Lab#3, NLP@CGU Spring 2023

This is due on 2023/03/20 16:00, commit to your github as a PDF (lab3.pdf) (File>Print>Save as PDF).

IMPORTANT: After copying this notebook to your Google Drive, please paste a link to it below. To get a publicly-accessible link, hit the *Share* button at the top right, then click "Get shareable link" and copy over the result. If you fail to do this, you will receive no credit for this lab!

LINK: paste your link here

https://colab.research.google.com/drive/1s4JEY5DijL2doRCTrh-40mbr_9kt3M2n?usp=sharing

---

Student ID:B0928013

Name:吳佳恩

Question 1 (100 points)

Implementing Yahoo Movies Crawler.

1. Design a Yahoo! Movie Crawler.
2. Crawl all the movie information listed in movie_intheaters page
3. The more movie data crawled, the higher the score

按兩下 (或按 Enter 鍵) 即可編輯

```
import requests
import re
from bs4 import BeautifulSoup

Y_MOVIE_URL = "https://movies.yahoo.com.tw/movie_intheaters.html"
response = requests.get(url=Y_MOVIE_URL)

soup = BeautifulSoup(response.text, 'lxml')
# YOUR CODE HERE!
# IMPLEMENTIG YAHOO MOVIES CRAWLER

class MovieCrawler(object):

        def __init__(self):
            info_items = soup.find_all('div', 'release_info')
        def get_movies(self, page_url):
            for item in info_items:
                chinese_name = item.find('div', 'release_movie_name').a.text.strip()
                english_name = item.find('div', 'en').a.text.strip()
                release_time = item.find('div', 'release_movie_time').text.split(': ')[-1].strip()
                moive_url = item.find('a',class_="gabtn")['href']
                movie_text = item.find('span', 'jq_text_overflow_180 jq_text_overflow_href_list').text

                #print('chname:{} ename: ({}) 上映日: {} moive_url:{} movie_text{}'.format(chinese_name, english_name, release_time, moive_ur

get_movies(self, page_url)
# # DO NOT MODIFY THE VARIABLES
crawler = MovieCrawler()
movies = crawler.get_movies(Y_MOVIE_URL)

# # THE RESULTS : AS THE FOLLOWING SECTION
# # {'ch_name', 'en_name', 'movie_url', 'release_date', 'intro'}
print(len(movies))
print(*movies, sep="\n")
```

```
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
<ipython-input-17-43bd994892c7> in <module>
     33 # # THE RESULTS : AS THE FOLLOWING SECTION
     34 # # {'ch_name', 'en_name', 'movie_url', 'release_date', 'intro'}
---> 35 print(len(movies))
     36 print(*movies, sep="\n")

TypeError: object of type 'NoneType' has no len()
```

SEARCH STACK OVERFLOW

```
import requests
from bs4 import BeautifulSoup

url = 'https://movies.yahoo.com.tw/movie_intheaters.html'
response = requests.get(url=url)

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

info_items = soup.find_all('div', 'release_info')

for item in info_items:
    name = item.find('div', 'release_movie_name').a.text.strip()
    english_name = item.find('div', 'en').a.text.strip()
    release_time = item.find('div', 'release_movie_time').text.split(': ')[-1].strip()
    #level = item.find('div', 'leveltext').span.text.strip()
    #moive_url = item.find('a','data-url' )
    moive_url = item.find('a',class_="gabtn")['href']
    #movie_text = item.find('div', 'release_text')
    movie_text = item.find('span', 'jq_text_overflow_180 jq_text_overflow_href_list').text

    print('chname:{} enname: ({}) 上映日：{} moive_url:{} movie_text:{}'.format(name, english_name, release_time, moive_url,movie_text ))
```

chname:配樂大師顏尼歐 enname: (Ennio: The Maestro) 上映日：2023-03-17 moive_url:https://movies.yahoo.com.tw/movieinfo_main/%E9%85%8D%E6%A8%82
★義大利奧斯卡大衛獎最佳紀錄片、最佳剪輯、最佳音效三項大獎
★義大利國家影評協會年度銀絲帶獎
★義大利巴里國際影展最佳導演
★加爾各答國際邪典電影節最佳紀錄片

他是最具影響力的音樂家，
2座奧斯卡獎得主，500多首令人難忘的電影配樂家…

《配樂大師顏尼歐》是全球最知名和受人喜愛的音樂大師顏尼歐莫利克奈的傳記電影。電影製作歷時5年、橫跨歐美進行拍攝，透過音樂和檔案片段講述了一代音

顏尼歐莫利克奈是20世紀最具影響力和最多產的音樂家之一，兩項奧斯卡獎的獲得者，以及500多首令人難忘的電影配樂的作曲家，本片則堪稱是他最完整肖像電

chname:熊蓋毒 enname: (Cocaine Bear) 上映日：2023-03-17 moive_url:https://movies.yahoo.com.tw/movieinfo_main/%E7%86%8A%E8%93%8B%E6%AF%92-coca
故事靈感來自發生在1985年的真實事件，一名毒販的飛機失事墜機，遺失了一批古柯鹼，結果被一隻黑熊吃掉。這部劇情既荒誕又瘋狂的喜

