

Hw4

資工三 4110056030 鄭詠謙

爬蟲任務:

- 1. 拿到電影海報
- 2. 取得10頁資訊
- 3. 整理成EXCEL

▼ chatgpt

• My question



chatgpt ans

\$

ChatGPT

當您使用爬蟲來從 https://ssr1.scrape.center/ 網站上獲取電影信息並將其存儲到 Excel 文件時,可以遵循以下步驟:

- 1. **匯入所需的庫**:首先,您需要匯入所需的Python庫。在這種情況下,我們將使用`requests`庫來 發送HTTP請求,`BeautifulSoup`庫來解析HTML內容,並使用`pandas`庫來操作數據和創建 Excel文件。
- 2. 發送HTTP請求:使用`requests.get()`函數發起HTTPGET請求,獲取網站的HTML內容。
- 3. 解析HTML內容:使用`BeautifulSoup`庫解析網頁的HTML內容,以便從中提取出電影信息。
- 4. **提取電影信息**:使用BeautifulSoup庫的方法和屬性,提取出電影的標題、類別、封面鏈接和評分等信息。
- 5. **保存電影信息**:將提取到的電影信息存儲到一個列表中,每個電影信息都以字典的形式存儲。然後,將列表轉換為`pandas`的DataFrame對象。
- 6. **將信息寫入Excel文件**:使用**DataFrame**的**`to_excel()**`方法將電影信息寫入**Excel**文件。您可以 指定文件名和路徑,也可以使用默認的文件名。
- 7. **執行整個流程**:將上述步驟封裝到兩個函數中,一個用於爬取電影信息,另一個用於將信息存儲到Excel文件中。最後,在`__main__`部分調用這兩個函數來執行整個流程。

通過執行以上步驟,您將能夠從 https://ssr1.scrape.center/ 網站上爬取電影信息,並將其存儲到 Excel文件中,方便您進行後續的數據分析和處之。

繼續問爬蟲怎麼做



You

爬蟲的步驟是什麼

這個爬蟲程式的步驟如下:

- 1. 設置環境:引入必要的庫,包括 `requests` 用於發送 HTTP 請求, `BeautifulSoup` 用於解析 HTML, `pandas` 用於數據處理和保存。
- 2. 定義目標:指定要爬取的網站 URL,以及網頁中包含目標信息的位置。
- 3. 設置請求標頭:設置 User-Agent 和 Referer 等請求標頭,以模擬瀏覽器行為,防止被網站擋掉。
- 4. **爬取數據**:通過迭代頁數,向目標網站的每個頁面發送請求,獲取HTML響應。
- 5. 解析HTML:使用 BeautifulSoup 將 HTML 響應解析為結構化的文檔對象,方便後續提取所需信息。
- 6. 提取信息:根據 HTML 結構,定位到目標信息所在的標籤,使用相應的方法(如 `find_all`)提取信息。
- 7. 處理信息:對提取到的信息進行必要的處理,如清理、轉換格式等。
- 8. 存儲數據:將爬取到的信息存儲到適當的數據結構中,這裡是存儲到一個字典中。
- 9. 數據分析:如果需要,可以對爬取到的數據進行分析處理,如統計、可視化等。
- 10. 保存數據:將爬取到的數據保存到本地文 (◆ 這裡是保存為 Excel 文件。

