파이썬 웹 크롤링

강사 장철원
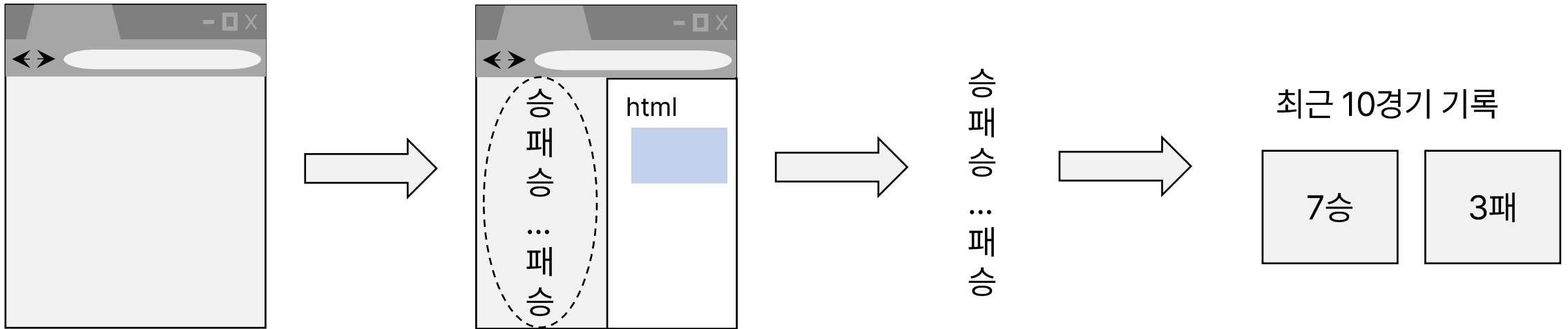
❑ 동적 크롤링 활용해서 해외 축구 데이터 크롤링하기

파이썬 웹 크롤링

Section 1. 해외 축구 데이터 크롤링

