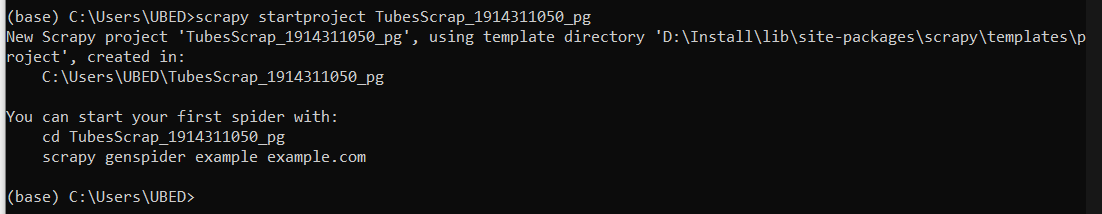
**Nama : Ubaidillah**

**NIM : 1914311050**

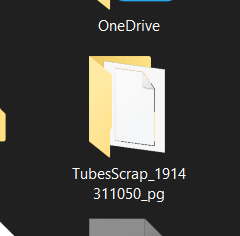
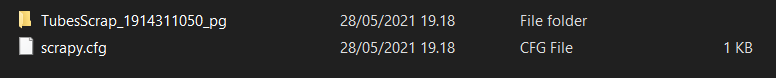
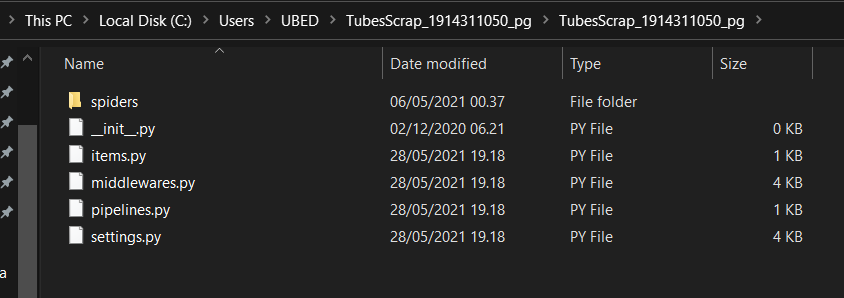
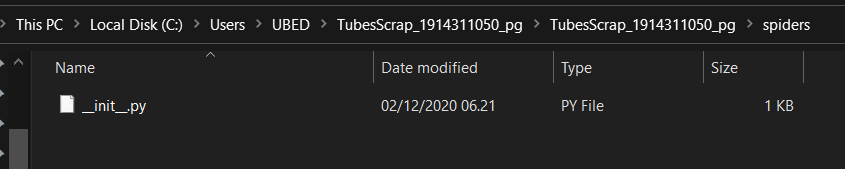
Soal No 2 Bsoup Link Code :

https://github.com/ubaidillah219/Tugas-Besar-Scrapy-NIM-Pagi-Ubhara-Surabaya

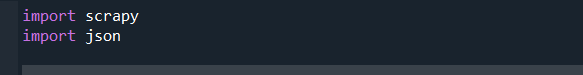
1. Setelah scrapy telah siap digunakan gunakan scrapy startproject TubesScrap\_1914311050\_pg untuk membuat project



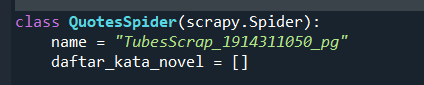
1. Lalu akan Otomatis terbentuk file pada directory tersebut sesuai nama file yang digunakan dan berisi sebagai berikut :



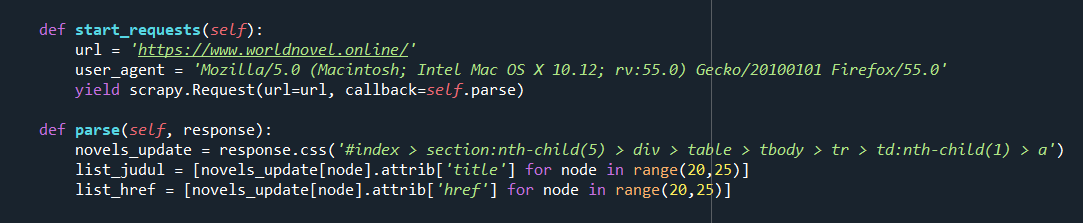
1. Lalu didalam folder spiders buatlah code seperti ini untuk mengscraping, pertama meng-import module yang digunakan



