

# 삼성 청년 SW 아카데미

Web Programming 기초

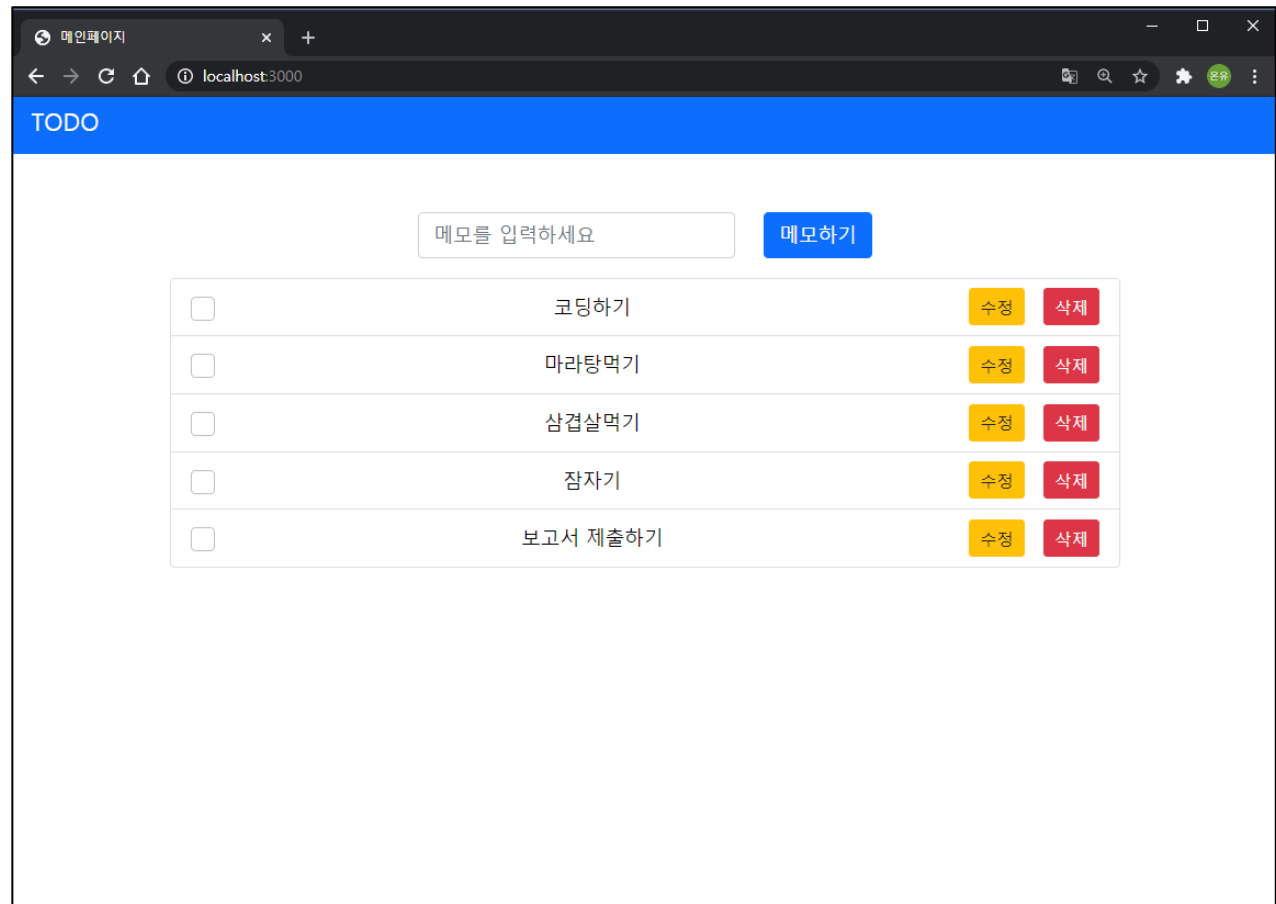
### <알림>

본 강의는 삼성 청년 SW아카데미의 콘텐츠로  
보안서약서에 의거하여  
강의 내용을 어떠한 사유로도 임의로 복사,  
촬영, 녹음, 복제, 보관, 전송하거나  
허가 받지 않은 저장매체를  
이용한 보관, 제3자에게 누설, 공개,  
또는 사용하는 등의 행위를 금합니다.

# TODO LIST 만들기

## TODO List 구현하기

- mysql 연동
- CRUD 구현해보기



# 웹 크롤링

## 챕터의 포인트

- 크롤링의 개요

조직적, 자동화된 방법으로 웹을 탐색하는 프로그램



검색엔진은 웹 크롤링을 진행한다.

A search bar with a magnifying glass icon on the left and keyboard and microphone icons on the right.

