UAS MULTIMEDIA RETRIVAL

Nama : M. Rizqika Brimadi. H

NIM 2014311035

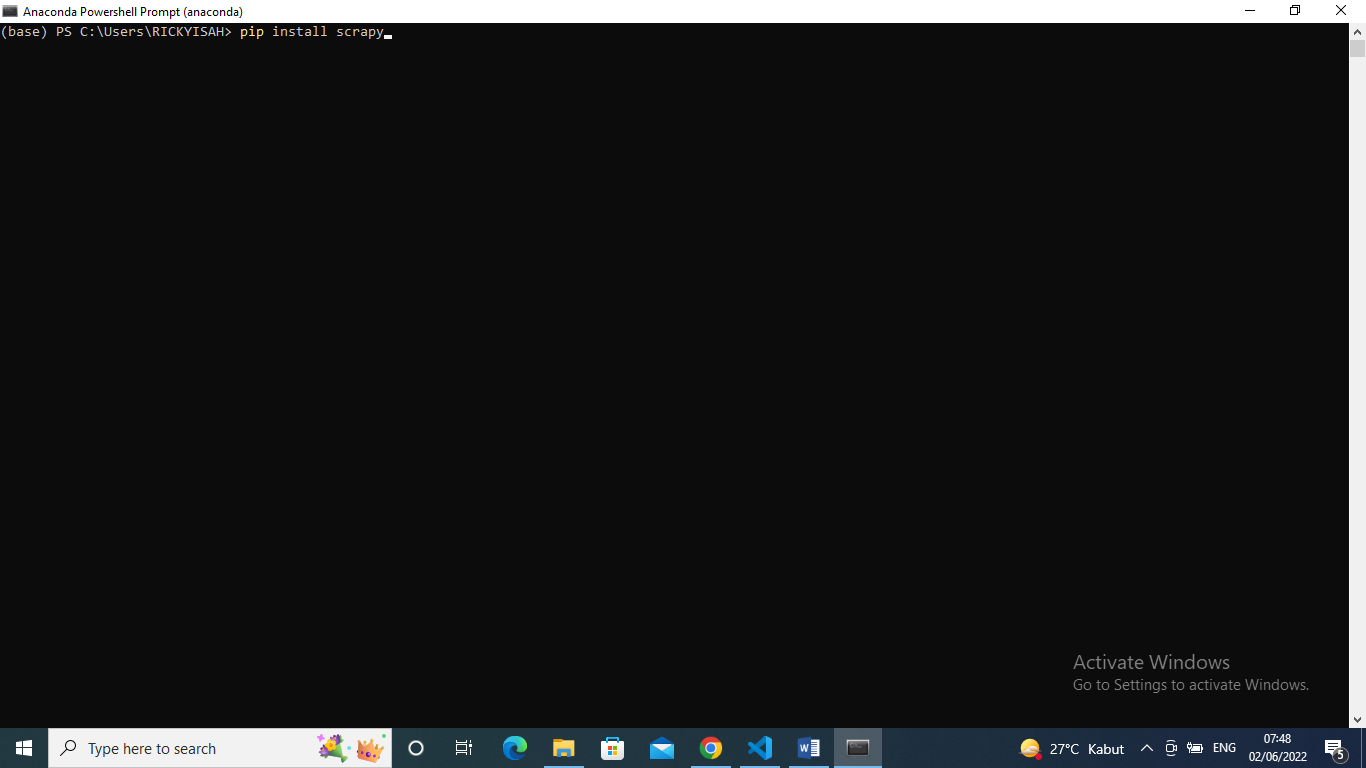
Kelas : Pagi

1. Mohon maaf nomor 2 tidak bisa di ambil datanya



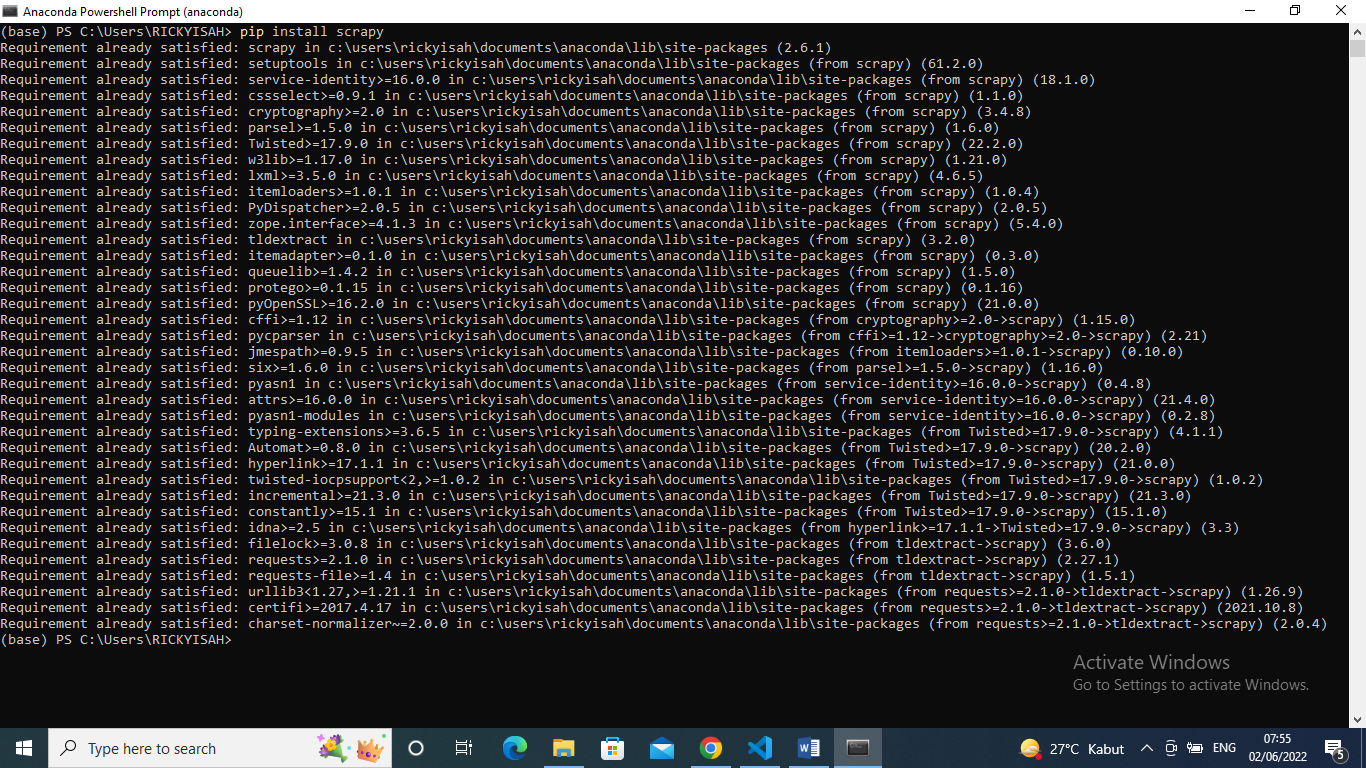
1. Novel NIM GENAP

Sebelum scraping novel, pertama install terlebih dahulu ‘ scrapy ‘ di cmd / terminal anaconda seperti gambar dibawah ini, lalu klik enter dan tunggu prosesnya dalam beberapa saat



Setelah scrapy sudah terinstall maka akan muncul tulisan ‘ already ‘ tiap penginstallannya

Seperti contoh gambar dibawah ini



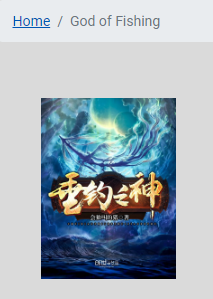
Selanjutnya membuat projek scrapy dan beri Nama pada projek anda sesuai dengan apa yang ingin dibuat

C:\Users\RICKYISAH\Pictures\Screenshots\Screenshot_2.png

Project tersebut akan membuat folder sesuai dengan letak yang dibuat

C:\Users\RICKYISAH\Pictures\Screenshots\Screenshot_4.png

Setelah itu pilih novel yang ingin kita scraping disini saya memilih 5 novel berikut ini, saya beri contoh salah satu novel yang saya ambil dari salah satu website, berikut contohnya seperti gambar dibawah ini



Kita ambil coding scraping yang ada pada website “ [Scrapy Tutorial — Scrapy 2.6.1 documentation](https://docs.scrapy.org/en/latest/intro/tutorial.html) ”

**import scrapy**

**class QuotesSpider**(scrapy.Spider): name = "quotes"

**def** start\_requests(self):

urls = [

'https://quotes.toscrape.com/page/1/', 'https://quotes.toscrape.com/page/2/',

]

