","@id":"https://www.unair.ac.id/tantangan-bonus-demografi-indonesia-dan-
pembangunan-berkelanjutan/#primaryimage","url":"https://www.unair.ac.id/wp-
content/uploads/2022/10/adw.png","contentUrl":"https://www.unair.ac.id/wp-
content/uploads/2022/10/adw.png","width":1345,"height":657,"caption":"Foto :

Dokumentasi Panitia Penulis : Satriyani Dewi Astuti"},
{"@type":"BreadcrumbList","@id":"https://www.unair.ac.id/tantangan-bonus-demografi-indonesia-dan-pembangunan-berkelanjutan/#breadcrumb","itemListElement":
[{"@type":"ListItem","position":1,"name":"Home","item":"https://www.unair.ac.id/"}
{"@type":"ListItem","position":2,"name":"Tantangan Bonus Demografi Indonesia dan Pembangunan Berkelanjutan\xa0"}]},
{"@type":"WebSite","@id":"https://www.unair.ac.id/#website","url":"https://www.una
 Airlangga Official Website","description":"Universitas Airlangga - Excellence with Morality","publisher":
{"@id":"https://www.unair.ac.id/#organization"},"potentialAction":
[{"@type":"SearchAction","target":
{"@type":"EntryPoint","urlTemplate":"https://www.unair.ac.id/?s={search_term_string}"},"query-input":"required
name=search_term_string}],"inLanguage":"id-ID"},
{"@type":"Organization","@id":"https://www.unair.ac.id/#organization","name":"Univ
 Airlangga Official Website","url":"https://www.unair.ac.id/","sameAs":[],"logo":
{"@type":"ImageObject","inLanguage":"id-
ID","@id":"https://www.unair.ac.id/#/schema/logo/image/","url":"https://www.unair.
content/uploads/2021/01/cropped-UNAIR_LOGO_SEAL_PRIMARY_72DPI-
1.png","contentUrl":"https://www.unair.ac.id/wp-content/uploads/2021/01/cropped-
UNAIR_LOGO_SEAL_PRIMARY_72DPI-
1.png","width":136,"height":136,"caption":"Universitas Airlangga Official
Website"},"image":{"@id":"https://www.unair.ac.id/#/schema/logo/image/"}},
{"@type":"Person","@id":"https://www.unair.ac.id/#/schema/person/d1f25b2f09db0607e
 NEWS","url":"https://www.unair.ac.id/author/unairnews/"}]]}</script>\n\t<!-- /
Yoast SEO plugin. -->\n\n\n<link rel=\'dns-prefetch\'
href=\'//fonts.googleapis.com\' />\n<link rel=\'dns-prefetch\'
href=\'//use.fontawesome.com\' />\n<link rel=\'dns-prefetch\' href=\'//s.w.org\'
/>\n<link rel="alternate" type="application/rss+xml" title="Universitas Airlangga
Official Website » Feed" href="https://www.unair.ac.id/feed/" />\n<link
rel="alternate" type="application/rss+xml" title="Universitas Airlangga Official

Scraping Beautifulsoup

```
url = rq.get("https://unair.ac.id/news")
```

+ Code   + Text

```
with open("index.html", 'w', encoding='utf_8') as nulis:
    nulis.write(url.text)


with open("index.html", 'r', encoding='utf_8') as htmll:
    soup = bs(htmll, 'html.parser')

listt = soup.find_all('h3' , class_='elementor-post__title')



data_judul = []
filter = []

for i in listt:
    splitted = re.split(r'[\n\t\f\v\r]+', i.find('a').get_text())

    for x in splitted[1:]:
```

```
        filter.append(x)

    for su in filter[0::2]:
        data_judul.append(su)
```

Save Hasil Scraping Headlines

```
df = pd.DataFrame(data_judul)
df.columns = ['Judul Berita']
df.to_csv('../scraping-judul.csv')

df.head()
```

	Judul Berita
0	Kontribusi Terhadap Keberlanjutan Lingkungan, ...
1	FK UNAIR Sambut 31 PPDS Program Hybrid TNI
2	Dekan FH UNAIR: Pendidikan Hukum Harus Berubah...
3	Jalani Hidup Sehat, Ini Yang Perlu Disiapkan
4	5 Mahasiswa UNAIR Ikuti Student Exchange dan K...

Crawling Beautifulsoup

```
urls = rq.get('https://unair.ac.id/news')


with open('index2.html', 'w', encoding= 'utf_8') as f:
    f.write(urls.text)

with open('index2.html', 'r') as duar:
    soup = bs(urls.text, 'html.parser')



all_links = soup.find_all('a', class_='elementor-post__read-more')
news_links = []

for i in all_links:
    ab = i['href']
    news_links.append(ab)


list_judul = []
list_date = []

for i in news_links:
```

```python
        buang = rq.get(i.lstrip("\'"))

        bakar = bs(buang.text, 'html.parser')
        judul = bakar.find_all('h3', class_='elementor-heading-title')
        date = bakar.find_all('span', class_='elementor-post-info__item--type-date')

        list_judul.append(judul[0].string)
        list_date.append(date[0].string)


raw_date = []
filtered_date = []

for so in list_date:
    splitted = re.split(r'[\n\t\f\v\r]+', so)

    for x in splitted[1:]:
        raw_date.append(x)

for su in raw_date[0::2]:
    filtered_date.append(su)
```

▼ Save Hasil Scraping Headlines dan Tanggal

```python
df2 = pd.DataFrame({'Judul' : list_judul, 'Date' : filtered_date})
df2.to_csv('../crawling-headlines.csv')

df2.head()
```

	Judul	Date
0	Kontribusi Terhadap Keberlanjutan Lingkungan, ...	Oktober 25, 2022
1	FK UNAIR Sambut 31 PPDS Program Hybrid TNI	Oktober 26, 2022
2	Dekan FH UNAIR: Pendidikan Hukum Harus Berubah...	Oktober 26, 2022
3	Jalani Hidup Sehat, Ini Yang Perlu Disiapkan	Oktober 26, 2022
4	5 Mahasiswa UNAIR Ikuti Student Exchange dan K...	Oktober 25, 2022

▼ Link GitHub

https://github.com/aryadgsp/Web-Scraping-and-Crawling

Note : File Crawling Play Station ada di GitHub

Colab paid products - Cancel contracts here