



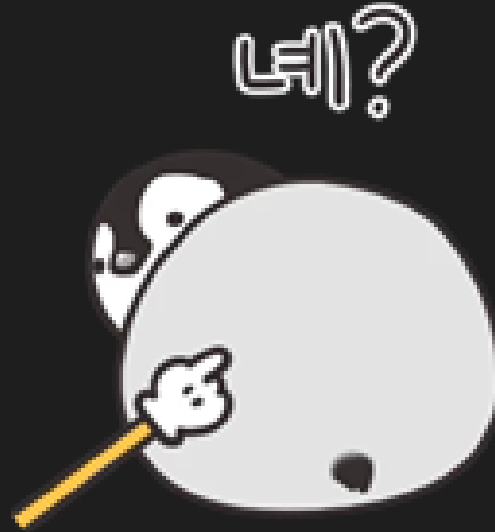
# Web Crawling 기초 및 실습

## First Session

엄펑 쓰앵님



이론따위... 별로 하지 않겠습니다..  
하나도 안 궁금해 하실 것 같거든요..



한 명씩 지목해서 가끔 **질문**합니다  
대답 못하시면 ... ㅎㅎ..  
**긴장** 바짝하고 **집중** 하셔야해용

## 수강생 소개



수원



진Aㅏ



어쩔다지5

가이니

부탁이 있어요



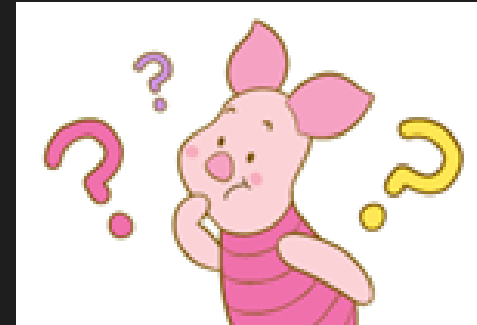
문짚왕 은진띠



쿼터 구



비된찌



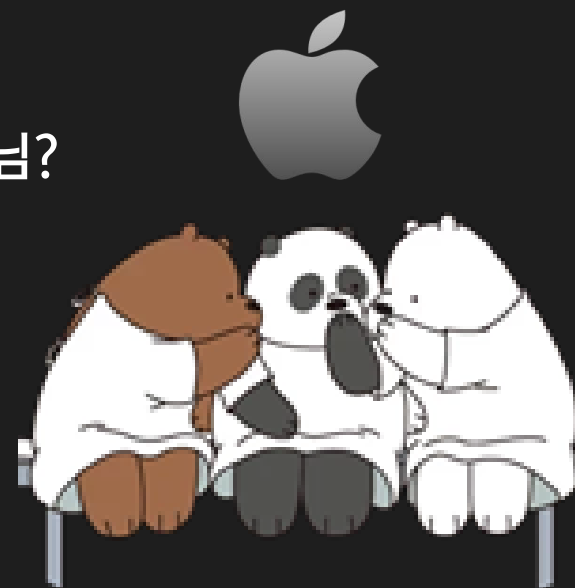
또글렛

그런데 ..... ..



엄펍 쌤님

누가 요새 맥북 아님?  
쑥덕쑥덕



쑥덕쑥덕

수강생 여러분

그래서 매우 걱정입니다만...

뿌이뿌이뿌이~



그럼 시작해 볼까요  
준비 되셨나요오오

# 오늘은 어떤 것을 배울까요옹

---

환경 셋팅, 설치 및 이론

교보문고 책 순위

멜론 플레이리스트

구글 이미지저장



# 환경 셋팅, 설치 및 이론

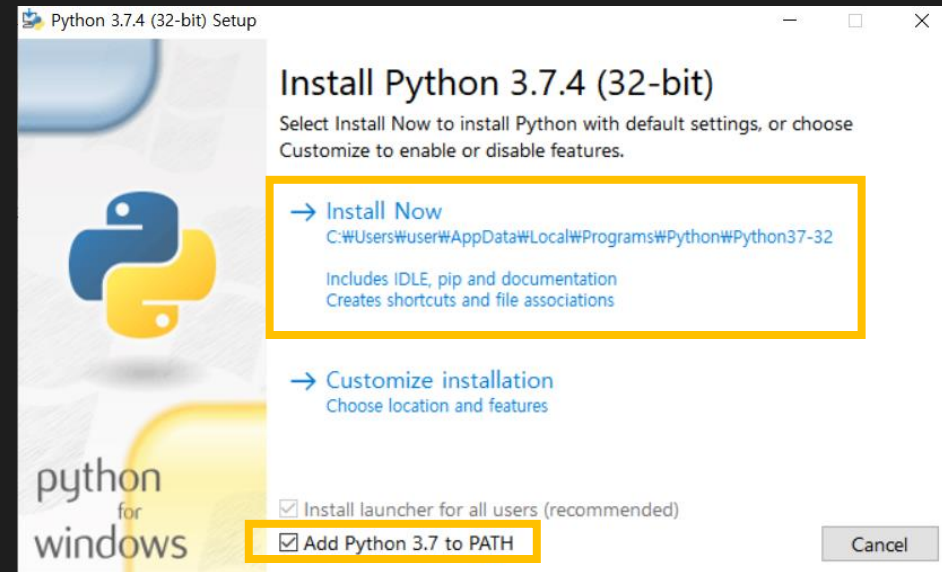
파이썬 홈페이지 접속 후 최신버전 다운



<https://www.python.org/downloads/>

여러분은 맥으로 다운 받으세요

PATH 부분 체크 후 install



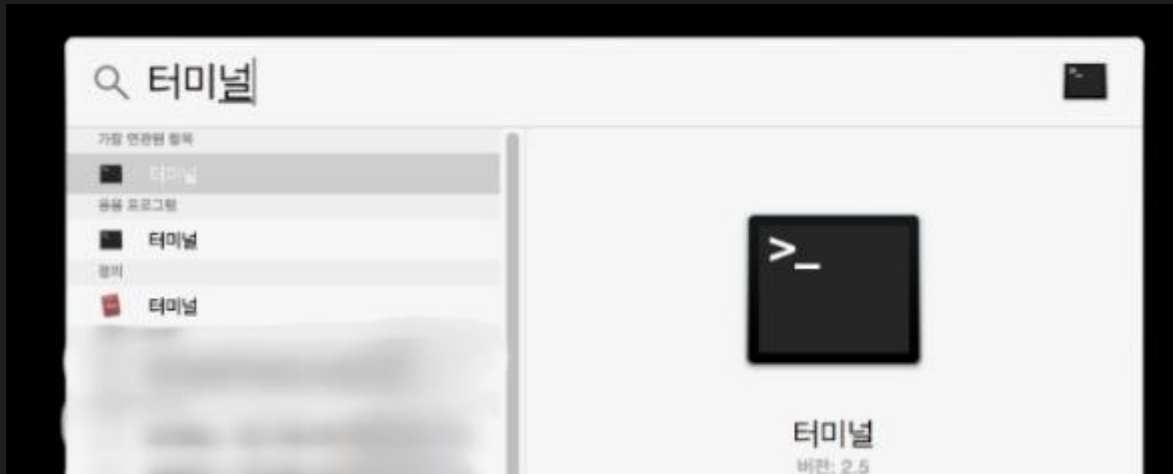
메모장에 다운로드 경로 하나 복사!



## 환경 셋팅, 설치 및 이론

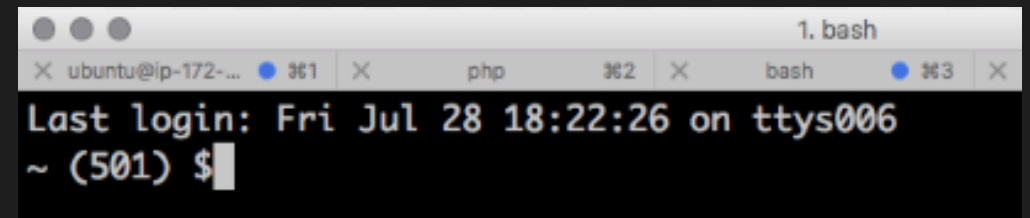
A white Spotlight search bar with a magnifying glass icon on the left and the text "Spotlight 검색" inside.

cmd + 스페이스로 spotlight 켜기



‘터미널’ 혹은 ‘Terminal’ 검색

※터미널 : 명령어 시행하는 부분



## 환경 셋팅, 설치 및 이론

```
1. bash
X ubuntu@ip-172-... 361 X php 362 X bash 363 X
Last login: Fri Jul 28 18:22:26 on ttys006
~ (501) $
```

### bash창에 명령어 입력

\$ sudo easy\_install pip

\$ python3 -m pip install --upgrade pip

**pip upgrade**

※ PIP is a package manager for Python packages, or modules if you like.