**for** url **in** urls:

**yield** scrapy.Request(url=url, callback=self.parse)

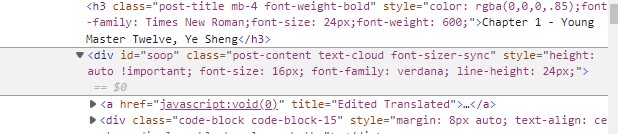
**def** parse(self, response):

terlebih dahulu ambil link novel yang mau di scraping contoh berikut ini

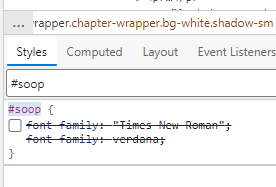
<https://www.worldnovel.online/novel-god-of-fishing/1/>

Sebelumnya ambil struktur novelnya yang mau kita ambil teksnya

kita ambil selector yang sekitarnya bisa membaca seluruh cerita novel yang kita ambil

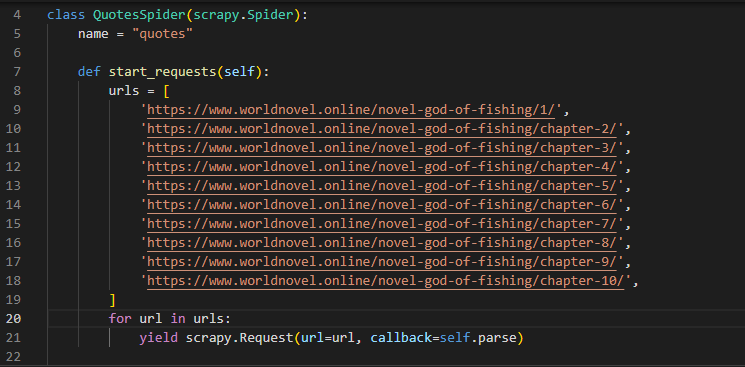


Ambil selector tersebut terus kita copy dan di paste di bagian “ filter “ maka akan muncul “ soop “ yang digunakan untuk memanggil “ respons.css “

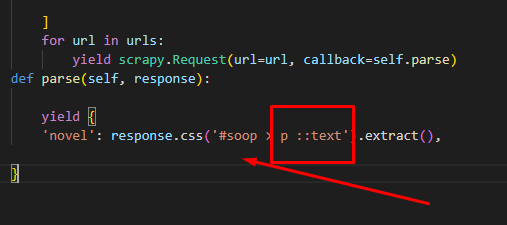


Lanjut ke VSCode

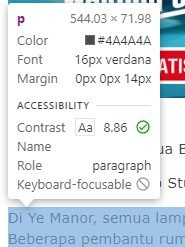
Link novel per chapter di copy dan di masukkan sesuai dengan gambar ini



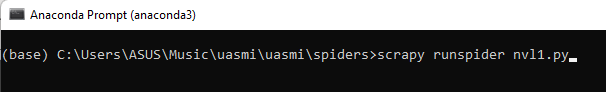
Selector ‘ #soop ‘ yang diambil tadi, di masukkan sesuai gambar ini dan ditambahkan variable “ p “ karena struktur dari teks novel tersebut berada di element “ p “



