# -\*- coding: utf-8 -\*-

"""

@author: Sagnik Das

"""

from bs4 import BeautifulSoup

import requests

import re

url = 'http://www.imdb.com/chart/top'

response = requests.get(url)

soup = BeautifulSoup(response.text, 'lxml')

movies = soup.select('td.titleColumn')

links = [a.attrs.get('href') for a in soup.select('td.titleColumn a')]

crew = [a.attrs.get('title') for a in soup.select('td.titleColumn a')]

ratings = [b.attrs.get('data-value') for b in soup.select('td.posterColumn span[name=ir]')]

votes = [b.attrs.get('data-value') for b in soup.select('td.ratingColumn strong')]

# create a empty list for storing

# movie information

list = []

# Iterating over movies to extract

# each movie's details

for index in range(0, len(movies)):

# Separating movie into: 'place',

# 'title', 'year'

movie\_string = movies[index].get\_text()

movie = (' '.join(movie\_string.split()).replace('.', ''))

movie\_title = movie[len(str(index))+1:-7]

year = re.search('\((.\*?)\)', movie\_string).group(1)

place = movie[:len(str(index))-(len(movie))]

data = {"movie\_title": movie\_title,

"year": year,

"place": place,

"star\_cast": crew[index],

"rating": ratings[index],

"vote": votes[index],

"link": links[index]}

list.append(data)

for movie in list:

print(movie['place'], '-', movie['movie\_title'], '('+movie['year'] +

') -', 'Starring:', movie['star\_cast'], movie['rating'])