\$ python3 -m pip install --user beautifulsoup4

\$ python3 -m pip install --user requests

\$ python3 -m pip install --user fake-useragent

**Beautiful soup**라는 라이브러리 다운

**requests**라는 라이브러리 다운

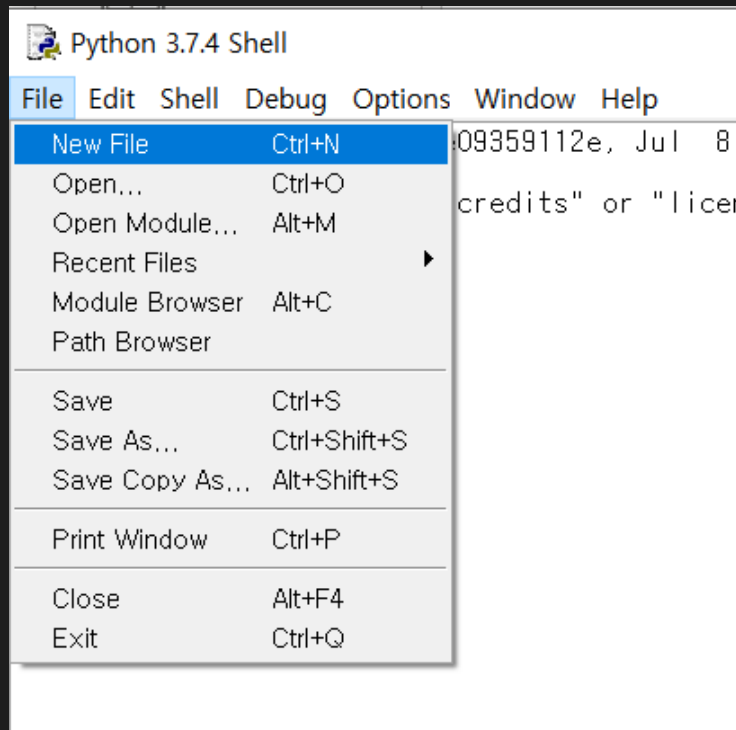
**Fake-useragent**라는 라이브러리 다운

# 환경 셋팅, 설치 및 이론

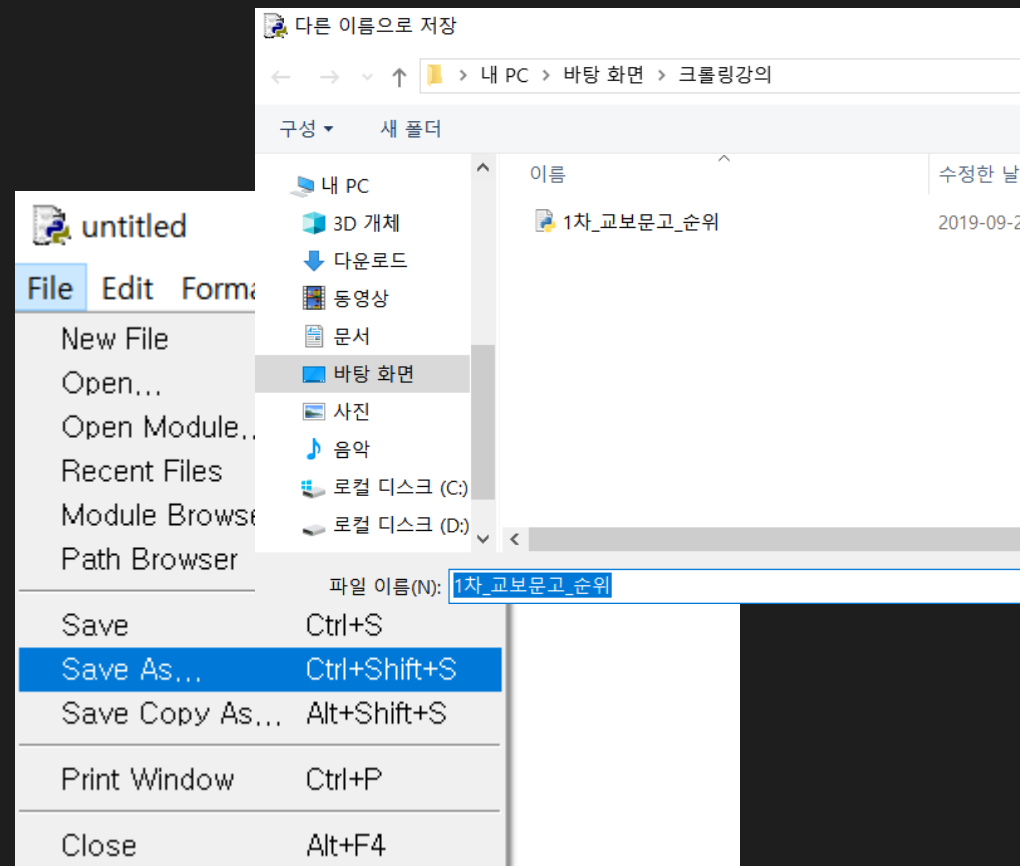
## Python IDLE 실행



## File → New File

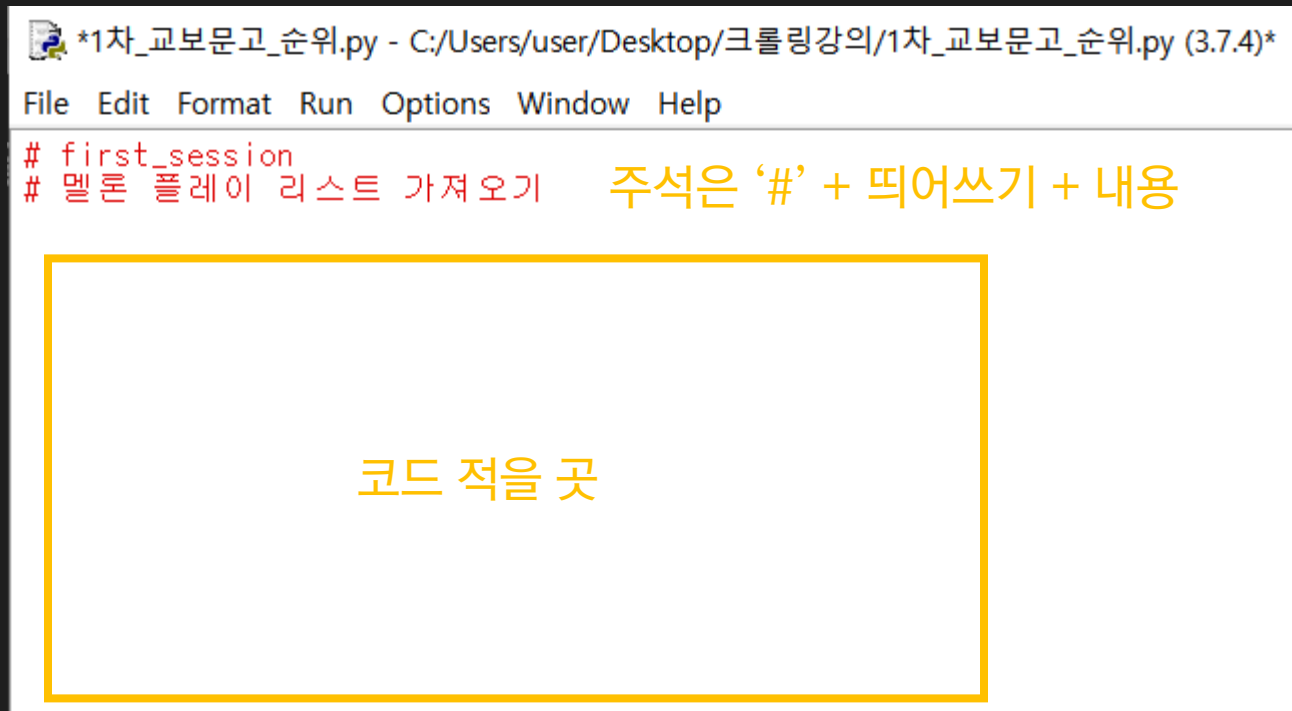


## 제목지정 후 다른 이름으로 저장

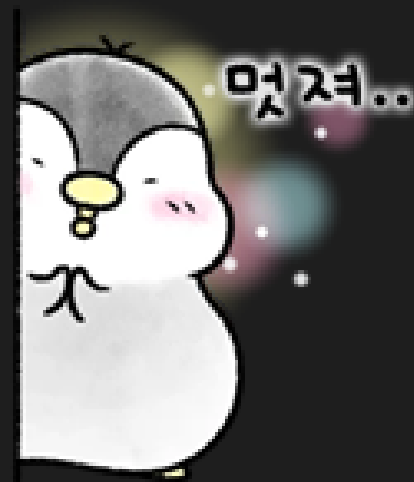


# 환경 셋팅, 설치 및 이론

## 코드 작성 전 주석 달기



※주석 : 코드 실행부분이 아닌 메모하는 부분



셋팅 완료!

이론을 안 한다고는 했지만..



크롤링이.. 뭘지는 알아야겠죠..??

짧고 쉽게 설명하고 가겠습니다!