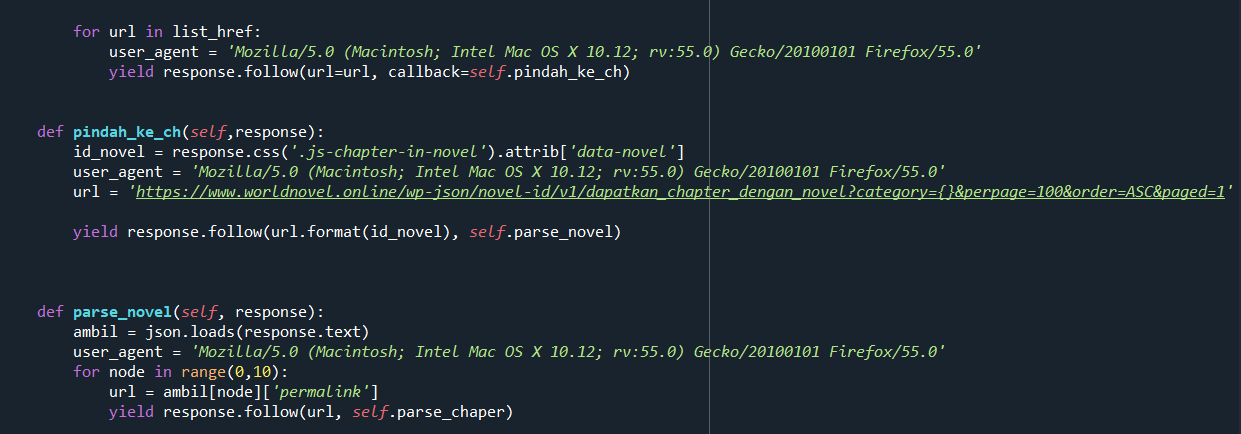
1. Memberikan nama pada code untuk nantinya digunakan untuk crawl pada cmd



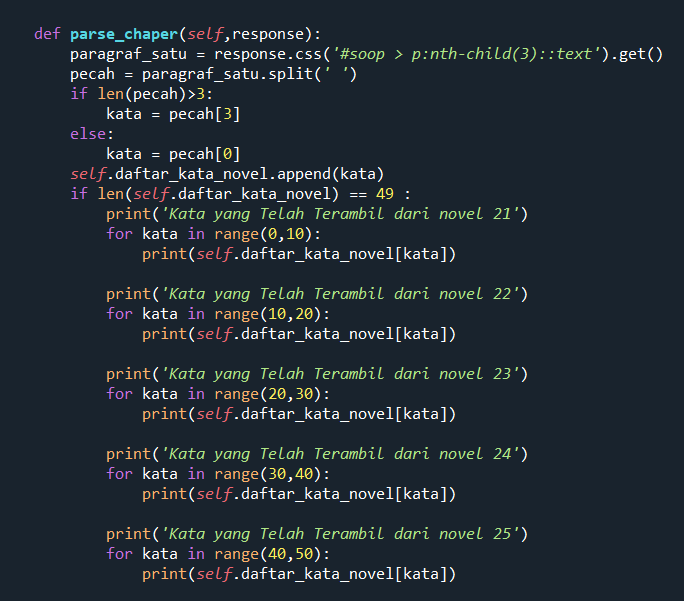
1. Memasukkan url yang akan di scrapy lalu memberikan user\_agent sebagai header untuk memberikan identitas kepada website, lalu mengambil inspect dari novel yang diinginkan tuntuk diambil href nya karena disini saya 5 terbawah jadi saya gunakan range 20 dan 25 untuk mengambil title dan href dari setiap novel tersebut



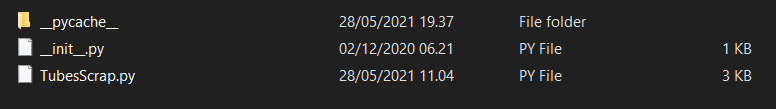
1. Lalu kita mengscraping tiap chapternya dengan menggunakan code dibawah ini

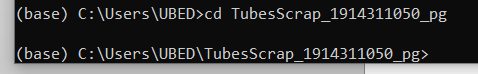


1. Kemudian untuk setiap chapternya kita ambil 1 kata per chapter dengan code dibawah



1. Setelah membuat code tersebut pada spiders dan di save pada folder tersebut kedmudian kita masuk kedalam file dengan cd TubesScrap\_1914311050\_pg lalu kita crawl pada cmd

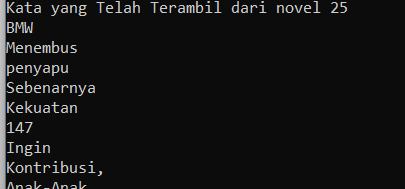






1. Maka ditemukan hasil sebagai berikut dan ini merupakan ss an recently added chapter nya :





Kesimpulan : 1 kata pada setiap 10 halaman chapter telah terambil namun pada saat dicoba untuk code tf idf selalu terjadi ke eror an dan tdak tahu untuk problem solve nya jadi hasil yang saya tunjukan hanya sampai di sini

Sekian tugas dari saya yang sudah saya kerjakan semampu saya.

**Terima Kasih**