

태그 (tag)

<a>,

HTML문서에서 기능을 정의

속성 (attribute)

태그를 사용할 때,
태그의 성격, 목표 등을 설정



괜찮아, 고3이야

greeny227

너와나의 리얼고3툰

★★★★★ 9.91

누르면 새로운 탭에 관련 페이지 뜸.
이때 소스 확인시,

<href에 저장되어 있는 주소>

<태그=a, 속성=href>

```
▼<div class="challengeInfo">  
  ▼<h6 class="challengeTitle">  
    ▼<a onclick="nclk_v2(event, 'tum.tit', '', '1')" href="/  
bestChallenge/list.nhn?titleId=643799"> == $0
```

```
title = clips[i].select_one("h6.challengeTitle a")  
title.attrs["href"]
```

/bestChallenge/list.nhn?titleId=643799

<페이지의 실제주소>

<https://comic.naver.com/bestChallenge/list.nhn?titleId=643799>

"https://comic.naver.com" + title.attrs["href"]

<https://comic.naver.com/bestChallenge/list.nhn?titleId=643799>

<태그=img, 속성=src>

img 202 x 290



```
▼<div class="mv_info_area" style="display: block;">  
▶<div class="mv_info">...</div>  
<!-- 포스터 -->  
▼<div class="poster">  
▼<a href="#" onclick="javascript:common.iwopen('136900');  
clickcr(this,'ifo.img','','',event);return false;">  
 == $0  
<span>포스터 크게보기</span>  
::after  
</a>  
</div>
```

```

```

파이썬 기본 라이브러리

파이썬에 기본적으로 설치된 라이브러리
(설치 불필요)
(import 필요)

