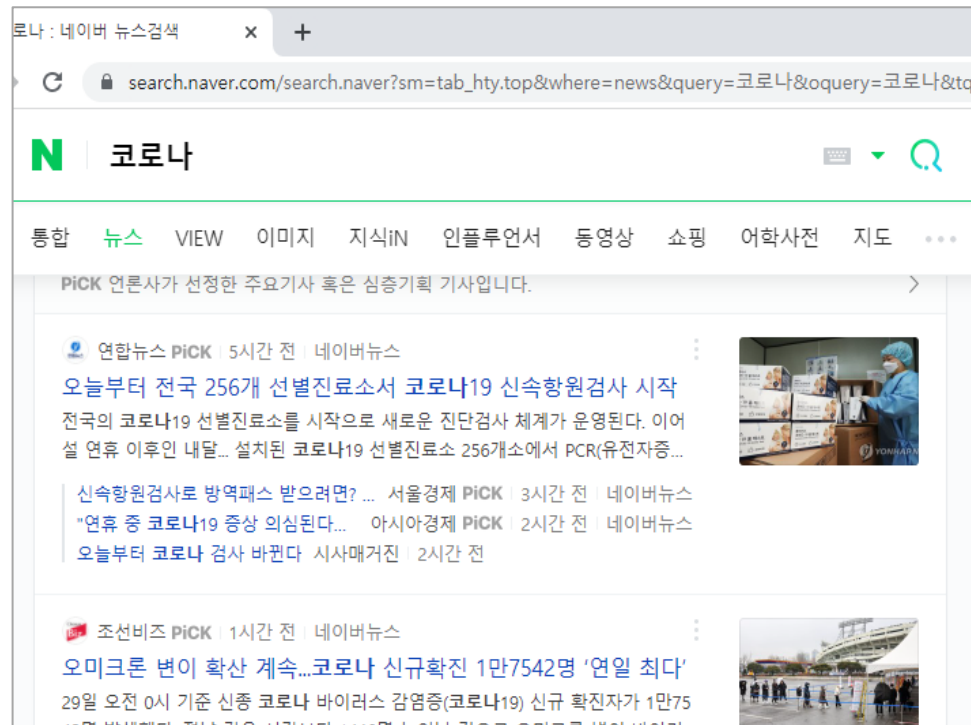


# Web crawling

# Web crawling

- Web crawling is the activity of collecting certain information within a web page in an automated way.



