

Web Crawling



학습 내용

1. 웹크롤링 기초
2. 정적 크롤링
3. 동적 크롤링

학습내용

크롤링 기초

정적 크롤링

동적 크롤링

- 인터넷 웹(Wide Wide Web)
- 웹서버와 DB
- Http와 소켓
- 브라우저
- 웹앱과 api
- 크롤링 주의 사항

학습내용

크롤링 기초

정적 크롤링

동적 크롤링

- requests
- xml, html, json, jsonp, binary
- regex(정규식)
- beautifulsoup
- css selector

학습내용

크롤링 기초

정적 크롤링

동적 크롤링

- browser driver
- selenium

웹 크롤링 기초

웹 크롤러(web crawler)란?

- 조직적, 자동화된 방법으로 **월드 와이드 웹**을 탐색하는 컴퓨터 프로그램 - 위키피디아 -
- 크롤링 vs 스크래핑 vs 파싱

크롤링 이해

홍길동의
홈페이지

gdhong.net



심청이의
홈페이지

ceshim.net



춘향이의
홈페이지

chlee.com



피터의
홈페이지

peter.com

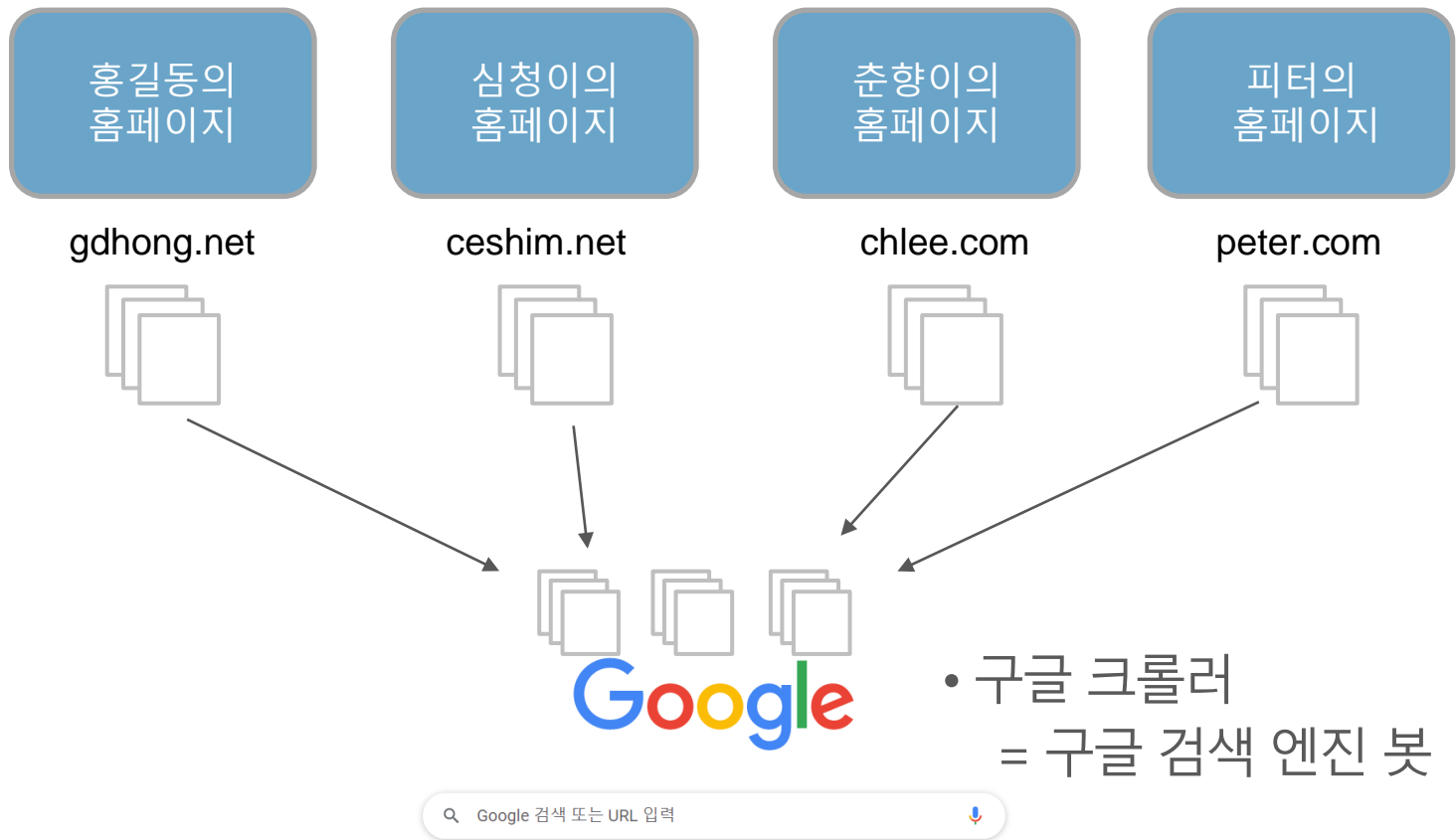


Google

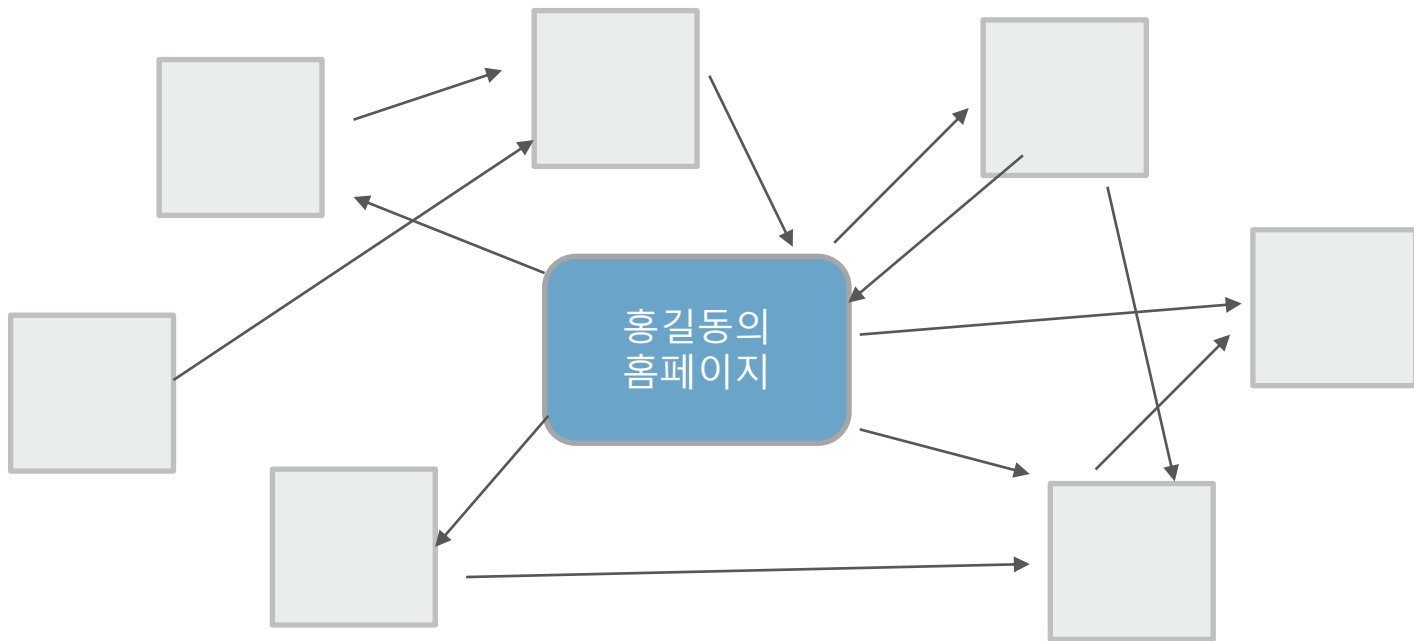
Google 검색 또는 URL 입력



크롤링 이해



크롤링 이해



- 크롤러 봇이 링크를 타고 다니며 검색어의 위치를 파악함

크롤링 vs 스크래핑 vs 파싱

- 크롤링(크롤러; crawler) : 조직적, 자동화된 방법으로 월드 와이드 웹을 탐색하는 것
- 스크래핑(scraping) : http를 통해 웹 사이트의 내용을 긁어와서 원하는 형태로 가공하는 것
- 파싱(파서; parser) : 어떤 페이지(문서, html 등)에서 내가 원하는 데이터를 특정 패턴이나 순서로 추출해 가공하는 것, 구문 분석 과정

크롤링(스크래핑) 시연

• 검색 상위 주식 종가 가져오기

<https://finance.naver.com/sise/lastsearch2.naver>

삼성전자 : 70,100 : 10,054,916
TIGER 200 : 39,705 : 640,469
KODEX 배당성장 : 15,930 : 2,968
한국조선해양 : 104,500 : 636,966
SK하이닉스 : 101,500 : 2,842,035
DL이앤씨우 : 77,300 : 5,052
현대공업 : 7,910 : 118,360
ARIRANG 고배당주 : 13,105 : 225,937
TIGER 200 IT : 34,615 : 140,269
카카오 : 128,500 : 1,436,107
현대차 : 213,500 : 602,212
DL이앤씨 : 135,000 : 90,805
TIGER 차이나전기차SOLACTIVE : 19,600 : 5,005,721
KBSTAR 미국나스닥100 : 13,600 : 78,960
세아제강 : 105,500 : 7,246
KINDEX 베트남VN30(합성) : 22,675 : 57,421
더블유게임즈 : 65,900 : 617,458
티에스이 : 64,300 : 50,230
코오롱인더우 : 41,700 : 13,811
일진파워 : 17,650 : 4,223,975
한국금융지주 : 88,000 : 182,728
KODEX 200 : 39,655 : 2,856,621
탐엔지니어링 : 7,300 : 12,524
동양에스텍 : 3,670 : 1,841,258
포스코강판 : 59,000 : 38,865
LG화학 : 825,000 : 146,364
에스씨디 : 1,875 : 188,607
우리기술투자 : 11,650 : 5,416,870
에스티아이 : 16,150 : 61,522
엔씨소프트 : 639,000 : 169,654