Google 검색

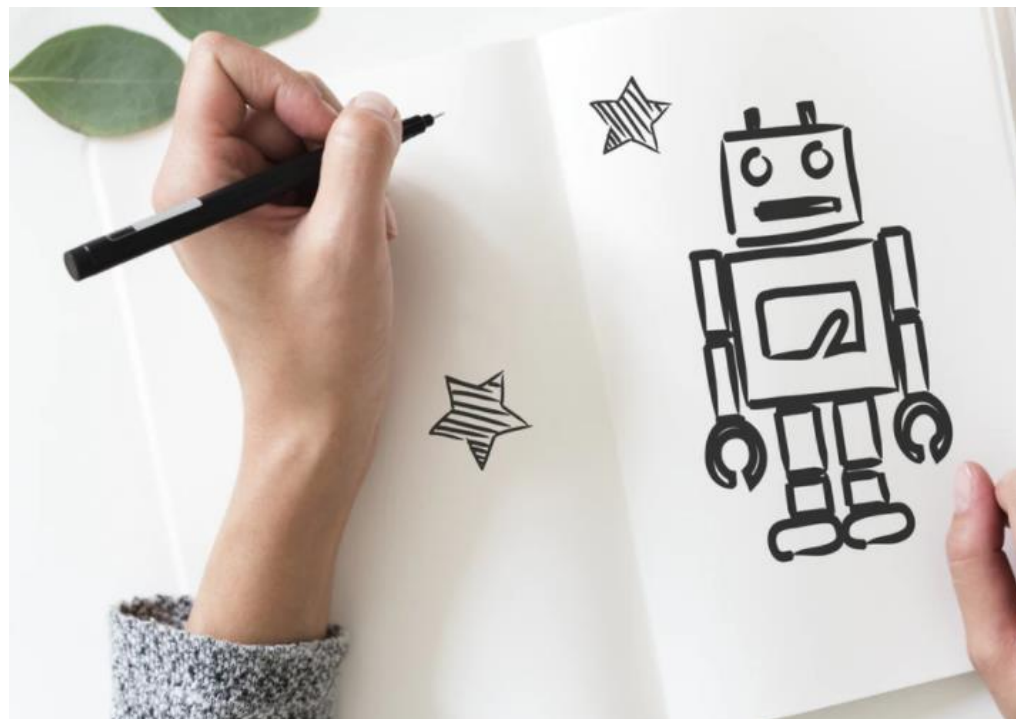
I'm Feeling Lucky

Google 제공 서비스: [English](#)



## robots.txt

- 각 홈페이지 뒤에 robots.txt 를 붙이면 조회할 수 있다. (ex: <https://www.google.com/robots.txt>)
- 웹 크롤러들이 접근하지 못하도록 제한하는 규약