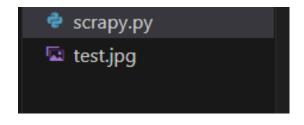
• chatgpt給一段程式碼

```
You
   給我code
    < 2/2 →
以下是您要求的程式碼:
                                                                           Copy code
     import requests
     from bs4 import BeautifulSoup
     import pandas as pd
     import re
     movie_info = {'電影名稱': [], '類型': [], '國家': [], '時長': [], '上映時間': [], '分數':
     header = {'Referer': 'https://scrape.center/',
     for page in range(1, 11):
         response = requests.get('https://ssri.scrape.center/page/%d' % page, headers=head
         soup = BeautifulSoup(response.text, 'html.parser')
         result = soup.find_all(name='div', class_='el-card item m-t is-hover-shadow')
         for i in range(len(result)):
             movie_info['電影名稱'].append(result[i].h2.string)
             btn_list = result[i].find_all(name='button', class_='el-button category el-bu
             movie_type = ''
             for btn in btn list:
```

```
for i in range(len(result)):
         movie_info['電影名稱'].append(result[i].h2.string)
         btn_list = result[i].find_all(name='button', class_='el-button category el-bu
         movie_type = ''
         for btn in btn_list:
             movie_type += btn.span.string + ','
         movie_info['類型'].append(movie_type)
         info_list = result[i].find_all(name='div', class_='m-v-sm info')
         span_list = info_list[0].find_all(name='span')
         movie_info['蒙'].append(span_list[0].string)
         movie_info['時長'].append(span_list[2].string)
         span_list = info_list[1].find_all(name='span')
         if len(span_list) > 0:
             movie_info['上映時間'].append(span_list[0].string)
             movie_info['上映時間'].append('')
         score = soup.find_all(name='p', class_='score m-t-md m-b-n-sm')
         movie_info['分數'].append(score[i].string.strip())
 data = pd.DataFrame(movie_info)
 data.to_excel('./movieinfo.xlsx', index=False)
這段程式碼將從網站 https://ssr1.scrape.center/ 中爬取電影信息,並將結果保存到名為
`movieinfo.xlsx`的Excel文件中。
$ C C $
```

▼ 拿到電影海報

Hw4





▼ 取得10頁資訊

• 先取一頁的資訊

```
import requests
from bs4 import BeautifulSoup
import pandas as pd
movie_info= {'電影名稱':[],'類型':[],'國家':[],'時長':[],'-
header = {'Referer':'https://scrape.center/',
          'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; W
response = requests.get('https://ssr1.scrape.center/pag
soup = BeautifulSoup(response.text, 'html.parser')
result = soup.find_all(name='div',class_='el-card item |
for i in range(len(result)):
    print(result[i].h2.string)
    btn_list = result[i].find_all(name='button',class_=
    movie type = ''
    for btn in btn_list:
        movie_type += btn.span.string + ','
    for btn in btn list:
        print(btn.span.string)
```

```
info_list = result[i].find_all(name='div',class
span_list = info_list[0].find_all(name='span')

span_list = info_list[1].find_all(name='span')

for info in info_list:
    span_list = info.find_all(name='span')
    for s in span_list:
        if s.string != ' / ':
            print(s.string)

score = soup.find_all(name='p',class_='score m-
print(score[i].string.strip())
print('--'*10)
```

```
霸王别姬 - Farewell My Concubine
剧情
爱情
中国内地、中国香港
171 分钟
1993-07-26 上映
9.5
这个杀手不太冷 - Léon
剧情
动作
犯罪
法国
110 分钟
1994-09-14 上映
9.5
肖申克的救赎 - The Shawshank Redemption
剧情
犯罪
美国
142 分钟
1994-09-10 上映
9.5
```

• 寫成loop獲取十頁的資料

```
import requests
from bs4 import BeautifulSoup
```

```
import pandas as pd
movie_info= {'電影名稱':[],'類型':[],'國家':[],'時長':[],'-
header = {'Referer':'https://scrape.center/',
          'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; W
for page in range(1,11):
    response = requests.get('https://ssr1.scrape.center.
    soup = BeautifulSoup(response.text, 'html.parser')
    result = soup.find all(name='div',class ='el-card i
   for i in range(len(result)):
        movie_info['電影名稱'].append(result[i].h2.string
        btn_list = result[i].find_all(name='button',cla
        movie type = ''
        for btn in btn list:
            movie_type += btn.span.string + ','
        movie_info['類型'].append(movie_type)
        info_list = result[i].find_all(name='div',class
        span list = info list[0].find all(name='span')
        movie_info['國家'].append(span_list[0].string)
        movie_info['時長'].append(span_list[2].string)
        span list = info list[1].find all(name='span')
        if len(span_list) > 0:
           movie_info['上映時間'].append(span_list[0].st
        else:
           movie_info['上映時間'].append('')
        score = soup.find_all(name='p',class_='score m-
        movie_info['分數'].append(score[i].string.strip(
print(movie info)
```

