

Program no -14

Aim: Implement a simple web crawler

Program

```
import requests
import lxml
from bs4 import BeautifulSoup
url = "https://www.rottentomatoes.com/top/bestofrt/"
headers = {
    'User-Agent': 'Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36
(KHTML,like Gecko) Chrome/63.0.3239.132 Safari/537.36 QIHU 360SE'
}
f = requests.get(url,headers = headers)
movies_lst = []
soup =BeautifulSoup(f.content,'html.parser')
movies =soup.find('table', {
    'class': 'table'
}).find_all('a')
print(movies)
num = 0
for anchor in movies:
    urls = 'https://www.rottentomatoes.com' + anchor['href']
    movies_lst.append(urls)
print(movies_lst)
num += 1
movie_url = urls
movie_f = requests.get(movie_url, headers = headers)
movie_soup = BeautifulSoup(movie_f.content,'lxml')
movie_content = movie_soup.find('div', {
    'class': 'movies_synopsis clamp clamp-6 js-clamp'
})
print(num, urls, '\n', 'Movie:' + anchor.string.strip())
print('Movie Info:' + movie_content.string.strip())
```

OUTPUT

```
C:\Users\ajcemca\AppData\Local\Programs\Python\Python39\python.exe C:/Users/ajcemca/PycharmProjects/pythonProject/pythonProject1/webcrawlers.py
[<a class="unstyled articleLink" href="/m/it_happened_one_night">
    It Happened One Night (1934)</a>, <a class="unstyled articleLink" href="/m/citizen_kane">
    Citizen Kane (1941)</a>, <a class="unstyled articleLink" href="/m/the_wizard_of_oz_1939">
    The Wizard of Oz (1939)</a>, <a class="unstyled articleLink" href="/m/modern_times">
    Modern Times (1936)</a>, <a class="unstyled articleLink" href="/m/black_panther_2018">
    Black Panther (2018)</a>, <a class="unstyled articleLink" href="/m/parasite_2019">
    Parasite (Gisaengchung) (2019)</a>, <a class="unstyled articleLink" href="/m/avengers_endgame">
    Avengers: Endgame (2019)</a>, <a class="unstyled articleLink" href="/m/1003707-casablanca">
    Casablanca (1942)</a>, <a class="unstyled articleLink" href="/m/knives_out">
    Knives Out (2019)</a>, <a class="unstyled articleLink" href="/m/us_2019">
    Us (2019)</a>, <a class="unstyled articleLink" href="/m/toy_story_4">
    Toy Story 4 (2019)</a>, <a class="unstyled articleLink" href="/m/lady_bird">
    Lady Bird (2017)</a>, <a class="unstyled articleLink" href="/m/mission_impossible_fallout">
    Mission: Impossible - Fallout (2018)</a>, <a class="unstyled articleLink" href="/m/blackklansman">
    BlackKlansman (2018)</a>, <a class="unstyled articleLink" href="/m/get_out">
    Get Out (2017)</a>, <a class="unstyled articleLink" href="/m/the_irishman">
    The Irishman (2019)</a>, <a class="unstyled articleLink" href="/m/godfather">
    The Godfather (1972)</a>, <a class="unstyled articleLink" href="/m/mad_max_fury_road">
    Mad Max: Fury Road (2015)</a>, <a class="unstyled articleLink" href="/m/spider_man_into_the_spider_verse">
    Spider-Man: Into the Spider-Verse (2018)</a>, <a class="unstyled articleLink" href="/m/moonlight_2016">
    Moonlight (2016)</a>, <a class="unstyled articleLink" href="/m/sunset_boulevard">
```

```
webcrawlers x
His Girl Friday (1940)</a>, <a class="unstyled articleLink" href="/m/portrait_of_a_lady_on_fire">
Portrait of a Lady on Fire (Portrait de la jeune fille en feu) (2020)</a>, <a class="unstyled articleLink" href="/m/coco_2017">
Coco (2017)</a>, <a class="unstyled articleLink" href="/m/dunkirk_2017">
Dunkirk (2017)</a>, <a class="unstyled articleLink" href="/m/star_wars_the_last_jedi">
Star Wars: The Last Jedi (2017)</a>, <a class="unstyled articleLink" href="/m/1015002-night_at_the_opera">
A Night at the Opera (1935)</a>, <a class="unstyled articleLink" href="/m/the_shape_of_water_2017">