순위	종목명	검색비율	현재가	전일비	동락률	거래량	시가	고가	저가	PER	ROE
1	삼성전자	0.39%	70,100	▼ 1,000	-1.41%	10,054,916	71,000	71,000	70,000	14.76	9.99
2	TIGER 200	0.06%	39,705	▼ 340	-0.85%	640,469	39,995	39,995	39,555	N/A	N/A
3	KODEX 배당성장	0.05%	15,930	▼ 115	-0.72%	2,968	16,080	16,080	15,915	N/A	N/A
4	한국조선해양	0.05%	104,500	▲ 500	+0.48%	636,966	105,000	108,500	103,500	-4.77	-7.42
5	SK하이닉스	0.05%	101,500	▼ 500	-0.49%	2,842,035	103,000	104,000	101,500	12.70	9.53
6	DL이앤씨우	0.04%	77,300	▲ 200	+0.26%	5,052	77,500	77,600	76,900	N/A	N/A
7	현대공업	0.04%	7,910	▼ 140	-1.74%	118,360	8,050	8,140	7,910	8.78	10.33
8	ARIRANG 고배당주	0.04%	13,105	▼ 45	-0.34%	225,937	13,150	13,150	13,050	N/A	N/A
9	TIGER 200 IT	0.04%	34,615	▼ 155	-0.45%	140,269	34,705	34,770	34,470	N/A	N/A
10	카카오	0.04%	128,500	▲ 1,000	+0.78%	1,436,107	128,000	129,000	126,500	119.76	2.70
11	현대차	0.04%	213,500	0	0.00%	602,212	214,500	215,000	212,500	15.47	2.04
12	DL이앤씨	0.03%	135,000	▼ 1,500	-1.10%	90,805	134,500	137,500	134,500	N/A	N/A
13	TIGER 차이나전기차SOLACTIVE	0.03%	19,600	▲ 460	+2.40%	5,005,721	19,105	19,750	19,055	N/A	N/A
14	KBSTAR 미국나스닥100	0.03%	13,600	▲ 35	+0.26%	78,960	13,570	13,605	13,510	N/A	N/A
15	세아제강	0.03%	105,500	▼ 1,000	-0.94%	7,246	106,500	106,500	105,000	6.18	5.67
16	KINDEX 베트남VN30(합성)	0.03%	22,675	▲ 405	+1.82%	57,421	22,495	22,720	22,305	N/A	N/A
17	더블유게임즈	0.03%	65,900	▲ 1,200	+1.85%	617,458	64,800	66,600	63,700	9.61	16.94
18	티에스이	0.03%	64,300	▲ 800	+1.26%	50,230	63,600	64,500	62,300	22.06	15.10
19	코오롱인더우	0.03%	41,700	▼ 650	-1.53%	13,811	42,300	42,300	41,300	8.35	N/A
20	일진파워	0.03%	17,650	▲ 150	+0.86%	4,223,975	17,100	18,150	17,000	20.60	10.85
21	한국금융지주	0.03%	88,000	▼ 1,700	-1.90%	182,728	89,300	89,300	87,700	4.18	16.27
22	KODEX 200	0.03%	39,655	▼ 335	-0.84%	2,856,621	39,865	39,930	39,515	N/A	N/A
23	탐엔지니어링	0.03%	7,300	▼ 40	-0.54%	12,524	7,340	7,370	7,290	26.74	-1.92
24	동양에스텍	0.03%	3,670	▲ 145	+4.11%	1,841,258	4,010	4,195	3,630	4.85	9.87
25	포스코강판	0.03%	59,000	▼ 900	-1.50%	38,865	60,000	60,200	58,600	6.93	2.88
26	LG화학	0.03%	825,000	▼ 5,000	-0.60%	146,364	827,000	832,000	811,000	22.22	2.93
27	에스씨디	0.03%	1,875	▼ 25	-1.32%	188,607	1,885	1,915	1,875	9.66	3.68
28	우리기술투자	0.03%	11,650	▼ 150	-1.27%	5,416,870	11,450	11,850	11,300	8.30	30.42
29	에스티아이	0.03%	16,150	▼ 350	-2.12%	61,522	16,500	16,550	16,050	34.96	6.42
30	엔씨소프트	0.03%	639,000	▲ 12,000	+1.91%	169,654	627,000	642,000	624,000	34.30	20.83

크롤링(스크래핑) 활용처

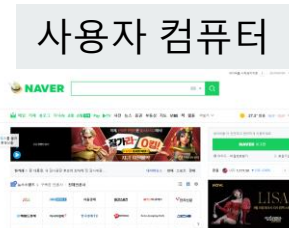
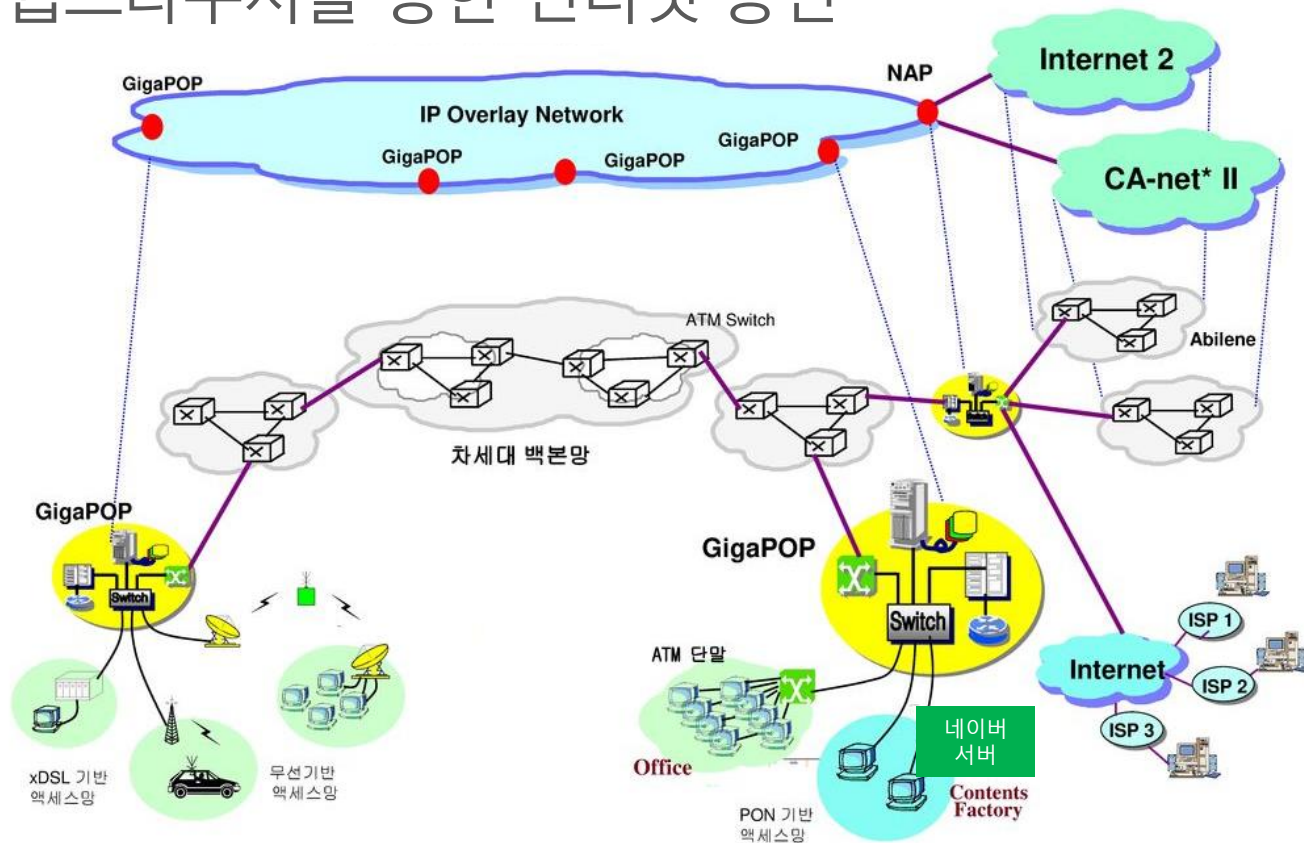
- 사이트에서 원하는 데이터 추출하기
- 업무 자동화하기
- 알림 받기

인터넷과 웹(www)

- 네이버, 구글 같은 사이트
- 그리고 그것을 보는 웹 브라우저(크롬, 에지)
- 네이버에 들어가면 무슨 일이 일어날까?

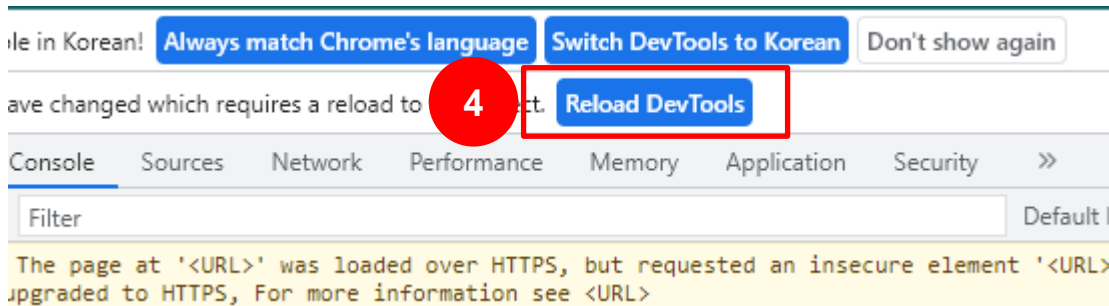
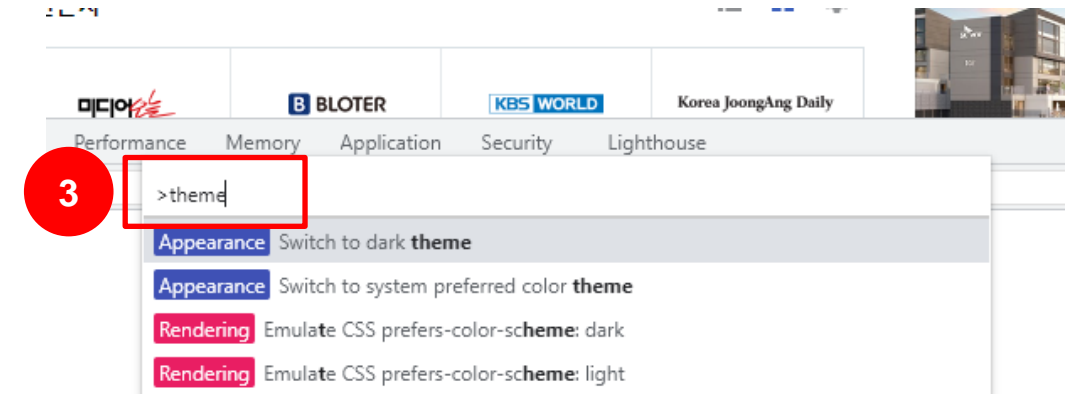
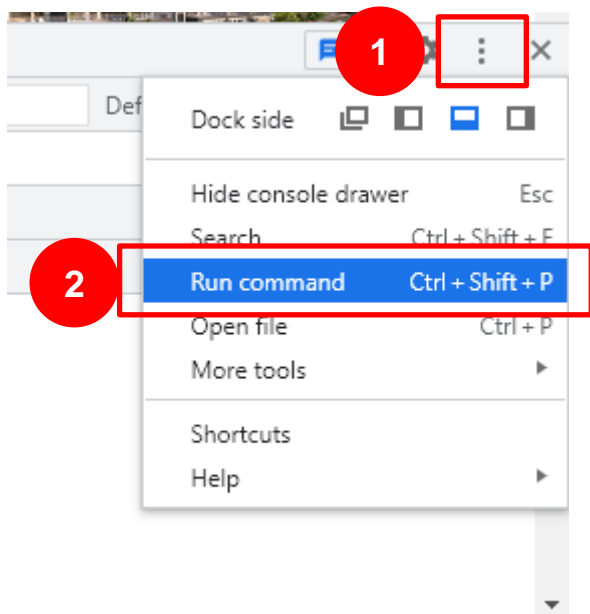
인터넷과 웹(WWW)

- 웹 브라우저를 통한 인터넷 통신



크롬 브라우저 - 개발자 도구(F12) - theme 변경하기

- 개발자 도구 theme 변경하기(어두운 모드로 변경)



크롬 브라우저 개발자 도구(F12)

The screenshot displays the Chrome Developer Tools Network tab. The top navigation bar includes tabs for Elements, Console, Source, Network, Performance, Memory, Application, Security, and Lighthouse. The Network tab is active, showing a list of network requests. The 'All' filter is selected, and the '314.png' request is highlighted. The bottom summary bar shows the following statistics:

- 189 requests
- 1.0 MB transferred
- 4.4 MB resources
- Finish: 49.50 s
- DOMContentLoaded: 2.40 s
- Load: 3.20 s

Name	Status	Type	Initiator	Size	Time	Waterfall
www.naver.com	200	document	Other	47.2 kB	165 ms	
pc.veta.core.min.js	200	script	(index)	(memory cache)	1 ms	
external.5d9e21ac.js?o=www	200	script	(index)	(memory cache)	11 ms	
gfp-core.js	200	script	(index)	(disk cache)	506 ms	
preload.59a14d58.js?o=www	200	script	(index)	(memory cache)	1 ms	
vendors~search.30b5d61b.js?o=www	200	script	(index)	(memory cache)	10 ms	
polyfill_async.942eb7d6.js?o=www	200	script	(index)	(disk cache)	325 ms	
search.6d032745.js?o=www	200	script	(index)	(memory cache)	2 ms	
vendors~more~nmain~sidebar_notice.031bbbfd.js?o=www	200	script	(index)	(memory cache)	0 ms	
more~nmain~sidebar_notice.3987c2c7.js?o=www	200	script	(index)	(memory cache)	0 ms	
nmain.3aead730.js?o=www	200	script	(index)	(memory cache)	0 ms	
055.png	200	png	(index)	(memory cache)	0 ms	
057.png	200	png	(index)	(memory cache)	0 ms	
314.png	200	png	(index)	(memory cache)	1 ms	

크롬 브라우저 개발자 도구(F12)

The screenshot shows the Chrome DevTools interface with the Network tab selected. The left sidebar lists network requests, with 'www.naver.com' highlighted. The main panel shows the 'Headers' tab for this request.

