Penjelasan

Inputan

import scrapy

import os

folder = "./Chapt"

class Tb3Spider(scrapy.Spider):

    name = 'TB3'

    allowed\_domains = ['worldnovel.online']

    start\_urls = ['http://worldnovel.online/']

    def start\_requests(self):

        urls = [

            # Super Detective in the Fictional World

            "https://www.worldnovel.online/super-detective-in-the-fictional-world/chapter-100-new-partner-new-case-and-new-star/",

            "https://www.worldnovel.online/super-detective-in-the-fictional-world/chapter-99-cash-over-promotion/",

            "https://www.worldnovel.online/super-detective-in-the-fictional-world/chapter-98-good-news-and-bad-news/",

            "https://www.worldnovel.online/super-detective-in-the-fictional-world/chapter-97-brocks-blessing-and-an-unexpected-offer/",

            "https://www.worldnovel.online/super-detective-in-the-fictional-world/chapter-96-extra-meal-and-pleasant-surprise/",

            "https://www.worldnovel.online/super-detective-in-the-fictional-world/chapter-95-bet-dinner-and-pick-me-up/",

            "https://www.worldnovel.online/super-detective-in-the-fictional-world/chapter-94-virtue-and-wit/",

            "https://www.worldnovel.online/super-detective-in-the-fictional-world/chapter-93-great-loot-and-bittersweet-ability/",

            "https://www.worldnovel.online/super-detective-in-the-fictional-world/chapter-92-patrol-and-harvest/",

            "https://www.worldnovel.online/super-detective-in-the-fictional-world/chapter-91-part-time-patrol-officers/",

            # I Have A Super USB Drive

            "https://www.worldnovel.online/super-usb-bahasa/chapter-486-verify/",

            "https://www.worldnovel.online/super-usb-bahasa/chapter-485-weeping-angel/",

            "https://www.worldnovel.online/super-usb-bahasa/chapter-484-secret-meeting/",

            "https://www.worldnovel.online/super-usb-bahasa/chapter-483-the-earth-federation-united-front/",

            "https://www.worldnovel.online/super-usb-bahasa/chapter-482-the-law-of-causality/",

            "https://www.worldnovel.online/super-usb-bahasa/chapter-481-wipe-out/",

            "https://www.worldnovel.online/super-usb-bahasa/chapter-480-transformation/",

            "https://www.worldnovel.online/super-usb-bahasa/chapter-479-appreciation/",

            "https://www.worldnovel.online/super-usb-bahasa/chapter-478-dimensions-child/",

            "https://www.worldnovel.online/super-usb-bahasa/chapter-477-to-live-in-the-face-of-death/",

            # Scholar’s Advanced Technological System

            "https://www.worldnovel.online/scholars-advanced-technological-system/chapter-1101-something-is-wrong/",

            "https://www.worldnovel.online/scholars-advanced-technological-system/chapter-1100-the-terrifying-fields-medalis/",

            "https://www.worldnovel.online/scholars-advanced-technological-system/chapter-1099-player-evaluation/",

            "https://www.worldnovel.online/scholars-advanced-technological-system/chapter-1098-a-taste-of-evil/",

            "https://www.worldnovel.online/scholars-advanced-technological-system/chapter-1097-second-closed-beta/",

            "https://www.worldnovel.online/scholars-advanced-technological-system/chapter-1096-handshake-between-two-giants/",

            "https://www.worldnovel.online/scholars-advanced-technological-system/chapter-1095-pigs-are-flying/",

            "https://www.worldnovel.online/scholars-advanced-technological-system/chapter-1094-price-is-just-one-of-the-reasons/",

            "https://www.worldnovel.online/scholars-advanced-technological-system/chapter-1093-perelmans-visi/",

            "https://www.worldnovel.online/scholars-advanced-technological-system/chapter-1092-winter-has-arrived-in-silicon-valley/",

            # Another World’s Versatile Crafting Master

           "https://www.worldnovel.online/another-worlds-versatile-crafting-master/chapter-1052-fleet/",