Section 1-1. 크롤링 웹사이트 확인
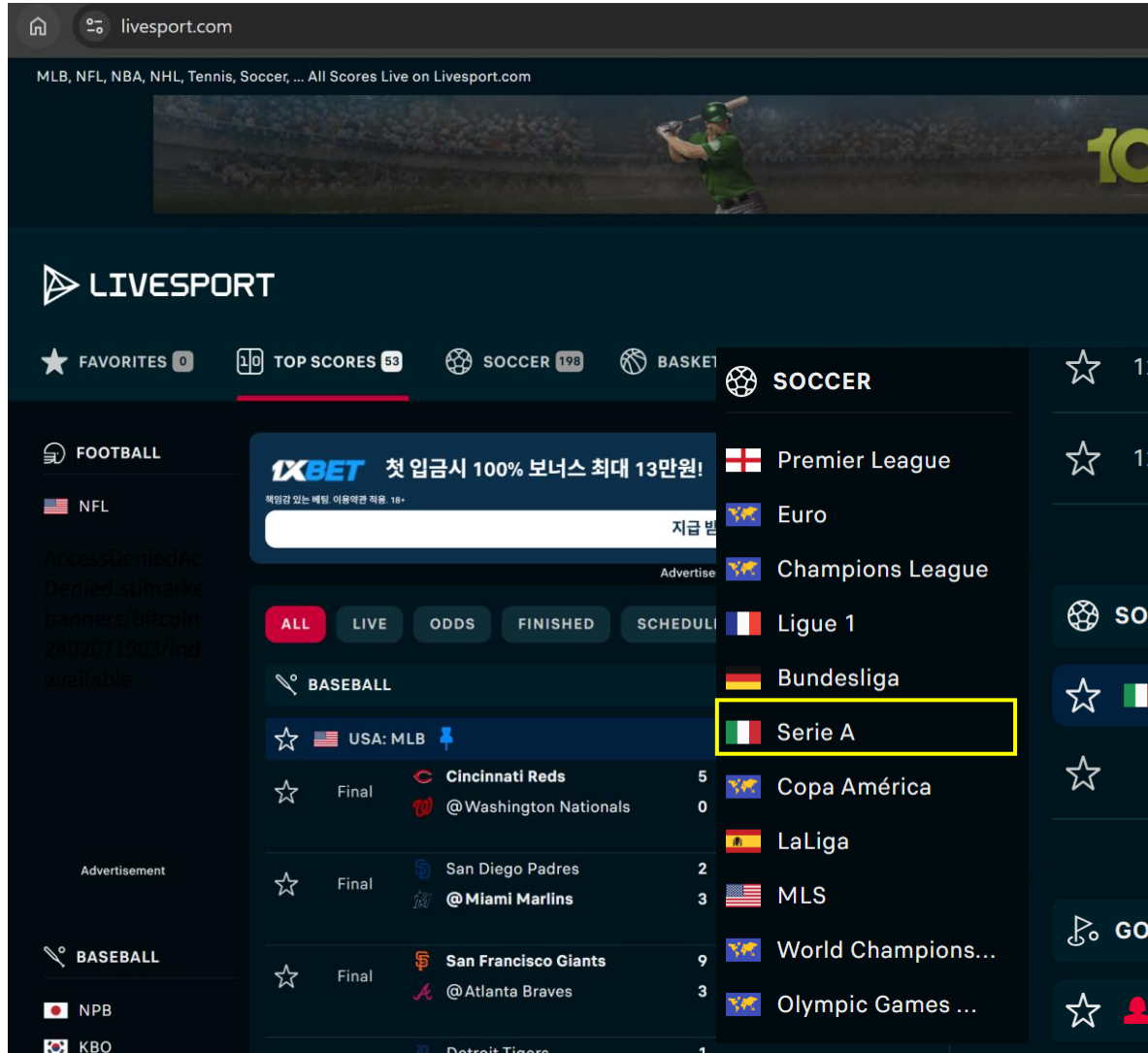
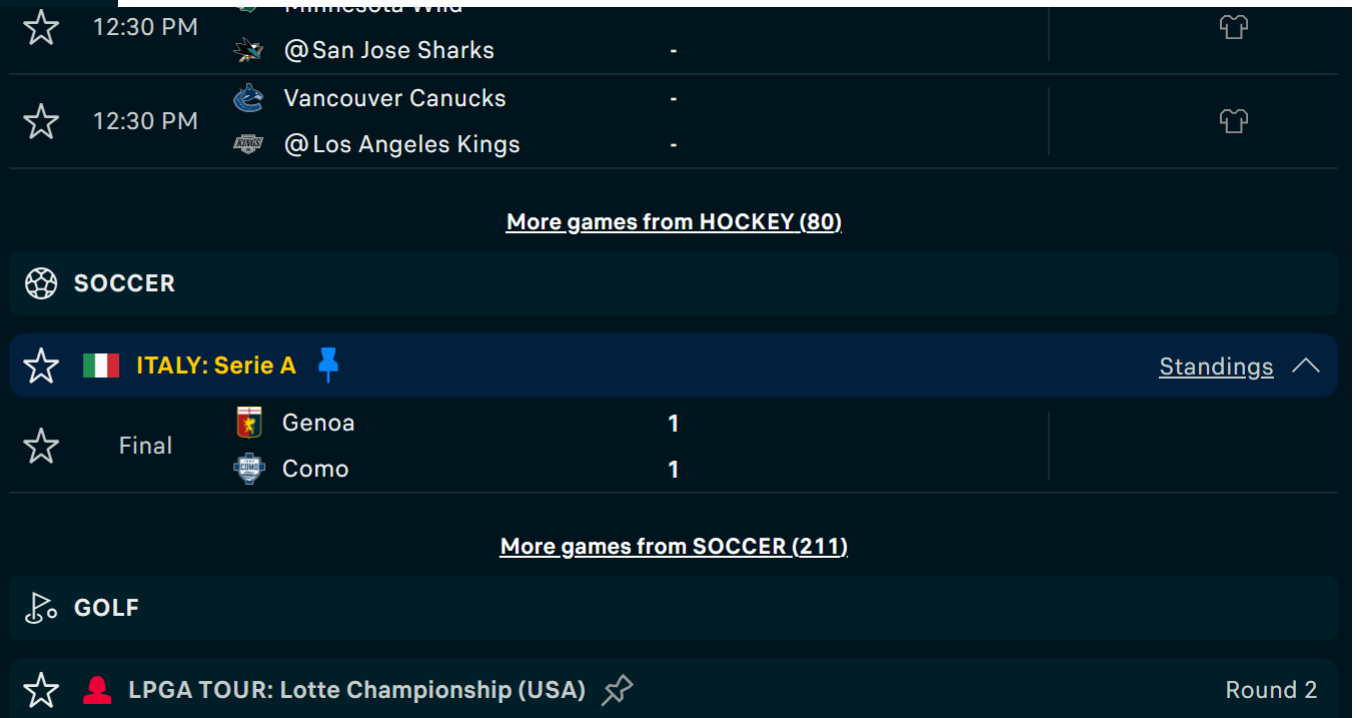# 크롤링 개요