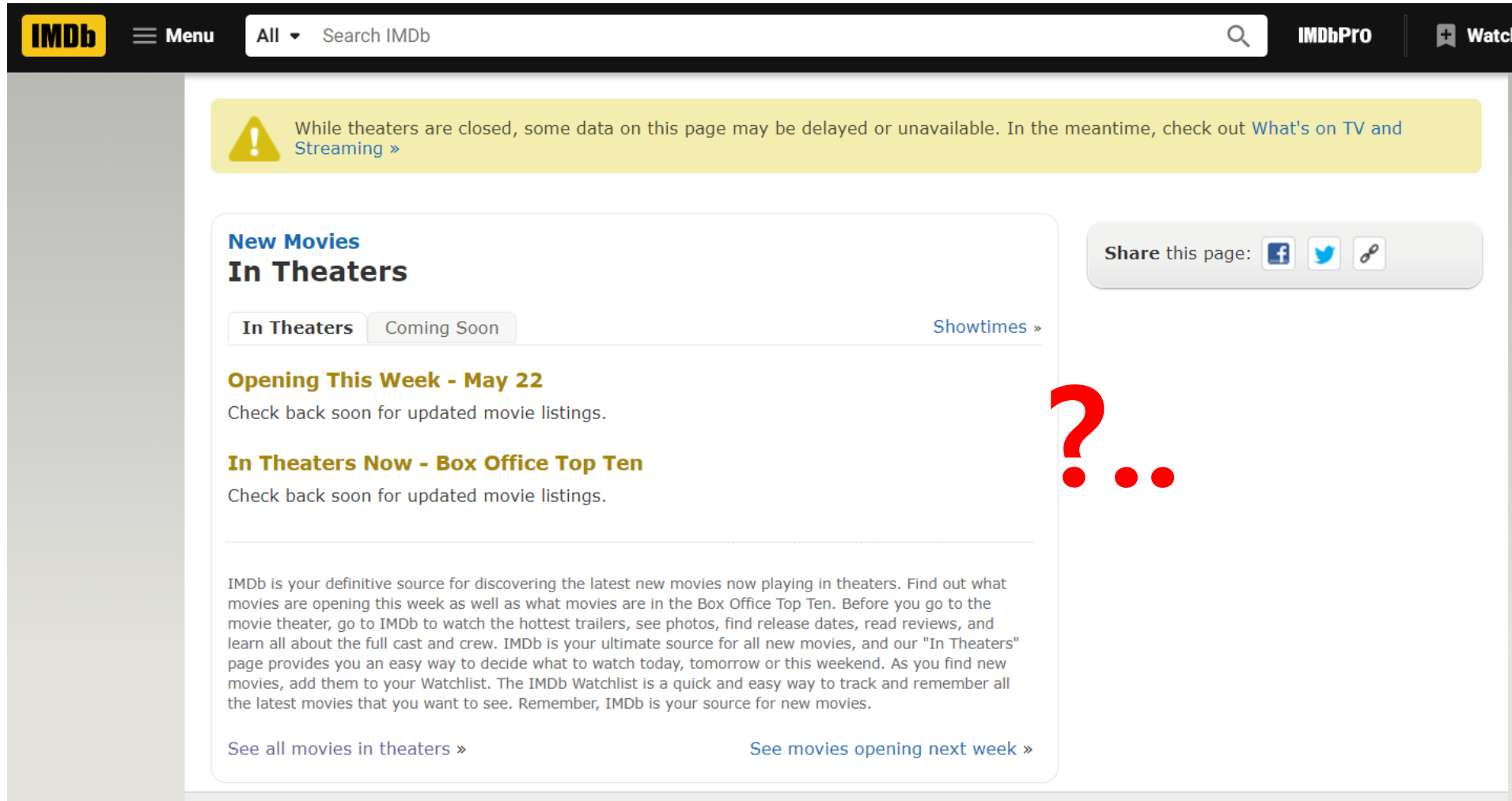
파이썬 외부 라이브러리

사람들이 만들어서 공유한 라이브러리
(설치 필요)
(import 필요)

urllib에 포함된 `urlretrieve`

```
from urllib.request import urlretrieve  
urlretrieve("https://movie-phinf.pstatic.net/20190417_250/1555465",  
           "어벤져스.png")
```

Challenge2_IMDb 포스터 저장하기



The screenshot shows the IMDb website interface. At the top, there is a navigation bar with the IMDb logo, a menu icon, a search bar with the text "All Search IMDb", and links for "IMDbPRO" and "Watchlist". Below the navigation bar, a yellow banner contains a warning icon and the text: "While theaters are closed, some data on this page may be delayed or unavailable. In the meantime, check out [What's on TV and Streaming](#) »".

The main content area is titled "New Movies In Theaters". It features two tabs: "In Theaters" (selected) and "Coming Soon". To the right of the tabs is a "Showtimes »" link. Below the tabs, there are two sections:

- Opening This Week - May 22**: "Check back soon for updated movie listings."
- In Theaters Now - Box Office Top Ten**: "Check back soon for updated movie listings."

At the bottom of the main content area, there is a paragraph of text: "IMDb is your definitive source for discovering the latest new movies now playing in theaters. Find out what movies are opening this week as well as what movies are in the Box Office Top Ten. Before you go to the movie theater, go to IMDb to watch the hottest trailers, see photos, find release dates, read reviews, and learn all about the full cast and crew. IMDb is your ultimate source for all new movies, and our 'In Theaters' page provides you an easy way to decide what to watch today, tomorrow or this weekend. As you find new movies, add them to your Watchlist. The IMDb Watchlist is a quick and easy way to track and remember all the latest movies that you want to see. Remember, IMDb is your source for new movies."

At the bottom of the main content area, there are two links: "See all movies in theaters »" and "See movies opening next week »".

On the right side of the page, there is a "Share this page:" button with icons for Facebook, Twitter, and a link icon. A large red question mark and three red dots are overlaid on the page, pointing towards the empty movie listings.

Challenge2_대체 <네이버 베스트웹툰 이미지 가져오기>

NAVER 만화 | 웹소설




제목/작가로 검색할 수 있습니다.

홈 웹툰 **베스트 도전** 도전만화 마이페이지 단행본만화 장르소설

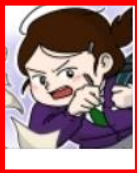



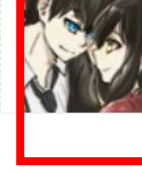
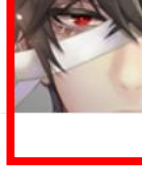
전체만화 에피소드 오니버스 스토리 | 일상 개그 판타지 액션 드라마 순정 감성 스릴러 시대극 스포츠

베스트 도전만화는 누구나 참여할 수 있는 창작만화(UCC) 게시판입니다.

오늘의 인기 베스트 2020.05.22

 낮잠 자는 사람들 31화 계약 (1) dwrumi ★★★★★ 9.91	 신과 인간 35. 윤시연 vs 예코 ino1998 ★★★★★ 9.88	 월오브포춘 9화 0331tt ★★★★★ 9.71
---	---	--

전체 업데이트순 조회순 별점순

 괜찮아, 고3이야 greeny227 너와나의 리얼고3툰 ★★★★★ 9.91	 천상대전 myjune210 천상대전 ★★★★★ 9.86	 고양이 kimyonari 취준생 고양이 ★★★★★ 9.91
 정신과심리만화 by... palhosquare 알고싶니 마음 ★★★★★ 9.75	 런어웨이 hama2294 초능력 소녀,그의 소... ★★★★★ 9.85	 여왕의 계절 imbobo 시대극판타지 웹툰 ★★★★★ 9.88

컨테이너 : div.challengeInfo
title 선택자 : h6.challengeTitle

```
<div class="challengeInfo">  
  <h6 class="challengeTitle">  
    <a onclick="nclk v2(event,'tum.tit','','1')" href="/bestChallenge/list.nhn?titleId=643799"> == $0  
  </h6>  
</div>
```

comic.naver.com/bestChallenge/list.nhn?titleId=643799

NAVER 만화 | 웹소설

제목/작가로 검색할 수 있습니다.