《熊蓋毒》一片的演員陣容包括凱莉羅素（諜報影集《冷戰諜夢》）、小歐席亞傑克森（《衝出康普頓》）、克利斯汀康佛瑞詹寧（奇幻影集《鹿角男孩》）、

《熊蓋毒》一片的導演是伊莉莎白班克斯（《霹靂嬌娃》、《歌喉讚2》），編劇是吉米沃頓（《辣手保姆2：女王蜂》），製片是奧斯卡獎得主菲爾洛德與克里

chname:若愛重來 enname: (Marriages) 上映日：2023-03-17 moive_url:https://movies.yahoo.com.tw/movieinfo_main/%E8%8B%A5%E6%84%9B%E9%87%8D%E4%BF
★賽博台客搖滾樂團 美秀集團 ★重磅合作電影主題曲〈戀人〉
★義大利奧斯卡金獎編劇《完美陌生人》導演 保羅克斯泰拉 最新力作

你是否曾經想過，自己多年的婚姻突然被宣判無效？在義大利一座小鎮上就有四對夫妻遇到這樣荒謬的事！因為他們的婚姻都是被一名不合法的牧師公證的。

在婚姻無效後，單親媽媽薩拉（雅柏拉安吉歐琳 飾）帶著兒子獨自生活，卻與前夫維多（法比歐佛洛 飾）再生愛苗？事業女強人寶拉（克勞迪婭潘多爾菲 飾）

當婚姻遇到不可抗力之因素，關係的天平即將失衡，四對夫妻會選擇再度牽起彼此的手嗎？還是以此為理由離開對方呢？

chname:無人相信的真相 enname: (La syndicaliste) 上映日：2023-03-17 moive_url:https://movies.yahoo.com.tw/movieinfo_main/%E7%84%A1%E4%BA%BA%E7
她是最大受害者 也是唯一嫌疑犯

本片改編自莫琳嘉內（伊莎貝雨蓓飾）的真實故事，她是法國一家跨國核電廠的工會代表。因揭發法國核工業的極密交易案而成為了吹哨者。她隻身與世界、政

然而某日，當她在家中遭到暴力襲擊後，生活開始天翻地覆。案件調查因為敏感議題而不斷遭到施壓，更讓原本的受害者莫琳嘉內頓時之間成為了自導自演的嫌

chname:闇黑對決 enname: (The Devil's Deal) 上映日：2023-03-17 moive_url:https://movies.yahoo.com.tw/movieinfo_main/%E9%97%87%E9%BB%91%E5%B0%8
★《我只是個計程車司機》製作人X《犯罪都市》製作團隊攜手打造犯罪鉅獻!
★《財閥家的小兒子》李星民X《無間對決》趙震雄鬥智耍狠，為名利不擇手段!
★ 第25屆加拿大奇幻電影節 正式入選
★ 第41屆美國夏威夷國際影展 正式入選

1992年，國會議員候選人海雄（趙震雄 飾）想結束跌至谷底的政治人生。海雄原本確信自己有望當選這屆的國會議員，卻慘遭能動搖政界的實權者順泰（李星民

為了推翻由順泰所策劃的選戰，海雄取得了有關釜山地區再開發計畫的機密文件。且在黑道老大弼道（金武烈 飾）的協助下，海雄籌備好選舉經費，以無黨籍的

而順泰得知了海雄擁有機密文件一事後，便決定將海雄逼上絕路……足以撼動大韓民國的祕密文件，推翻這場選戰的傢伙將顛覆整個國家!

chname:噩夢輓歌 4K數位修復版 enname: (Requiem For A Dream) 上映日：2023-03-17 moive_url:https://movies.yahoo.com.tw/movieinfo_main/%E5%99%A9%
★ 《黑天鵝》《我的鯨魚老爸》金獎名導成名代表作
★ 《大法師》奧斯卡影后艾倫鮑絲汀再創演技新巔峰
★ 《魔比斯》傑瑞德雷托、《捍衛戰士：獨行俠》珍妮佛康納莉超尺度演出
★ 2001 奧斯卡金像獎 最佳女主角 入圍
★ 2001 金球獎 戲劇類電影最佳女主角 入圍
★ 2001 美國電影學會獎 年度十大電影
★ 獲得全球近40個獎項肯定，網路流傳「十大禁片」之一

✓ 0 秒 完成時間：下午3:52