MODUL

BIG DATA ANALYTIC



Disusun oleh :

Sri Redjeki, S.Si, M.Kom.

Maria Mediatrix S., S.Kom., M.Eng.

SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER

AKAKOM

YOGYAKARTA

2020

# KATA PENGANTAR

# DAFTAR ISI

[KATA PENGANTAR ii](#_Toc15905388)

[DAFTAR ISI iii](#_Toc15905389)

[MODUL 1 NAMA POKOK BAHASAN 1](#_Toc15905390)

# MODUL 5 COLLECTING DATA

|  |  |
| --- | --- |
|  | CAPAIAN PEMBELAJARAN |
| 1. Mampu melakukan *Web Scraping* | |

|  |  |
| --- | --- |
|  | KEBUTUHAN ALAT/BAHAN/SOFTWARE |
| 1. Python (Anaconda / Miniconda) 2. Jupyter Notebook 3. BeautifulSoup4 4. Lxml 5. Requests | |

|  |  |
| --- | --- |
|  | DASAR TEORI |

|  |  |
| --- | --- |
|  | PRAKTIK |

1. Masuk ke folder c:\python
2. Install module pandas dan xlrd, dengan menjalankan :

C:\python>conda install beautifulsoup4

C:\python>conda install lxml

C:\python>conda install requests

1. Untuk memastikan installasi berhasil, masuk ke REPL python

(base) C:\Users\ASUS>python

Python 3.7.3 (default, Apr 24 2019, 15:29:51) [MSC v.1915 64 bit (AMD64)] :: Anaconda, Inc. on win32

Type "help", "copyright", "credits" or "license" for more information.

>>> import bs4

>>> import lxml

>>> import requests

*Jika tidak ada pesan Error, berarti installasi pandas dan xlrd sukses.*

1. Bukalah Jupyter Notebook, ketikkan script berikut ini :



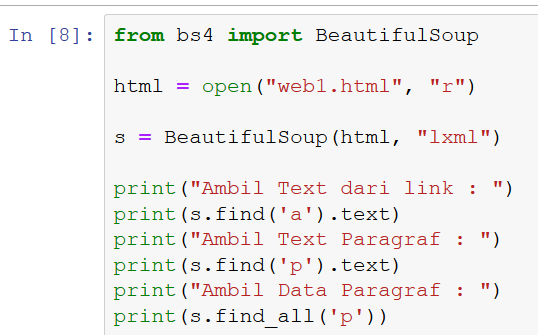
Jalankan dan amati hasilnya! Jelaskan fungsi .text dan .strip()

1. Buatlah file baru dan copykan script program diatas, kemudian modifikasilah seperi dibawah ini :



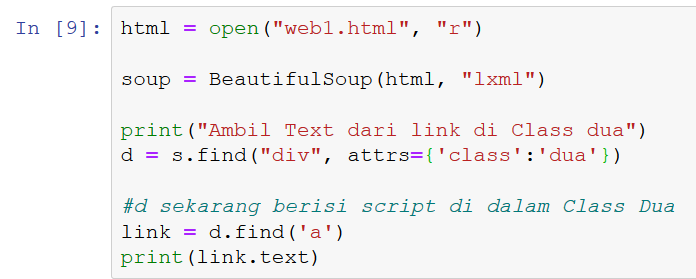
Jalankan dan amati hasilnya! Jelaskan setiap perintah print()

1. Gunakan file web1.html, kemudian buatlah file baru dan ketikkan perintah seperti dibawah ini :



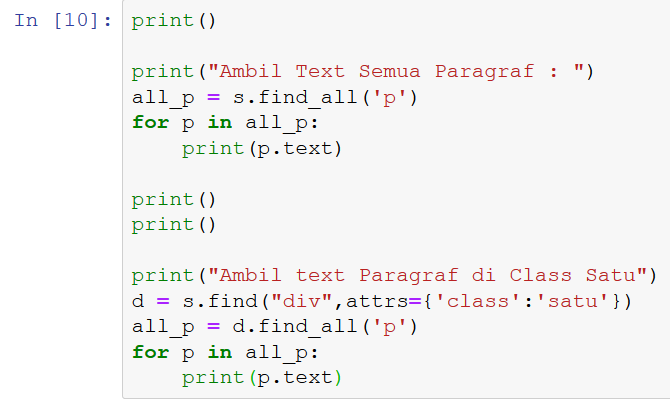
Jalankan dan amati hasilnya! Jelaskan perbedaan fungsi dan hasil dari find() dan find\_all()

