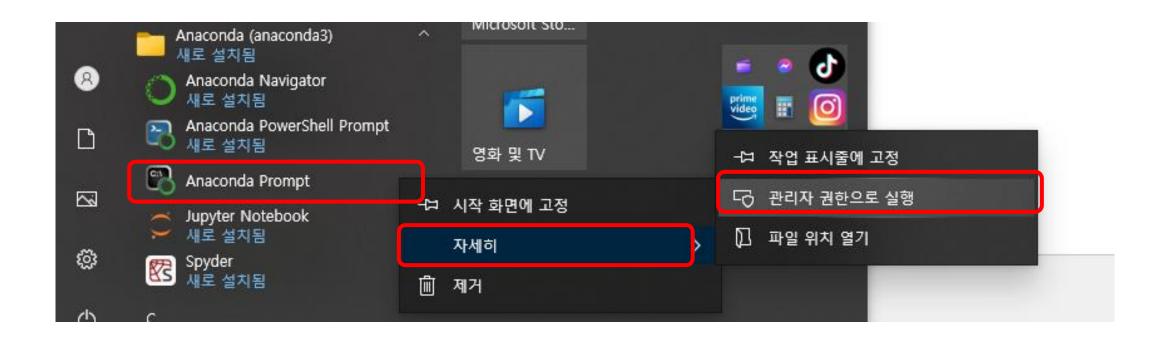
네이버 영화에서 범죄도시4의 영화 정보를 수집 합니다



■ 아래를 입력 합니다



■ 아래 명령을 입력 합니다



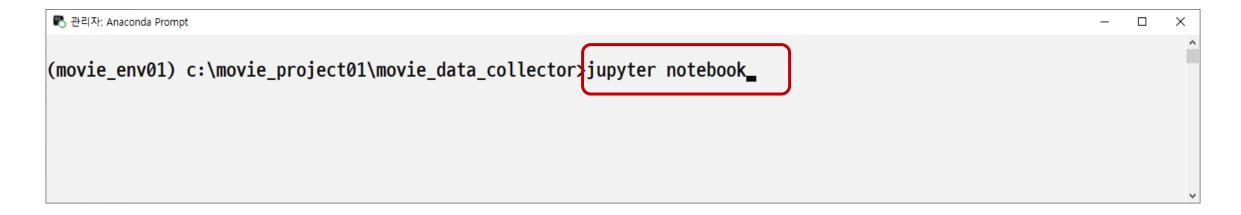
- 이 명령어의 의미:
 - conda: Anaconda라는 프로그램에서 사용되는 명령어입니다.
 - activate: 특정 가상 환경을 사용하도록 전환하는 명령입니다.
 - movie_env01: 이전에 만들어둔 가상 환경의 이름입니다.
- 가상 환경을 활성화하는 것은 프로젝트 전용 작업 공간에 들어가는 것과 같습니다.movie_env01은 특정 작업(예: 영화 프로젝트)을 위한 개인 작업 공간입니다.이 공간 안에서는 다른 프로젝트와 관련 없는 도구나 설정을 사용할 걱정 없이 해당 프로젝트에 최적화된 환경에서 작업할 수 있습니다.즉, 이 작업 공간은 영화 프로젝트를 위한 "전용 책상" 같은 역할을 합니다.