해외 스포츠 기록 사이트 확인

https://www.livesport.com/

스크롤 내려서 SOCCER 메뉴의 Serie A 확인

해외 스포츠 기록 사이트 확인   https://www.livesport.com/

Azerbaijan

Bahrain

Bangladesh

Barbados

Belarus

Belgium

Benin

Bermuda

Bhutan

OVERALL   HOME   AWAY

# ▲	TEAM		GP	W	T	L	G	GD	PTS	FORM					
1.		Napoli	11	8	1	2	18:8	10	25	?	L	W	W	W	W
2.		Inter	11	7	3	1	25:13	12	24	?	W	W	T	W	W
3.		Atalanta	11	7	1	3	29:14	15	22	?	W	W	W	W	W
4.		Fiorentina	11	6	4	1	22:9	13	22	?	W	W	W	W	W
5.		Lazio	11	7	1	3	24:14	10	22	?	W	W	W	L	W
6.		Juventus	11	5	6	0	19:7	12	21	?	W	T	T	W	T
7.		AC Milan 클릭	10	5	2	3	17:11	6	17	?	W	L	W	L	W
8.		Udinese	11	5	1	5	14:16	-2	16	?	L	L	W	L	W
9.		Bologna	10	3	6	1	12:11	1	15	?	W	W	T	T	T
10.		Torino	11	4	2	5	15:16	-1	14	?	L	L	W	L	L

해외 스포츠 기록 사이트 확인

최근 전적 확인

| | 🏠 | livesport.com/team/ac-milan/8Sa8HInO/ | |

해외 스포츠 기록 사이트 확인

https://www.livesport.com/

해외 스포츠 기록 사이트 확인 https://www.livesport.com/

```
▼<button class="wcl-badgeform_yYFgV wcl-win_9cMNW wcl-clickable_BbSlo
    formIcon__lastMatches" title="Win" data-testid="wcl-badgeForm-win">
    flex
    <span class="wcl-simpleText_Asp-0 wcl-scores-simpleText-01_pV2Wk"
      data-testid="wcl-scores-simpleText-01">W</span> == $0
  </button>
</div>
```

span 태그 사용 및
wcl-simpleText_Asp-0 wcl-scores-simpleText-01_pV2Wk
클래스 사용 확인

해외 스포츠 기록 사이트 확인 https://www.livesport.com/

```
[1]: from urllib.request import urlopen
     from bs4 import BeautifulSoup
```

```
[2]: url = "https://www.livesport.com/team/ac-milan/8Sa8HInO/"
     html = urlopen(url)
     bs_obj01 = BeautifulSoup(html, "html.parser")
     win01 = bs_obj01.find_all("span", {"class":"wcl-simpleText_Asp-0 wcl-scores-simpleText-01_pV2Wk"})
```

```
[3]: print(win01)
```

    []   아무것도 안나온다?

해외 스포츠 기록 사이트 확인   https://www.livesport.com/

```
[4]:  print(bs_obj01)          bs_obj01은 잘 나오는거 같은데?

<!DOCTYPE html>

<html lang="en-US">
<head>
<meta charset="utf-8"/>
<title>Livesport: AC Milan - results, fixtures, Liverpool - AC Milan live</title>
<script charset="UTF-8" data-document-language="true" data-domain-script="be4cd6ae-414e-4784-a9f8-29c9eb574714" defer="" src="https://cdn.cookielaw.org/scripttemplates/otSDKStub.js" type="text/javascript">
</script>
<link crossorigin="" href="https://500.ds.lsapp.eu/pq_graphql" rel="preconnect"/>
<link crossorigin="" href="https://500.flashscore.ninja" rel="preconnect"/>
<link as="font" crossorigin="" href="https://static.flashscore.com/res/font/LivesportFinderLatin-Regular_Static.woff2" rel="preload" type="font/woff2"/>
<link as="font" crossorigin="" href="https://static.flashscore.com/res/font/LivesportFinderLatin-Bold_Static.woff2" rel="preload" type="font/woff2"/>
<noscript>
<meta content="0;url=https://www.flashscore.mobi/" http-equiv="refresh">
</meta></noscript>
<meta content="AC Milan scores and fixtures - follow AC Milan livescore, results, fixtures and match details on Livesport.com." name="description"/>
<meta content="Copyright (c) 2012-2025 Livesport s.r.o." name="copyright"/>
<meta content="index,follow" name="robots">
<meta content="width=device-width, initial-scale=1" name="viewport"/>
```
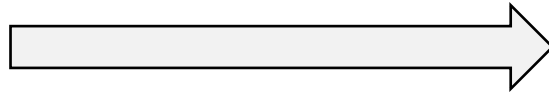
wcl-simpleText_Asp-0 wcl-scores-simpleText-01_pV2Wk 클래스 검색하면 안나옴