진Aㅏ



쿼터 구



Q. 크롤링이 뭔가요?

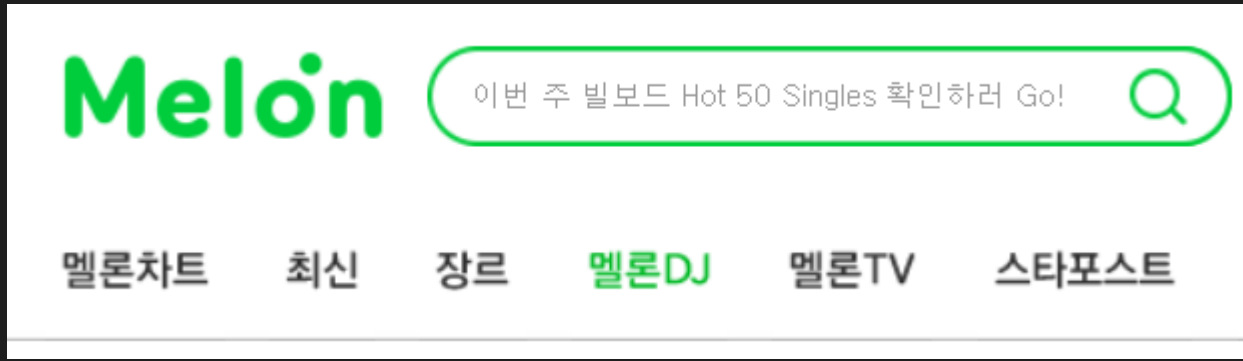
# Crawling

Web scraping, web harvesting, or web data extraction is data scraping used for extracting data from websites.

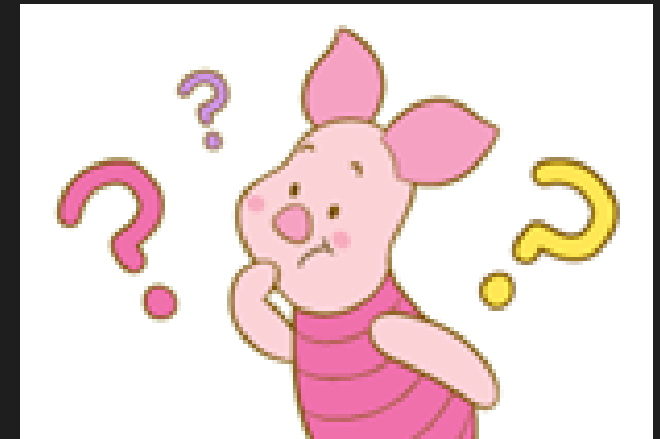
쉽게 말해, 웹 사이트에서 원하는 데이터를 추출하는 것!



어디에 쓰느냐...



어느 날 토글렛은 생각합니다



멜론 플레이리스트를  
저장해 두고 싶당

9월 27일 오늘은 뭘 듣지?!



2019년 꼭 들어야 할 HOT TRENDY POP






설애.

♡ 419 | 240곡

#팝 #분위기









수록곡 (240)

▶ 듣기 + 담기 다운로드 FLAC 선물

번호	곡정보	앨범
1	 Better Apart (Feat. Dresage) Jai Wolf	Better Apart (Feat. Dresage...)
2	 Sharpest Weapon (Feat. Cameron ... tyDi & MORTEN	Sharpest Weapon
3	 Better With You (feat. Iselin) 3Lau, Justin Caruso	Better With You (feat. Ise...)
4	 Happy Now (BEAUZ Remix) (Feat. El... Zedd	Happy Now (Remixes)
5	 Here We Go Again (Feat. Louisa) Sigma, Louisa	Here We Go Again



어디에 쓰느냐...

<input type="checkbox"/>	47		 Melodies Madison Beer
<input type="checkbox"/>	48		 POP/STARS (여자)아이들, Madison Beer, Jaira Burns... ▾
<input type="checkbox"/>	49		 Fa La La (Feat. Boyz II Men) Justin Bieber
<input type="checkbox"/>	50		 September Campsite Dream

으 아닛 너무 많네



어디에 쓰느냐...

Ctrl + C

Ctrl + V

Ctrl + C

Ctrl + V

Ctrl + V

Ctrl + C

Ctrl + C

Ctrl + V

Ctrl + V

대세는 노가다

Ctrl + C

Ctrl + C

Ctrl + C

Ctrl + V

Ctrl + V



어디에 쓰느냐...

엄펑 쓰앵님



자네.. 이거 한 번 써보겠나?



이거슨...  
크롤링??!!!

어디에 쓰느냐...

눈떠보니 50개가 뽑히다니..!



```
1위
title : Better Apart (Feat. Dresage)
artist : Jai WolfJai Wolf
-----
2위
title : Sharpest Weapon (Feat. Cameron Walker)
artist : tyDi & MORTENTyDi & MORTEN
-----
3위
title : Better With You (feat. Iselin)
artist : 3Lau, Justin Caruso3Lau, Justin Caruso
-----
4위
title : Happy Now (BEAUZ Remix) (Feat. Elley Duhé)
artist : ZeddZedd
-----
5위
title : Here We Go Again (Feat. Louisa)
artist : Sigma, LouisaSigma, Louisa
-----
6위
title : Do It Yourself
artist : ILIRAILIRA
-----
7위
title : With You
artist : OysangOysang
-----
8위
title : Crown
artist : Camila Cabello, GreyCamila Cabello, Grey
-----
9위
title : Vertigo (Spitfya Remix)
artist : Rob Gasser, Laura BrehmRob Gasser, Laura Brehm
-----
10위
title : Play (ft. Mangoo)
```

## 실습 1. 멜론 플레이리스트

```
# first_session
# 멜론 플레이리스트 가져오기

import urllib.parse as rep
import urllib.request as req
from bs4 import BeautifulSoup
from fake_useragent import UserAgent
```

```
# Header 정보 초기화
opener = req.build_opener()
# User-Agent 정보
opener.addheaders = [('User-agent', UserAgent().ie)]
# Header 정보 삽입
req.install_opener(opener)
```

`from` 모듈 이름(A) `import` module 요소(B) `as` 별명(C)

————→ A로부터 B를 가져오고 C라고 부를 것이다.

User-agent 값을 새로 셋팅하여  
브라우저라 속이고  
데이터를 요청해야 함

## 실습 1. 멜론 플레이리스트

# 멜론 플레이리스트 URL \*추출하고 싶은 사이트 url 변수에 저장

```
url = "https://www.melon.com/mymusic/dj/mymusicdjplaylistview_inform.htm?plylstSeq=465199375"
```

# 요청 URL 확인

```
print(url)
```

print(내용) → 내용을 출력한다.

# Request

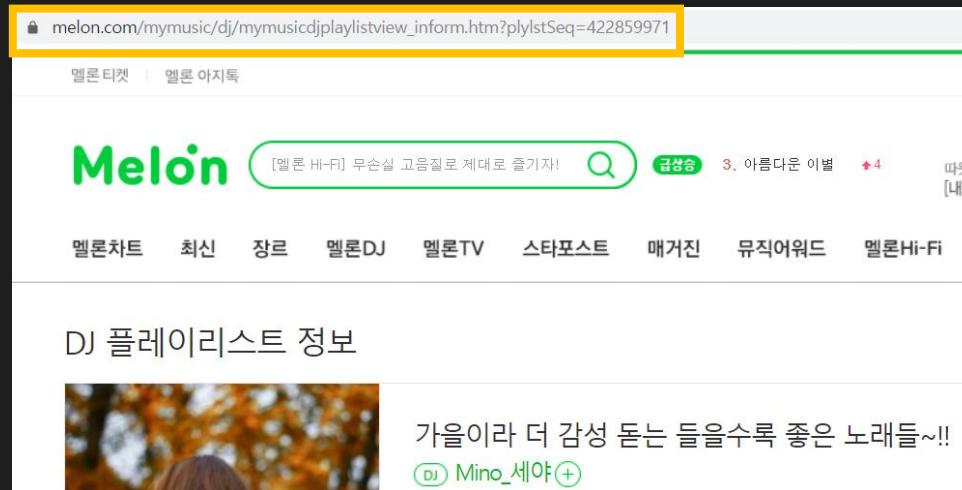
```
res = req.urlopen(url)
```

```
print(res)
```

\*데이터 요청 후 url열어서 가져오기

# bs4 초기화

```
soup = BeautifulSoup(res, "html.parser")
```



Python 3.7.4 Shell

File Edit Shell Debug Options Window Help

Python 3.7.4 (tags/v3.7.4:e09359112e, Jul 8 2019, 19:29:22) [MSC v.1916 32 bit (Intel)] on win32  
Type "help", "copyright", "credits" or "license()" for more information.

>>>

===== RESTART: C:\Users\User\Desktop\크롤링강의\1차\_멜론\_플레이리스트.py =====

https://www.melon.com/mymusic/dj/mymusicdjplaylistview\_inform.htm?plylstSeq=465199375

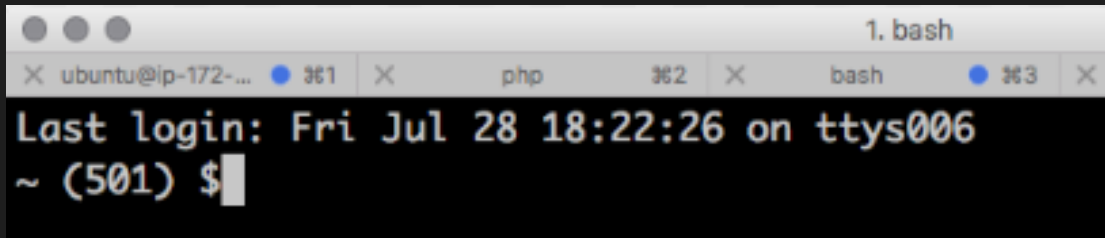
<http.client.HTTPResponse object at 0x03CEAE10>

## 실습 1. 멜론 플레이리스트

```
# first_session  
# 멜론 플레이리스트 가져오기
```

```
import urllib.parse as rep  
import urllib.request as req  
from bs4 import BeautifulSoup  
from fake_useragent import UserAgent
```

```
ua = UserAgent(verify_ssl=False)
```

 ※ 오류 발생시 한줄 추가A screenshot of a terminal window titled '1. bash'. The window shows the login message 'Last login: Fri Jul 28 18:22:26 on ttys006' and a shell prompt '~ (501) \$' with a cursor. The terminal has three tabs: 'ubuntu@ip-172-...', 'php', and 'bash'.