■ 아래를 입력 합니다



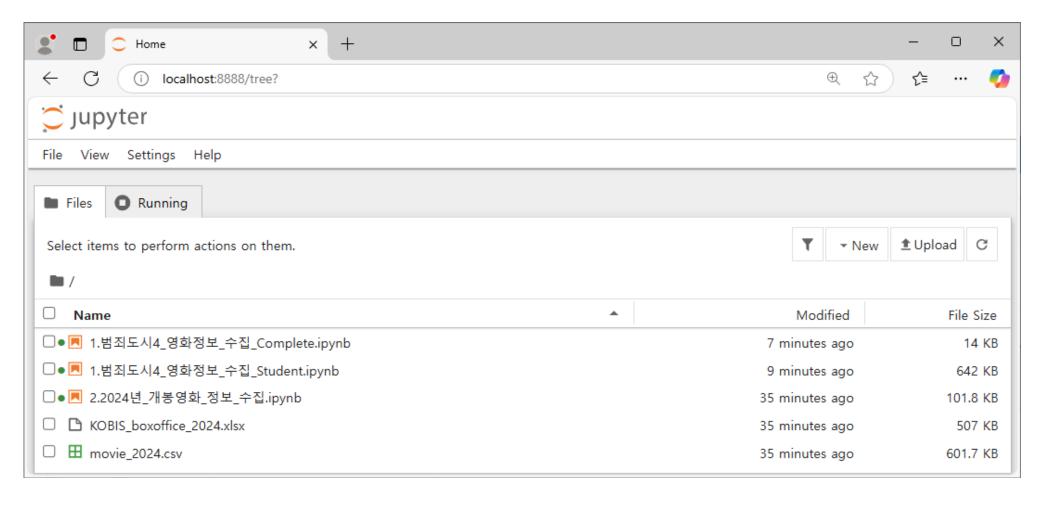
■ 프로젝트가 다운로드된 movie_data_collector 디렉토리로 이동합니다

■ 아래를 입력 합니다

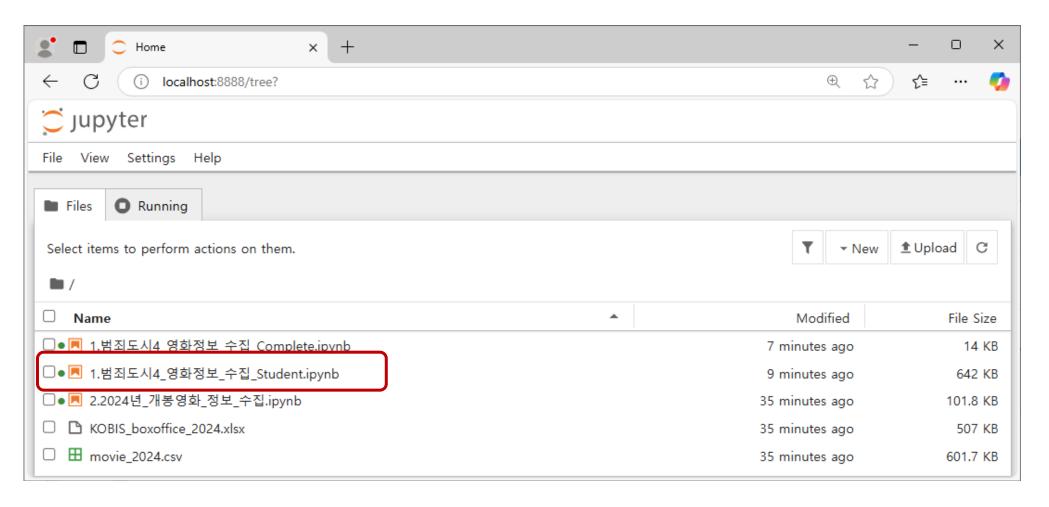


jupyter notebook은 Jupyter Notebook을 실행하라는 명령입니다.명령을 실행하면 컴퓨터의 웹 브라우저(크롬, 엣지 등)에서 Jupyter Notebook 화면이 열립니다.