정적 페이지 vs 동적 페이지

서버

환영합니다, **고객**님
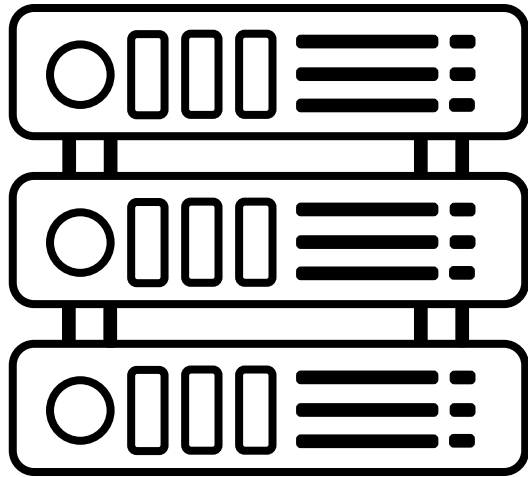
클라이언트

정적 페이지 vs 동적 페이지

데이터베이스

서버

환영합니다, **길동**님
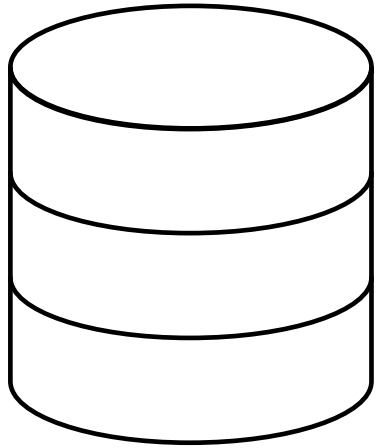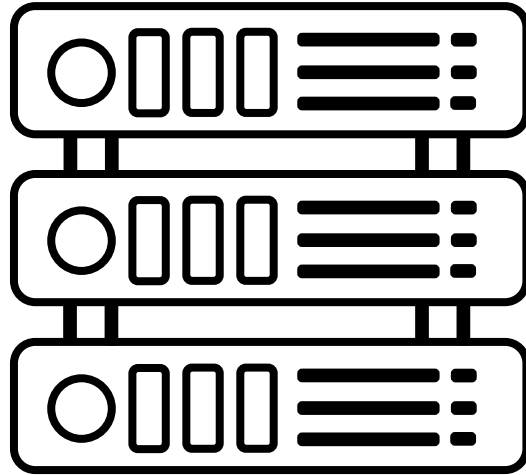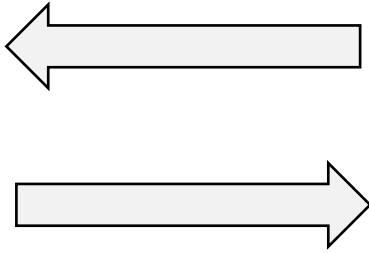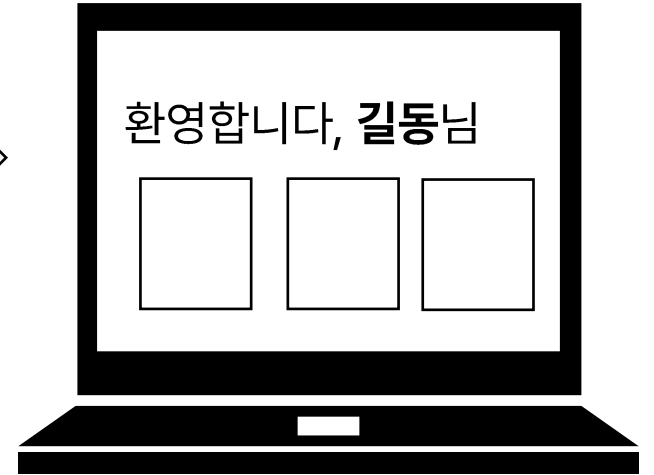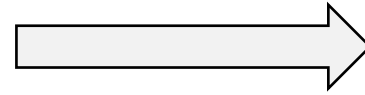
클라이언트