※ 터미널에 아래 명령어 붙여넣기

```
$ open /Applications/Python\ 3.7/Install\ Certificates.command
```

## 실습 1. 멜론 플레이리스트

# select 사용

```
songs = soup.select("td:nth-child(5) > div > div")
```

\*원하는 내용 가져오기

```
print(songs)
```

```
[<div class="wrap_song_info">
<div class="ellipsis rank01"><span>
<a href="javascript:melon.play.playSong('67220101',31700955);" title="Better Apart (Feat. Dresage) 재생">Be
tter Apart (Feat. Dresage)</a>
</span></div><br/>
<div class="ellipsis rank02">
<a href="javascript:melon.link.goArtistDetail('1066619');" title="Jai Wolf - 페이지 이동">Jai Wolf</a><span
class="checkEllipsis" style="display:none"><a href="javascript:melon.link.goArtistDetail('1066619');" title
="Jai Wolf - 페이지 이동">Jai Wolf</a></span>
</div>
</div>, <div class="wrap_song_info">
<div class="ellipsis rank01"><span>
<a href="javascript:melon.play.playSong('67220101',106860677);" title="Sharpest Weapon (Feat. Cameron Walke
r) 재생">Sharpest Weapon (Feat. Cameron Walker)</a>
</span></div><br/>
<div class="ellipsis rank02">
<a href="javascript:melon.link.goArtistDetail('2122758');" title="tyDi & MORTEN - 페이지 이동">tyDi &
p; MORTEN</a><span class="checkEllipsis" style="display:none"><a href="javascript:melon.link.goArtistDetail
('2122758');" title="tyDi & MORTEN - 페이지 이동">tyDi & MORTEN</a></span>
</div>
</div>, <div class="wrap_song_info">
<div class="ellipsis rank01"><span>
<a href="javascript:melon.play.playSong('67220101',31649773);" title="Better With You (feat. Iselin) 재생">
Better With You (feat. Iselin)</a>
</span></div><br/>
<div class="ellipsis rank02">
<a href="javascript:melon.link.goArtistDetail('771493');" title="3Lau - 페이지 이동">3Lau</a>, <a href="jav
ascript:melon.link.goArtistDetail('982969');" title="Justin Caruso - 페이지 이동">Justin Caruso</a><span cl
ass="checkEllipsis" style="display:none"><a href="javascript:melon.link.goArtistDetail('771493');" title="3
Lau - 페이지 이동">3Lau</a>, <a href="javascript:melon.link.goArtistDetail('982969');" title="Justin Caruso
- 페이지 이동">Justin Caruso</a></span>
</div>
```



# 실습 1. 멜론 플레이리스트

# select 사용

```
songs = soup.select("td:nth-child(5) > div > div")
```

※ 엄청 중요



'개발자 도구'

F12

or

cmd + alt + i

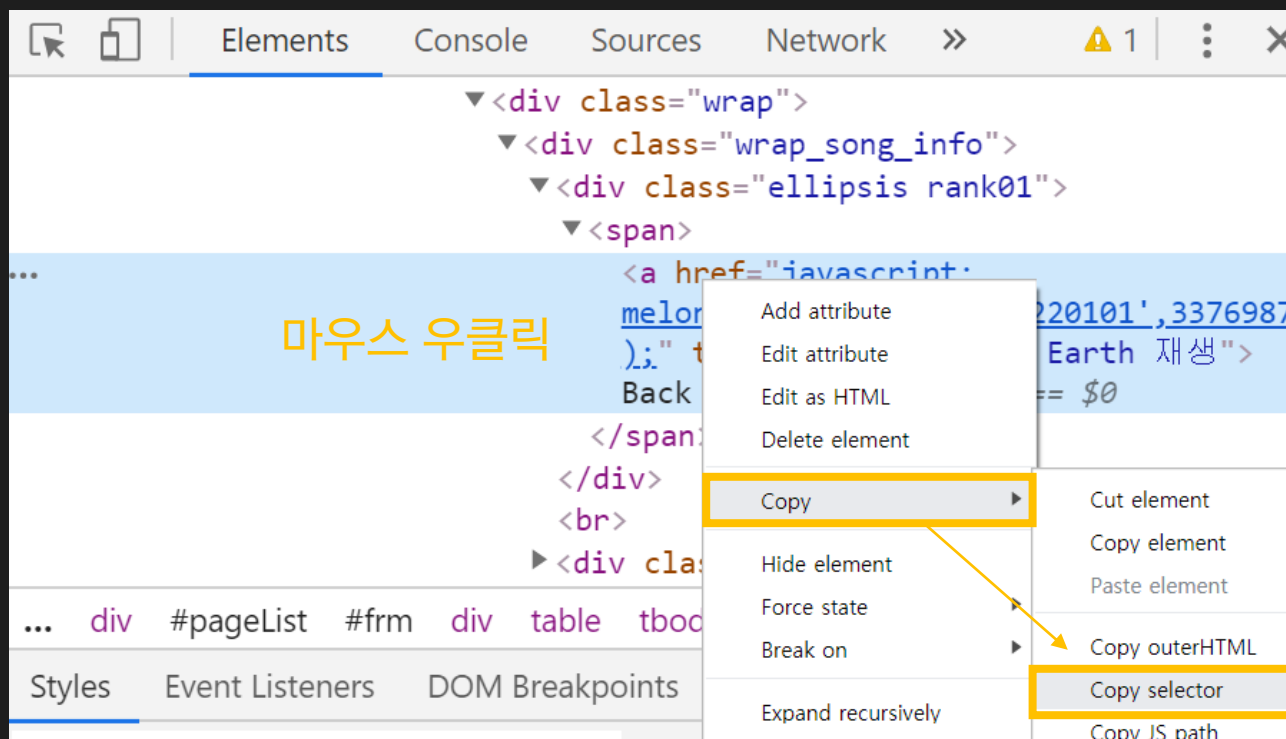


누른 후

원하는 부분 마우스 대보기

# Css selector

In CSS, selectors are patterns used to select the element(s) you want to style.



#frm > div > table > tbody > tr:nth-child(1) > td:nth-child(5) > div > div > div.ellipsis.rank01 > span > a



# Css Selector 정리

*	· 모든 요소	:root	· Document의 루트 요소
태그명	· 태그명이 일치하는 요소(예)	:nth-child(n)	· 동위 요소 중 n 번째 위치한 요소
.클래스명	· 클래스 속성 값이 일치하는 요소		
#id명	· id 속성의 값이 일치하는 요소	:nth-last-child(n)	· 동위 요소 중 뒤에서 n 번째 위치한 요소

요소[att]	· 특정 속성을 가지는 요소를 선택
요소[att = 'val']	· att 속성의 값이 val인 요소, val값이 전체 일치해야 함

선택자, 선택자	· 열거된 복수의 선택자 (예) h1, h2
선택자 선택자	· 하위 계층의 후손 요소 (예) div h1
선택자 > 선택자	· 바로 아래 계층의 자식 요소 (예) div > h1

## 실습 1. 멜론 플레이리스트



수원



비똥찌



문째왕 은진띠

Q. 어떤 요소를 선택하는 것일까?

```
#frm > div > table > tbody > tr:nth-child(1) > td:nth-child(5) > div > div > div.ellipsis.rank01 > span > a
```

## 실습 1. 멜론 플레이리스트

#frm > div > table > tbody > tr:nth-child(1) > td:nth-child(5) > div > div > div.ellipsis.rank01 > span > a

```
▼<form id="frm" name="frm">
  ▼<div class="service_list_song d_song_list">
    <h3 class="none"></h3>
    <!-- 곡리스트 테이블 -->
    ▶<div class="wrap_btn_tb top">...</div>
    ▼<table border="1" style="width:100%">
      ▶<caption>...</caption>
      ▶<colgroup>...</colgroup>
      ▶<thead>...</thead>
      ▼<tbody>
        ▼<tr>
          ▶<td>...</td>
          ▶<td>...</td>
          ▶<td>...</td>
          ▶<td>...</td>
          ▼<td>
            ▼<div class="wrap">
              ▼<div class="wrap_song_info">
                ▼<div class="ellipsis rank01">
                  ▼<span>
                    <a href="javascript:melon.play.playSong('67220101',3376987);"
                      title="Back Down To Earth 재생">Back Down To Earth</a>
                  </span>
```