■ 웹 브라우저 주소 표시줄에 <u>http://localhost:8888</u> 을 입력 합니다



■ 범죄도시4_영화정보_수집_Student.ipynb 를 클릭 합니다



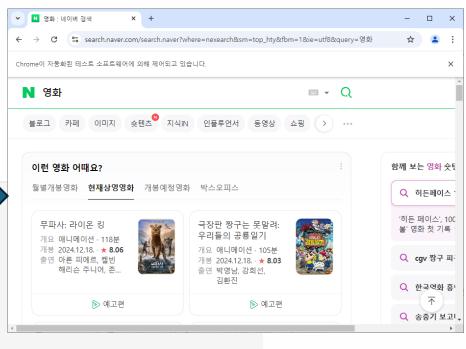
8

```
[1]: # 기존에 설치된 numpy(수학 계산 라이브러리) 버전을 제거합니다.
    # 시스템에서 호환성을 위해 특정 버전이 필요할 수 있기 때문입니다.
    !pip uninstall numpy -y
                                                      Shift+Enter 입력해서 실행시킴
    Found existing installation: numpy 1.23.5
    Uninstalling numpy-1.23.5:
      Successfully uninstalled numpy-1.23.5
[2]: # numpy 라이브러리를 특정 버전(1.23.5)으로 다시 설치합니다.
                                                      Shift+Enter 입력해서 실행시킴
    # 특정 기능이나 호환성을 위해 이 버전이 필요합니다.
    !pip install numpy==1.23.5
    Collecting numpy==1.23.5
      Using cached numpy-1.23.5-cp310-cp310-win_amd64.whl.metadata (2.3 kB)
    Using cached numpy-1.23.5-cp310-cp310-win_amd64.whl (14.6 MB)
    Installing collected packages: numpy
    Successfully installed numpy-1.23.5
                                                       Shift+Enter 입력해서 실행시킴
[3]: # selenium(웹 브라우저 자동화 라이브러리)을 설치합니다.
    # 이 라이브러리를 사용하면 사람이 브라우저에서 수행하는 작업을 코드로 자동화할 수 있습니다.
    !pip install selenium
```

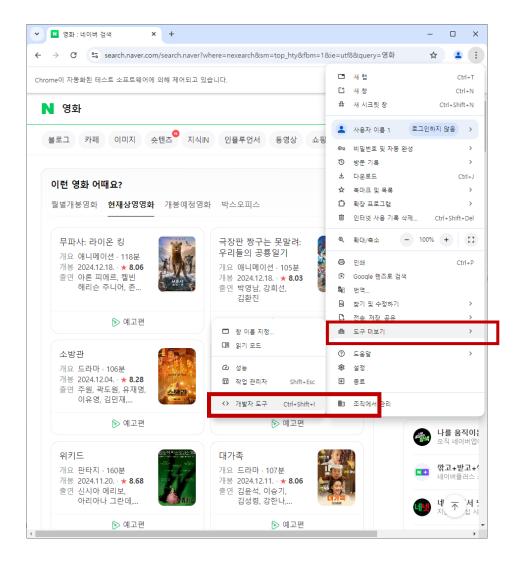
```
[4]: # 필요한 라이브러리 불러오기 단계
# selenium: 브라우저 자동화를 위한 라이브러리
from selenium import webdriver
# BeautifulSoup: HTML 분석을 위한 라이브러리
from bs4 import BeautifulSoup
# selenium의 By: HTML 요소를 찾는 데 사용됩니다.
from selenium.webdriver.common.by import By
# selenium의 Options: 브라우저 옵션 설정을 위한 라이브러리
from selenium.webdriver.chrome.options import Options
# time: 코드 실행 중 지연을 추가하기 위한 라이브러리
import time
# re: 정규 표현식을 사용하여 문자열을 처리하기 위한 라이브러리
import re
```