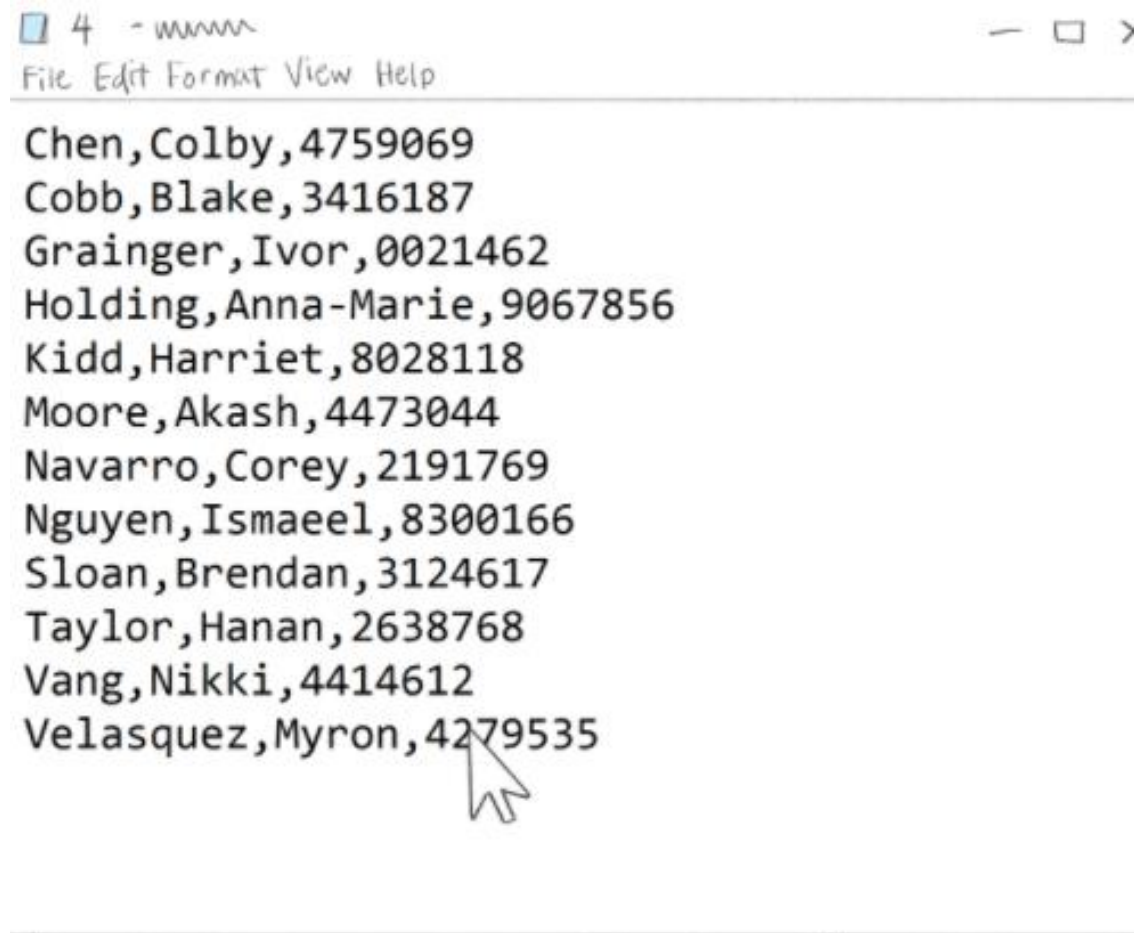


# 파일 파싱하기

## 챕터의 포인트

- csv 파싱
- xlsx 파싱

- 행렬구조의 데이터를 표현/저장 하기 위한 포맷



A screenshot of a text editor window titled "4 - www". The window contains a list of names and IDs, separated by commas, representing CSV data. The text is as follows:

```
Chen,Colby,4759069  
Cobb,Blake,3416187  
Grainger,Ivor,0021462  
Holding,Anna-Marie,9067856  
Kidd,Harriet,8028118  
Moore,Akash,4473044  
Navarro,Corey,2191769  
Nguyen,Ismaeel,8300166  
Sloan,Brendan,3124617  
Taylor,Hanan,2638768  
Vang,Nikki,4414612  
Velasquez,Myron,4279535
```

A mouse cursor is visible pointing at the end of the last line of text.

## npm | csv-parser

- csv 형식을 가져올 때 활용

```
data-movie.csv ×  result.csv  JS  xlsxParsing.js  JS  add_to_sheet.js  JS  axiosCh
lecture > csv > data-movie.csv
1  알라딘,https://movie.naver.com/movie/bi/mi/basic.nhn?code=163788
2  위대한 쇼맨,https://movie.naver.com/movie/bi/mi/basic.nhn?code=106360
3 1917,https://movie.naver.com/movie/bi/mi/basic.nhn?code=187321
4 어바웃 타임,https://movie.naver.com/movie/bi/mi/basic.nhn?code=92075
5 비긴어게인,https://movie.naver.com/movie/bi/mi/basic.nhn?code=96379
6 라라랜드,https://movie.naver.com/movie/bi/mi/basic.nhn?code=134963
7
```

- <https://support.spatialkey.com/spatialkey-sample-csv-data/>
  - CSV 파일을 다운로드 받는다.
  - county가 MIAMI DADE COUNTY 이면서 construction가 wood인 data만 뽑아내기
  - console.log로 출력

	A	B	C	D	E	F	G	H	I	
1	policyID	statecode	county	eq_site_lin	hu_site_lin	fl_site_limi	fr_site_limi	tiv_2011	tiv_2012	e
2	119736	FL	CLAY COUNTY	498960	498960	498960	498960	498960	792148.9	
3	448094	FL	CLAY COUNTY	1322376	1322376	1322376	1322376	1322376	1438164	
4	206893	FL	CLAY COUNTY	190724.4	190724.4	190724.4	190724.4	190724.4	192476.8	
5	333743	FL	CLAY COUNTY	0	79520.76	0	0	79520.76	86854.48	
6	172534	FL	CLAY COUNTY	0	254281.5	0	254281.5	254281.5	246144.5	
7	785275	FL	CLAY COUNTY	0	515035.6	0	0	515035.6	884419.2	
8	995932	FL	CLAY COUNTY	0	19260000	0	0	19260000	20610000	
9	223488	FL	CLAY COUNTY	328500	328500	328500	328500	328500	348374.3	
10	433512	FL	CLAY COUNTY	315000	315000	315000	315000	315000	265821.6	
11	142071	FL	CLAY COUNTY	705600	705600	705600	705600	705600	1010843	
12	253816	FL	CLAY COUNTY	831498.3	831498.3	831498.3	831498.3	831498.3	1117791	
13	894922	FL	CLAY COUNTY	0	24059.09	0	0	24059.09	33952.19	
14	422834	FL	CLAY COUNTY	0	48115.94	0	0	48115.94	66755.39	
15	582721	FL	CLAY COUNTY	0	28869.12	0	0	28869.12	42826.99	

## 엑셀파일을 파싱해보자



## npm | xlsx

- xlsx형식의 파일 파싱 라이브러리

H13		
	A	B
1	제목	링크
2	알라딘	<a href="https://movie.naver.com/movie/bi/mi/basic.nhn?code=163788">https://movie.naver.com/movie/bi/mi/basic.nhn?code=163788</a>
3	위대한 쇼맨	<a href="https://movie.naver.com/movie/bi/mi/basic.nhn?code=106360">https://movie.naver.com/movie/bi/mi/basic.nhn?code=106360</a>
4	1917	<a href="https://movie.naver.com/movie/bi/mi/basic.nhn?code=187321">https://movie.naver.com/movie/bi/mi/basic.nhn?code=187321</a>
5	어바웃 타임	<a href="https://movie.naver.com/movie/bi/mi/basic.nhn?code=92075">https://movie.naver.com/movie/bi/mi/basic.nhn?code=92075</a>
6	비긴어게인	<a href="https://movie.naver.com/movie/bi/mi/basic.nhn?code=96379">https://movie.naver.com/movie/bi/mi/basic.nhn?code=96379</a>
7	라라랜드	<a href="https://movie.naver.com/movie/bi/mi/basic.nhn?code=134963">https://movie.naver.com/movie/bi/mi/basic.nhn?code=134963</a>



# 웹 크롤링 기초

## 챕터의 포인트

- axios + cheerio 활용하기

- vue, react에서 권고하는 ajax 라이브러리

# Axios

페이지를 파싱해서 필요한 부분의 정보를 가져올 수 있다.

