Laporan Tugas Besar Multimedia Retrieval Information System Semester 4



Nama: Wendi Rosalia Pratiwi

Nim: 1914311055

UNIVERSITAS BHAYANGKARA SURABAYA

TEKNIK INFORMATIKA

2020/2021

Tugas No.3

SCRAPPY

Pertama, ambil 5 bagian atas recently added chapter.

Title	Release	Author	Time
God Of Soul System	Chapter 601 - Sternritter	Ye Nan Ting Feng	3 hours ago
Another World's Versatile Crafting Master	Chapter 1059 - Aid	Zhuang Bifan	3 hours ago
I Have Countless Legendary Swords!	Chapter 752 - The Fight for the Sovereign Position	Ren Wo Xiao	3 hours ago
I Might Be A Fake Cultivator	Chapter 1783 - Destroying This World	Bright Moonlight Frost	3 hours ago
Scholar's Advanced Technological System	Chapter 1531 - Float Out of Water	Morning Star LL, 晨星LL	3 hours ago
Picking Up Attributes From Today	Chapter 428 - Destined To Buy Credits, Endless Darkness	Panda Reunion, 团圆	3 hours ago
Plague Doctor	Chapter 459 - Faceless God with a Thousand Faces	Wallee The Robot, 机	4 hours ago
Forty Millenniums of Cultivation	Chapter 2698 - Super Jump Point!	The Enlightened Mast	4 hours ago
I Have A Super USB Drive	Chapter 495 - : Destroy the Evidence	Darkness Black Bear,	4 hours ago
A Record of a Mortal's Journey to Immortality	Chapter 2216 Black Gourd City	Wang Yu	4 hours ago
Worlds Apocalypse Online	Chapter 1334 - Return of a fallen star	Yan Huo Cheng Cheng	4 hours ago

Kemudian, membuat project baru.

Shell chapter.

Lalu scrapy halaman.

```
od Supplierc/Ms\nc/a>',
'va here'"https://www.worldnovel.online/novel/legendary-sword.god/" title="Legendary Sword God">\ncfigure class="novel-cover">\ncfigure class="lazy" data-src-"https://www.worldnovel.online/novel.online/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/novel.forline/
```

Lalu mengambil link:

Masukkan sc.

```
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
🕞 🔒 🖺 📭 🦰 😘 📤 | 🕹 🐚 🛍 | Þ 🕒 C | ## 🗽 | 🔍 🔍 🎮 🖼 🗔 🗔 1 🗐 🗜 🖫 💋 😉 🥙 🕒 🗉 🗩

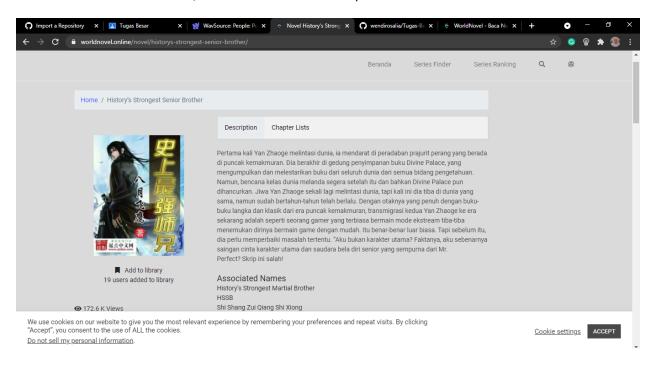
    yt_spiders.py 

     class webnovel(scrapy.Spider):
          name = "yt"
          url = "https://www.worldnovel.online/"
          page = urlopen(url)
          html = page.read().decode("utf-8")
          soup = bs(html, "html.parser")
          soup.find_all("url")
          def start_requests(self):
    urls = [
 17
18
                  'https://www.worldnovel.online/'
              for url in urls:
                 yield scrapy.Request(url=url, callback=self.parse)
            return soup.find("a", text="Recently Added Chapters")
         def extract_links(soup):
             return [a['title'] for a in soup.select("#mw-pages li a")]
        with requests.Session() as session:
             content = session.get(url).content
 30
             soup = bs(content, 'lxml')
 32
             links = extract links(soup)
             next_link = Recently(soup)
 36
          while next_link is not None: # while there is a Next Page link
             url = urljoin(url, next link['href'])
              content = session.get(url).content
 39
              soup = bs(content, 'lxml')
              links += extract_links(soup)
                               Python file
```

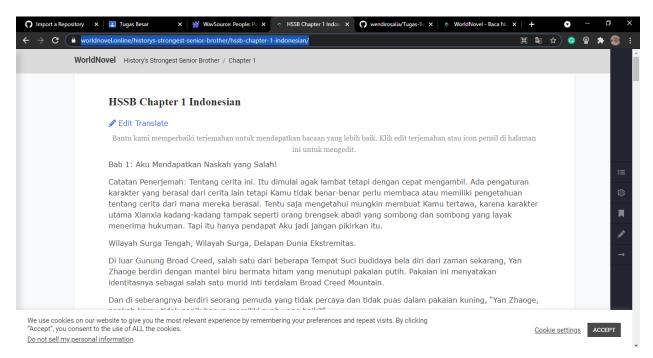
Berikutnya lakukan crawl.



Setelah itu pada laman web yang kita pilih 5 dari yang sebelumnya disourcecode scrapy crawl, kita buka salah satu link novel tersebut, dan buka di laman web seperti dibawah ini:



Lalu kita buka Chapter 1 – 10, dengan contoh saya buka Chapter 1 seperti dibawah ini:



Kita copy link tersebut, kemudian kita kembali ke cmd anaconda, kita melakukan scrapy shell "scrapy shell https://www.worldnovel.online/historys-strongest-senior-brother/hssb-chapter-1-indonesian/" seperti dibawah ini :

```
Package (Property (asconda)): stopy shell https://www.worldnovel.online/historys-trongest-senior-brother/hisb-chapter-l-indonesian/
Press ENTER to continue...
in [16]
in [16]
to you really want to exit ([y]/n)? y

(base) C. Useraluser\Decoments\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\TugasBesar\RTS\T
```

Lalu kita ketikan "response.css('p::text')" untuk menampilkan halaman teksnya seperti dibawah ini:

```
Accords Frompt (nacronda) scropy shell those, Jestell ()

[5] response.cs([milect.]) getall ()

[5] response.cs([milect.]) getall ()

[5] Sab 1: Aku Mendapatkan Naskah yang Salah!,

(Catatan Penerjenah: Tentang cerita ini. Itu dimulai agak lambat tetapi dangan cepat mengambil. Ada pengaturan karakter yang berasal dari cerita lain tetapi Kamu tidak berar-benar perlu membaca daru memiliki pengetahuan tentang cerita dari mana mereka berasal. Tentu saja mengetahui mungkin membuut kamu tetawa, karena karakter utawa kerataan Penerjenah: Verangan pengan penju membaca daru memiliki pengetahuan tentang cerita dari mana mereka berasal. Tentu saja mengetahui mungkin membuut kamu tetawa, karena karakter utawa kerataan pengetahun mengan pengetahun mengan pengan penju membaca hita dari pengan pengan pengan pengan kelipah Sungan pengan pengan pengan kelipah Sungan pengan pen
```



Note:	
Untuk repository github berada dibawah ini	
No.3:	

https://github.com/wendirosalia/Tugas-Besar-Scrapy-1914311055-Pagi-Ubhara-Surabaya