	A	B	C	D	E	F	G	H
1	크롤링 날짜	제목	링크					
2	2022-01-29	오늘부터 전국 256개 선별진료소서 코로나19 신속항원검사 시작	<a href="http://yna.kr/AKR20220128185500530?did=1195m">http://yna.kr/AKR20220128185500530?did=1195m</a>					
3	2022-01-29	오미크론 변이 확산 계속...코로나 신규확진 1만7542명 '연일 최대'	<a href="https://biz.chosun.com/topics/topics_social/2022/01/29/MWCQWPQ2G">https://biz.chosun.com/topics/topics_social/2022/01/29/MWCQWPQ2G</a>					
4	2022-01-29	유재석, 코로나19 음성... "스케줄 정상 진행"	<a href="http://www.newsis.com/view/?id=NISX20220129_0001742456&amp;clD=106">http://www.newsis.com/view/?id=NISX20220129_0001742456&amp;clD=106</a>					
5	2022-01-29	설 연휴 첫 날, 코로나19 신규확진 1만7542명... "또" 역대 최대	<a href="http://www.fnnews.com/news/202201291002240932">http://www.fnnews.com/news/202201291002240932</a>					
6	2022-01-29	코로나19 국내유입 2년...누적 확진자 80만명 넘겨	<a href="https://news.imaeil.com/page/view/2022012910003117060">https://news.imaeil.com/page/view/2022012910003117060</a>					
7	2022-01-29	코로나 검사, 오늘부터 바뀐다...PCR-신속항원검사 병행	<a href="http://www.newsis.com/view/?id=NISX20220128_0001742456&amp;clD=102">http://www.newsis.com/view/?id=NISX20220128_0001742456&amp;clD=102</a>					
8	2022-01-29	[브리핑] 정부 "3일부터 지정 동네병원서 코로나 검사-처방-재택치료"	<a href="https://news.sbs.co.kr/news/endPage.do?news_id=N1006622078&amp;pln1">https://news.sbs.co.kr/news/endPage.do?news_id=N1006622078&amp;pln1</a>					
9	2022-01-29	직접 코로나19 검사	<a href="http://yna.kr/PYH20220129016700013?did=1196m">http://yna.kr/PYH20220129016700013?did=1196m</a>					
10	2022-01-29	2월3일부터 동네병원에서도 코로나19 검사치료 받는다	<a href="https://www.khan.co.kr/national/health-welfare/article/202201281104">https://www.khan.co.kr/national/health-welfare/article/202201281104</a>					
11	2022-01-29	코로나 증상에도 제주도 여행한 '강남모녀'...역대 손백소 승소	<a href="http://news.mt.co.kr/mtview.php?no=2022012817353794023">http://news.mt.co.kr/mtview.php?no=2022012817353794023</a>					
12	2022-01-29	3일부터 전국 호흡기전담클리닉서 코로나 검사치료까지(종합)	<a href="http://yna.kr/AKR20220128074151530?did=1195m">http://yna.kr/AKR20220128074151530?did=1195m</a>					
13	2022-01-29	'부스터샷' 맞은 조세호, 코로나 확진...유재석 또 PCR 검사	<a href="https://www.chosun.com/national/welfare-medical/2022/01/28/H31PC">https://www.chosun.com/national/welfare-medical/2022/01/28/H31PC</a>					

# Web crawl procedure

- Web crawling leverages requests, the BeautifulSoup selenium package, and Chrome developer tools
- Requires basic knowledge of Hyper Text Markup Language (HTML) tags and Cascading Style Sheets (CSS) and Javascript

1) HTML 문서 요청 및 추출하기	requests 또는 selenium 패키지 활용하여 HTML 문서 추출
2) HTML 문서 분석하기	크롬 개발자 도구, HTML 태그, CSS 선택자를 활용해 분석
3) 원하는 데이터 추출하기	requests, selenium, BeautifulSoup 활용, 원하는 위치의 데이터 추출
4) 찾은 데이터 가공 및 저장하기	추출한 데이터를 가공하고 엑셀 또는 텍스트 파일로 저장

# Crawling HTML source

- **Installing the requests module**

```
pip install requests
```

- **Crawling HTML source**

- **GET method**

- When requesting resources such as HTML pages from a specific site, parameters are added after the URL.
- GET is mainly used when a web browser requests data from a web server