Network Tab:

- Filter: ☐ Hide data URLs **All** Fetch/XHR JS CSS Img Media Font Doc WS Wasm Manifest Other ☐ Has blocked cookies
- Request list:
 - www.naver.com** (highlighted)
 - pc.veta.core.min.js
 - external.5d9e21ac.js?o=www
 - preload.59a14d58.js?o=www
 - vendors~search.30b5d61b.js?
 - search.6d032745.js?o=www
 - vendors~more~nmain~sideb
 - more~nmain~sidebar_notice.
 - nmain.3aead730.js?o=www
 - 055.png
 - 005.png
 - gfp-core.js
 - polyfill_async.942eb7d6.js?o=
 - sp_main_57f073.png
 - sp_weather_time_5f2bbb.png
 - 214.png
 - 117.png
 - 330.png

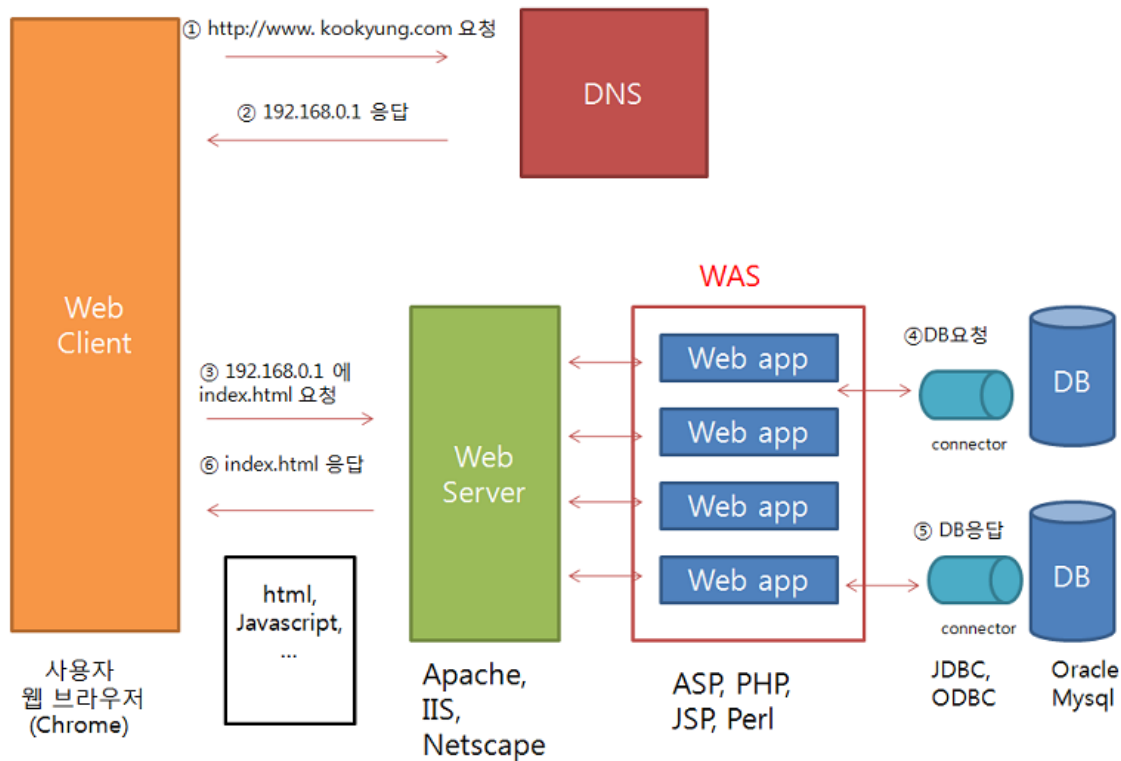
Headers Panel:

- General**
 - Request URL: `https://www.naver.com/`
 - Request Method: `GET`
 - Status Code: 200
 - Remote Address: `23.51.28.215:443`
 - Referrer Policy: `unsafe-url`
- Response Headers**
 - `cache-control: no-cache, no-store, must-revalidate`
 - `content-encoding: gzip`
 - `content-type: text/html; charset=UTF-8`
 - `date: Tue, 14 Sep 2021 07:20:52 GMT`
 - `p3p: CP="CAO DSP CURa ADMa TAIa PSAa OUR LAW STP PHY ONL UNI PUR FIN COM NAV INT DEM STA PRE"`
 - `pragma: no-cache`
 - `referrer-policy: unsafe-url`
 - `server: NWS`
 - `strict-transport-security: max-age=63072000; includeSubdomains`
 - `vary: Accept-Encoding`
 - `x-frame-options: DENY`

181 requests | 278 kB transferred

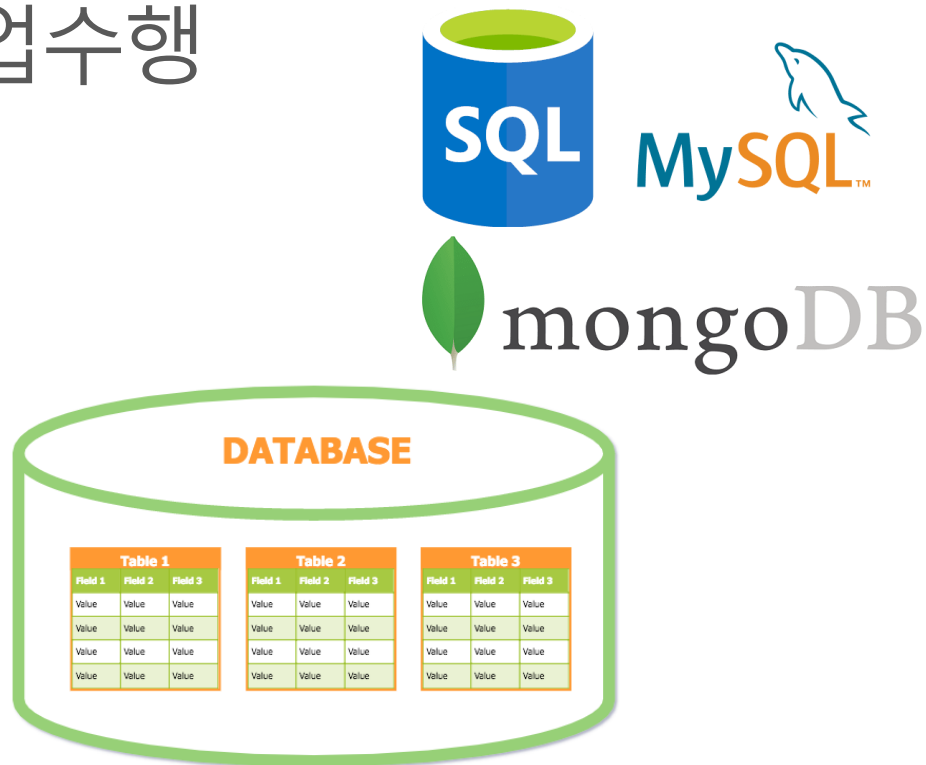
Web Server와 DB

- 웹서버의 역할
- DB의 역할



DB 역할

- Data의 CRUD 작업수행
- Create(Input)
- Read
- Update
- Delete



크롤링 목적

- 크롤링으로 얻으려는 데이터는?
- DB의 데이터가 아닌, DB에서 커리 된 사용자 경험의 데이터를 얻으려는 것.
- PC 버전, mobile 버전에 접속해서 크롤링함.

http 통신 프로토콜

- http(hypertext transfer protocol)
- 인터넷에서 hypertext를 전송하기 위해 사용되는 통신 규약
- 인터넷에서 웹 서버와 사용자 컴퓨터에 설치된 웹 브라우저 사이에 문서를 전송하기 위한 통신 규약
- 기본 port : 80
- 보안 취약점을 개선 : https 프로토콜

http, https

- 통신 프로토콜(규약)
 - http:// : Hypertext Transfer Protocol
 - https:// : Hypertext Transfer Protocol Secure

Request Headers

:authority: www.naver.com

:method: GET

:path: /

:scheme: https

Accept:

text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image/webp,image/apng,*/*;q=0.8,application/signed-exchange;v=b3;q=0.9

Accept-encoding: gzip, deflate, br

Accept-language: ko,ko-KR;q=0.9,en-US;q=0.8,en;q=0.7

Response Headers

cache-control: no-cache, no-store, must-revalidate

content-encoding: gzip

content-type: text/html; charset=UTF-8

date: Tue, 14 Sep 2021 07:20:52 GMT

p3p: CP="CAO DSP CURa ADMa TAIA PSAa OUR LAW STP PHY ONL UNI PUR FIN COM NAV INT DEM STA PRE"