개발자 도구 vs 페이지 소스

개발자 도구 - 웹 서버에서 전달 받은 이후 가공한 후의 최종적으로 브라우저가 그린 결과물

페이지 소스 보기 - 웹 서버에서 최초로 전달 받은 내용이 그대로 나타남. beautifulsoup을 이용해 파싱할 때는 페이지 소스를 파싱하므로 가공 이후의 내용은 표시되지 않음(자바스크립트 실행되기 이전 원본 HTML)
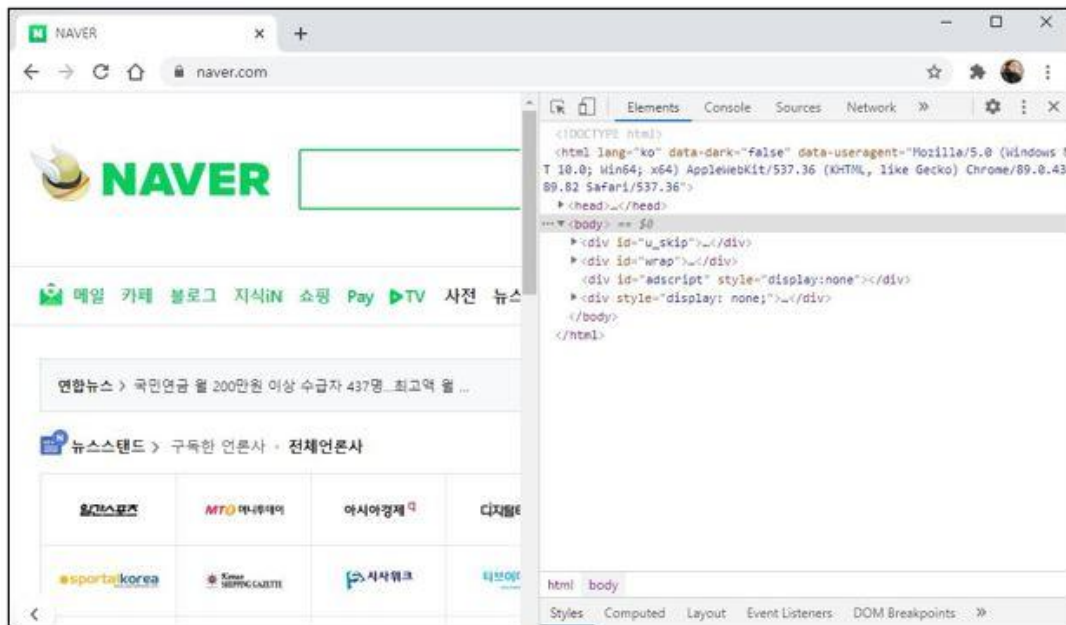
개발자 도구
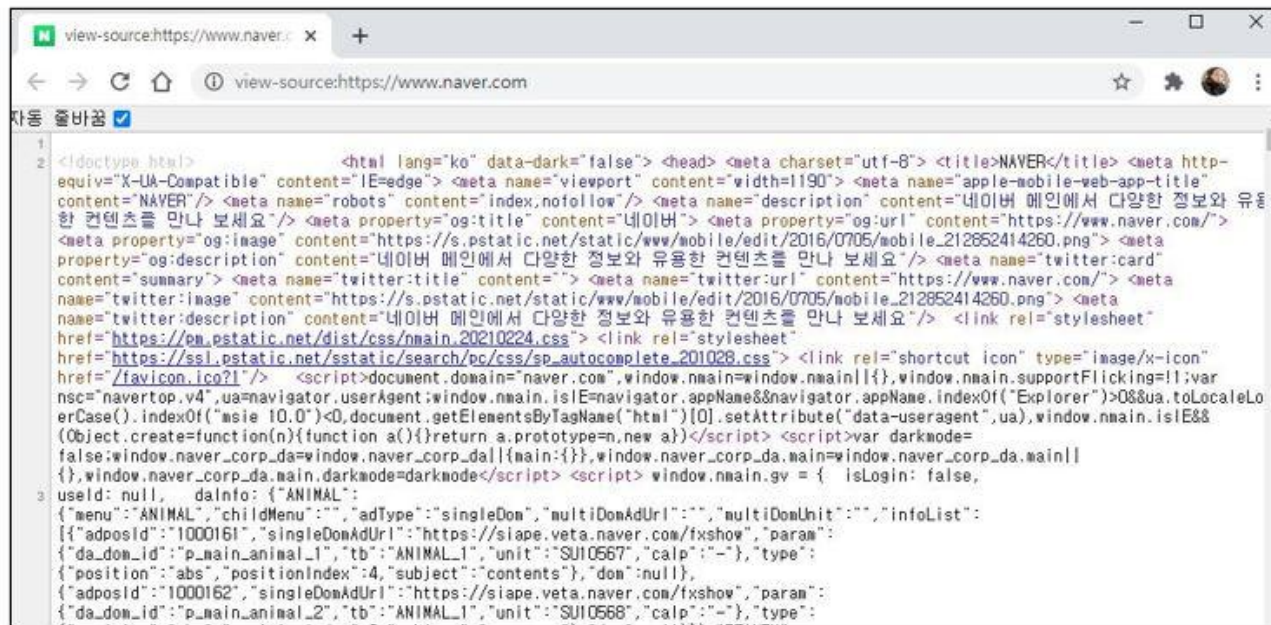
페이지 소스