## 실습 1. 멜론 플레이리스트


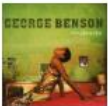



\*원하는 내용 가져오기

# select 사용

```
songs = soup.select("td:nth-child(5) > div > div.wrap_song_info")
print(songs)
```

수록곡 (39)

▶ 듣기 + 담기 다운로드 FLAC 선물

번호	곡정보	앨범
1	 Back Down To Earth Michelle Shaprow	Purple
2	 <b>div.wrap_song_info</b> Color #666666 Font 12px "맑은 고딕", "Malgun Gothic", 돋움, ... Margin 0px 0px 0px 12px	Irrepla
3	 Just You And Me (캐논 IXUS 광고 ... Zee Avi	Zee Av
4	 If I Die Young The Band Perry	The Ba
5	 A Step You Can't Take Back Keira Knightley	Begin

※ 공통점 찾기!

```
> <td>...</td>
> <td>...</td>
> <td>...</td>
> <td>...</td>
▼ <td>
  ▼ <div class="wrap">
    ▼ <div class="wrap_song_info">
      ▼ <div class="ellipsis rank01">
        ▼ <span>
          <a href="javascript:melon.play.playSong('67220101',3376987);"
            title="Back Down To Earth 재생">Back Down To Earth</a>
        </span>
```

Select!

```
[<div class="wrap_song_info">
  <div class="ellipsis rank01"><span>
    <a href="javascript:melon.play.playSong('67220101',3606165);" title="가슴아 심장
    아 재생">가슴아 심장아</a>
  </span></div><br/>
  <div class="ellipsis rank02">
    <a href="javascript:melon.link.goArtistDetail('556740');" title="BoM - 페이지 이
    동">BoM</a><span class="checkEllipsis" style="display:none"><a href="javascript:
    melon.link.goArtistDetail('556740');" title="BoM - 페이지 이동">BoM</a></span>
  </div>
</div>, <div class="wrap_song_info">
  <div class="ellipsis rank01"><span>
    <a href="javascript:melon.play.playSong('67220101',1898154);" title="그녀를 찾아
    주세요 재생">그녀를 찾아주세요</a>
```

## 실습 1. 멜론 플레이리스트

#반복해서 데려오기

```
for i,song in enumerate(songs, 1):
    title = song.select_one('div.ellipsis.rank01').text.strip()
    artist = song.select_one('div.ellipsis.rank02').text.strip()

    # 순위출력
    print(str(i) + '위')
    print('title : ' + title)
    print('artist : ' + artist)
    print('-----')
```

for i, A in enumerate(B, 시작 숫자):  
 띄어쓰기네칸반복해서 할 내용(C)

→ B에 있는 것들을 하나씩 A로 가져오면서 시작숫자부터 i로 번호를 먹이며 C를 수행한다.

```
1위
title : Better Apart (Feat. Dresage)
artist : Jai WolfJai Wolf
-----
2위
title : Sharpest Weapon (Feat. Cameron Walker)
artist : tyDi & MORTENTyDi & MORTEN
-----
3위
title : Better With You (feat. Iselin)
artist : 3Lau, Justin Caruso3Lau, Justin Caruso
-----
4위
title : Happy Now (BEAUZ Remix) (Feat. Elley Duhé)
artist : ZeddZedd
-----
5위
title : Here We Go Again (Feat. Louisa)
artist : Sigma, LouisaSigma, Louisa
-----
6위
title : Do It Yourself
artist : ILIRAILIRA
-----
7위
title : With You
artist : OysangOysang
-----
8위
title : Crown
artist : Camila Cabello, GreyCamila Cabello, Grey
-----
9위
title : Vertigo (Spitfya Remix)
artist : Rob Gasser, Laura BrehmRob Gasser, Laura Brehm
-----
10위
title : Play (ft. Mangoo)
```

반복!

```
x = ('apple', 'banana', 'cherry') [(0, 'apple'), (1, 'banana'), (2, 'cherry')]
```

## 실습 1. 멜론 플레이리스트

#반복해서 데려오기

```
for i,song in enumerate(songs, 1):
```

```
    title = song.select_one('div.ellipsis.rank01').text.strip()
    artist = song.select_one('div.ellipsis.rank02').text.strip()
```

# 순위출력

```
    print(str(i) + '위')
```

```
    print('title : ' + title)
```

```
    print('artist : ' + artist)
```

```
    print('-----')
```

`.select_one(선택자)` → 선택자에 해당하는 것 하나를 가져온다.

`.text.strip()` → 태그의 텍스트요소를 가져온 후 공백을 제거한다.

```
1위
title : Better Apart (Feat. Dresage)
artist : Jai WolfJai Wolf
-----
2위
title : Sharpest Weapon (Feat. Cameron Walker)
artist : tyDi & MORTENTyDi & MORTEN
-----
3위
title : Better With You (feat. Iselin)
artist : 3Lau, Justin Caruso3Lau, Justin Caruso
-----
4위
title : Happy Now (BEAUZ Remix) (Feat. Elley Duhé)
artist : ZeddZedd
-----
5위
title : Here We Go Again (Feat. Louisa)
artist : Sigma, LouisaSigma, Louisa
-----
6위
title : Do It Yourself
artist : ILIRAILIRA
-----
7위
title : With You
artist : OysangOysang
-----
8위
title : Crown
artist : Camila Cabello, GreyCamila Cabello, Grey
-----
9위
title : Vertigo (Spitfya Remix)
artist : Rob Gasser, Laura BrehmRob Gasser, Laura Brehm
-----
10위
title : Play (ft. Mangoo)
```

반복!



## 실습 1. 멜론 플레이리스트

```
# bs4 초기화
```

```
soup = BeautifulSoup(res, "html.parser")
```

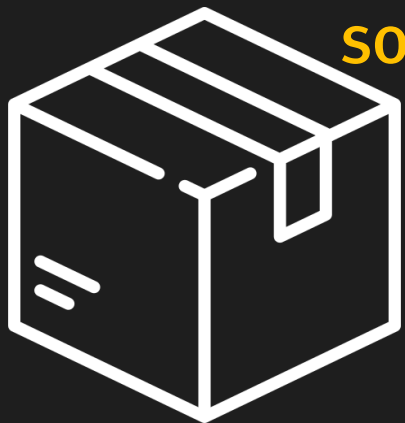
```
# select 사용
```

```
songs = soup.select("td:nth-child(5) > div > div.wrap_song_info")  
print(songs)
```

```
#반복해서 데려오기
```

```
for i,song in enumerate(songs, 1):
```

```
    title = song.select_one('div.ellipsis.rank01').text.strip()  
    artist = song.select_one('div.ellipsis.rank02').text.strip()
```



Html 문서 전체



원하는 태그들



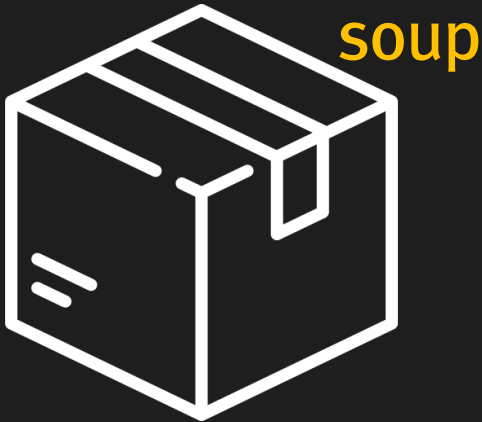
원하는 태그

artist / title



원하는 요소

# 실습 1. 멜론 플레이리스트



Html 문서 전체



원하는 태그들

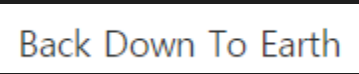
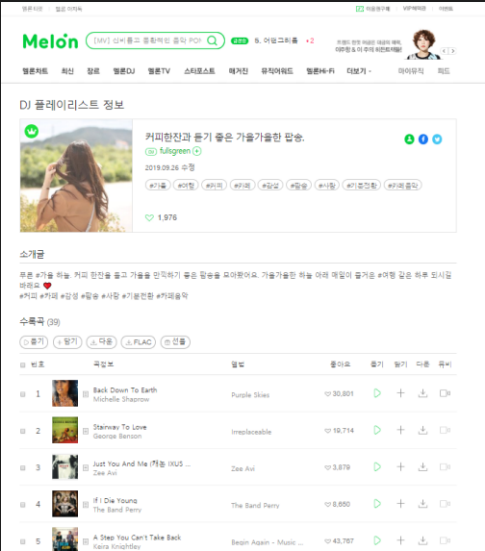


원하는 태그

artist / title



원하는 요소



## 환경 셋팅, 설치 및 이론

#반복해서 데려오기

```
for i,song in enumerate(songs, 1):  
    title = song.select_one('div.ellipsis.rank01').text.strip()  
    artist = song.select_one('div.ellipsis.rank02').text.strip()
```