pragma: no-cache

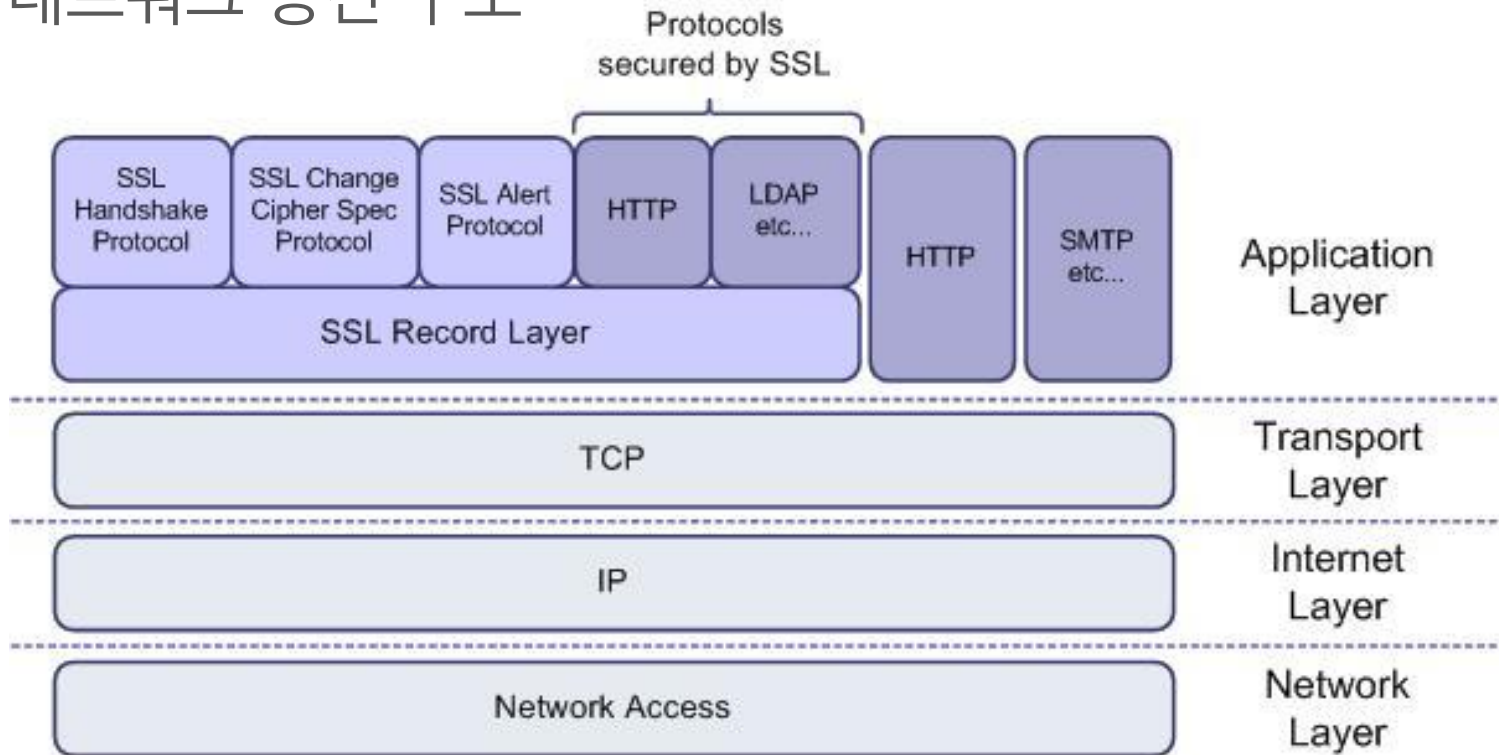
referrer-policy: unsafe-url

server: NWS

strict-transport-security: max-age=63072000; includeSubdomains

http, https

- 네트워크 통신 구조



http vs https



Helen

HTTP

http://www.example.com

password: abc123



Without password encryption

Hacker see "abc123"



Carol

HTTPS

https://www.example.com

password: abc123



With password encryption

Hacker see "xyaerXzabc"

HTTP Example login

- GET /money HTTP/1.1
 - Web 에러 code 의미

HTTP/1.1 400 Unauthorized

100~199 정보전달용

200~299 성공

300~399 리다이렉션

400~499 클라이언트 에러

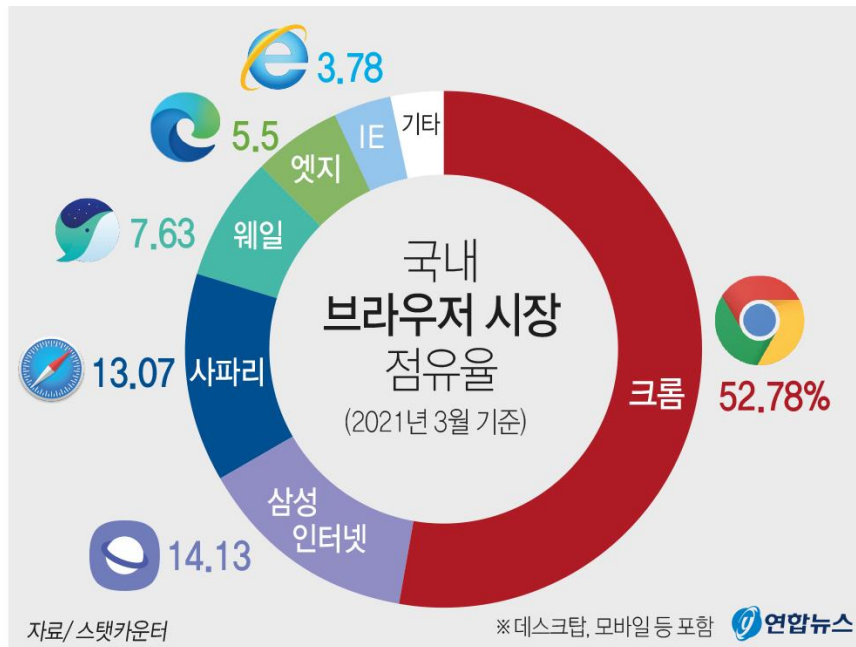
500~599 서버 에러

에러코드 확인하기(URL)

[http.cat/401](http://cat/401)

브라우저의 역할

- UI Rendering
 - HTML 파일로 변경
- 편의성
 - 북마킹, 저장
- 보안
- 스크립트 실행
 - 동적기능



장예진 기자 / 20210427 트위터 @yonhap_graphics 페이스북 tuneey.kr/LeYN1

크롤링 주의 사항(네티켓)

- 크롤링은 불법행위?
- 사람인 vs. 잡코리아
 - 기사 : <http://news.bizwatch.co.kr/article/mobile/2017/09/27/0023>
- 여기어때 vs. 야놀자
 - 기사 : <http://www.zdnet.co.kr/view/?no=20200211153634>

Robots.txt

- 공식 페이지 <https://www.robotstxt.org/robotstxt.html>
- robots.txt : 크롤링 관련된 규칙 설명
- 대부분의 커뮤니티의 콘텐츠 소유는 각자 올린 사람의 소유

크롤링 가능한지 확인하는 방법 :

<https://www.daum.net/robots.txt>

<https://www.naver.com/robots.txt>

<https://www.google.com/robots.txt>

<https://www.danawa.com/robots.txt>

<https://www.coupang.com/robots.txt>

정적 크롤링 vs. 동적 크롤링

	정적 크롤링	동적 크롤링
크롤링 속도	빠르고 간단함	느리고 복잡함
개발 편의성	처음엔 쉽지만 고도화 어려움	처음엔 손이 많이 가지만 나중엔 편리함
디버깅 편의성	테스트 쉬움	테스트 어려움
오류 취약점	상대적으로 낮음	상대적으로 높음
대표적인 도구	Beautifulshoup4	selenium

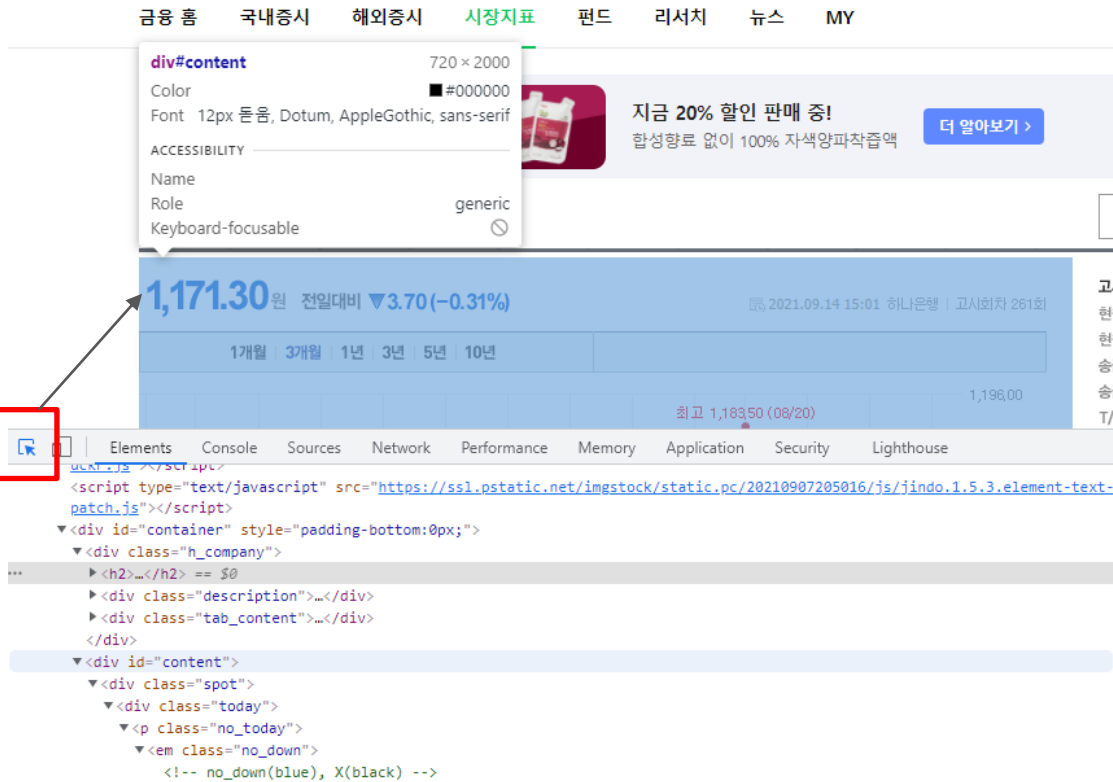
—

정적 크롤링

크롬 개발자 도구

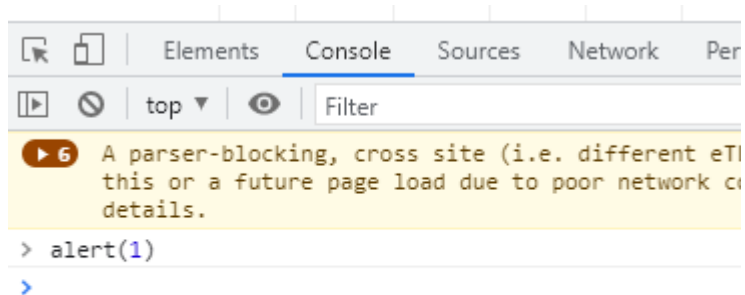
• 크롬 브라우저 개발자 도구 활용하기

클릭하고 뷰포트의
보고싶은 부분을
클릭하면 바로
이동



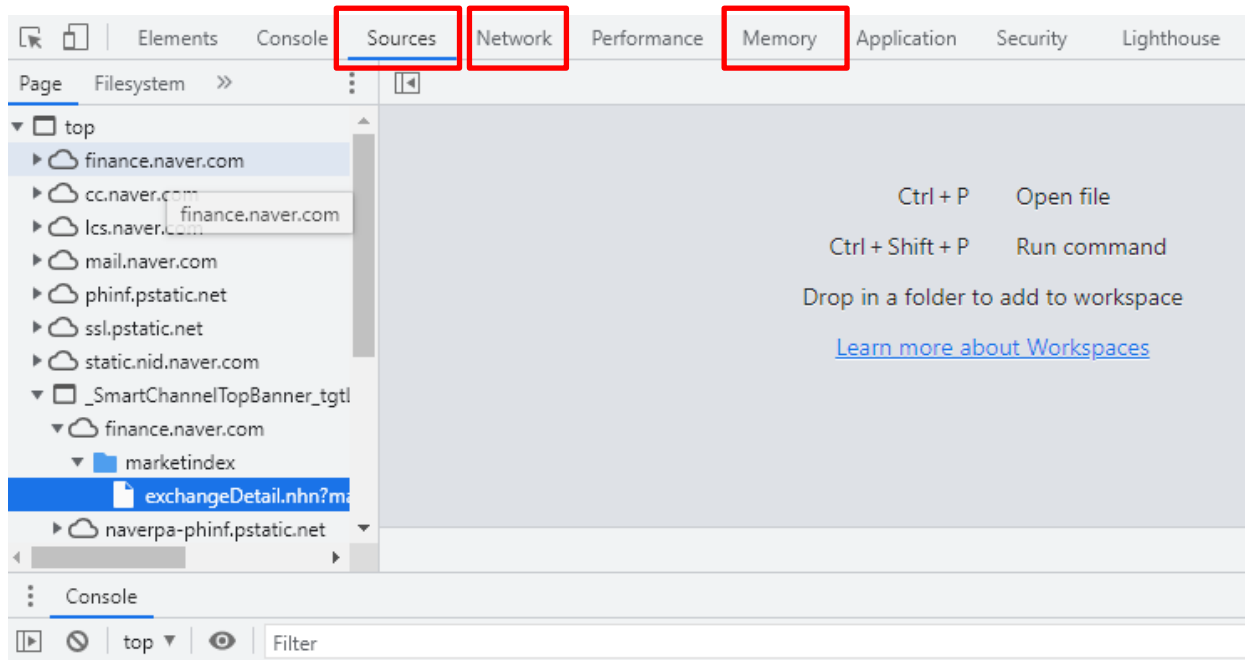
크롬 개발자 도구

- console



크롬 개발자 도구

- Sources, Network, Memory



Python request 설치

- `python -m pip install requests`

<https://docs.python-requests.org/en/latest/user/quickstart/>

Requests 주요 함수

- `get()` : 요청, 값을 가져옴
- `post()` : 생성, 액션
- `put()` : 수정, 덮어쓰기
- `delete()` : 삭제
- `text`
- `json()`

분석할 데이터 구조 타입

XML 타입

- XML(Extensible Markup Language)

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<CAT> <!-- node -->
```

```
  <NAME>Izzy</NAME>
```

```
  <BREED>Siamese</BREED>
```

```
  <AGE>6</AGE>
```

```
  <ALTERED>yes</ALTERED>
```

```
  <DECLAWED>no</DECLAWED>
```

```
  <LICENSE>Izz138bod</LICENSE>
```

```
  <OWNER>Colin Wilcox</OWNER>
```

```
</CAT>
```

HTML 타입

- 태그가 정해져 있음.