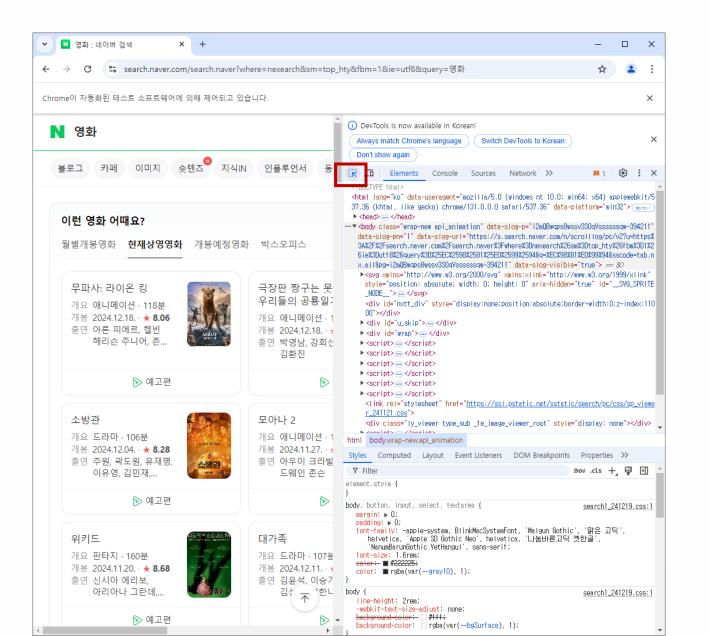
크롬 브라우저가 실행되고 네이버 영화가 검색됨

```
[51: # 크롬 드라이버의 옵션을 설정하는 객체 생성
    chrome_options = Options()
                                                                               무파사: 라이온 킹
    # 브라우저가 자동으로 닫히지 않도록 설정
                                                                               개요 애니메이션 - 118분
                                                                               개봉 2024.12.18. - ★ 8.06
                                                                               출연 아론 피에르, 켈빈
    chrome options.add experimental option("detach", True)
                                                                                 해리슨 주니어, 존.
    # 크롬 드라이버에 옵션 설정 전용
                                                                                    ▶ 예고편
    # (브라우저를 띄우고 자동화를 시작합니다.)
    driver = webdriver.Chrome(options=chrome_options)
    # 네이버 영화 검색 사이트 URL 설정
    url = 'https://search.naver.com/search.naver?where=nexearch&sm=top_hty&fbm=1&ie=utf8&query=%EC%98%81%ED%99%94'
    # 페이지가 로딩될 때 최대 10초까지 기다리도록 설정
    # 브라우저가 느린 경우를 대비한 안전 장치입니다.
    driver.implicitly_wait(10)
    # 네이버 영화 검색 페이지를 엽니다.
    driver.get(url)
    # 1초간 대기하여 페이지 로딩이 완료되도록 설정
    # (로딩 시간이 너무 짧으면 일부 요소가 표시되지 않을 수 있습니다.)
    time.sleep(1)
```