▼ 整理成EXCEL

```
# ...上述程式碼
data = pd.DataFrame(movie_info)
data.to_excel('./movieinfo.xlsx',index='False')
```

這是在vscode用excel viewer查看

Market							
	A T	В Т	C T	D '	r E T	FT	
1	電影名稱	類型	國家	時長	上映時間	分數	
2	霸王别姬 - Farewell My Concub	剧情,爱情,	中国内地、中国香港	171 分钟	1993-07-26 上版	9.5	
3	这个杀手不太冷 - Léon	剧情,动作,犯罪,	法国	110 分钟	1994-09-14 上版	9.5	
4	肖申克的救赎 - The Shawshank	剧情,犯罪,	美国	142 分钟	1994-09-10 上版	9.5	
5	泰坦尼克号 - Titanic	剧情,爱情,灾难,	美国	194 分钟	1998-04-03 上版	9.5	
6	罗马假日 - Roman Holiday	剧情,喜剧,爱情,	美国	118 分钟	1953-08-20 上版	9.5	
7	唐伯虎点秋香 - Flirting Scholar	喜剧,爱情,古装,	中国香港	102 分钟	1993-07-01 上版	9.5	
8	乱世佳人 - Gone with the Wind	剧情,爱情,历史,战	美国	238 分钟	1939-12-15 上版	9.5	
9	喜剧之王 - The King of Comedy	剧情,喜剧,爱情,	中国香港	85 分钟	1999-02-13 上版	9.5	
10	楚门的世界 - The Truman Show	剧情,科幻,	美国	103 分钟		9.0	
11	狮子王 - The Lion King	动画,歌舞,冒险,	美国	89 分钟	1995-07-15 上版	9.0	
12	V字仇杀队 - V for Vendetta	剧情,动作,科幻,恢	美国、英国、德国	132 分钟	2005-12-11 上既	8.9	
13	少年派的奇幻漂流 - Life of Pi	剧情,奇幻,冒险,	美国、台湾、英国、加拿大	127 分钟	2012-11-22 上版	8.9	
14	美丽心灵 - A Beautiful Mind	剧情,传记,	美国	135 分钟	2001-12-13 上版	8.8	
15	初恋这件小事 - สิ่งเล็กเล็กที่เรียก	喜剧,爱情,	泰国	118 分钟	2012-06-05 上版	8.9	
16	借东西的小人阿莉埃蒂 - 借りぐ	动画,奇幻,冒险,	日本	94 分钟	2010-07-17 上版	8.8	