개발자 도구 vs 페이지 소스

개발자 도구
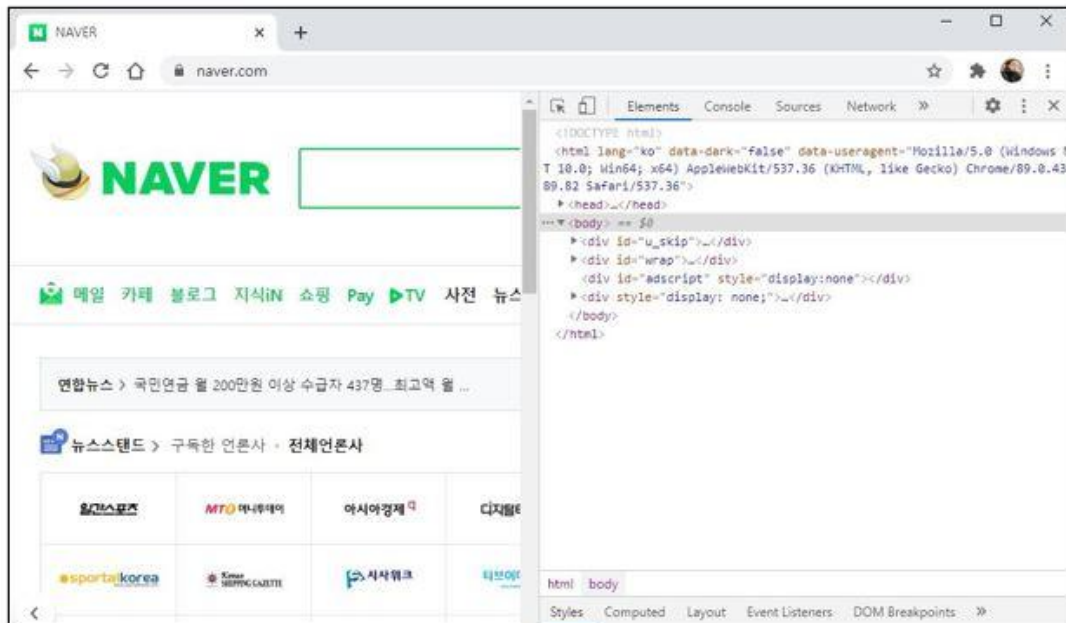
페이지 소스

웹 브라우저가 문서를 읽고,
자바스크립트도 실행해서 화면에 보여준 최종 상태

원본 스크립트만 존재

개발자 도구 - 키워드 나옴

페이지 소스 - 키워드 안나옴

해외 스포츠 기록 사이트 확인 https://www.livesport.com/

```
[5]: from selenium import webdriver
```
webdriver: 브라우저를 제어하기 위한 다양한 도구를 포함하고 있는 모듈

```
[ ]: #options = webdriver.ChromeOptions()
     #options.add_argument("--headless")
```
크롬 브라우저 설정 변경을 위한 객체 생성

headless 옵션: 브라우저 창이 열리지 않고 백그라운드에서 실행하는 모드

```
[6]: driver = webdriver.Chrome()
```
크롬 브라우저 제어를 위한 객체 생성

```
[7]: driver.implicitly_wait(3)
```
웹 페이지가 로딩될때까지 기다리는 시간, 단위는 초(second)

```
[8]: driver.get("https://www.livesport.com/team/ac-milan/8Sa8HInO/")
```
해당 웹 페이지로 이동