# Crawling HTML source

```
import requests                                # requests 모듈 불러오기
res = requests.get("http://www.naver.com", verify = False)        # get으로 네이버 홈페이지 가져오기
print(res.content)                                출력결과                # 내용 출력
```

```
b'<!doctype html>                <html lang="ko" data-dark="false"> <head> <meta charset="utf-8"> <title>NAVER</title> <
meta http-equiv="X-UA-Compatible" content="IE=edge"> <meta name="viewport" content="width=1190"> <meta name="apple-mobile-web-app-ti
tle" content="NAVER"/> <meta name="robots" content="index,nofollow"/> <meta name="description" content="#xe4#xa4#x9d#xb4#xeb
#xb2#x84 #xe9#xa9#x94#xec#x9d#xb8#xec#x97#x90#xec#x84#x9c #xeb#xb#xa4#xec#x96#x91#xed#x95#x9c #xec#xa0#x95#xeb#xb3#xb4#xec#x99#x80
#xec#x9c#xa0#xec#x9a#xa9#xed#x95#x9c #xec#xb#xa8#xed#x85#x90#xec#xb8#xa0#xeb#xa5#x8c #xeb#xa7#x8c#xeb#x82#x98 #xeb#xb3#xb4#xec#x84
#xb8#xec#x9a#x94"/> <meta property="og:title" content="#xe4#xa4#xec#x9d#xb4#xeb#xb2#x84"> <meta property="og:url" content="http
s://www.naver.com/"> <meta property="og:image" content="https://s.pstatic.net/static/www/mobile/edit/2016/0705/mobile_212852414260.p
ng"> <meta property="og:description" content="#xe4#xa4#xec#x9d#xb4#xeb#xb2#x84 #xeb#xa9#x94#xec#x9d#xb8#xec#x97#x90#xec#x84#x9c
#xeb#xb#xa4#xec#x96#x91#xed#x95#x9c #xec#xa0#x95#xeb#xb3#xb4#xec#x99#x80 #xec#x9c#xa0#xec#x9a#xa9#xed#x95#x9c #xec#xb#xa8#xed#x85
#x90#xec#xb8#xa0#xeb#xa5#x8c #xeb#xa7#x8c#xeb#x82#x98 #xeb#xb3#xb4#xec#x84#xb8#xec#x9a#x94"/> <meta name="twitter:card" content="sum
mary"> <meta name="twitter:title" content=""> <meta name="twitter:url" content="https://www.naver.com/"> <meta name="twitter:image"
content="https://s.pstatic.net/static/www/mobile/edit/2016/0705/mobile_212852414260.png"> <meta name="twitter:description" content
="#xe4#xa4#xec#x9d#xb4#xeb#xb2#x84 #xeb#xa9#x94#xec#x9d#xb8#xec#x97#x90#xec#x84#x9c #xeb#xb#xa4#xec#x96#x91#xed#x95#x9c #xec#xa
0#x95#xeb#xb3#xb4#xec#x99#x80 #xec#x9c#xa0#xec#x9a#xa9#xed#x95#x9c #xec#xb#xa8#xed#x85#x90#xec#xb8#xa0#xeb#xa5#x8c #xeb#xa7#x8c#xeb
```

# Crawling HTML source

<https://finance.naver.com/item/sise.nhn?code=005930>

```
parameters = {"code" : "005930"}
res = requests.get("https://finance.naver.com/item/sise.nhn", params = parameters, verify = False)
print (res.url)
print (res.text)
```

출력결과

```
<meta property="og:title" content="네이버 금융"/>
<meta property="og:image" content="https://ssl.pstatic.net/static/m/stock/im/2016/08/og_stock-200.png"/>
<meta property="og:url" content="https://finance.naver.com"/>
<meta property="og:description" content="국내 해외 증시 지수, 시장지표, 펀드, 뉴스, 증권사 리서치 등 제공"/>
```

```
<meta property="og:type" content="article"/>
<meta property="og:article:thumbnailUrl" content=""/>
<meta property="og:article:author" content="네이버금융"/>
<meta property="og:article:author:url" content="http://FINANCE.NAVER.COM"/>
```

```
<link rel='stylesheet' type='text/css' href='https://ssl.pstatic.net/imgstock/static.pc/20210803162128/css/finance_header.css'>
```

# Crawling HTML source

## ➤ **POST method**

- When requesting resources such as HTML pages from a specific site, parameters are not added to the URL.
- Request by putting it in a separate item, request body
- Since parameters are not exposed in the URL, it is relatively advantageous for security
- `<form>` When receiving information from users using tags, or when the amount of data to be transmitted to the server is large, Using the POST method
- The code below is the code that gets the Samsung Electronics market, and the execution result is the same as `requests.get()`.