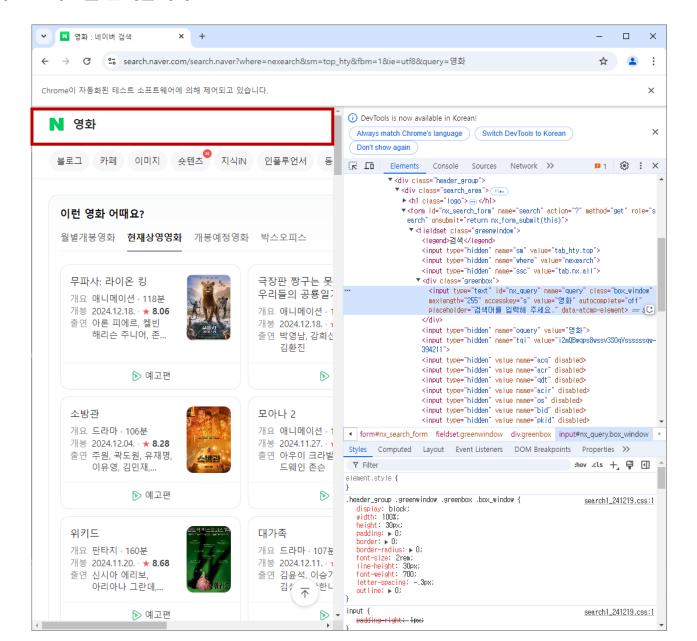


개발자 도구 선택

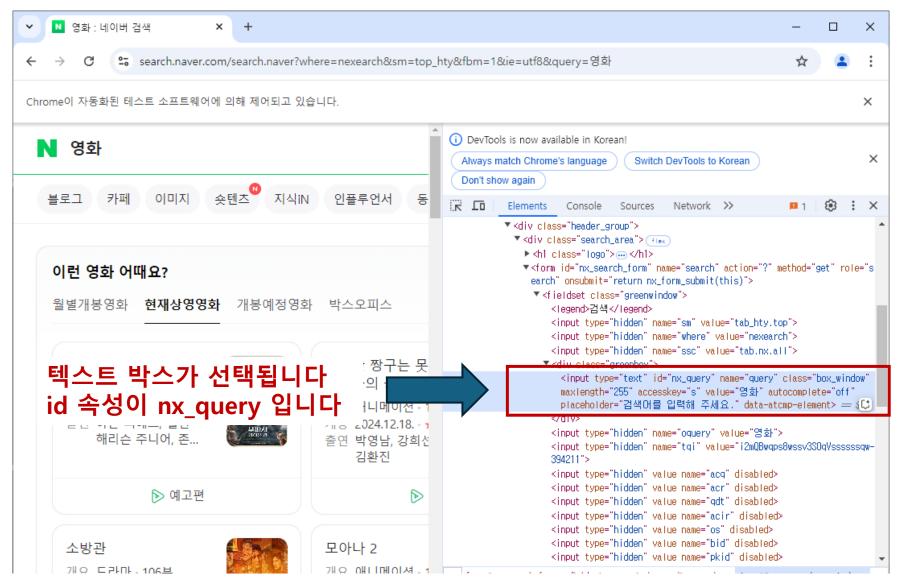


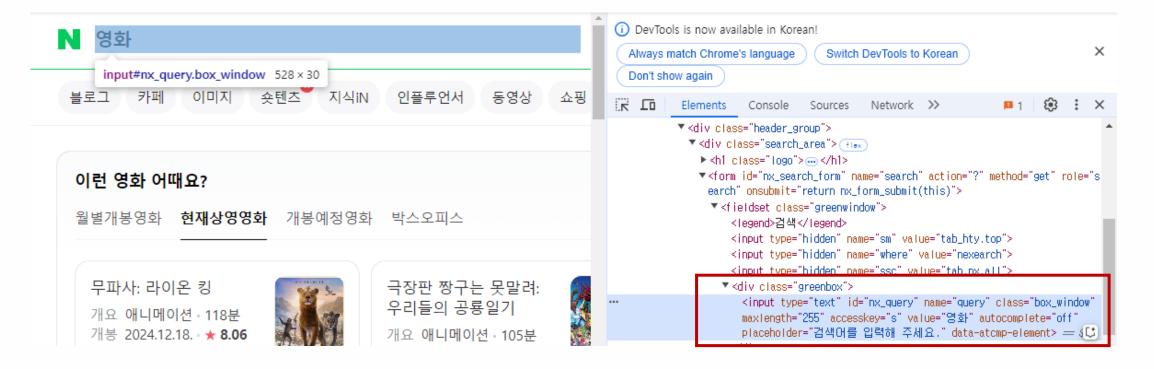


■ 영화 텍스트 박스를 선택합니다



■ 영화 텍스트 박스를 선택합니다



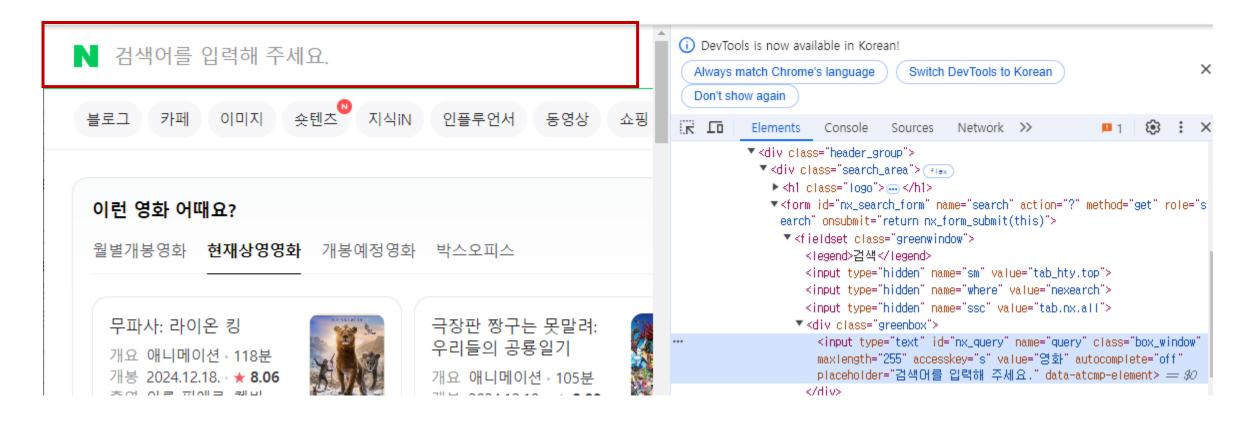


```
[6]: # 검색하려는 영화 제목
movie = "범죄도시 4"

#id 속성이 nx_query 인 객체(검색창)를 search_box에 대입
# 검색창 요소 찾기
search_box = driver.find_element(By.ID, 'nx_query')
#기존에 입력한 메시지 삭제
search_box.clear()
```

구현 후