            "https://www.worldnovel.online/another-worlds-versatile-crafting-master/chapter-1051-undercurrent/",

            "https://www.worldnovel.online/another-worlds-versatile-crafting-master/chapter-1050-reincarnation/",

            "https://www.worldnovel.online/another-worlds-versatile-crafting-master/chapter-1049-world-sword/",

            "https://www.worldnovel.online/another-worlds-versatile-crafting-master/chapter-1048-golden/",

            "https://www.worldnovel.online/another-worlds-versatile-crafting-master/chapter-1047-meditation-ground/",

            "https://www.worldnovel.online/another-worlds-versatile-crafting-master/chapter-1046-wind-and-thunder-beast-king//",

            "https://www.worldnovel.online/another-worlds-versatile-crafting-master/chapter-1045-vipers-poison/",

            "https://www.worldnovel.online/another-worlds-versatile-crafting-master/chapter-1044-viper/",

            "https://www.worldnovel.online/another-worlds-versatile-crafting-master/chapter-1043-prehistoric-times/",

            # Everlasting

            "https://www.worldnovel.online/everlasting/chapter-1152-the-dark-kirin-appears/",

            "https://www.worldnovel.online/everlasting/chapter-1151-hidden-black-hand/",

            "https://www.worldnovel.online/everlasting/chapter-1150-mid-grade-divine-artifact/",

            "https://www.worldnovel.online/everlasting/chapter-1149-tomb/",

            "https://www.worldnovel.online/everlasting/chapter-1148-graveyard-of-divine-dragons/",

            "https://www.worldnovel.online/everlasting/chapter-1147-reinforcements-arrive/",

            "https://www.worldnovel.online/everlasting/chapter-1146-spatial-gate/",

            "https://www.worldnovel.online/everlasting/chapter-1145-entering-the-battlefield/",

            "https://www.worldnovel.online/everlasting/chapter-1144-breaking-out-from-the-ambush/",

            "https://www.worldnovel.online/everlasting/chapter-1143-dispatched/",

        ]

        for url in urls:

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

    def parse(self, response):

        judul = response.url.split("/")[-3]

        chapter = response.url.split("/")[-2]

        if not os.path.exists(folder):

            os.makedirs(folder)

        filename = os.path.join(folder, f"{judul}-{chapter}.html")

        with open(filename, "wb") as f:

            f.write(response.body)

            self.log(f"Saved file {filename}")

setelah proses download saya masukkan filenya ke folder **CHAPT** yang nantinya akan di proses

TF-IDF Code Program

from bs4 import BeautifulSoup as bs

import os

import re

import pandas as pd

import numpy as np

from sklearn.feature\_extraction.text import TfidfVectorizer

folder = "/content/Tugas-Besar-Scrapy-1914311022-Pagi-Ubhara-Surabaya/Tubes3/Chapt"

for filename in os.listdir(folder):

    if filename.endswith(".html"):

        fname = os.path.join(folder, filename)

        print("Filename: {}".format(fname))

def teks\_utuh(fname):

    with open(fname, "r", encoding="utf8") as f:

        soup = bs(f.read(), "html.parser")

        isi = soup.find\_all("div",{"id": "asli"}, "p")

        pola = re.compile('<.\*?>')

        teks\_awal = re.sub(pola, '', str(isi))

        new = teks\_awal[1:teks\_awal.find('&lt; Chapter')]

        teks\_bersih = new.replace(". ", ".").replace(" , ", " ").replace("www.worldnovel.online", "").replace("\xa0", " ")

        return teks\_bersih

def nilai\_tfidf(doc):

    feature\_names = tfidf.get\_feature\_names()

    for col in doc.nonzero()[1]:

        print (feature\_names[col], ' - ', doc[0, col])

def top10(doc):

    feature\_array = np.array(tfidf.get\_feature\_names())

    tfidf\_sorting = np.argsort(doc.toarray()).flatten()[::-1]

    n = 10

    top\_n = feature\_array[tfidf\_sorting][:n]

    print(top\_n)

Output :