## cheerio

1.0.0-rc.3 • Public • Published a year ago

[Readme](#)[Explore](#)[BETA](#)[6 Dependencies](#)[11,194 Dependents](#)[60 Versions](#)

## cheerio

Fast, flexible & lean implementation of core jQuery designed specifically for the server.

build passing coverage 99% gitter join chat backers 22 sponsors 73

```
const cheerio = require('cheerio')
const $ = cheerio.load('<h2 class="title">Hello world</h2>')
```

```
$('h2.title').text('Hello there!')
$('h2').addClass('welcome')
```

```
$.html()
```

```
//=> <html><head></head><body><h2 class="title welcome">Hello there!</h2></body></html>
```

### Install

```
> npm i cheerio
```

### Weekly Downloads

4,444,884

### Version

1.0.0-rc.3

### License

MIT

### Unpacked Size

112 kB

### Total Files

15



- axios + cheerio를 활용한 크롤링
  - 간단한 페이지 크롤링
  - ajax가 적용된 사이트 들을 파싱 할 때 어려움이 있다





## data-movie.xlsx 파일 읽어 오기

- 주소를 기반으로 axios + cheerio 를 활용해서 네티즌 평점 가져오기

1917  
1917, 2019

관람객  9.04    기자·평론가  7.67

**네티즌  8.86**    내 평점  등록 >

개요   드라마, 전쟁   미국   119분   2020.02.19 개봉

감독   샘 멘데스

출연   조지 맥케이(스코필드), 던-찰스 채프먼(블레이크) [더보기](#) >

등급   [국내] 15세 관람가

성별·나이별 관람추이



### • 상영등급 가져오기

#### 라라랜드 상영중

La La Land, 2016

관람객 ★★★★★ 8.91    기자·평론가 ★★★★★ 8.34

네티즌 ★★★★★ 8.60    내 평점 ★★★★★ 등록>

개요    드라마, 뮤지컬, 멜로/로맨스 | 미국 | 127분  
2020.03.25재개봉, 2017.12.08재개봉, 2016.12.07개봉

감독    데이미언 셔젤

출연    라이언 고슬링(세바스찬), 엠마 스톤(미아) [더보기](#)

등급    [국내] 12세 관람가 [해외] PG-13 ?

성별·나이별 관람추이





- 도전 및 과제 제출시간(20분)
- <https://www.shakeshack.com/food-and-drink/>
  - 다이어트를 해야하는 애인을 위한 프로그램을 제작하자.
  - 메뉴별 칼로리를 크롤링
  - 가장 칼로리가 높은 음식부터
  - 낮은 칼로리의 음식까지 정렬하여 출력한다.

### Chicken

Our real, white-meat chicken is fresh, never frozen, hand-breaded, and cooked to order every time with no antibiotics—ever.



#### Chick'n Shack

Antibiotic-free, crispy chicken breast with lettuce, pickles, and buttermilk herb mayo on a non-GMO Martin's Potato Roll.

590 cal

#### Chick'n Bites

Antibiotic-free, crispy, all white-meat chicken bites, hand breaded + served with BBQ or honey mustard sauce.

300 cal

#### Chicken Dog

Chicken sausage with sage and apple—no antibiotics, ever.