■ 검색창의 텍스트가 삭제 됩니다





№ 영화 범죄도시 4

관심사를 반영한 컨텍스트 자동완성 ?

자동완성 끄기 - 도움말 - 신고

이런 영화 어때요?

월별개봉영화 현재상영영화 개봉예정영화 박스오피스

무파사: 라이온 킹 개요 애니메이션 - 118분 개봉 2024.12.18. ★ 8.06 출연 아론 피에르, 켈빈

해리슨 주니어, 존...



극장판 짱구는 못말려: 우리들의 공룡일기

개요 애니메이션 105분 개봉 2024.12.18. ★ 8.03 출연 박영남, 강희선, 김환진



▶ 예고편

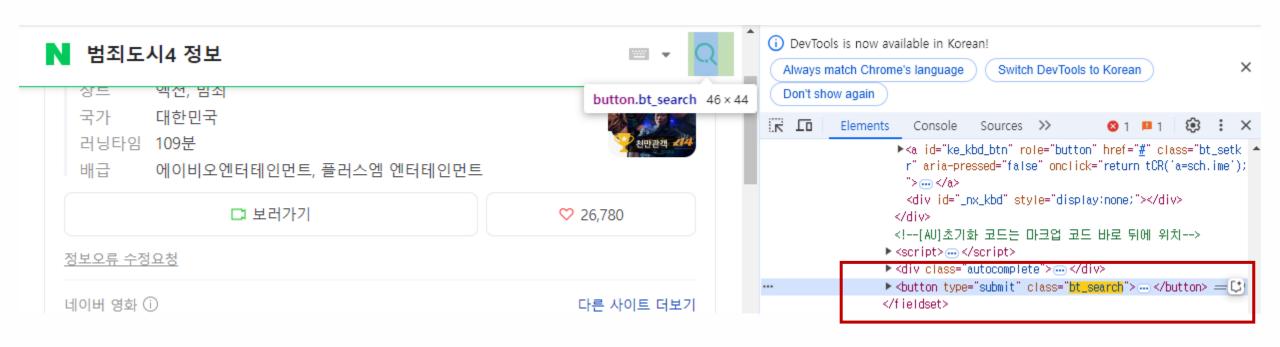
▶ 예고편

[7]: #**1**초 대기

time.sleep(1)

#id 속성이 nx_query인 객체(검색창)에 영화 영화 제목 입력 search_box.send_keys('영화', movie)