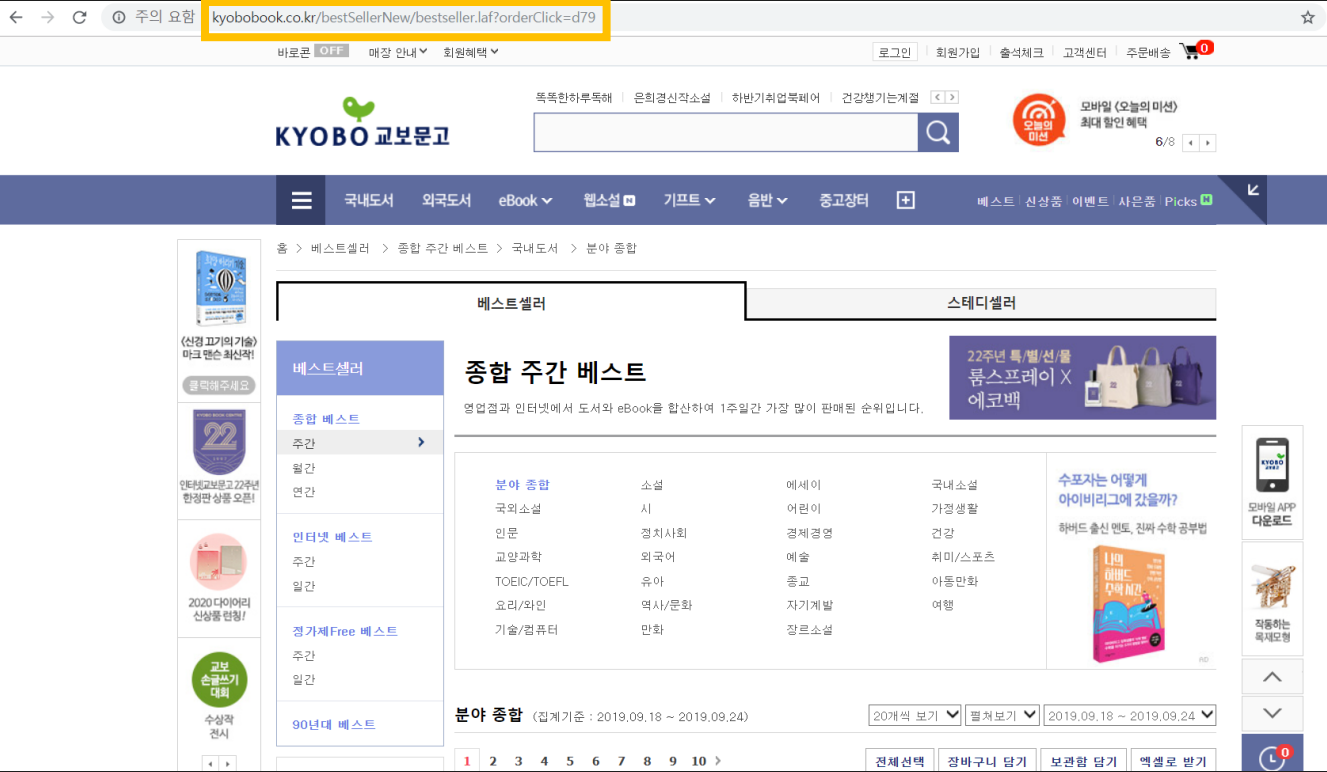
1. 이 부분이 왜 이렇게 되는지 이해하기
2. 지금까지 배운 내용 복습하기

다했으면 쉬는 타임!



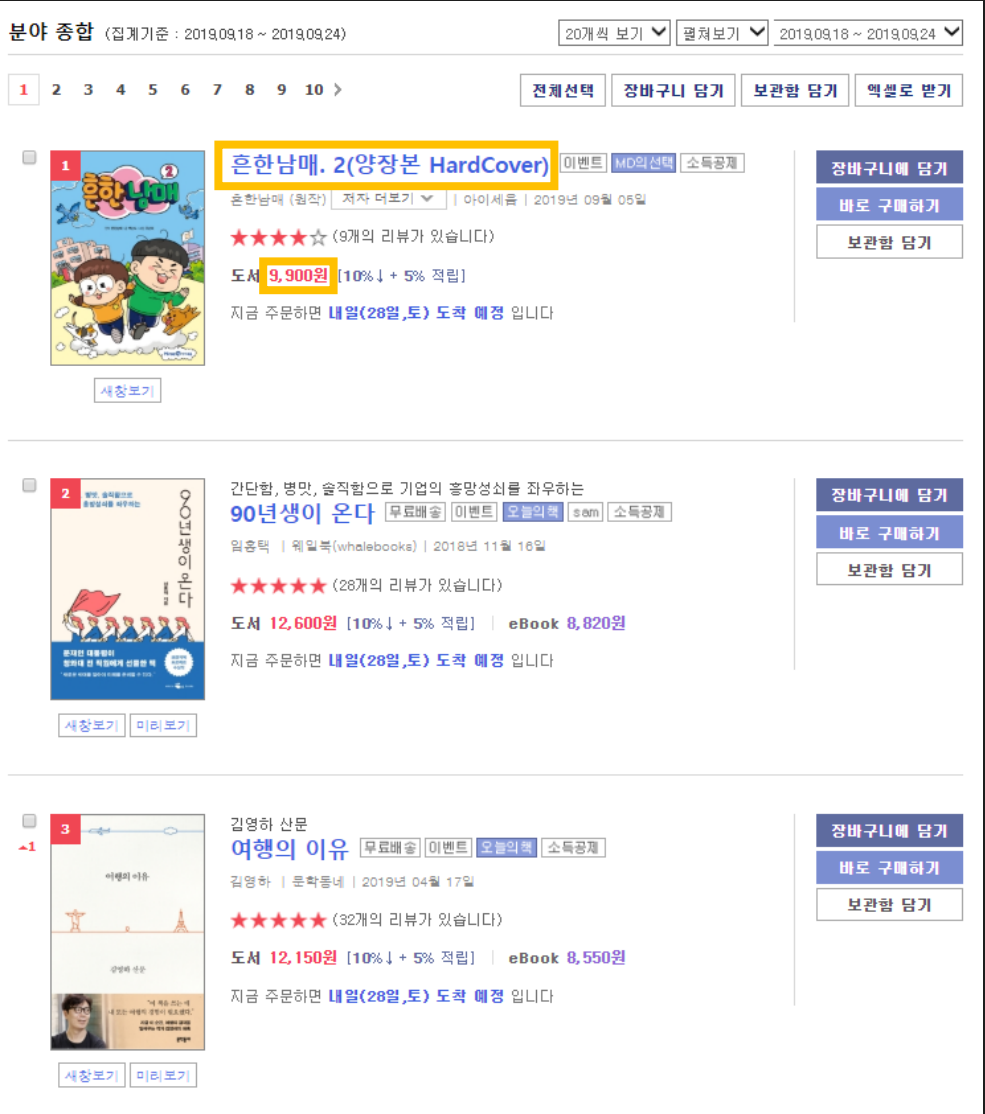
# 실습 2. 교보문고 책 순위

## 긐어오고 싶은 페이지의 URL 복사



http://www.kyobobook.co.kr/bestSellerNew/bestseller.laf?orderClick=d79

## 어떤 부분을 가져올 것인지 생각



## 실습 2. 교보문고 책 순위

새 파일 만들고 주석 달기

```
*1차_교보문고_순위.py - C:\Users\User\Desktop\크롤링강의\1차_교보문고_...
File Edit Format Run Options Window Help
# first_session
# 교보문고 순위 가져오기
|
```

라이브러리 가져오고 user 정보 셋팅하기

```
# first_session
# 교보문고 순위 가져오기

import urllib.parse as rep
import urllib.request as req
from bs4 import BeautifulSoup
from fake_useragent import UserAgent

# Header 정보 초기화
opener = req.build_opener()
# User-Agent 정보
opener.addheaders = [('User-agent', UserAgent().ie)]
# Header 정보 삽입
req.install_opener(opener)
```

## 실습 2. 교보문고 책 순위

### 웹페이지 전체 html 가져오기

```
# 교보문고 종합순위 URL
```

```
url = "http://www.kyobobook.co.kr/bestSellerNew/bestseller.laf?orderClick=d79"
```

```
# 요청 URL 확인
```

```
print(url)
```

```
# Request
```

```
res = req.urlopen(url)
```

```
print(res)
```

```
# bs4 초기화
```

```
soup = BeautifulSoup(res, "html.parser")
```

```
http://www.kyobobook.co.kr/bestSellerNew/bestseller.laf?orderClick=d79  
<http.client.HTTPResponse object at 0x0447ADF0>
```