310 cal





# 웹 크롤링 활용

## 챕터의 포인트

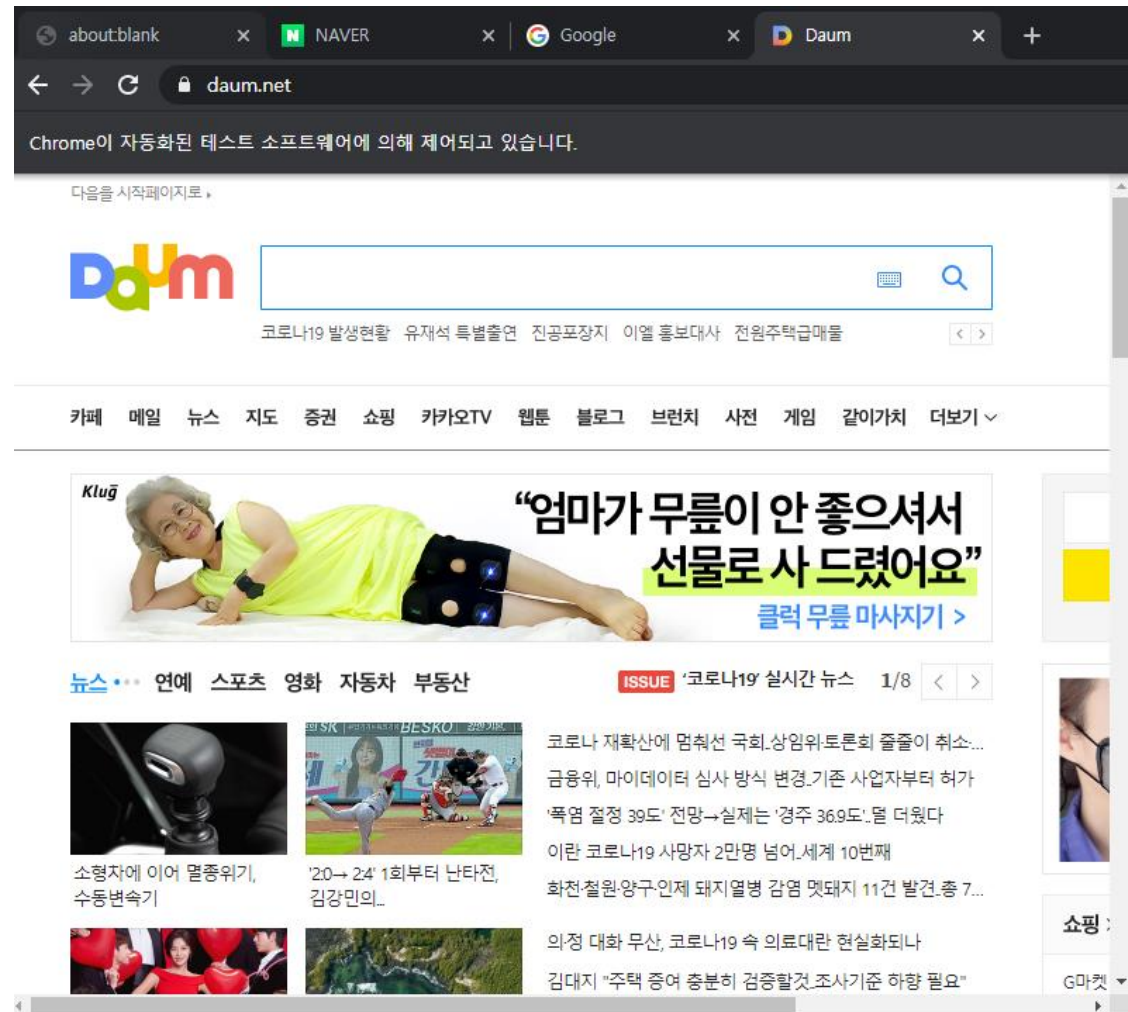
- puppeteer 개요
- puppeteer 크롤링

## • 구글 크롬에서 공개한 Node.js 크롤링 라이브러리

- 크로미움 기반
- headless 로 동작 가능
- PDF 생성
- 스크린샷 찍기
- 키보드입력 자동화
- 데이터 저장
- SPA 크롤링



## puppeteer을 통한 다중 페이지 제어



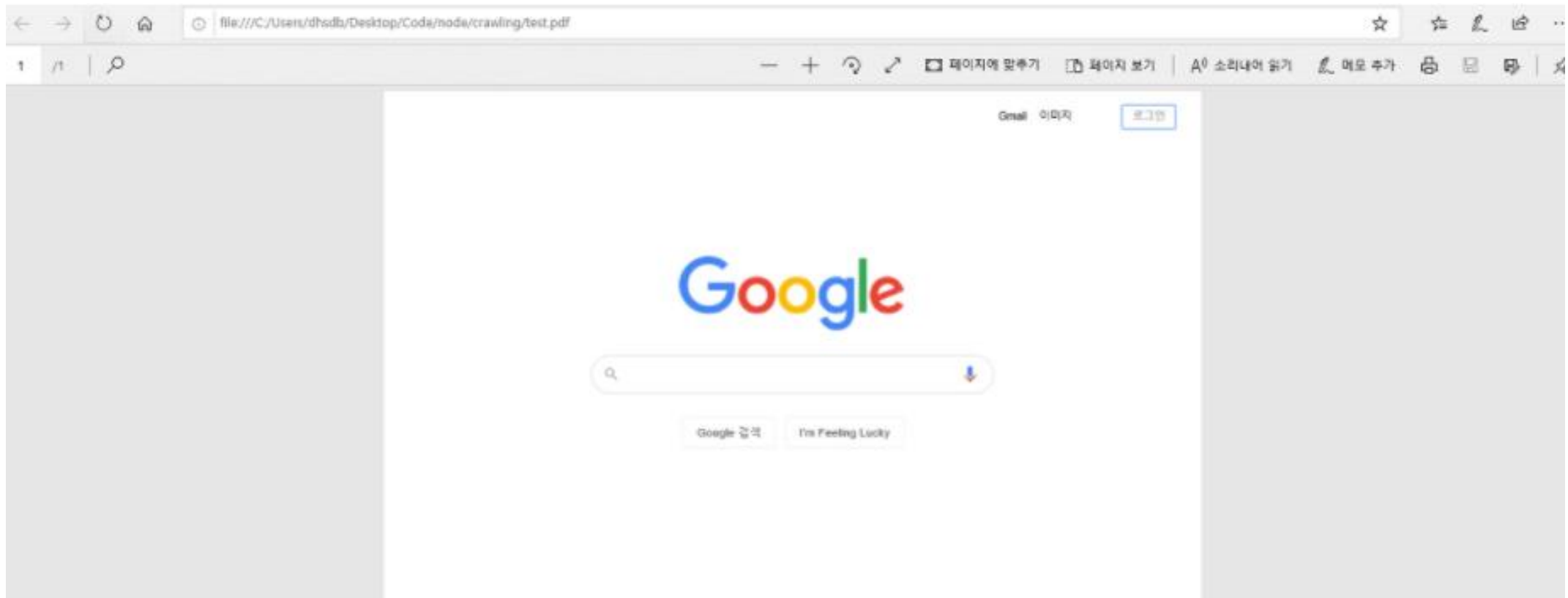
## 스크린샷 찍기

The Google logo is centered on the page, featuring its characteristic multi-colored letters.A standard Google search bar with a magnifying glass icon on the left, a text input field, and icons for keyboard shortcuts and voice search on the right.