1. Buatlah file baru dan ketikkan script berikut ini



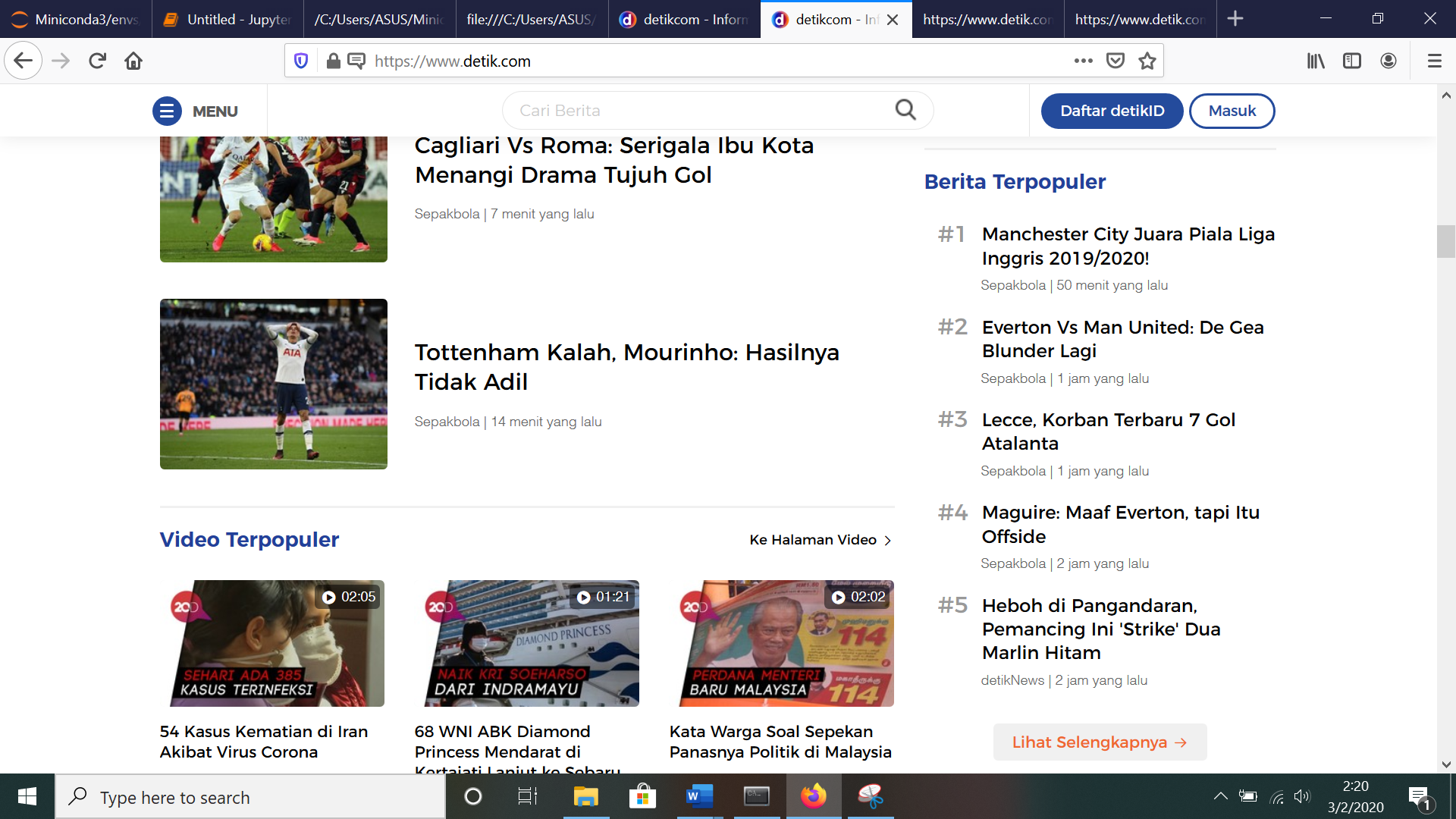
Jalankan dan amati hasilnya.

1. Kemudian modifikasilah file dari no.7, tambahkan perintah seperti berikut ini



Jalankan dan amati hasilnya.

1. Bukalah Alamat Web: <https://www.detik.com>, inspect Element pada bagian “Terpopuler” :



1. Klik kanan pada halaman tersebut, kemudian View Source



1. Buatlah file baru dan ketikkan perintah berikut ini



Jalankan dan amati hasilnya.