<http://www.kyobobook.co.kr/bestSellerNew/bestseller.laf?orderClick=d79>

## 환경 셋팅, 설치 및 이론

진Aㅏ



쿼터 구



## 실습 2. 교보문고 책 순위

```
# select 사용
books = soup.select(" ? ")
print(books)
```

\*원하는 내용 가져오기

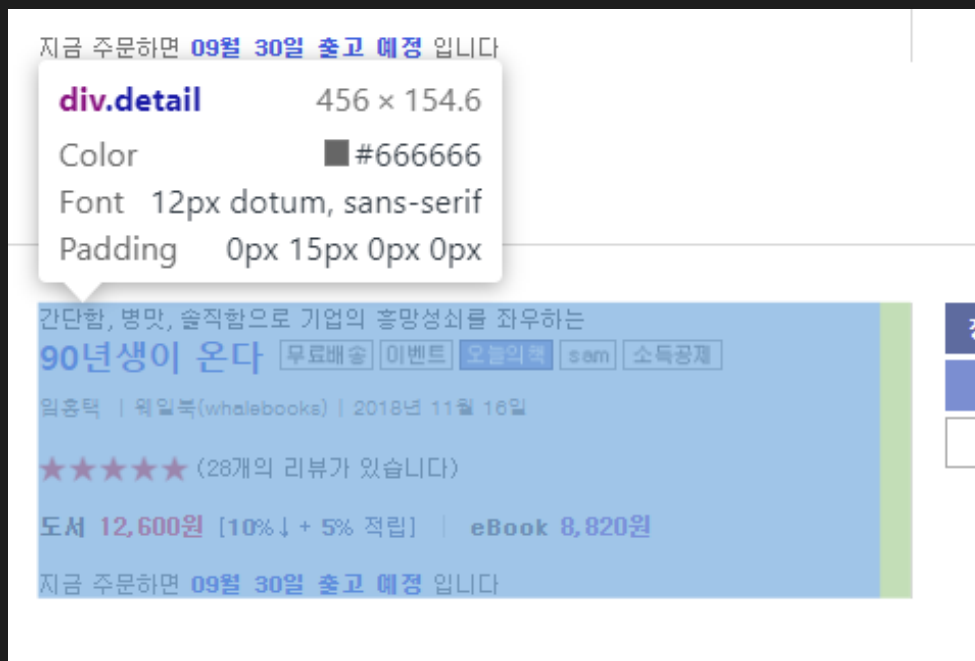
'개발자 도구'

F12  
or  
cmd + alt + i



누른 후

원하는 부분 마우스 대보기





## 실습 2. 교보문고 책 순위

```
# select 사용
books = soup.select("div.detail")
print(books)
```

\*원하는 내용 가져오기

'개발자 도구'

F12  
or  
cmd + alt + i



누른 후

원하는 부분 마우스 대보기



## 실습 2. 교보문고 책 순위

#반복해서 데려오기

```
for i, book in enumerate(books, 1):  
    title = book.select_one(' ? ').text.strip()  
    price = book.select_one(' ? ').text.strip()
```

# 순위출력

```
print(str(i) + '위')  
print('책제목 : ' + title)  
print('가격 : ' + price)  
print('-----')
```

```
▼<div class="detail"> == $0  
    <div class="subtitle">  
  
    </div>  
    ▶<div class="title">...</div>  
    ▶<div class="author">...</div>  
    ▶<div class="review">...</div>  
    ▶<div class="price">...</div>  
    ▶<div class="info">...</div>  
</div>
```

```
    </div>  
▼<div class="title">  
    ▼<a href="http://www.kyobobook.co.kr/product/detailViewKor.laf?  
    mallGb=KOR&ejkGb=KOR&barcode=9791164132454">  
        <strong>흔한남매. 2(양장본 Hardcover)</strong> == $0  
    </a>
```

```
▼<div class="price"> == $0  
    "  
  
    도서  
    "  
    <strong class="book_price">9,900원</strong>
```

## 실습 2. 교보문고 책 순위

#반복해서 데려오기

```
for i,book in enumerate(books, 1):  
    title = book.select_one('div.title strong').text.strip()  
    price = book.select_one('div.price > strong').text.strip()
```

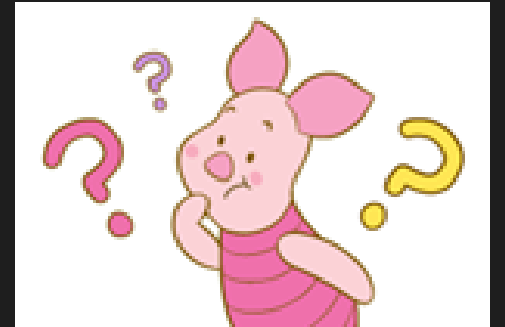
# 순위출력

```
print(str(i) + '위')  
print('책제목 : ' + title)  
print('가격 : ' + price)  
print('-----')
```

### Q. 둘의 차이는?



가이니



또글렛

```
</div>  
▼<div class="title">  
  ▼<a href="http://www.kyobobook.co.kr/product/detailViewKor.laf?mallGb=KOR&ejkGb=KOR&barcode=9791164132454">  
    <strong>흔한남매. 2(양장본 HardCover)</strong> == $0  
  </a>
```

```
▼<div class="price"> == $0  
  "  
  도서  
  "  
  <strong class="book_price">9,900원</strong>
```

## 실습 3. 구글 이미지 저장


← → ↻ google.com/search?q=드래곤길들이기&source=lnms&tbm=isch&sa=X&ved=0ahUKEwiku6zr5fPkAhUMyosBHc55B5IQ\_AUIEigB&biw=1536&bih=754

Google 드래곤길들이기


📷 🔊 🔍

🔍 All 🖼️ Images 📺 Videos 📍 Maps 📰 News ⋮ More Settings Tools


투스리스 theqoo youtube train 영화 히컵 how to train your dragon 드림웍스 장난감




드래곤 길들이기 3 메인 예고편 - YouTube  
youtube.com




드래곤 길들이기 3> 리브  
brunch.co.kr




드래곤 길들이기 3 후기...  
creativesilver.tistory.com




드래곤 길들이기 3 - 끝이 아닌 또 다른 시작이 될 ...  
m.post.naver.com




드래곤 길들이기 2 - 나...  
namu.wiki




드래곤 길들이기에 숨겨진 "불편한 진실"  
blog.naver.com




감성의 피날레!..드래곤 ...  
news.chosun.com



How To Train Your Dragon ...  
yes24.com



드래곤 길들이기 3 1차 예고편 - YouTube  
youtube.com



여친 보자마자 첫눈에 반한 투슬리  
insight.co.kr

## 실습 3. 구글 이미지 저장

새 파일 만들고 주석 달기

1차\_구글\_이미지다운.py - C:\Users\User\Desktop\크롤링강의

File Edit Format Run Options Window Help

```
# first_session  
# 구글 검색 후 이미지 다운로드
```

라이브러리 가져오고 user 정보 셋팅하기

```
import os  
import urllib.parse as rep  
import urllib.request as req  
from bs4 import BeautifulSoup  
from fake_useragent import UserAgent  
  
# Header 정보 초기화  
opener = req.build_opener()  
# User-Agent 정보  
opener.addheaders = [('User-agent', UserAgent().ie)]  
# Header 정보 삽입  
req.install_opener(opener)
```

## 실습 3. 구글 이미지 저장

### 검색어를 포함한 URL 셋팅

# 구글 이미지 기본 URL(크롬 개발자 도구)

```
base1 = "https://www.google.com/search?q="
```

```
base2 = "&source=lnms&tbm=isch&sa=X&ved=0ahUKEwiku6zr5fPkAhUMyosBHc55B5IQ_AUIEigB&biw=1536&bih=754"
```

# 검색어

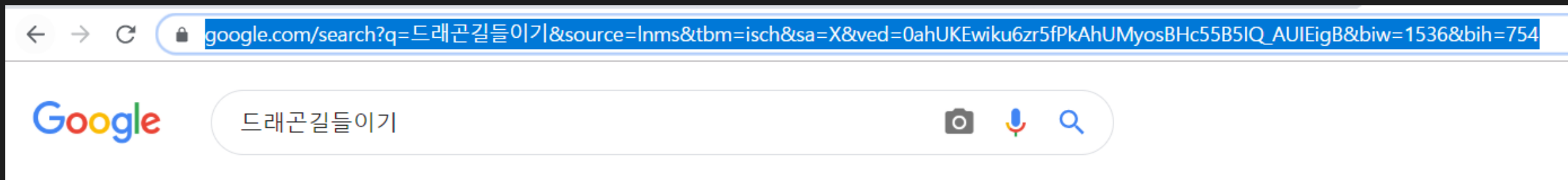
```
quote = rep.quote_plus("드래곤 길들이기")
```

# URL 완성

```
url = base1 + quote + base2
```

# Request

```
res = req.urlopen(url)
```



<https://www.google.com/search?q=>

base1

+

드래곤길들이기

quote

+

&source=lnms&tbm=isch&sa=X&ved=0ahUKEwiku6zr5fPkAhUMyosBHc55B5IQ\_AUIEigB&biw=1536&bih=754

base2



## 실습 3. 구글 이미지 저장

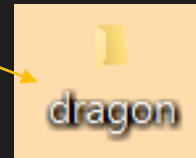
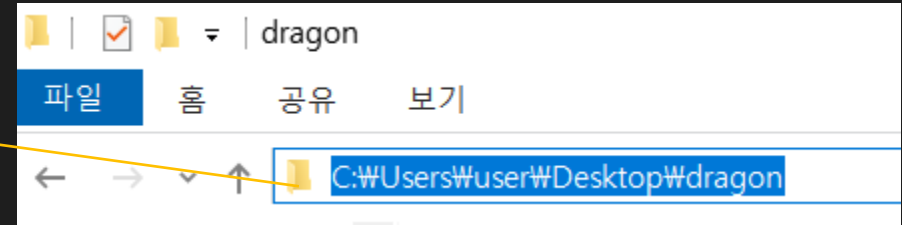
### 폴더 만들기, 이미지 저장 경로 설정

```
# 이미지 저장 경로)
savePath = "C:\\Users\\user\\Desktop\\dragon\\" # C:\\imagedown\\

# 폴더 생성 예외처리 (문제 발생 시 프로그램 종료)
try:
    # 기존 폴더가 있는지 체크
    if not (os.path.isdir(savePath)):
        # 없으면 폴더 생성
        os.makedirs(os.path.join(savePath))
except OSError as e:
    # 에러 내용
    print("folder creation failed!")
    print("folder name : {}".format(e.filename))

    # 런타임 에러 raise
    raise RuntimeError('System Exit!')
else:
    # 폴더 생성 정상일 경우 실행
    print('folder is created!')
```