Google 검색





I'm Feeling Lucky

## pdf 파일 만들기



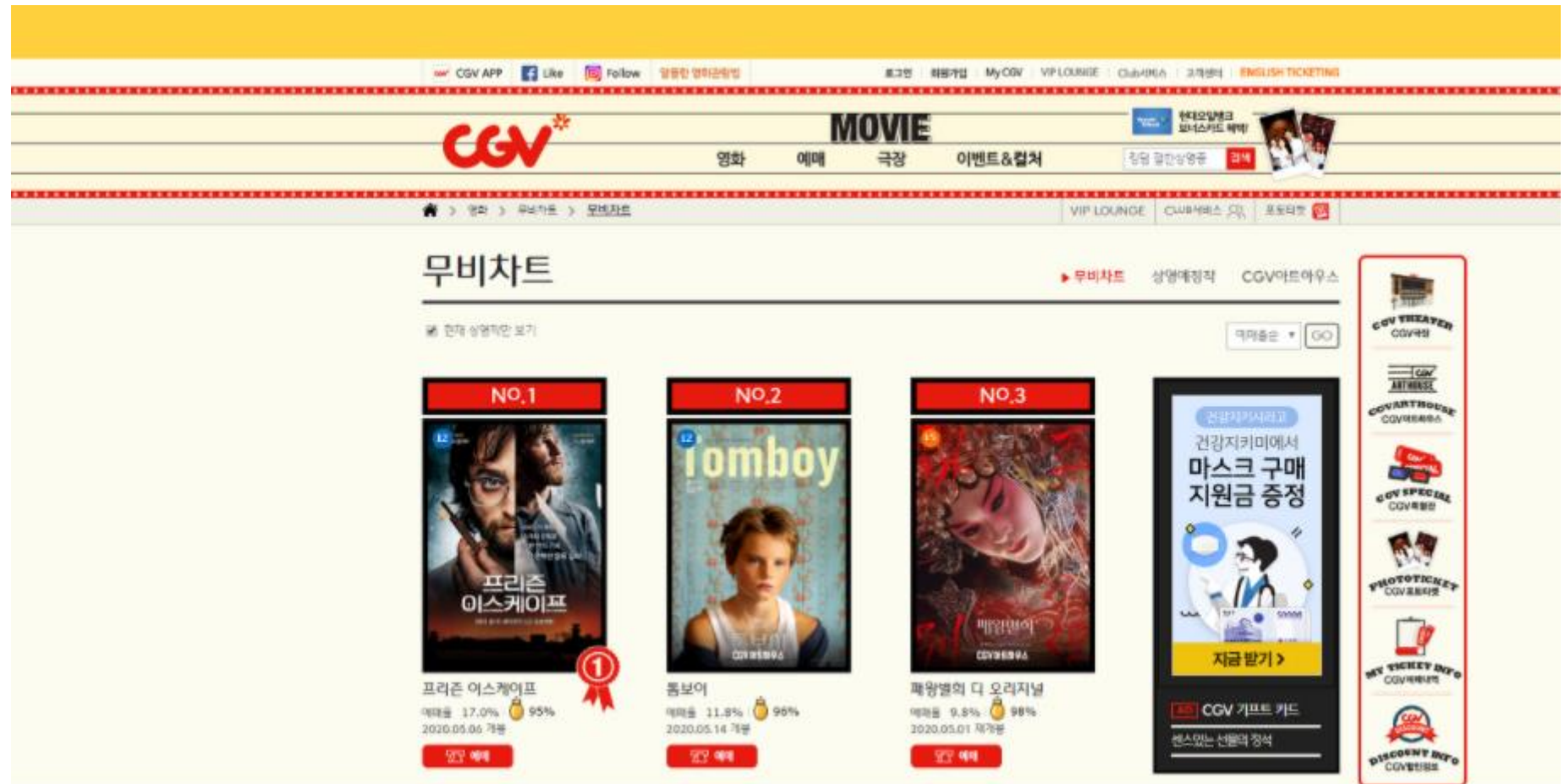
## 다이어트를 하고있는 애인에게 치킨을 선물하고자 한다. 각 치킨 시리즈를 캡처하는 크롤링 프로그램 제작

- csv의 각 메뉴의 링크를 저장한다.
- 각 링크의 접속하여 캡처를 하는 크롤러를 제작한다.
- 교촌 반반시리즈는 pdf로도 저장한다.