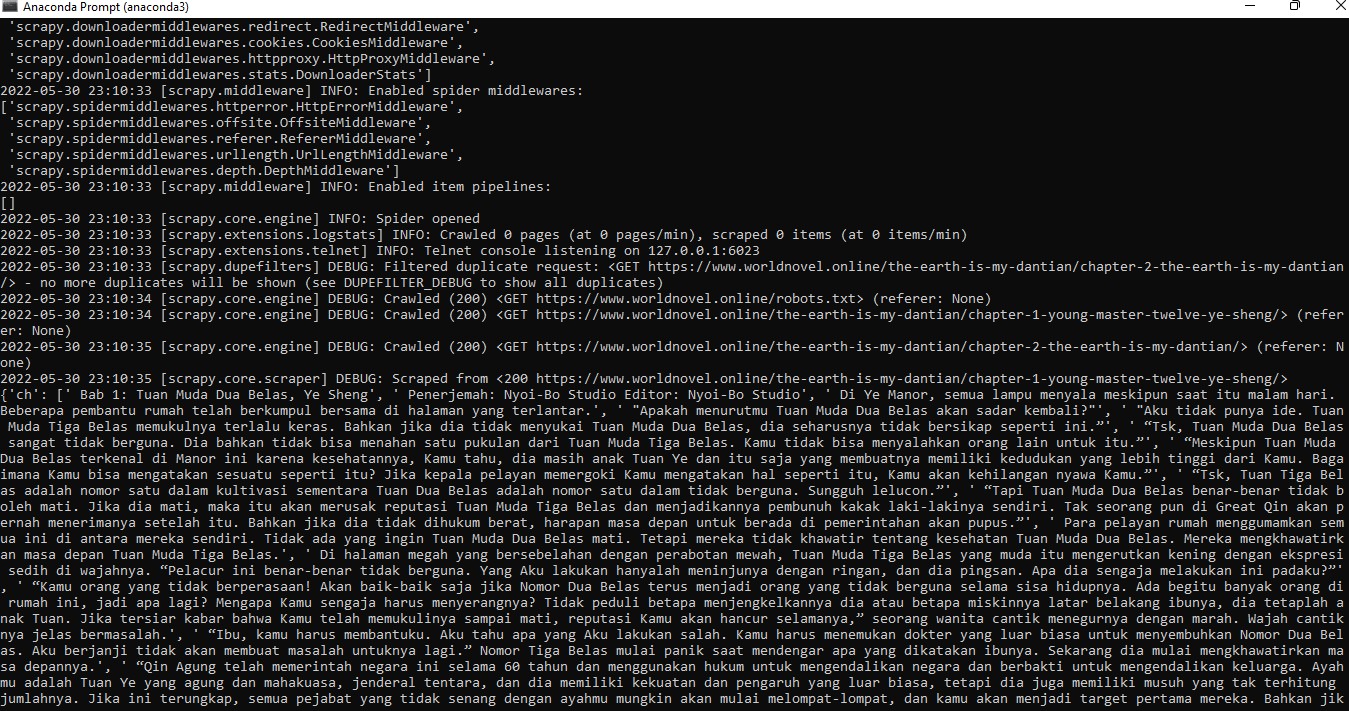
Element “ p “ pada inscpeck terdapa semua teks novel



Jalan kan python tersebut dengan command “ scrapy runspider ( namafile.py ) “

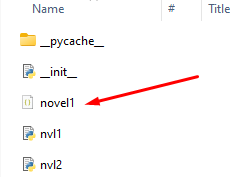


Apabila berhasil hasil dari scrapy novel tersebut akan muncul pada CMD / Terminal yang dijalankan

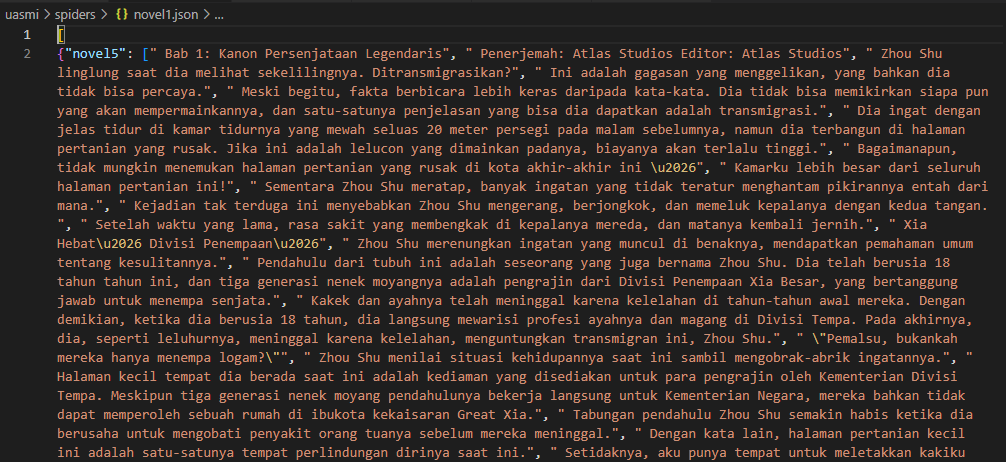


Setelah itu di save dalam bentuk json / csv, dengan command “ scrapy crawl (sipdername) -o (namafile.csv/json) “

Jika berhasil makan file hasil scraping tadi akan tersimpan di file sesuai dengan letak lokasi projek yang dibuat sebelumya



Hasil scraping yang tersimpan dalam bentuk .json



Link Github : <https://github.com/rifaldzikir/rifaldzikir_.git>