홈 웹툰 **베스트 도전** 도전만화 마이페이지 단행본만화 장르소설

전체만화 에피소드 오니버스 스토리 | 일상 개그 판타지 액션 드라마 순정 감성 스릴러 시대극 스포츠

베스트 도전만화는 누구나 참여할 수 있는 창작만화(UCC) 게시판입니다.

 **괜찮아, 고3이야** greeny227
도대체 고3이 뭐길래!!! 로맨스도 화려한 액션도 없는 노잼현실을 살아가는 진짜 고3의 이야기

관심웹툰 | 첫화보기 | 작가의 다른 작품 | 9,928

1. 소스 코드 수집하기

```
import requests
from bs4 import BeautifulSoup
from urllib.request import urlretrieve

raw = requests.get(
    "https://comic.naver.com/genre/bestChallenge.nhn",
    headers={
        'User-Agent': 'Mozilla/5.0 (X11; Linux x86_64)'
    })
html = BeautifulSoup(raw.text, "html.parser")
clips = html.select("div.challengeInfo")
```

컨테이너 : div.challengeInfo
title 선택자 : h6.challengeTitle

2. url 생성하기

```
for i in range(len(clips)):
    title = clips[i].select_one("h6.challengeTitle a")
    url = title.attrs["href"]
```

<결과창>

```
/bestChallenge/list.nhn?titleId=643799
/bestChallenge/list.nhn?titleId=690673
/bestChallenge/list.nhn?titleId=641817
/bestChallenge/list.nhn?titleId=729599
/bestChallenge/list.nhn?titleId=550073
/bestChallenge/list.nhn?titleId=655184
/bestChallenge/list.nhn?titleId=694186
/bestChallenge/list.nhn?titleId=671295
```

#url 생성하기

```
url = "https://comic.naver.com" + title.attrs["href"]
raw_each = requests.get(url, headers={"User-Agent": "Mozilla/5.0"})
html_each = BeautifulSoup(raw_each.text, 'html.parser')
```

<결과창>

```
https://comic.naver.com/bestChallenge/list.nhn?titleId=643799
https://comic.naver.com/bestChallenge/list.nhn?titleId=690673
https://comic.naver.com/bestChallenge/list.nhn?titleId=641817
https://comic.naver.com/bestChallenge/list.nhn?titleId=729599
https://comic.naver.com/bestChallenge/list.nhn?titleId=550073
https://comic.naver.com/bestChallenge/list.nhn?titleId=655184
https://comic.naver.com/bestChallenge/list.nhn?titleId=694186
https://comic.naver.com/bestChallenge/list.nhn?titleId=671295
```

img 98 x 79



```

... <div class="thumb"> == $0
  <a onclick="nclk_v2(event,'inf.img')" href="/bestChallenge/
  list.nhn?titleId=643799">
    
    <span class="mask"></span>
  </a>
</div>

```

이미지 선택자 : div.thumb img

html body #wrap div#container div#content div.comicinfo div.thumb a img

div.thumb img

```

#웹툰 이미지 출력하기
poster=html_each.select_one("div.thumb img")
poster_src=poster.attrs["src"]
print(poster_src)

```

<결과창>

- https://shared-comic.pstatic.net/thumb/user_contents_data/challenge_comic/2020/03/16/greeny227/thumbnail_98x79aca1b08f_d88e_48a4_92e9_2bd45a5ccbf7_00000009.JPEG
- https://shared-comic.pstatic.net/thumb/user_contents_data/challenge_comic/2020/04/30/myjune210/title_thumbnail_004227_98x79.jpg
- https://shared-comic.pstatic.net/thumb/user_contents_data/challenge_comic/2016/05/28/kimyonari/thumbnail_title_kimyonari_143232.jpg
- https://shared-comic.pstatic.net/thumb/user_contents_data/challenge_comic/2020/03/12/palhosquare/thumbnail_98x7984a23498_ae11_47a9_8765_ea5fb217671e_00006206.JPEG
- https://shared-comic.pstatic.net/thumb/user_contents_data/challenge_comic/2015/11/01/hama2294/thumbnail_title_hama2294_141826.jpg
- https://shared-comic.pstatic.net/thumb/user_contents_data/challenge_comic/2018/05/20/imbobo/thumbnail_98x7920795a04_2347_48df_b0d5_3c063cefd96c_00001003.JPEG

웹툰제목_3글자.png 로 이미지 저장하기

```
urlretrieve(postersrc, "wposter/"+title.text.strip()[:3]+".png")
```