以下是excel完整內容

	A	В	С	D	Е	F
1	電影名稱	類型	國家	時長	上映時間	分數
2	霸王别姬 - Farewell My	剧情,爱情,	中国内地、中国香港	171 分钟	1993-07-26 上映	9.5
3	这个杀手不太冷 - Léon	剧情,动作,犯罪,	法国	110 分钟	1994-09-14 上映	9.5
4	肖申克的救赎 - The Sha	剧情,犯罪,	美国	142 分钟	1994-09-10 上映	9.5
5	泰坦尼克号 - Titanic	剧情,爱情,灾难,	美国	194 分钟	1998-04-03 上映	9.5
6	罗马假日 - Roman Holic	剧情,喜剧,爱情,	美国	118 分钟	1953-08-20 上映	9.5
7	唐伯虎点秋香 - Flirting	喜剧,爱情,古装,	中国香港	102 分钟	1993-07-01 上映	9.5
8	乱世佳人 - Gone with th	剧情,爱情,历史,战争,	美国	238 分钟	1939-12-15 上映	9.5
9	喜剧之王 - The King of	剧情,喜剧,爱情,	中国香港	85 分钟	1999-02-13 上映	9.5
10	楚门的世界 - The Trum	剧情,科幻,	美国	103 分钟		9.0
11	狮子王 - The Lion King	动画,歌舞,冒险,	美国	89 分钟	1995-07-15 上映	9.0
12	V字仇杀队 - V for Vend	剧情,动作,科幻,惊悚,	美国、英国、德国	132 分钟	2005-12-11 上映	8.9
13	少年派的奇幻漂流 - Li	剧情,奇幻,冒险,	美国、台湾、英国、加拿大	127 分钟	2012-11-22 上映	8.9
14	美丽心灵 - A Beautiful	剧情,传记,	美国	135 分钟	2001-12-13 上映	8.8
15	初恋这件小事 - สิ่งเล็กเ	喜剧,爱情,	泰国	118 分钟	2012-06-05 上映	8.9
16	借东西的小人阿莉埃蒂	动画,奇幻,冒险,	日本	94 分钟	2010-07-17 上映	8.8
17	Yi yi: A One and	剧情,爱情,家庭,	台湾、日本	173 分钟	2000-05-15 上映	8.8
18	美丽人生 - La vita è bel	战争,剧情,爱情,	意大利	116 分钟	2020-01-03 上映	9.1
19	海上钢琴师 - La leggen	剧情,爱情,音乐,	意大利	126 分钟	2019-11-15 上映	9.1
20	千与千寻 - 千と千尋の	动画,冒险,奇幻,家庭,	日本	125 分钟	2019-06-21 上映	9.1
21	迁徙的鸟 - The Travelli	纪录片,	法国、德国、意大利、西班牙、	98 分钟	2001-12-12 上映	9.1
22	黄金三镖客 - Il buono, i	西部,冒险,	意大利、西班牙、西德	161 分钟	1966-12-23 上映	9.1
23	海洋 - Océans	纪录片,	法国、瑞士、西班牙、美国、阿里	104 分钟	2011-08-12 上映	9.1
24	我爱你 - 그대를 사랑협	剧情,爱情,	韩国	118 分钟	2011-02-17 上映	9.1
25	阿飞正传 - Days of Beir	剧情,爱情,犯罪,	中国香港	94 分钟	2018-06-25 上映	9.1
26	7号房的礼物 - 7번방의	剧情,喜剧,家庭,	韩国	127 分钟	2013-01-23 上映	8.8
27	爱·回家 - 집으로	剧情,家庭、	韩国	80 分钟	2002-04-05 上映	9.1

	A	В	С	D	Е	F G
28	龙猫 - となりのトトロ	动画,冒险,奇幻,家庭,	日本	86 分钟	2018-12-14 上映	9.1
29	七武士 - 七人の侍	剧情,动作,冒险,	日本	207 分钟	1954-04-26 上映	8.8
30	美国往事 - Once Upon a Time in Am	剧情,犯罪,	意大利、美国	229 分钟	2015-04-23 上映	8.8
31	完美的世界 - A Perfect World	剧情,犯罪,	美国	138 分钟	1993-11-24 上映	8.8
32	上帝之城 - Cidade de Deus	剧情,犯罪,	巴西、法国	130 分钟		8.8
33	辩护人 - 변호인	剧情,	韩国	127 分钟	2013-12-18 上映	8.8
34	忠犬八公物语 - ハチ公物語	剧情,	日本	107 分钟	1987-08-01 上映	8.8
35	海豚湾 - The Cove	纪录片,	美国	92 分钟	2009-07-31 上映	8.8
36	英雄本色 - A Better Tomorrow	剧情,动作,犯罪,	中国香港	95 分钟	2017-11-17 上映	8.8
37	恐怖直播 - 더 테러 라이브	剧情,悬疑,犯罪,	韩国	97 分钟	2013-07-31 上映	8.8
38	窃听风暴 - Das Leben der Anderen	剧情,悬疑,	德国	137 分钟	2006-03-23 上映	8.8
39	时空恋旅人 - About Time	喜剧,爱情,奇幻,	英国	123 分钟	2013-09-04 上映	8.8
40	穿条纹睡衣的男孩 - The Boy in the	剧情,战争,	英国、美国	94 分钟	2008-08-28 上映	8.8
41	教父 - The Godfather	剧情,犯罪,	美国	175 分钟	2015-04-18 上映	8.8
42	萤火之森 - 蛍火の杜へ	剧情,爱情,动画,奇幻,	日本	45 分钟	2011-09-17 上映	8.8
43	素媛 - 소원	剧情,	韩国	123 分钟	2013-10-02 上映	8.8
44	小鞋子 - بچههای آسمان	剧情,家庭,	伊朗	89 分钟		8.8
45	熔炉 - 도가니	剧情,	韩国	125 分钟	2011-09-22 上映	8.8
46	大话西游之大圣娶亲 - A Chinese O	喜剧,爱情,奇幻,	中国香港、中国力	110 分钟	2014-10-24 上映	8.9
47	新龙门客栈 - New Dragon Gate Inn	动作,爱情,武侠,古装,	中国香港、中国力	88 分钟	2012-02-24 上映	8.9
48	触不可及 - Intouchables	剧情,喜剧,	法国	112 分钟	2011-11-02 上映	8.9
49	钢琴家 - The Pianist	剧情,音乐,传记,历史,战	法国、德国、英国	150 分钟	2002-05-24 上映	8.9
50	本杰明·巴顿奇事 - The Curious Case	剧情,爱情,奇幻,	美国	166 分钟	2008-12-25 上映	8.9
51	倩女幽魂 - A Chinese Ghost Story	爱情,奇幻,武侠,古装,	中国香港	98 分钟	2011-04-30 上映	8.9
52	哈利·波特与死亡圣器(下) - Harr	剧情,悬疑,奇幻,冒险,	英国、美国	130 分钟	2011-08-04 上映	8.9
53	甜蜜蜜 - Comrades: Almost a Love St	剧情,爱情,	中国香港	118 分钟	2015-02-13 上映	8.9
54	蝙蝠侠:黑暗骑士崛起 - The Dark	剧情,动作,科幻,惊悚,犯	美国、英国	165 分钟	2012-08-27 上映	8.9