폴더 자동 생성 부분



헛갈리고 어렵다 느끼는 사람은  
수동으로 폴더 만든 후 경로만 저장

### 실습 3. 구글 이미지 저장

```
# bs4 초기화
soup = BeautifulSoup(res, "html.parser")
```

```
# select 사용
img_list = soup.select("img")
print(img_list)
```

Image 요소 select

```
for i, v in enumerate(img_list):
```

```
    try:
        # 저장 파일명 및 경로
        fullFileName = os.path.join(savePath, savePath + str(i) + '.png')
        # 다운로드 요청(URL, 저장경로)
        req.urlretrieve(v['src'], fullFileName)
    except:
        #파일이 없으면 넘어감
        pass
    else:
        # 다운로드 완료 시 출력
        print("download succeeded!")
```

`os.path.join(경로(A), 경로 + 파일이름(B))`

→ A에 B이름으로 경로를 지정한다

`req.urlretrieve(A[B], C)`

→ A의 B속성을 C에 다운한다

```
try:
    시도할 내용(A)
except:
    pass
else:
    시도 성공 후 명령(B)
```

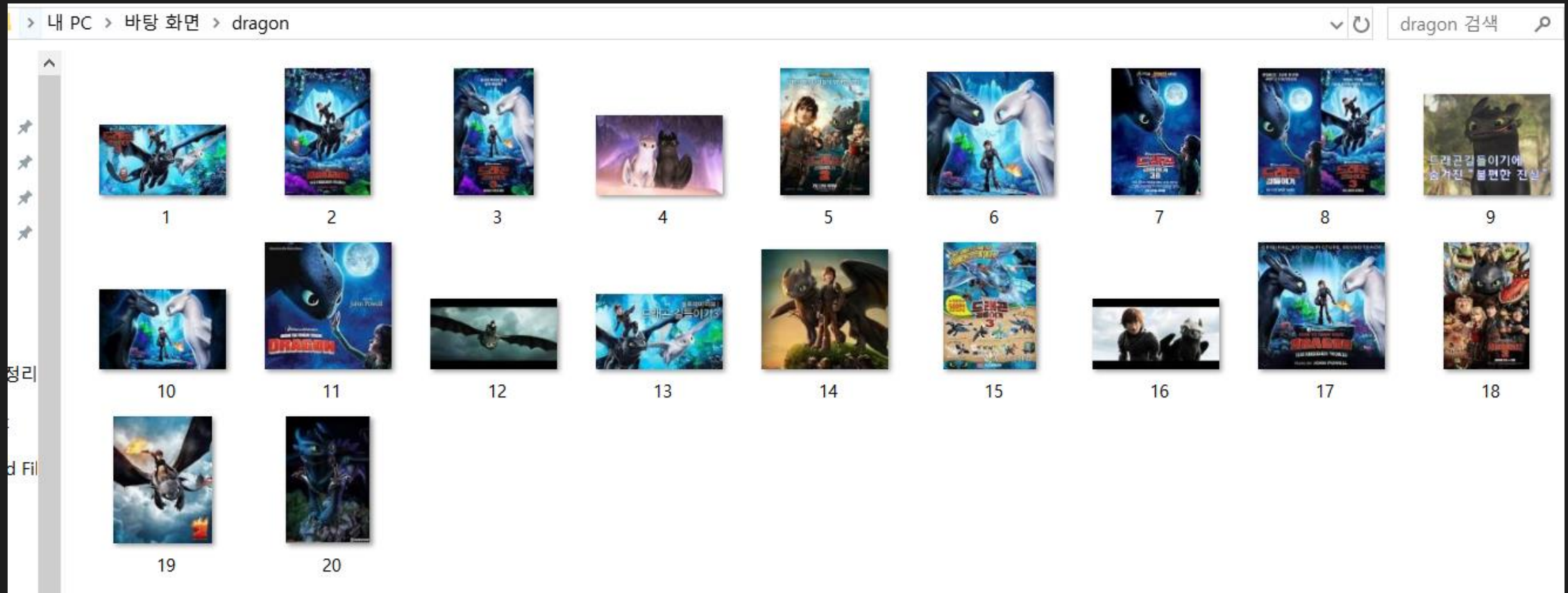
→ A를 시도하는데 실패 시 넘어간다  
A를 성공하면 B를 수행한다.



## 실습 3. 구글 이미지 저장

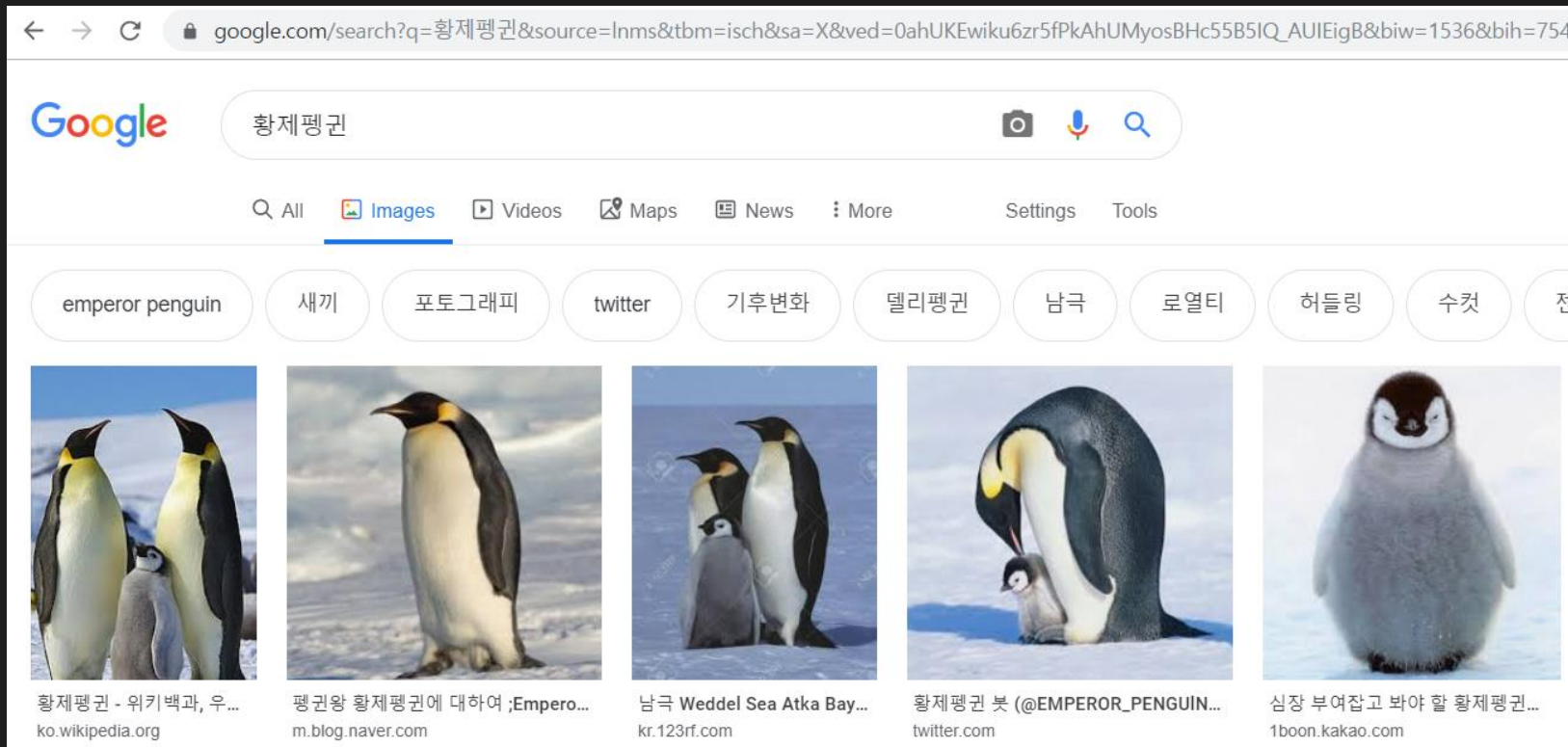


download succeeded!  
download succeeded!  
download succeeded!  
download succeeded!  
download succeeded!



## 실습 3. 구글 이미지 저장

다른 검색어로 해보기!



# 이미지 저장 경로)

savePath = "C:\\Users\\user\\Desktop\\penguin\\"

# 검색어

quote = rep.quote\_plus("펭귄")



수고 하셨습니다