```
<!DOCTYPE html>
<html lang="ko">
  <head>...</head>
  <body class="finance"> == $0
    <div id="wrap">
      <div id="kakaowrap">
        <div id="wrapMinidaum">
          <div id="minidaum">
            <strong class="minidaum_logo">
              <a href="http://www.daum.net/?nil_profile=mini&nil_src=daum" target="_top" class="img_logo minidaum_tiara_click"
                um logo" data-tiara-action-name="미니다음_로고">Daum</a>
            </strong>
            <div class="minidaum_gnb">
              <h2 class="screen_out">서비스 바로가기</h2>
              <div id="minidaumUser">...</div>
              <div id="minidaumService">...</div>
            </div>
          </div>
        </div>
      </div>
    </div>
    <script>
```

JSON 타입



로그인

금융 부동산

종목명, 종목코드를 입력해주세요

홈 국내 해외 뉴스 투자정보 토론 MY 환율



코스피 지수 KOSPI

3,148.83

코스닥 지수 KOSDAQ

1,037.74

오늘

09.1

Elements Console Sources Network Performance Memory Application Security Lighthouse

Filter ☐ Hide data URLs ☐ Disable cache ☐ Disable throttling ☐ Fetch/XHR ☐ CSS ☐ Img ☐ Media ☐ Font ☐ Doc ☐ WS ☐ Wasm ☐ Manifest ☐ Other ☐ Has blocked cookies ☐ Blocked Requests ☐

5000 ms 10000 ms 15000 ms 20000 ms 25000 ms 30000 ms 35000 ms 40000 ms 45000 ms 50000 ms 55000 ms 60000 ms

Name ☐ news_top ☐ ranks?limit=10 ☐ today?type=DOMESTIC ☐ ?pagination=true&perPage=. ☐ business_dates?exchangeCou ☐ sectors?market=KOSPI&chan ☐ sectors?market=KOSDAQ&ch ☐ recents ☐ banner?id=00872&containeri

Headers Preview Response Initiator Timing Cookies

```
{data: [...]}
data: [...]
```

▶ 0: {rank: 1, rankChange: 0, symbolCode: "A035720", code: "KR7035720002", name: "카카오", tradePrice: 124000,...}

▶ 1: {rank: 2, rankChange: 0, symbolCode: "A005930", code: "KR7005930003", name: "삼성전자", tradePrice: 76600,...}

▶ 2: {rank: 3, rankChange: 0, symbolCode: "A323410", code: "KR7323410001", name: "카카오뱅크", tradePrice: 69700,...}

▶ 3: {rank: 4, rankChange: 1, symbolCode: "A112040", code: "KR7112040001", name: "위메이드", tradePrice: 74400,...}

▶ 4: {rank: 5, rankChange: -1, symbolCode: "A247540", code: "KR7247540008", name: "에코프로비엠",...}

▶ 5: {rank: 6, rankChange: 2, symbolCode: "A302440", code: "KR7302440003", name: "SK바이오사이언스",...}

▶ 6: {rank: 7, rankChange: -1, symbolCode: "A068270", code: "KR7068270008", name: "셀트리온",...}

▶ 7: {rank: 8, rankChange: 1, symbolCode: "A035420", code: "KR7035420009", name: "NAVER",...}

▶ 8: {rank: 9, rankChange: 0, symbolCode: "A352910", code: "KR7352910004", name: "오비고", tradePrice: 19800,...}

▶ 9: {rank: 10, rankChange: -3, symbolCode: "A028300", code: "KR7028300002", name: "에이치엘비",...}

11 / 62 requests 7.8 kB / 161 kB

JSON 타입

▼ {data: Array(10)} ⓘ

▼ data: Array(10)

- ▶ 0: {rank: 1, rankChange: 0, symbolCode: 'A035720', code: 'KR7035720002', name: '카카오', ...}
- ▶ 1: {rank: 2, rankChange: 0, symbolCode: 'A005930', code: 'KR7005930003', name: '삼성전자', ...}
- ▶ 2: {rank: 3, rankChange: 0, symbolCode: 'A323410', code: 'KR7323410001', name: '카카오뱅크', ...}
- ▶ 3: {rank: 4, rankChange: 1, symbolCode: 'A112040', code: 'KR7112040001', name: '위메이드', ...}
- ▶ 4: {rank: 5, rankChange: -1, symbolCode: 'A247540', code: 'KR7247540008', name: '에코프로비엠', ...}
- ▶ 5: {rank: 6, rankChange: 2, symbolCode: 'A302440', code: 'KR7302440003', name: 'SK바이오사이언스', ...}
- ▶ 6: {rank: 7, rankChange: -1, symbolCode: 'A068270', code: 'KR7068270008', name: '셀트리온', ...}
- ▶ 7: {rank: 8, rankChange: 1, symbolCode: 'A035420', code: 'KR7035420009', name: 'NAVER', ...}
- ▶ 8: {rank: 9, rankChange: 0, symbolCode: 'A352910', code: 'KR7352910004', name: '오비고', ...}
- ▶ 9: {rank: 10, rankChange: -3, symbolCode: 'A028300', code: 'KR7028300002', name: '에이치엘비', ...}

length: 10

▶ [[Prototype]]: Array(0)

▶ [[Prototype]]: Object

정적 크롤링 실습

프로젝트 진행하며 배울 내용

- requests 활용한 기초적인 파싱(http 파싱)
- find(), split() 등을 활용한 기초적인 문자열 파싱
- 정규식(regex)을 활용한 패턴 검색
- 쿼리스트링에 대한 이해
- beautifulsoup을 활용한 편리한 html 파싱
- css selector를 활용한 손쉬운 파싱

파이썬 패키지 설치

(VENV) ~\$ pip install requests

프로젝트 1 : 환율 계산기

환 율 계 산 기

미 국 USD : 1,115.00
일 본 JPY(100엔) : 1,014.33
유 럽 연 합 EUR : 1,357.40
중 국 CNY : 175.22
달 러 /일 본 엔 : 109.9800
유 로 /달 러 : 1.2179
영 국 파 운 드 /달 러 : 1.4175
달 러 인 텍 스 : 89.9900
WTI : 66.32
휘 발 유 : 1551.03
국 제 금 : 1902.5
국 내 금 : 67837.5

달 러 로 바 꾸 길 원 하 는 금 액 (원)을 입 력 해 주 세 요 : 100000
89 달 러 환 전 되 었 습 니 다

문자열 조작 파싱

- HTML문서를 일반 문자열로 취급하여
- 원하는 문자열의 위치 등을 이용해
- 원하는 데이터를 추출하는 방법

문자열 조작 파싱

- `find()` : 문자열의 위치와 길이를 기반으로 작업
- `split()` : 문자열을 배열로 나눠서 작업

네이버 금융 환율 공시 정보 가져오기(exchange_rate.py)

```
import requests as req

res = req.get('https://finance.naver.com/marketindex/?tabSel=exchange#tab_section')
html = res.text
#pos = html.find('미국 USD') # 데이터가 있는지 찾음.
#print(pos)

s = html.split('<span class="value">')[1].split('</span>')[0]
print(s)
```


정규식(Regular Expression)

- regex, regexp
- 패턴 검색 기법
- seriali[sz]e : serialize, serialise
- color?r : color, colour

정규식 표현 실습 : <https://regexr.com/>

정규식(Regular Expression) 기호

기호	의미
()	캡처
[]	이 중 아무거나
.	그 위치에 아무거나
*	바로 이전 것이 0 개 이상
+	바로 이전 것이 1 개 이상
?	바로 이전 것이 없을 수도 있음
\	\ 다음의 특수 기호 무효화

정규식(Regular Expression) 메소드

import re

관련 메소드

- match()
- search()
- findall()
- compiler()

```
import re
s= 'hay'
print(re.match(r'..', s))
print(re.match(r'hay1*', s))
s= 'color'
print(re.match(r'colou?r', s))
s= 'how are you?'
print(re.match(r'how are you\W?', s))
```

```
#A, B, C, D
s= '이 고기는 A등급 입니다.'
print(re.match(r'이 고기는 [ABCD]등급 입니다\W.', s))
```

```
s= '이 고기는 A등급 입니다.'
print(s.split('이 고기는 ')[1].split('등급 입니다.')[0])
p = re.findall(r'이 (..)는 (.)등급 입니다\W.', s)
print(p)
p = re.findall(r'이 (..)는 ([ABCD])등급 입니다\W.', s)
print(p)
```

정규식(Regular Expression)

▼ `<h3 class="h_1st">`

`미국 USD`

`</h3>`

▼ `<div class="head_info point_dn">`

`1,171.50 == $0`

▶ `...`

` 3.50`

`하락`

Html 데이터의 패턴 찾기

```
▼<h3 class="h_1st">
  <span class="blind">미국 USD</span>
</h3>
▼<div class="head_info point_dn">
  <span class="value">1,171.50</span> == $0
  ▶<span class="txt_krw">...</span>
  <span class="change"> 3.50</span>
  <span class="blind">하락</span>

▼<div class="data">
  ▼<ul class="data_1st" id="worldExchangeList">
    ▼<li class>
      ▼<a href="/marketindex/worldExchangeDetail.nhn?marketindexCd=FX_USDJPY" class="head_jpy_usd"
        2.jpyut', '', '', event);">
        ▼<h3 class="h_1st">
          <span class="blind">달러/일본 엔</span> == $0
          </h3>
          ▼<div class="head_info point_dn">
            <span class="value">109.6500</span>
            ▼<span class="txt_jpy">
              <span class="blind">엔</span>
              </span>
```

환율계산기 만들기(regex_exchange_rate.py)

```
import requests as req
import re

url = 'https://finance.naver.com/marketindex/?tabSel=exchange#tab_section'
res = req.get(url)
body = res.text
r = re.compile(r'h_lst.*?blindW">(.*?)</span>.*?valueW">(.*?)</', re.DOTALL)
captuers = r.findall(body)
#print(captuers)

print("-----")
print("환율 계산기")
print("-----")
print("")
```

환율계산기 만들기(regex_exchange_rate.py)