구현 후



```
[8]: #클래스 속성이 bt_search인 객체 (검색 버튼) 리턴
search_button = driver.find_element(By.CLASS_NAME,'bt_search')
#버튼 클릭
search_button.click()
Shift+Enter 입력해서 실행시킴
```





[9]: #1초 대기 time.sleep(1) #클래스 속성이 tab_list인 객체 (정보 카테고리 탭)리턴 tab_list = driver.find_element(By.CLASS_NAME, 'tab_list') #클래스 속성이 tab _tab (클래스 속성이 tab _tab) 인 객체들 (여러개) 중 인덱스 1인 객체 (두 번째 탭 선택) 리턴 tab = tab_list.find_elements(By.CSS_SELECTOR,".tab._tab")[1] #선택한 탭 클릭 tab.click()

```
[10]: #1초 대기
time.sleep(1)

#HTML 페이지 소스를 가져옵니다.
html = driver.page_source
#페이지의 정보를 추출 할 soup 객체 생성
soup = BeautifulSoup(html, 'html.parser')
```

구현 후

```
[11]: # 영화 제목 수집
#클래스 속성이 area_text_title인 객체(영화제목) 의 문자열 리턴
title = soup.select_one(".area_text_title").text
print("title=",title)
구현 후

title= 범죄도시4
```



```
57 UTV2
▼ <div class="cm_content_wrap">
 <div class="cm_content_area_cm_content_area_info">
    ▼ <div class="cm_info_box">

▼ <div class="detail info"> = 30

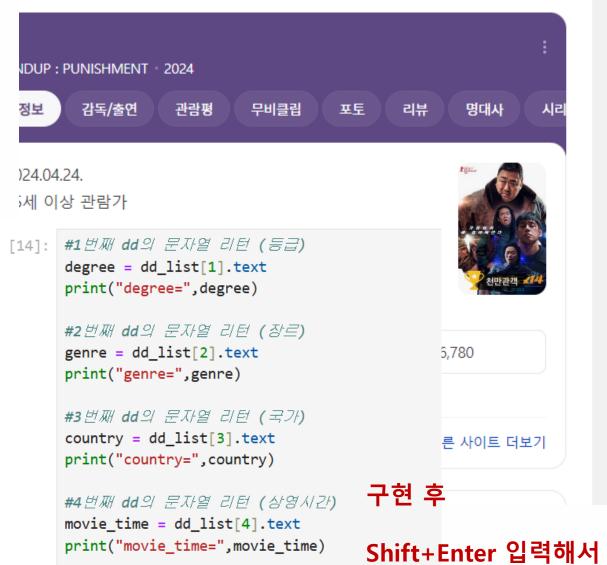
        <a nocr oncrick= return gouthertk(this, a=nco_xua+A.tab)</p>
         oster&r=1&i=1800009D_000001BB8D57&u=' + urlencode(this.hi
         ef));" href="?where=nexearch&sm=tab_etc&mra=bkEw&pkid=68{
         os=29068631&qvt=0&query=%EB%B2%94%EC%A3%84%EB%8F%84%EC%8
         B%9C4%20%ED%8F%AC%ED%86%AO" class="thumb_litem">
           ::before
           <img src="https://search.pstatic.net/common?type=o&siz/</pre>
           =176x264&quality=85&direct...2F20240425_256%2F1714007350
           O223JK9r_JPEG%2Fmovie_image.jpg%3Ftype%3Dw640_2" width:
            "88" height="132" alt="범죄도시4" class="_img">
         ▶ <div class="area badge 1000">.... </div>
         </a>
        ▶ <dl class="info txt_4">... </dl>
         :::after
       </div>

    Additional ages "button ages" >
```

```
[12]: # 영화 포스터 URL 수집
#클래스 속성이 detail_info인 객체 리턴
detail_info = soup.select_one(".detail_info")
#클래스 속성이 detail_info인 객체에 포함된 객체 중 클래스 속성이 _img인 객체 리턴
img = detail_info.select_one("._img")
#src 속성의 값 (영화 포스터의 url) 리턴
poster = img['src']
#영화 포스터의 url 출력
print("poster=",poster)
Shift+Enter 입력해서 실행시킴
```



```
[13]:
# 영화의 세부 정보를 수집 (개봉일, 등급, 장르, 국가, 상영시간)
#클래스 속성이 info txt_4 (2개 클래스 적용) 인 객체 리턴
info_group = soup.select_one(".info.txt_4")
#dd 엘레먼트 리턴 (여러개)
dd_list = info_group.select("dd")
#0번째 dd의 문자열 리턴 (개봉일)
open_date = dd_list[0].text
print("open_date=",open_date)
```



```
degree= 15세 이상 관람가
genre= 액션, 범죄
country= 대한민국
movie_time= 109분
```