해외 스포츠 기록 사이트 확인 https://www.livesport.com/

```
[9]:  from bs4 import BeautifulSoup
```

```
[10]: page = driver.page_source
```
현재 브라우저에서 로드된 전체 HTML 소스 코드 변환

```
[11]: bs_obj = BeautifulSoup(page, "html.parser")
      print(bs_obj)
```

```
<html lang="en-US"><head>
<meta charset="utf-8"/>
<title>Livesport: AC Milan - results, fixtures, Liverpool - AC Milan live</title>
<script async="" src="https://www.googletagmanager.com/gtag/destination?id=AW-999393459&a
1509157~103116026~103200004~103233427~104684208~104684211" type="text/javascript"></scrip
web-vitals.iife.js" type="text/javascript"></script><script async="" src="https://www.goo
7m0v71314122za200&amp;tag_exp=101509157~103116026~103200004~103233427~104684208~104684211
ogletagmanager.com/gtm.js?id=GTM-PWJ3NQ"></script><script charset="UTF-8" data-document-l
b574714" defer="" src="https://cdn.cookielaw.org/scripttemplates/otSDKStub.js" type="text
</script>
<link crossorigin="" href="https://500.ds.lsapp.eu/pq_graphql" rel="preconnect"/>
```

해외 스포츠 기록 사이트 확인 https://www.livesport.com/

```
[12]:  win = bs_obj.find_all("span", {"class":"wcl-simpleText_Asp-0 wcl-scores-simpleText-01_pV2Wk"})
```

```
[13]:  print(win)
```

```
[<span class="wcl-simpleText_Asp-0 wcl-scores-simpleText-01_pV2Wk" data-testid="wcl-scores-simple
scores-simpleText-01_pV2Wk" data-testid="wcl-scores-simpleText-01">Bracket</span>, <span class="w
-testid="wcl-scores-simpleText-01">Standings</span>, <span class="wcl-simpleText_Asp-0 wcl-scores
-01"></span>, <span class="wcl-simpleText_Asp-0 wcl-scores-simpleText-01_pV2Wk" data-testid="wcl-
ext_Asp-0 wcl-scores-simpleText-01_pV2Wk" data-testid="wcl-scores-simpleText-01">Standings</span>
t-01_pV2Wk" data-testid="wcl-scores-simpleText-01">W</span>, <span class="wcl-simpleText_Asp-0 we
mpleText-01">L</span>, <span class="wcl-simpleText_Asp-0 wcl-scores-simpleText-01_pV2Wk" data-te
ss="wcl-simpleText_Asp-0 wcl-scores-simpleText-01_pV2Wk" data-testid="wcl-scores-simpleText-01">
mpleText-01_pV2Wk" data-testid="wcl-scores-simpleText-01">Standings</span>, <span class="wcl-sim
="wcl-scores-simpleText-01">W</span>, <span class="wcl-simpleText_Asp-0 wcl-scores-simpleText-01_
<span class="wcl-simpleText_Asp-0 wcl-scores-simpleText-01_pV2Wk" data-testid="wcl-scores-simple
scores-simpleText-01_pV2Wk" data-testid="wcl-scores-simpleText-01">Bracket</span>, <span class="w
-testid="wcl-scores-simpleText-01">W</span>, <span class="wcl-simpleText_Asp-0 wcl-scores-simple
ndings</span>, <span class="wcl-simpleText_Asp-0 wcl-scores-simpleText-01_pV2Wk" data-testid="wcl
Text_Asp-0 wcl-scores-simpleText-01_pV2Wk" data-testid="wcl-scores-simpleText-01">W</span>]
```

```
[14]:  len(win)
```

```
[14]:  19
```