```
for c in captuers:
```

```
    print(c[0] + " : " + c[1])
```

```
print("")
```

```
#usd =captuers[0][1].replace(",","")
```

```
usd = float(captuers[0][1].replace(",",""))
```

```
won = input("달러로 바꾸길 원하는 금액(원)을 입력해 주세요. : ")
```

```
won = int(won)
```

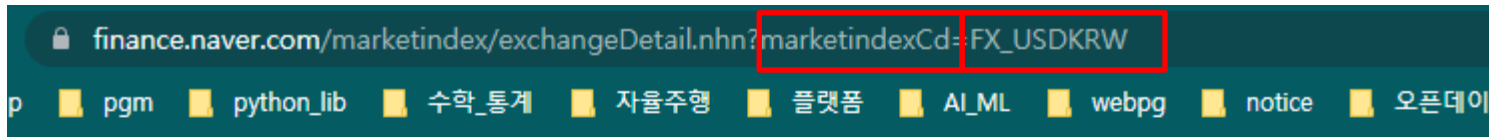
```
dollar = won / usd
```

```
dollar = int(dollar)
```

```
print(f"{dollar} 달러 환전되었습니다.")
```

쿼리 스트링(Query string)

- 웹요청시 보내는 추가 인자 값



미국 USDKRW [환율>](#)

1,172.90 원 전일대비 ▲1.4

1개월 | 3개월 | 1년 | 3년



finance.naver.com/marketindex/worldExchangeDetail.nhn?marketindexCd=FX_EURUSD

pgm python_lib 수학_통계 자율주행 플랫폼 AI_ML webpg notice

유로/달러 EURUSD [환율>](#)

1유로 = 1.1822 달러 전일대비 ▲0.0012 (+0.10%)

2021.09.14 모

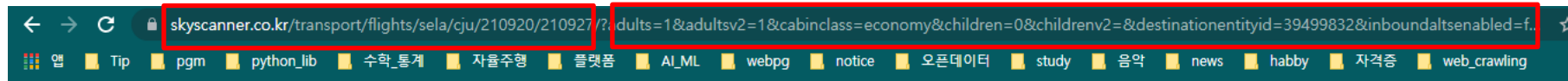
1개월 | 3개월 | 1년 | 3년 | 5년 | 10년

최고 1,2126 (06/15)




쿼리 스트링(Query string)

- 인원수, 날짜, 목적지 등 다양한 정보로 쿼리



도움말

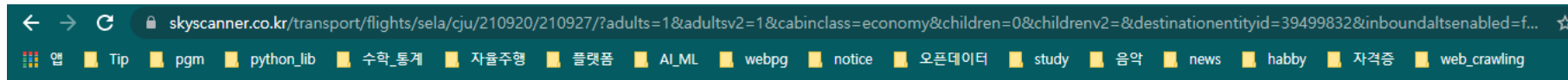
한국어  대한민국 ₩ KRW

✈ 항공권

🏨 호텔

🚗 렌터카

쿼리스트링 형식



Skyscanner

도움말

한국어 대한민국 ₩ KRW

→ 항공권

호

? : url과 query 구분

url?querystring

값이름=값

Adults=2

Children=0

& : query 항목들 구분

Adults=2&children=0

배열은

arr[]=1

arr[]=2

arr[]=3

-> arr[1,2,3]

쿼리스트링

- Query에 검색어의 경우

search.daum.net/search?w=tot&DA=YZR&t_nil_searchbox=btn&sug=&sugo=&sq=&o=&q=삼성전자



삼성전자



search.naver.com/search.naver?where=nexearch&sm=top_h ty&fbm=1&ie=utf8&query=삼성전자



삼성전자



통합

VIEW

이미지

지식IN

동영상

쇼핑

뉴스

어학사전

지도

책



www.samsung.com

삼성전자

쿼리스트링

- 주소창에서는 감자로 보이지만 주소를 직제로 복사해보면

https://search.daum.net/search?w=tot&DA=YZR&t__nil_searchbox=btn&sug=&sugo=&sq=&o=&q=%EC%82%BC%EC%84%B1%EC%A0%84%EC%9E%90

https://search.daum.net/search?w=tot&DA=YZR&t__nil_searchbox=btn&sug=&sugo=&sq=&o=&q=삼성전자

정적 크롤링
Beautifulsoup

Beautifulsoup

- requests : http 파싱을 편하게
- BeautifulSoup : html 파싱을 편하게

- 설치

(VENV) ~\$ pip install BeautifulSoup4

Beautifulsoup4 공식 문서

<https://www.crummy.com/software/BeautifulSoup/bs4/doc/>

Beautifulsoup 주요기능

주요기능	설명
html 문자열 파싱 가능	-
html 요소 인식 및 편리한 기능들	-
parent, children, contents, descendants, sibling	DOM 객체 트레이싱의 편리함
string, strings, stripped_strings, get_text()	요소인식 함수 제공
prettify	html의 중첩구조를 시각적으로 보여줌
html attribute	요소의 속성 접근이 가능함.

BeautifulSoup 실습1

- 네이버 환율 계산 하기
- 달러 환전하기

```
-----  
환 율 계 산 기  
-----  
  
미 국 USD : 1,115.00  
일 본 JPY(100엔) : 1,014.33  
유 럽 연 합 EUR : 1,357.40  
중 국 CNY : 175.22  
달 러 /일 본 엔 : 109.9800  
유 로 /달 러 : 1.2179  
영 국 파 운 드 /달 러 : 1.4175  
달 러 인 텍 스 : 89.9900  
WTI : 66.32  
휘 발 유 : 1551.03  
국 제 금 : 1902.5  
국 내 금 : 67837.5  
  
달 러 로 바 꾸 길 원 하 는 금 액 (원 )을 입 력 해 주 세 요 : 100000  
89 달 러 환 전 되 었 습 니 다
```


BeautifulSoup 실습1

- html 구조 확인

iframe#frame_ex1 720 x 1378

2021.09.15 11:19 허나은행 기준 | 고시회차 103회

통화명	매매기준율	환율		송금		미화환산율
		사실 때	파실 때	보내실 때	받으실 때	
미국 USD	1,172.50	1,193.01	1,151.99	1,183.90	1,161.10	1.000
유럽연합 EUR	1,384.49	1,412.04	1,356.94	1,398.33	1,370.65	1.181
일본 JPY (100엔)	1,069.56	1,088.27	1,050.85	1,080.04	1,059.08	0.912
중국 CNY	182.05	191.15	172.95	183.87	180.23	0.155
홍콩 HKD	150.75	153.71	147.79	152.25	149.25	0.129
대만 TWD	42.33	47.87	39.37	0.00	0.00	0.036

Elements Console Sources Network Performance Memory Application Security Lighthouse

```
<div class="section_exchange">
  <div id="section_ex1">
    <h3 id="exchange1" class="blind">환전 고시 환율</h3>
    <div class="exchange_info">
      <iframe id="frame_ex1" title="환전 고시 환율" src="/marketindex/exchangelist.nhn" width=100% height="1378px" frameborder="0" scrolling="no"
        marginwidth="0"> == $0
    </div>
  </div>
  <div class="tbl_area">
    <table border="1" class="tbl_exchange" summary="환전 고시 환율 리스트">
      <caption>환전 고시 환율</caption>
      <colgroup>
        <thead>
          <tr>
            <td class="tit">
              <a href="/marketindex/exchangeDetail.nhn?marketindexCd=FX_USDKRW" target="_parent" onclick="parent.clickcr(this, 'ex1.exlist
                nt);"> 미국 USD </a>
            </td>
            <td class="sale">1,172.50</td>
            <td>1,193.01</td>
```

BeautifulSoup 실습

- 네이버 환율 계산 하기
- 금액, 환전국가 선택 => 결과 반환하기

```
-----  
환율 계산기  
-----
```

```
미국 USD : 1,115.00  
일본 JPY(100엔) : 1,014.33  
유럽연합 EUR : 1,357.40  
중국 CNY : 175.22  
달러/일본 엔 : 109.9800  
유로/달러 : 1.2179  
영국 파운드/달러 : 1.4175  
달러인덱스 : 89.9900  
WTI : 66.32  
휘발유 : 1551.03  
국제 금 : 1902.5  
국내 금 : 67837.5
```

```
달러로 바꾸길 원하는 금액(원)을 입력해주세요 : 100000  
89 달러 환전되었습니다
```

다양한 CSS Selector 살펴보기

CSS 셀렉터

- CSS 셀렉터를 활용하는 방법

마우스를 위에
놓기만 함.

통화명 >	매매기준율	사실 때	파실 때
미국 USD	1,172.50	1,193.01	1,151.99
유럽연합 EUR			1,356.94
일본 JPY (100엔)			1,050.85
중국 CNY			172.95
홍콩 HKD			147.79
대만 TWD			39.37
영국 GBP			1,586.18
오만 OMR			2,862.80
캐나다 CAD			905.62
스위스 CHF			

td.sale 92.95 × 30

Color ■ #444444

Font 12px tahoma

Background □ #F5FBF6

Padding 0px 12px 0px 0px

ACCESSIBILITY

Contrast Aa 9.28 ✓

Name 1,172.50

Role gridcell

Keyboard-focusable ⓧ

Elements Console Sources Network Performance Memory Application Security Lighthouse

```
nt);"> 미국 USD </a>
</td>
<td class="sale">1,172.50</td> == $0
<td>1,193.01</td>
<td>1,151.99</td>
<td>1,183.90</td>
<td>1,161.10</td>
<td>1.000</td>
</tr>
```

CSS 셀렉터

- CSS selector가 크롤링, 파싱에 유리한 이유
 - 웹을 만들 때 CSS selector를 직접 활용해서 이름을 붙임
 - Human readable, 사람이 읽기 좋고, 이해하기 편한 이름

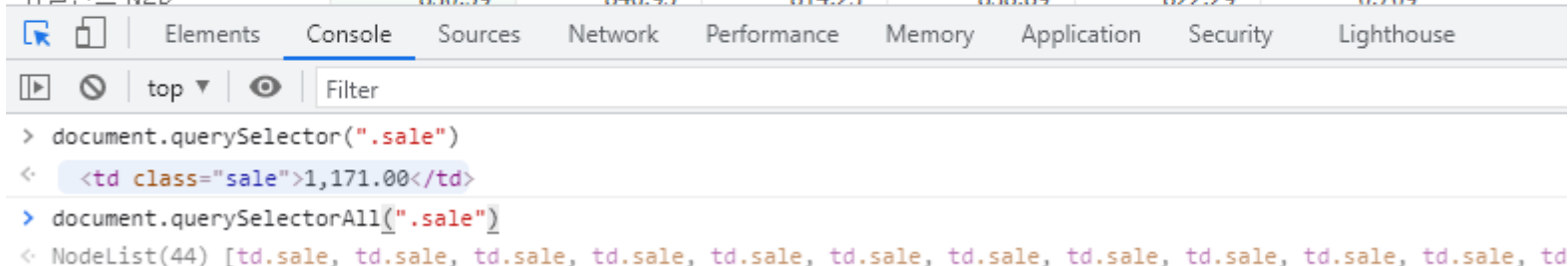
CSS 셀렉터

- css selector – HTML element type 기반
 - HTML element Type을 구분해 부르는 방식
 - p, h, div, span, 등
 - 크롤링 할 때 매우 유용함.

CSS 셀렉터

- 크롬 개발자 도구
- document.querySelector()
- document.querySelectorAll()

스위스 CHF	1,273.66	1,298.75	1,248.57	1,286.39	1,260.93	1.088
스웨덴 SEK	135.97	139.30	132.64	137.32	134.62	0.116
호주 AUD	857.41	874.30	840.52	865.98	848.84	0.732
뉴질랜드 NZD	830.59	846.95	814.23	838.89	822.29	0.709



CSS 셀렉터

- <https://finance.naver.com//marketindex/exchangeList.nhn>

← → ↺ 🔒 finance.naver.com//marketindex/exchangeList.nhn

통화명 ▾	매매기준율	현할		송금		미화환산율
		사실 때	파실 때	보내실 때	받으실 때	
미국 USD	1,171.00	1,191.49	1,150.51	1,182.40	1,159.60	1.000
유럽연합 EUR	1,382.83	1,410.34	1,355.32	1,396.65	1,369.01	1.181
일본 JPY (100엔)	1,068.33	1,087.02	1,049.64	1,078.79	1,057.87	0.912
중국 CNY	181.87	190.96	172.78	183.68	180.06	0.155
홍콩 HKD	150.54	153.50	147.58	152.04	149.04	0.129
대만 TWD	42.29	47.82	39.33	0.00	0.00	0.036
영국 GBP	1,616.74	1,648.58	1,584.90	1,632.90	1,600.58	1.381
오만 OMR	3,041.64	3,312.34	2,859.15	0.00	0.00	2.598
캐나다 CAD	923.06	941.24	904.88	932.29	913.83	0.788
스위스 CHF	1,273.66	1,298.75	1,248.57	1,286.39	1,260.93	1.088
스웨덴 SEK	135.97	139.30	132.64	137.32	134.62	0.116
호주 AUD	857.41	874.30	840.52	865.98	848.84	0.732
뉴질랜드 NZD	830.59	846.95	814.23	838.89	822.29	0.709

Elements

Console

Sources

Network

Performance

Memory

Application

Security

Lighthouse

