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