해외 스포츠 기록 사이트 확인 https://www.livesport.com/

```python
for i in range(0, len(win)):
    print(win[i])
```

```python
for i in range(0, len(win)):
    print(win[i].text)
```

```
<span class="wcl-simpleText_Asp-0 wcl-scores-simpleText-01_pV2Wk" data-testid="wcl-scores-simpleText-01"></span>
<span class="wcl-simpleText_Asp-0 wcl-scores-simpleText-01_pV2Wk" data-testid="wcl-scores-simpleText-01">Bracket</span>
<span class="wcl-simpleText_Asp-0 wcl-scores-simpleText-01_pV2Wk" data-testid="wcl-scores-simpleText-01">Standings</span>
<span class="wcl-simpleText_Asp-0 wcl-scores-simpleText-01_pV2Wk" data-testid="wcl-scores-simpleText-01"></span>
<span class="wcl-simpleText_Asp-0 wcl-scores-simpleText-01_pV2Wk" data-testid="wcl-scores-simpleText-01">L</span>
<span class="wcl-simpleText_Asp-0 wcl-scores-simpleText-01_pV2Wk" data-testid="wcl-scores-simpleText-01">Standings</span>
<span class="wcl-simpleText_Asp-0 wcl-scores-simpleText-01_pV2Wk" data-testid="wcl-scores-simpleText-01">W</span>
<span class="wcl-simpleText_Asp-0 wcl-scores-simpleText-01_pV2Wk" data-testid="wcl-scores-simpleText-01">L</span>
<span class="wcl-simpleText_Asp-0 wcl-scores-simpleText-01_pV2Wk" data-testid="wcl-scores-simpleText-01">Bracket</span>
<span class="wcl-simpleText_Asp-0 wcl-scores-simpleText-01_pV2Wk" data-testid="wcl-scores-simpleText-01">L</span>
<span class="wcl-simpleText_Asp-0 wcl-scores-simpleText-01_pV2Wk" data-testid="wcl-scores-simpleText-01">Standings</span>
<span class="wcl-simpleText_Asp-0 wcl-scores-simpleText-01_pV2Wk" data-testid="wcl-scores-simpleText-01">W</span>
<span class="wcl-simpleText_Asp-0 wcl-scores-simpleText-01_pV2Wk" data-testid="wcl-scores-simpleText-01">W</span>
<span class="wcl-simpleText_Asp-0 wcl-scores-simpleText-01_pV2Wk" data-testid="wcl-scores-simpleText-01">W</span>
<span class="wcl-simpleText_Asp-0 wcl-scores-simpleText-01_pV2Wk" data-testid="wcl-scores-simpleText-01">Bracket</span>
<span class="wcl-simpleText_Asp-0 wcl-scores-simpleText-01_pV2Wk" data-testid="wcl-scores-simpleText-01">W</span>
<span class="wcl-simpleText_Asp-0 wcl-scores-simpleText-01_pV2Wk" data-testid="wcl-scores-simpleText-01">Standings</span>
<span class="wcl-simpleText_Asp-0 wcl-scores-simpleText-01_pV2Wk" data-testid="wcl-scores-simpleText-01">L</span>
<span class="wcl-simpleText_Asp-0 wcl-scores-simpleText-01_pV2Wk" data-testid="wcl-scores-simpleText-01">W</span>
```

```
Bracket
Standings

L
Standings
W
L
Bracket
L
Standings
W
W
W
Bracket
W
Standings
L
W
```

해외 스포츠 기록 사이트 확인

https://www.livesport.com/

```python
for i in range(0, len(win)):
    if win[i].text == 'W':
        result['win'] += 1
    elif win[i].text == 'D':
        result['draw'] += 1
    elif win[i].text == 'L':
        result['lose'] += 1
    print(result)
```

```
{'win': 0, 'draw': 0, 'lose': 0}
{'win': 0, 'draw': 0, 'lose': 0}
{'win': 0, 'draw': 0, 'lose': 0}
{'win': 0, 'draw': 0, 'lose': 0}
{'win': 0, 'draw': 0, 'lose': 1}
{'win': 0, 'draw': 0, 'lose': 1}
{'win': 1, 'draw': 0, 'lose': 1}
{'win': 1, 'draw': 0, 'lose': 2}
{'win': 1, 'draw': 0, 'lose': 2}
{'win': 1, 'draw': 0, 'lose': 3}
{'win': 1, 'draw': 0, 'lose': 3}
{'win': 2, 'draw': 0, 'lose': 3}
{'win': 3, 'draw': 0, 'lose': 3}
{'win': 4, 'draw': 0, 'lose': 3}
{'win': 4, 'draw': 0, 'lose': 3}
{'win': 5, 'draw': 0, 'lose': 3}
{'win': 5, 'draw': 0, 'lose': 3}
{'win': 5, 'draw': 0, 'lose': 4}
{'win': 6, 'draw': 0, 'lose': 4}
```

[17]:
```python
result = {'win':0, 'draw':0, 'lose':0}
```

[18]:
```python
keys = list(result.keys())
print(keys)
```

```
['win', 'draw', 'lose']
```

[20]:
```python
print(result)
```

```
{'win': 6, 'draw': 0, 'lose': 4}
```

감사합니다.

Q & A