```
<html lang="ko">
  <head>...</head>
  ...<body> == $0
    <div class="tbl_area">...</div>
    <script type="text/javascript" language="javascript">...</script>
    <script type="text/javascript"> </script>
  </body>
</html>
```


CSS 셀렉터

← → ↻ 🔒 finance.naver.com//marketindex/exchangeList.nhn

통화명	매매기준율	현찰		송금		미화환산율
		사실 때	파실 때	보내실 때	받으실 때	
미국 USD	1,171.00	1,191.49	1,150.51	1,182.40	1,159.60	1.000
유럽연합 EUR	1,382.83	1,410.34	1,355.32	1,396.65	1,369.01	1.181
일본 JPY (100엔)	1,068.33	1,087.02	1,049.64	1,078.79	1,057.87	0.912
중국 CNY	181.87	190.96	172.78	183.68	180.06	0.155
홍콩 HKD	150.54	153.50	147.58	152.04	149.04	0.129
대만 TWD	42.29	47.82	39.33	0.00	0.00	0.036
영국 GBP	1,616.74	1,648.58	1,584.90	1,632.90	1,600.58	1.381
오만 OMR	3,041.64	3,312.34	2,859.15	0.00	0.00	2.598
캐나다 CAD	923.06	941.24	904.88	932.29	913.83	0.788
스위스 CHF	1,273.66	1,298.75	1,248.57	1,286.39	1,260.93	1.088
스웨덴 SEK	135.97	139.30	132.64	137.32	134.62	0.116
호주 AUD	857.41	874.30	840.52	865.98	848.84	0.732
뉴질랜드 NZD	830.59	846.95	814.23	838.89	822.29	0.709

45 × 14

Elements Console Sources Network Performance Memory Application Security Lighthouse

top Filter

```
> document.querySelector("a")
```

```
< a href="#" onclick="javascript:changeOrder('exchange'); return false;">...</a>
```

CSS 셀렉터 - id

- CSS selector – id : #id_name
 - HTML 한 페이지 내에 유일한 속성값을 가지는 것
 - HTML 상에서 Element의 유일성을 보장하기 위함

Selector 지정

국내증시

주요시세정보

코스피 | 코스닥 | 선물
코스피200 | 코넥스

시가총액 | 배당
업종 | 테마 | 그룹사
ETF | ETN
상승 | 보합 | 하락

코스피

3,161.49 ▲12.66 +0.40%

거래량(천주) 447,945 거래대금(백만) 8,571,645

장중최고 3,163.23 장중최저 3,138.80

60초최고 3,166.08 60초최저 3,166.03

Elements Console Sources Network Performance Memory Application Security Lighthouse

```
<!-- // global include -->
<script type="text/javascript" src="https://ssl.pstatic.net/imgstock/static.pc/20210907205016/js/thirdparty/html2canvas.min.js"></script>
<script language="javascript" src="https://ssl.pstatic.net/imgstock/static.pc/20210907205016/js/recent_code.js"></script>
<div id="newarea">
  <!-- leftmenu -->
  <div class="snb_snb2"></div>
  <!-- //leftmenu -->
  <div id="contentarea">
    <!-- 콘텐츠 -->
    <div id="contentarea_left">
      <!-- 로케이터 -->
```

선언부

CSS 셀렉터 - class

- CSS selector – class
 - 여러 HTML Element에 동일한 스타일 적용시 사용
 - HTML Element의 그룹을 만드는 의미



[부트스트랩 예시]

```
<!-- Standard button -->
<button type="button" class="btn btn-default">Default</button>

<!-- Provides extra visual weight and identifies the primary action in a set of buttons -->
<button type="button" class="btn btn-primary">Primary</button>

<!-- Indicates a successful or positive action -->
<button type="button" class="btn btn-success">Success</button>

<!-- Contextual button for informational alert messages -->
<button type="button" class="btn btn-info">Info</button>

<!-- Indicates caution should be taken with this action -->
<button type="button" class="btn btn-warning">Warning</button>

<!-- Indicates a dangerous or potentially negative action -->
<button type="button" class="btn btn-danger">Danger</button>

<!-- Deemphasize a button by making it look like a link while maintaining button behavior -->
<button type="button" class="btn btn-link">Link</button>
```

CSS 셀렉터 - class

- selector : .class_name
- 클래스 순서 중요하지 않음

Selector 지정



선언부

```
<div class="thumb_box _NM_NEWSSTAND_THUMB _NM_NEWSSTAND_THUMB_press_valid" data-pid="327">...</div>
<div class="thumb_box _NM_NEWSSTAND_THUMB _NM_NEWSSTAND_THUMB_press_valid" data-pid="031">
  ...</img>
</div>
<div class="popup_wrap">...</div>
</div>
<div class="thumb_box _NM_NEWSSTAND_THUMB _NM_NEWSSTAND_THUMB_press_valid" data-pid="930">...</div>
<div class="thumb_box _NM_NEWSSTAND_THUMB _NM_NEWSSTAND_THUMB_press_valid" data-pid="079">...</div>
<div class="thumb_box _NM_NEWSSTAND_THUMB _NM_NEWSSTAND_THUMB_press_valid" data-pid="308">...</div>
```

CSS 셀렉터 - attribute

- CSS selector – attribute 기반
- HTML Element는 다양하고 고유한, 속성을 가짐
attribute 이름 = 속성 값
- id, class도 attribute 중 하나임



CSS 셀렉터 - attribute

- html element attributes 레퍼런스 참조
- https://www.w3schools.com/tags/ref_attributes.asp

HTML Reference

- HTML by Alphabet
- HTML by Category
- HTML Browser Support
- HTML Attributes**
- HTML Global Attributes
- HTML Events
- HTML Colors
- HTML Canvas
- HTML Audio/Video
- HTML Character Sets
- HTML Doctypes
- HTML URL Encode
- HTML Language Codes
- HTML Country Codes
- HTTP Messages
- HTTP Methods
- PX to EM Converter
- Keyboard Shortcuts

HTML Attribute Reference

[< Previous](#) [Next >](#)

HTML Attribute Reference

The table below lists all HTML attributes and what elements they can be used within:

Attribute	Belongs to	Description
accept	<input>	Specifies the types of files that the server accepts (only for type="file")
accept-charset	<form>	Specifies the character encodings that are to be used for the form submission
accesskey	Global Attributes	Specifies a shortcut key to activate/focus an element
action	<form>	Specifies where to send the form-data when a

CSS 셀렉터 - attribute

- css selector – attribute 예시
- *= : 포함
- ^= : ~으로 시작
- \$= : ~으로 끝남
- alt *= '살림백서'
- alt ^= '살림백서'
- alt \$= '살림백서'

```
document.querySelector("img[alt='살림백서 오픈 바디로션']")

document.querySelector("img[alt*='살림백서']")

document.querySelector("img[alt^='살림백서']")

document.querySelector("img[alt~='살림백서']")

document.querySelector("img[alt$='살림백서']")

```

CSS 셀렉터 - 한정자

한정자	설명
*	모든 노드들
div, p	div와 p 노드들
div > p	div 바로 안에 있는 p노드들
div ~ p	p 바로 앞에 있는 div 노드들
div + p	div 바로 뒤에 있는 p 노드들