전체메뉴	SET	교촌신화시리즈	교촌시리즈	교촌하니시리즈	교촌레드시리즈	교촌반반시리즈	교촌후라이드시리즈	교촌샐살시리즈	교촌E-쿠폰
									
<b>교촌스틱</b> 마늘과 간장 소스의 풍부한 맛과 다리부위의 담백한 맛이 어우러진 치킨	<b>교촌순살(R)</b> 부드럽고 바삭한 정육 순살에 감칠맛나는 마늘 간장 소스가 어우러진 순살치킨	<b>교촌오리지날</b> 교촌만의 차별화된 마늘과 간장 소스의 풍부한 맛이 어우러진 한 마리 치킨	<b>교촌콤보</b> 마늘과 간장소스의 풍부한 맛에 가장 인기있는 부위인 날개와 다리를 함께 즐길 수 있는 메뉴						
관장소비자가격 17,000 원	관장소비자가격 18,000 원	관장소비자가격 15,000 원	관장소비자가격 17,000 원						
<a href="#">자세히보기</a>	<a href="#">자세히보기</a>	<a href="#">자세히보기</a>	<a href="#">자세히보기</a>						

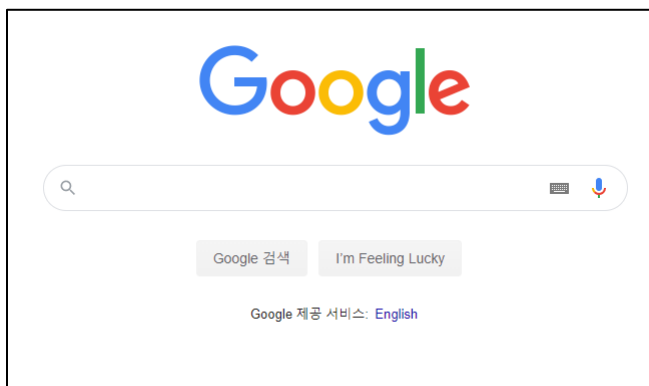


- puppeteer를 활용한 페이지 복제

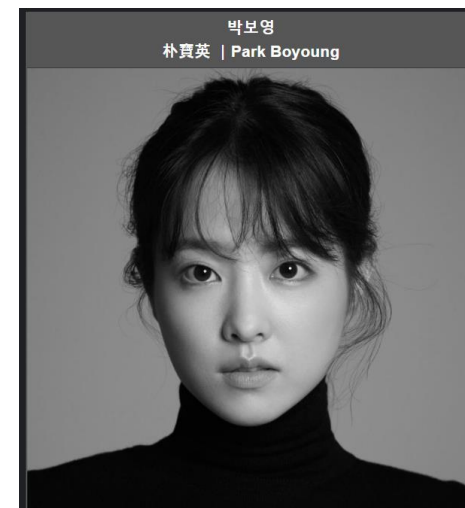
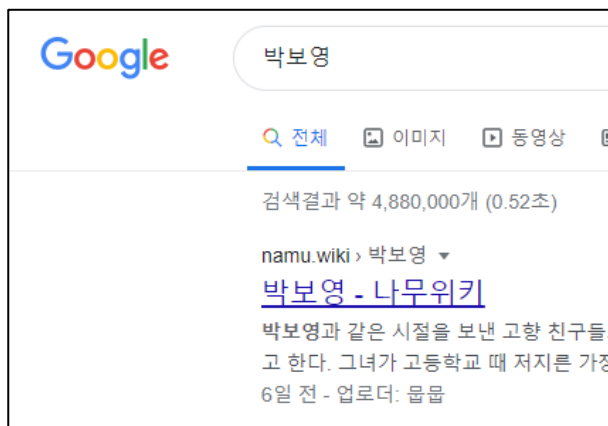