The Shape of Water (2017)</a>, <a class="unstyled articleLink" href="/m/thor_ragnarok_2017">
Thor: Ragnarok (2017)</a>, <a class="unstyled articleLink" href="/m/spotlight_2015">
Spotlight (2015)</a>, <a class="unstyled articleLink" href="/m/the_farewell_2019">
The Farewell (2019)</a>, <a class="unstyled articleLink" href="/m/selma">
Selma (2014)</a>, <a class="unstyled articleLink" href="/m/the_third_man">
The Third Man (1949)</a>, <a class="unstyled articleLink" href="/m/1017289-rear_window">
Rear Window (1954)</a>, <a class="unstyled articleLink" href="/m/et_the_extraterrestrial">
E.T. The Extra-Terrestrial (1982)</a>, <a class="unstyled articleLink" href="/m/seven_samurai_1956">
Seven Samurai (Shichinin no Samurai) (1956)</a>, <a class="unstyled articleLink" href="/m/la_grande_illusion">
La Grande illusion (Grand Illusion) (1938)</a>, <a class="unstyled articleLink" href="/m/arrival_2016">
Arrival (2016)</a>, <a class="unstyled articleLink" href="/m/singin_in_the_rain">
Singin' in the Rain (1952)</a>, <a class="unstyled articleLink" href="/m/the_favourite_2018">
The Favourite (2018)</a>, <a class="unstyled articleLink" href="/m/double_indemnity">
Double Indemnity (1944)</a>, <a class="unstyled articleLink" href="/m/1000642-all_quiet_on_the_western_front">
All Quiet on the Western Front (1930)</a>, <a class="unstyled articleLink" href="/m/1048445-snow_white_and_the_seven_dwarfs">
Snow White and the Seven Dwarfs (1937)</a>, <a class="unstyled articleLink" href="/m/mariage_story_2019">
Zootopia (2016)</a>, <a class="unstyled articleLink" href="/m/alien">
Alien (1979)</a>, <a class="unstyled articleLink" href="/m/1011615-king_kong">
King Kong (1933)</a>, <a class="unstyled articleLink" href="/m/1018688-shadow_of_a_doubt">
Shadow of a Doubt (1943)</a>, <a class="unstyled articleLink" href="/m/call_me_by_your_name">
Call Me by Your Name (2018)</a>, <a class="unstyled articleLink" href="/m/psycho">
Psycho (1960)</a>, <a class="unstyled articleLink" href="/m/1917_2019">
1917 (2020)</a>, <a class="unstyled articleLink" href="/m/la_confidential">
L.A. Confidential (1997)</a>, <a class="unstyled articleLink" href="/m/the_florida_project">
The Florida Project (2017)</a>, <a class="unstyled articleLink" href="/m/war_for_the_planet_of_the_apes">
War for the Planet of the Apes (2017)</a>, <a class="unstyled articleLink" href="/m/paddington_2">
Paddington 2 (2018)</a>, <a class="unstyled articleLink" href="/m/beatles_a_hard_days_night">
A Hard Day's Night (1964)</a>, <a class="unstyled articleLink" href="/m/widows_2018">
Widows (2018)</a>, <a class="unstyled articleLink" href="/m/never_rarely_sometimes_always">
Never Rarely Sometimes Always (2020)</a>, <a class="unstyled articleLink" href="/m/baby_driver">
Baby Driver (2017)</a>, <a class="unstyled articleLink" href="/m/spider_man_homecoming">
Spider-Man: Homecoming (2017)</a>, <a class="unstyled articleLink" href="/m/godfather_part_ii">
The Godfather, Part II (1974)</a>, <a class="unstyled articleLink" href="/m/the_battle_of_algiers">
The Battle of Algiers (La Battaglia di Algeri) (1967)</a>]
['https://www.rottentomatoes.com/m/it_happened_one_night', 'https://www.rottentomatoes.com/m/citizen_kane', 'https://www.rottentomatoes.com/m/the_wizard_of_oz
1 https://www.rottentomatoes.com/m/the_battle_of_algiers
Movie:The Battle of Algiers (La Battaglia di Algeri) (1967)
Movie Info:Paratrooper commander Colonel Mathieu (Jean Martin), a former French Resistance fighter during World War II, is sent to 1950s Algeria to reinforce

Process finished with exit code 0
```