Hw4

	A	В	С	D	Е	F
55	鬼子来了 - Devils on the Doorstep	剧情,战争,	中国大陆	139 分钟	2000-05-13 上映	8.8
56	无敌破坏王 - Wreck-It Ralph	喜剧,动画,奇幻,冒险,	美国	101 分钟	2012-11-06 上映	8.8
57	致命魔术 - The Prestige	剧情,悬疑,惊悚,	美国、英国	130 分钟	2006-10-17 上映	8.8
58	神偷奶爸 - Despicable Me	喜剧,动画,冒险,	美国、法国	95 分钟	2010-06-20 上映	8.8
59	断背山 - Brokeback Mountain	剧情,爱情,家庭,	美国、加拿大	134 分钟	2005-09-02 上映	8.8
		剧情,喜剧,爱情,	美国	90 分钟	2010-07-26 上映	8.8
61	驯龙高手 - How to Train Your Drago	喜剧,动画,奇幻,冒险,	美国	98 分钟	2010-05-14 上映	8.8
62	飞屋环游记 - Up	剧情,喜剧,动画,冒险,	美国	96 分钟	2009-08-04 上映	8.8
63	黑客帝国3:矩阵革命 - The Matrix	动作,科幻,	美国、澳大利亚	129 分钟	2003-11-05 上映	8.8
64	速度与激情5 - Fast Five	动作,犯罪,	美国	130 分钟	2011-05-12 上映	8.9
65	勇敢的心 - Braveheart	剧情,动作,传记,历史,战	美国	177 分钟	1995-05-18 上映	8.9
66	三傻大闹宝莱坞 - 3 Idiots	剧情,喜剧,爱情,歌舞,	印度	171 分钟	2011-12-08 上映	8.9
67	闻香识女人 - Scent of a Woman	剧情,	美国	157 分钟	1992-12-23 上映	8.9
68	末代皇帝 - The Last Emperor	剧情,传记,历史,	英国、意大利、中	163 分钟	1987-10-23 上映	8.9
69	风之谷 - 風の谷のナウシカ	动画,奇幻,冒险,	日本	117 分钟		8.9
70	大话西游之月光宝盒 - A Chinese O	喜剧,爱情,奇幻,古装,	中国香港、中国力	87 分钟	2014-10-24 上映	8.9
71	放牛班的春天 - Les choristes	剧情,音乐,	法国、德国、瑞Ⅎ	97 分钟	2004-10-16 上映	8.9
72	当幸福来敲门 - The Pursuit of Happy	剧情,家庭,传记,	美国	117 分钟	2008-01-17 上映	8.9
73	幽灵公主 - もののけ姫	动画,奇幻,冒险,	日本	134 分钟	1998-05-01 上映	8.9
74	十二怒汉 - 12 Angry Men	剧情,	美国	96 分钟	1957-04-13 上映	8.9
75	搏击俱乐部 - Fight Club	剧情,动作,悬疑,惊悚,	美国、德国	139 分钟	1999-09-10 上映	8.9
76	疯狂原始人 - The Croods	喜剧,动画,冒险,	美国	98 分钟	2013-04-20 上映	8.9
77	阿凡达 - Avatar	动作,科幻,冒险,	美国、英国	162 分钟	2010-01-04 上映	8.9
78	哈尔的移动城堡 - ハウルの動く城	动画,奇幻,冒险,	日本	119 分钟	2004-09-05 上映	8.9
79	盗梦空间 - Inception	剧情,科幻,悬疑,冒险,	美国、英国	148 分钟	2010-09-01 上映	8.9
80	忠大八公的故事 - Hachi: A Dog's Ta	剧情,	美国、英国	93 分钟	2009-06-13 上映	8.9
81	拯救大兵瑞恩 - Saving Private Ryan	剧情,历史,战争,	美国	169 分钟	1998-11-13 上映	8.9