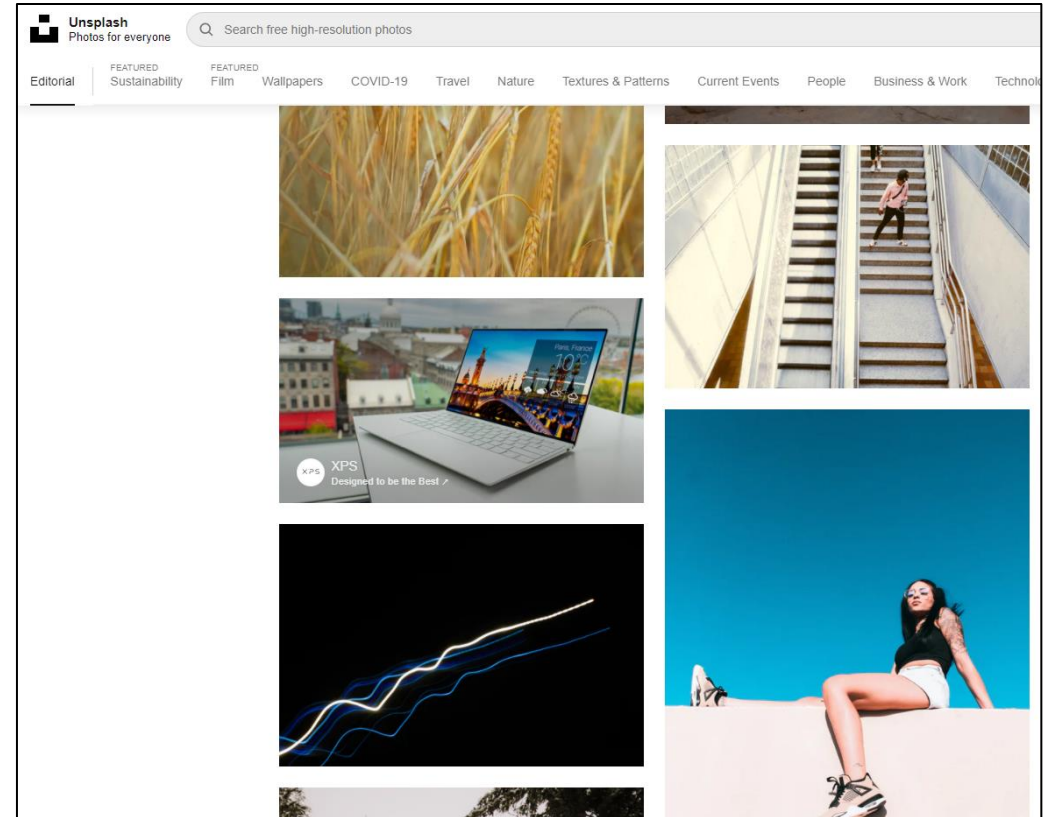
## 좋아하는 연예인을 검색하여 첫번째 사이트에 접속하기



박보영 자동입력 후 검색

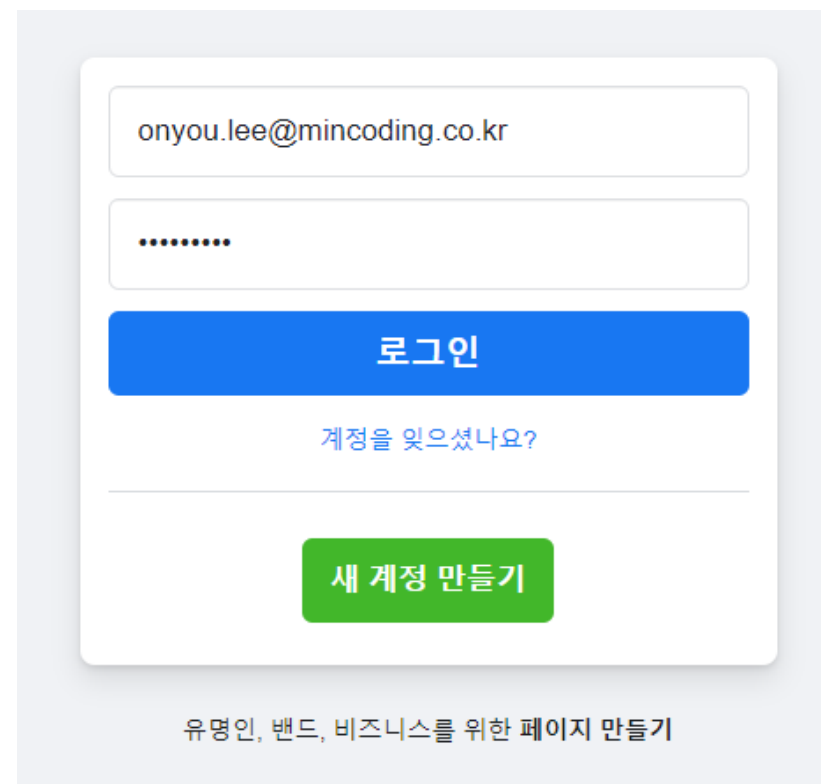
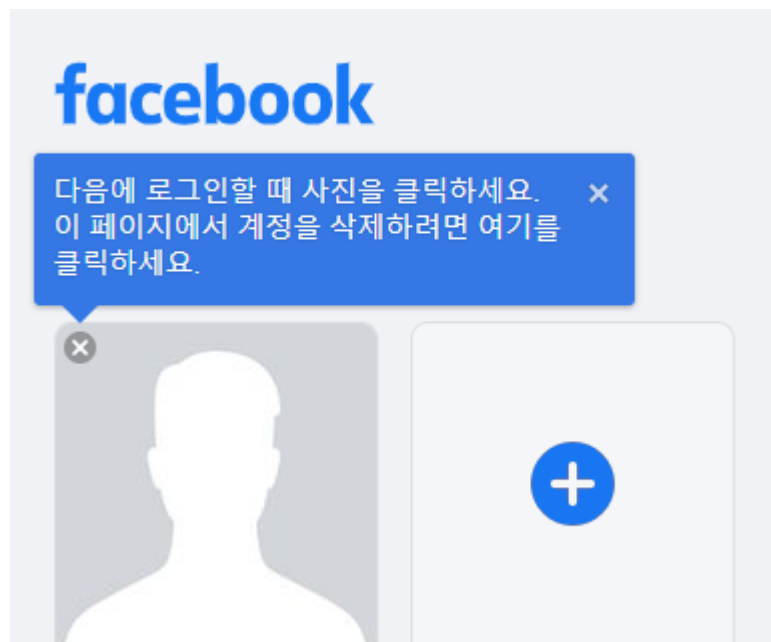


## 무한 스크롤 이미지를 크롤링해오자



## 테스트 계정을 만들고 페이스북 로그인 기능을 구현하자.

- 주의) 무리한 크롤링은 테스트 계정 정지를 불러온다.



# 내일 방송에서 만나요!

삼성 청년 SW 아카데미