```
<h1></h1>
<div>
  <div>
    <p>p1<span></span></p>
  </div>
  <h1></h1>
  <p>p2</p>
  <div>
    <h1></h1>
  </div>
  <p>p2</p>
</div>
```


CSS 셀렉터 - 한정자

[illegible]

CSS 고급 - 고급 한정자

- 일반 한정자에서 더 세밀하게 select 할 수 있음
- 예시) "input:disabled"

한정자	설명		
:enabled	활성화 된 상태	:first-of-type	해당 타입의 첫번째의 노드
:checked	체크 된 상태	:last-of-type	해당 타입의 마지막의 노드
:disabled	비활성화 된 상태	:hover	마우스가 올라간 상태
:empty	값이 비어 있는 상태	:not	다음 조건이 거짓일 경우
:first-child	첫번째 자식	:nth-child()	N번째 자식(시작1부터)
:last-child	마지막 자식	:nth-of-type()	n번째 타입

정적 크롤링 프로젝트

BeautifulSoup 활용 프로젝트2

• 인상 상승 종목 주가, 거래량 가져오기

금융 홈 **국내증시** 해외증시 시장지표 펀드 리서치 뉴스 MY



진공블렌더 최대 37% 할인
59만원 진공블렌더를 30만원대에 구매 가능!

지금 구매하기 >

AD

국내증시

주요시세정보

코스피 | 코스닥 | 선물

코스피200 | 코넥스

시가총액 | 배당

업종 | 테마 | 그룹사

ETF | ETN

상승 | 보합 | 하락

상한가 | 하한가

급등 | 급락

거래상위 | 급등 | 급락

투자자별매매동향

외국인매매 | 기관매매

프로그램매매동향

종시자금동향

신규상장

금융홈 > 국내증시 > 검색상위 종목

검색상위 종목

항목을 자유롭게 변경하실 수 있습니다. | 최대 6개까지 설정 가능합니다.

<input checked="" type="checkbox"/> 거래량	<input type="checkbox"/> 매수호가	<input type="checkbox"/> 거래대금(백만)	<input type="checkbox"/> 시가총액(억)	<input type="checkbox"/> 영업이익(억)	<input checked="" type="checkbox"/> PER(배)
<input checked="" type="checkbox"/> 시가	<input type="checkbox"/> 매도호가	<input type="checkbox"/> 전일거래량	<input type="checkbox"/> 자산총계(억)	<input type="checkbox"/> 영업이익증가율	<input checked="" type="checkbox"/> ROE(%)
<input checked="" type="checkbox"/> 고가	<input type="checkbox"/> 매수종전량	<input type="checkbox"/> 외국인비율	<input type="checkbox"/> 부채총계(억)	<input type="checkbox"/> 당기순이익(억)	<input type="checkbox"/> ROA(%)
<input checked="" type="checkbox"/> 저가	<input type="checkbox"/> 매도종전량	<input type="checkbox"/> 상장주식수(천주)	<input type="checkbox"/> 매출액(억)	<input type="checkbox"/> 주당순이익(원)	<input type="checkbox"/> PBR(배)
		<input type="checkbox"/> 상장주식수(천주)	<input type="checkbox"/> 매출증가율	<input type="checkbox"/> 보통주배당금(원)	<input type="checkbox"/> 유보율(%)
<div>적용하기 초기 항목으로</div>					

순위	종목명	검색비율	현재가	전일비	등락률	거래량	시가	고가	저가	PER	ROE
1	삼성전자	1.78%	77,000	▲ 400	+0.52%	12,500,473	77,400	77,400	76,400	16.21	9.99
2	카카오	1.40%	122,500	▼ 1,500	-1.21%	9,028,112	123,500	127,500	122,000	114.17	2.70
3	NAVER	0.93%	400,500	▼ 2,000	-0.50%	706,020	403,000	409,500	400,000	3.99	15.22
4	한국반도체	0.76%	2,020	▲ 465	+29.90%	509,200	1,555	2,020	1,555	7.80	-379.56
5	한국비엔씨	0.67%	43,100	▲ 4,200	+10.80%	5,439,254	39,700	43,650	39,350	-365.25	4.41

삼성전자 : 77,000 : 12,500,473
카카오 : 122,500 : 9,028,112
한국반도체 : 2,020 : 509,200
삼성엔지니어링 : 26,850 : 19,609,743
셀트리온 : 272,500 : 570,117
NAVER : 400,500 : 706,020
엔씨소프트 : 596,000 : 151,818
메스디바이오센서 : 47,900 : 2,333,593
이노덱스 : 23,900 : 2,766,792
씨아이에스 : 16,700 : 48,990,560
LG화학 : 718,000 : 449,681
카카오뱅크 : 69,000 : 5,036,295
POSCO : 361,500 : 449,945
한국비엔씨 : 43,100 : 5,439,254
하나마이크론 : 17,150 : 6,940,561
한국조선해양 : 117,500 : 232,315
SK하이닉스 : 107,500 : 2,797,024
SK바이오사이언스 : 280,000 : 824,744
멜파스 : 3,255 : 9,784,840
한국가스공사 : 40,800 : 1,971,262
HMM : 38,400 : 2,020,277

BeautifulSoup 활용 프로젝트3

• 거래 상위 종목, 주가, 등락율, 거래량 가져오기

국내증시

주요시세정보

코스피 | 코스닥 | 선물

코스피200 | 코넥스

시가총액 | 배당

업종 | 테마 | 그룹사

ETF | ETN

상승 | 보합 | 하락

상한가 | 하한가

급등 | 급락

거래상위 | 급등 | 급락

투자자별매매동향

외국인매매 | 기관매매

프로그램매매동향

종시자금동향

신규상장

외국인보유

장외시세

IPO

투자자보호

관리종목

거래장지종목

시장경보종목

금융홈 > 국내증시 > 거래상위

거래상위

항목을 자유롭게 변경하실 수 있습니다. | 최대 7개까지 설정 가능합니다.

<input checked="" type="checkbox"/> 거래량	<input checked="" type="checkbox"/> 매수호가	<input checked="" type="checkbox"/> 거래대금(백만)	<input checked="" type="checkbox"/> 시가총액(억)	<input type="checkbox"/> 영업이익(억)	<input checked="" type="checkbox"/> PER(배)
<input type="checkbox"/> 시가	<input checked="" type="checkbox"/> 매도호가	<input type="checkbox"/> 전일거래량	<input type="checkbox"/> 자산총계(억)	<input type="checkbox"/> 영업이익증가율	<input checked="" type="checkbox"/> ROE(%)
<input type="checkbox"/> 고가	<input type="checkbox"/> 매수증잔량	<input type="checkbox"/> 외국인비율	<input type="checkbox"/> 부채총계(억)	<input type="checkbox"/> 당기순이익(억)	<input type="checkbox"/> ROA(%)
<input type="checkbox"/> 저가	<input type="checkbox"/> 매도증잔량	<input type="checkbox"/> 상장주식수(천주)	<input type="checkbox"/> 매출액(억)	<input type="checkbox"/> 주당순이익(원)	<input type="checkbox"/> PBR(배)
			<input type="checkbox"/> 매출액증가율	<input type="checkbox"/> 보통주배당금(원)	<input type="checkbox"/> 유보율(%)

[적용하기](#) [초기 항목으로](#)

코스피		코스닥									
N	종목명	현재가	전일비	등락률	거래량	거래대금	매수호가	매도호가	시가총액	PER	ROE
1	KODEX 200선물 인버스2X	2,220	▲ 40	+1.83%	101,691,866	225,075	2,215	2,220	21,834	N/A	N/A
2	삼성중공업	6,450	▲ 500	+8.40%	37,750,649	242,411	6,450	6,460	40,635	-2.83	-33.06
3	세원이앤씨	1,080	▲ 168	+18.42%	33,581,353	34,631	1,075	1,080	1,572	-6.59	-15.32
4	KODEX 인버스	4,050	▲ 40	+1.00%	22,825,257	92,258	4,045	4,050	11,312	N/A	N/A
5	KODEX 코스닥 150선물인버스	4,250	▲ 20	+0.47%	22,672,876	96,420	4,250	4,255	4,645	N/A	N/A
6	현대두산인프라 코어	10,600	▲ 1,280	+13.73%	22,123,374	241,206	10,600	10,650	8,430	8.20	6.47
7	HMM	26,900	▼ 2,500	-8.50%	20,099,789	536,178	26,850	26,900	109,051	17.04	8.93
8	대원화성	5,720	▲ 60	+1.06%	17,666,194	108,374	5,720	5,730	2,359	-143.00	-2.80
9	우리금융	1,000	▼ 5	-0.50%	16,212,181	16,379	999	1,000	8,742	10.64	14.05
10	쌍방울	748	▼ 40	-5.08%	15,676,441	11,881	748	749	1,964	15.58	-8.46

KODEX 200선물인버스2X : 2,220 : +1.83% : 101,691,866
삼성중공업 : 6,450 : +8.40% : 37,750,649
세원이앤씨 : 1,080 : +18.42% : 33,581,353
KODEX 인버스 : 4,050 : +1.00% : 22,825,257
KODEX 코스닥150선물인버스 : 4,250 : +0.47% : 22,672,876
현대두산인프라코어 : 10,600 : +13.73% : 22,123,374
HMM : 26,900 : -8.50% : 20,099,789
대원화성 : 5,720 : +1.06% : 17,666,194
우리금융 : 1,000 : -0.50% : 16,212,181
쌍방울 : 748 : -5.08% : 15,676,441
KODEX 레버리지 : 23,795 : -1.78% : 15,581,449
웰바이오텍 : 1,900 : +6.44% : 14,979,245
이아이지 : 399 : +0.25% : 14,901,930
KEC : 3,335 : -0.45% : 13,559,295
두산중공업 : 25,700 : -1.53% : 13,140,001
IHQ : 1,820 : +4.00% : 12,123,384
신원 : 3,380 : -3.57% : 11,786,621
SH에너지화학 : 1,095 : +2.34% : 11,777,156
삼성전자 : 70,100 : -1.41% : 10,054,916
국보 : 2,260 : -9.24% : 9,949,179
대한전선 : 2,315 : -1.91% : 7,369,985
한화투자증권 : 5,850 : +1.39% : 7,265,173
대우에이텍 : 1,840 : -1.08% : 6,720,079
대우플러스 : 1,470 : -1.67% : 6,607,784
한신기계 : 4,675 : -2.60% : 5,866,685
팬오션 : 6,240 : -0.79% : 5,863,201
대원전선 : 1,995 : -1.24% : 5,397,138

BeautifulSoup 활용 프로젝트3 (스스로 하기)

• 야후 해외 인기 상승 종목 가져오기

<input type="checkbox"/> Symbol	Name	Price (Intraday)	Change	% Change	Volume	Avg Vol (3 month)
<input type="checkbox"/> FCEL	FuelCell Energy, Inc.	6.44	+0.82	+14.59%	213.616M	17.24M
<input type="checkbox"/> GSAT	Globalstar, Inc.	1.7500	-0.5300	-23.25%	117.979M	43.995M
<input type="checkbox"/> AAPL	Apple Inc.	148.12	-1.43	-0.96%	109.296M	78.577M
<input type="checkbox"/> AMC	AMC Entertainment Holdings, Inc.	47.30	-4.39	-8.49%	65.415M	104.941M
<input type="checkbox"/> SDC	SmileDirectClub, Inc.	5.76	-0.22	-3.68%	60.35M	7.309M
<input type="checkbox"/> CMCSA	Comcast Corporation	55.59	-4.38	-7.30%	54.749M	13.274M
<input type="checkbox"/> BAC	Bank of America Corporation	39.84	-1.10	-2.69%	53.82M	47.593M
<input type="checkbox"/> LVS	Las Vegas Sands Corp.	38.71	-4.18	-9.75%	51.516M	8.339M
<input type="checkbox"/> F	Ford Motor Company	12.86	-0.13	-1.00%	46.911M	60.657M
<input type="checkbox"/> AMD	Advanced Micro Devices, Inc.	105.73	+0.93	+0.89%	40.528M	62.596M
<input type="checkbox"/> WISH	ContextLogic Inc.	6.33	-0.28	-4.24%	41.115M	64.682M
<input type="checkbox"/> UBER	Uber Technologies, Inc.	39.01	-1.06	-2.65%	36.341M	21.502M
<input type="checkbox"/> WFC	Wells Fargo & Company	46.05	+0.29	+0.63%	35.978M	27.805M
<input type="checkbox"/> PLTR	Palantir Technologies Inc.	26.26	+0.43	+1.66%	33.898M	42.254M
<input type="checkbox"/> BBD	Banco Bradesco S.A.	4.0600	-0.0500	-1.22%	30.886M	26.199M
<input type="checkbox"/> NIO	NIO Inc.	37.89	-0.50	-1.30%	30.634M	52.584M
<input type="checkbox"/> OPEN	Opendoor Technologies Inc.	16.90	-0.86	-4.84%	30.49M	11.183M
<input type="checkbox"/> ITUB	Itaú Unibanco Holding S.A.	5.49	-0.05	-0.90%	28.566M	36.209M

FCEL : 6.44 : +14.59%
GSAT : 1.7500 : -23.25%
AAPL : 148.12 : -0.96%
AMC : 47.30 : -8.49%
SDC : 5.76 : -3.68%
CMCSA : 55.59 : -7.30%
BAC : 39.84 : -2.69%
LVS : 38.71 : -9.75%
F : 12.86 : -1.00%
AMD : 105.73 : +0.89%
WISH : 6.33 : -4.24%
UBER : 39.01 : -2.65%
WFC : 46.05 : +0.63%
PLTR : 26.26 : +1.66%
BBD : 4.0600 : -1.22%
NIO : 37.89 : -1.30%
OPEN : 16.90 : -4.84%
ITUB : 5.49 : -0.90%
ZNGA : 8.19 : -1.68%
CPNG : 29.87 : -2.13%
T : 27.33 : -0.69%
EDU : 2.0000 : -4.31%
VALE : 17.83 : -1.44%
C : 68.78 : -2.47%
ET : 9.30 : -0.43%

http request 테스트

- HTTP GET vs. POST
- GET : 리소스 얻기
- POST : 리소스 생성

https://webhook.site/

REQUESTS (2/500) Newest First

GET #b4b29 39.117.205.175
09/15/2021 9:06:19 PM

GET #5c535 39.117.205.175
09/15/2021 9:05:32 PM

Request Details

[Permalink](#) [Raw content](#) [Export as ▼](#) [Delete](#)

GET https://webhook.site/c399c2a0-1092-4d25-8d9d-99d2a591b617

Host39.117.205.175 whois

Date09/15/2021 9:05:32 PM (a minute ago)

Size0 bytes

ID5c535a4e-553e-4c11-a1aa-75315b494760

Files

Headers

connectionclose

accept*/*

accept-encodinggzip, deflate

user-agentpython-requests/2.26.0

hostwebhook.site

content-length

content-type

Query strings

(empty)

Form values

(empty)

No content