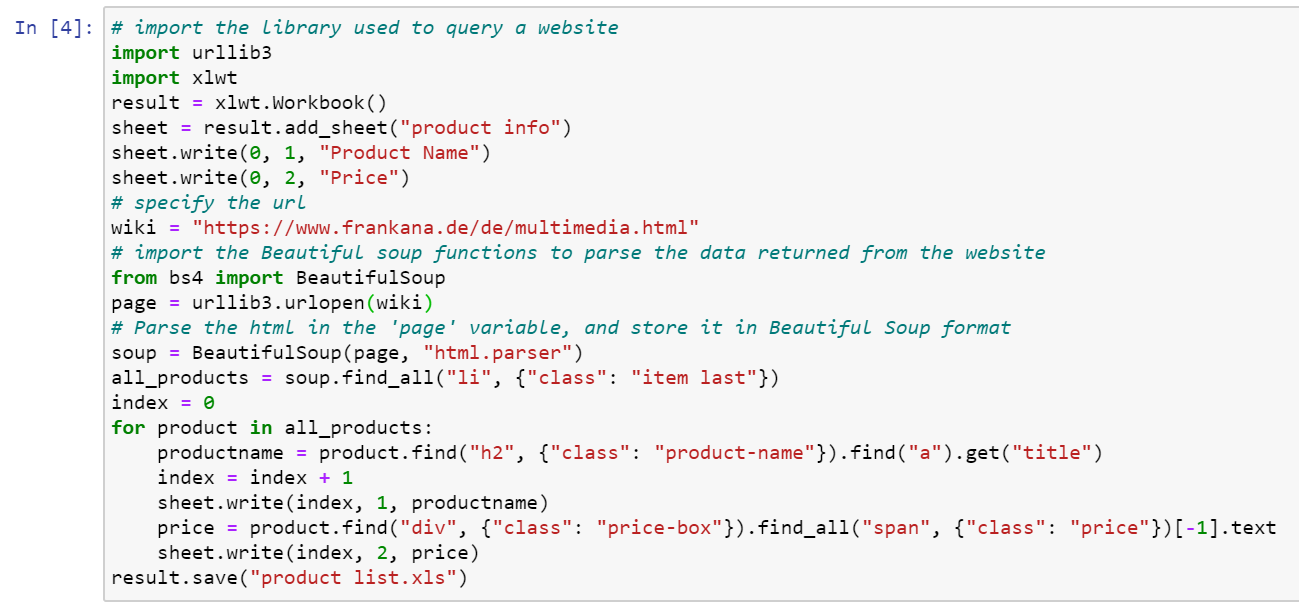
1. Gunakan file tabel.html lalu buat file baru, kemdian ketikkan script berikut ini

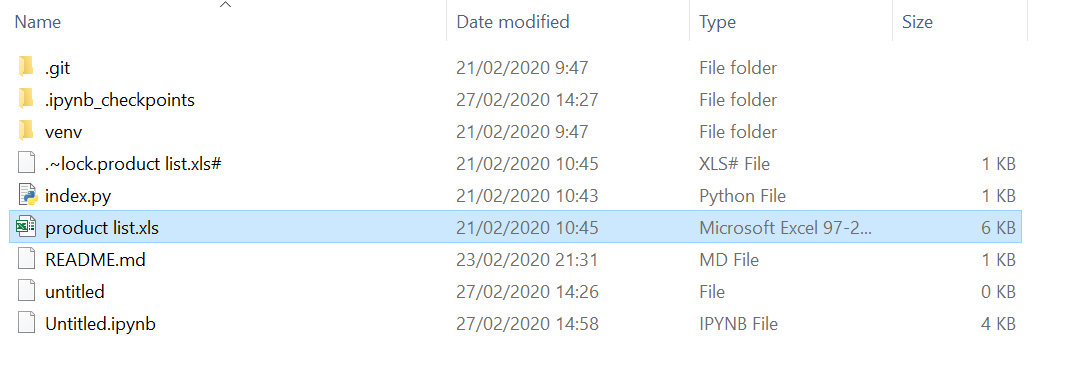


Jalankan dan amati hasilnya

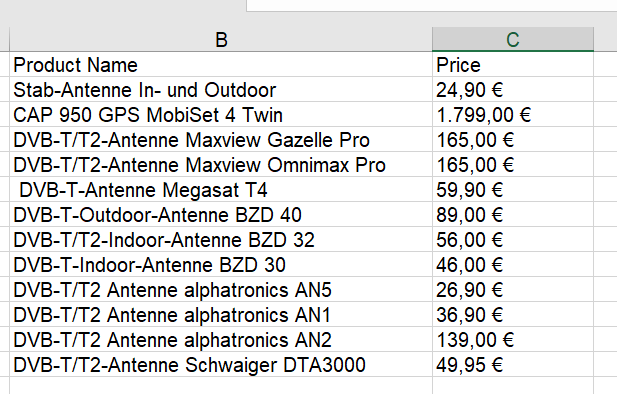
**Scraping data Online Shop**

* + - 1. Kita buka contoh online shop <https://www.frankana.de/>
      2. Ketikkan perintah berikut ini:

Hasil akan disimpan dengan nama file product list dalam bentuk excel. File disimpan dalam satu folder sama dengan nama file python kita.



* + - 1. Pastikan isi dari file excel tersebut



|  |  |
| --- | --- |
|  | LATIHAN |

|  |  |
| --- | --- |
| tugas.png | TUGAS |

* + - 1. Bukalah Web <https://www.akakom.ac.id/>, kemudian lakukan scraping untuk “Berita Terbaru” (yang ditampilkan hanya **Judul** dan **Tanggal**).

|  |  |
| --- | --- |
|  | REFERENSI |