TUGAS BESAR_SCRAPY MRIS



Oleh:

Joko Susilo 1914321012

UNIVERSITAS BHAYANGKARA SURABAYA TEKNIK INFORMATIKA 2021

Ambil data dari https://www.worldnovel.online/ Dengan Scrapy

Link Github

 $\underline{https://github.com/jokosby123/Tugas-Besar-Scrapy-1914321012-Malam-Ubhara-Surabaya-\underline{.git}$

Langkah-Langkah

- 1. Membuat project scrapy dengan anaconda prompt
- 2. crawling atau di ambil datanya seperti gambar di bawah ini urls = ['https://www.worldnovel.online/super-gene-optimization-fluid/chapter-501-a-hero-who-cares/']
- 3. Membuat variable limit untuk membatasi halaman yang ingin di crawling

 limit = 9
- 4. Setelah itu mengambil data di laman yang dituju setelah selesai mengambil data saya membuat agar bisa mengambil data yang sebelumnya atau prev chapter seperti gambar di bawah

```
def parse(self, response):
    for next_page in response.css('a.prevchap'):
        if self.limit > 0 :
            yield response.follow(next_page, self.parse)
            self.limit -= 1
```

5. Setelah itu mengabil data Judul Novel , Isi Novel,dan Link Novel

```
for title in response.css('.post-title.mb-4.font-weight-bold'):
    yield {'title': title.css('::text').get()}

for link in response.css('div.chapter-fill'):|
    yield {'isi': link.css('p::text').getall()}

for link in response.css('a.prevchap'):
    yield {'link': link.css('::attr(href)').get()}
```

6. Setelah itu saya jalankan source code tersebut dengan anaconda prompt

Anaconda Prompt (anaconda3)

```
(base) C:\Users\Joko Susilo>cd uas
(base) C:\Users\Joko Susilo\UAS>scrapy crawl crawling -o Ambil_data.jl
```

Dan hasil crawling saya simpan di Ambil_data.jl

7. Setelah saya RUN atau dijalankan

Macros Prompt personal more paling [costs. Scolab-olah dia tidak perlu melihet atas berpikir untuk menemban titik yang bias dia merfantum.", 'Bahban Xia Fei tidak mengharaphan hasil ini. Ketta menendijari ka merfantum.', "Tidak dia sengharaphan hasil ini. Ketta menendijari ka merfantum.' "Tidak menemban tidak menemban t

8. Data yang telah saya simpan di Ambil_data.jl