4	A	В	С	D	Е	F
82	活着 - To Live	剧情,家庭,历史,	中国大陆、中国都	132 分钟	1994-05-17 上映	9.0
83	机器人总动员 - WALL·E	喜剧,科幻,动画,	美国	98 分钟	2008-06-27 上映	9.0
84	天堂电影院 - Nuovo Cinema Paradiso	剧情,爱情,	意大利、法国	155 分钟	1988-11-17 上映	9.0
85	指环王2:双塔奇兵 - The Lord of th	剧情,动作,奇幻,冒险,	美国、新西兰	179 分钟	2003-04-25 上映	9.0
86	指环王1:护戒使者 - The Lord of th	剧情,动作,奇幻,冒险,	新西兰、美国	178 分钟	2002-04-04 上映	9.0
87	射雕英雄传之东成西就 - The Eagle	喜剧,奇幻,武侠,古装,	中国香港	113 分钟	1993-02-05 上映	9.0
88	蝙蝠侠:黑暗骑士 - The Dark Knigh	剧情,动作,科幻,惊悚,犯	美国、英国	152 分钟	2008-07-14 上映	9.0
89	无间道 - Infernal Affairs	剧情,悬疑,犯罪,	中国香港	101 分钟	2003-09-05 上映	9.0
90	教父2 - The Godfather: Part II	剧情,犯罪,	美国	202 分钟	1974-12-12 上映	9.0
91	加勒比海盗 - Pirates of the Caribbean	动作,奇幻,冒险,	美国	143 分钟	2003-11-21 上映	9.0
92	哈利·波特与魔法石 - Harry Potter ar	奇幻,冒险,	美国、英国	152 分钟	2002-01-26 上映	9.0
93	指环王3:王者无敌 - The Lord of th	剧情,动作,奇幻,冒险,	美国、新西兰	201 分钟	2004-03-15 上映	9.0
94	黑客帝国 - The Matrix	动作,科幻,	美国、澳大利亚	136 分钟	2000-01-14 上映	9.0
95	剪刀手爱德华 - Edward Scissorhands	剧情,爱情,奇幻,	美国	105 分钟	1990-12-06 上映	9.0
96	春光乍泄 - Happy Together	剧情,爱情,	中国香港、日本、	96 分钟	1997-05-17 上映	9.0
97	大闹天宫 - The Monkey King	动画,奇幻,	中国大陆	114 分钟	1965-12-31 上映	9.0
98	天空之城 - 天空の城ラピュタ	动画,奇幻,冒险,	日本	125 分钟	1992-05-01 上映	9.0
99	音乐之声 - The Sound of Music	剧情,爱情,歌舞,传记,	美国	174 分钟	1965-03-02 上映	9.0
100	辛德勒的名单 - Schindler's List	剧情,历史,战争,	美国	195 分钟	1993-11-30 上映	9.5
101	魂断蓝桥 - Waterloo Bridge	剧情,爱情,战争,	美国	108 分钟	1940-05-17 上映	9.5
100						

Hw4