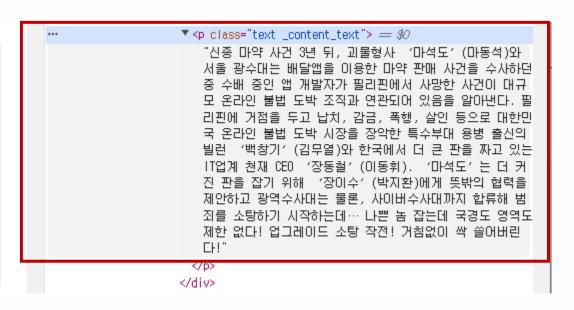
```
▼ <dl class="info txt_4">
 ▼ <div class="info_group">

▶ <dt> --- </dt>

     <dd>2024.04.24.</dd>
   </div>
 ▼ <div class="info_group">
   ▶ <dt class> --- </dt>
     <dd>15세 이상 관람가</dd> = 30
   </div>
 ▼ <div class="info_group">
   <dd>액션, 범죄</dd>
   </div>
 ▼ <div class="info_group">
   ▶ <dt>.... </dt>
     <dd>대한민국</dd>
   </div>
 ▼ <div class="info_group">
   ▶ <dt>.... </dt>
     <dd>109분</dd>
   </div>
 ▼ <div class="info_group">
   <dd>에이비오엔터테인먼트, 플러스엠 엔터테인먼트</dd>
   </div>
 </d1>
 ::after
</div>
```

#text 629.11 × 189

신종 마약 사건 3년 뒤, 괴물형사 '마석도'(마동석)와 서울 광수대는 배달앱을 이용한 마약판매 사건을 수사하던 중 수배 중인 앱 개발자가 필리핀에서 사망한 사건이 대규모 온라인불법 도박 조직과 연관되어 있음을 알아낸다. 필리핀에 거점을 두고 납치, 감금, 폭행, 살인등으로 대한민국 온라인불법 도박 시장을 장악한 특수부대 용병출신의 빌런 '백창기'(김무열)와 한국에서 더 큰 판을 짜고 있는 IT업계 천재 CEO '장동철'(이동휘). '마석도'는 더 커진 판을 잡기 위해 '장이수'(박지환)에게 뜻밖의 협력을 제안하고 광역수사대는 물론, 사이버수사대까지 합류해 범죄를 소탕하기 시작하는데... 나쁜 놈 잡는데 국경도 영역도 제한없다! 업그레이드 소탕 작전! 거침없이 싹 쓸어버린다!



[15]: #클래스 속성이 text _content_text (2개 클래스 적용) 인 객체 문자열(줄거리) 리턴
synopsis = soup.select_one(".text._content_text").text
print("synopsis=",synopsis)
구현 후

Shift+Enter 입력해서 실행시킴 synopsis= 신종 마약 사건 3년 뒤, 괴물형사 '마석도'(마동석)와 서울 광수대는 배달앱을 이용한 건이 대규모 오라의 북번 도반 조지과 여과되어 이유은 아이네다. 찌리피에 길찍은 드고 ''현' 그'