# BeautifulSoup

- **Converting HTML Source to Data**

- **What is BeautifulSoup?**

- Python package to convert complex HTML documents into structured data types
    - The process of converting HTML to data is called parsing, which refers to the conversion of Python into a form accessible through code.
    - Import the web page with the requests module, parse it with the BeautifulSoup module, and extract the desired data

```
pip install bs4
```



# BeautifulSoup

- Use HTML tags to find the attribute you're looking for
  - Use BeautifulSoup's find() and find\_all() functions to find the desired tag in an html page
- Parsing (transforming) HTML sources into data for analysis
  - Load the BeautifulSoup module and enter html code and text into html\_txt variables to create data for analysis
  - Use the BeautifulSoup() function to parse (transform) html sources into data for analysis
  - If you parse soup = BeautifulSoup (html\_txt,"html.parser") code, the result is stored in the soup variable.

```
from bs4 import BeautifulSoup
```

```
# BeautifulSoup 모듈 불러오기
```

```
# html_txt 변수에 html 코드와 텍스트 입력
```

```
html_txt = "<p class = 'wether' id = 'tw'>오늘의 날씨</p> <h1> 한때 소나기가 내리겠습니다. </h1>"
```

```
soup = BeautifulSoup(html_txt , "html.parser") # 파싱하기, 분석 가능한 데이터 형태로 변환
```

# BeautifulSoup

## ➤ Find the desired property with the find() function

- You can use various functions in soup, and the function to find tags is find()
- print (tag) shows all attribute values in the p tag found by the find() function.

```
tag = soup.find("p")
print (tag)
print(tag.name)
print( tag.attrs )
print(tag.attrs["class"])
print(tag.attrs["id"])
print(tag.text)
```

```
# p 태그를 찾아 tag에 저장
# tag 출력
# tag 명 출력
# tag 속성 출력
# tag 속성 중 class만 출력
# tag 속성 중 id만 출력
# tag 내 텍스트 출력
```

```
<p class="wether" id="tw">오늘의 날씨</p>
p
{'class': ['wether'], 'id': 'tw'}
['wether']
tw
오늘의 날씨
```

# BeautifulSoup

- select and select\_one function

```
from bs4 import BeautifulSoup                                # BeautifulSoup 모듈 불러오기
html_txt = """                                              # html 소스 생성
<html>
<head><title>BS page</title></head>
<body>

<h1 class="portal_cls">검색포털</h1>
<p>
<a href="http://www.daum.net">다음 바로가기</a><br>
<a href="http://www.naver.com">네이버 바로가기</a>
</p>
<a href="http://www.google.com" class="alink_cls">구글</a>
<a target="_테스트_">테스트</a>
<p class="footage_cls" id="company"> 2021, ABC Company </p>
<p class="footage_cls" id="addr">Korea</p>
</body>
</html>
"""

soup = BeautifulSoup(html_txt , "html.parser")# 파싱하기, 분석 가능한 데이터 형태로 변환
```

# BeautifulSoup

## ➤ Find a specific tag

- `<h1>` Code that finds tags as `select_one()` and outputs only text
- Can be used similarly to the `find()` function
- `select_one()` gets only one tag that meets the condition

```
tag = soup.select_one("h1")  
print (tag.text)
```

```
# <h1> </h1> 태그 추출  
# tag 안에 있는 텍스트만 출력
```

[검색포털](#)

- The following example `<h1>` uses the tag `select()` function
- If you extract the tag with the `select()` function, it will be returned as list-type data, so you need to use loops to get the data one by one.

```
tag_list = soup.select("h1")  
for tag in tag_list:  
    print (tag.text)
```

```
# <h1> ~ </h1> 태그 모두 찾기  
# tag_list 내 데이터를 하나씩 읽어오기  
# 텍스트만 출력
```

[검색포털](#)

# BeautifulSoup

## ➤ Find child/child tags

- When expressing child tags, express them with > symbols
- The `soup.select("body p > a")` code <body> means to extract all the a tags under the p tag of
- When expressing child tags, express them as white space
- e.g. the `soup.select("body p a")` code extracts the A tag under the <body> p tag, and extracts all the tags under that

```
tag_list = soup.select("body p > a")          # <body><p> 태그 하위의 <a> 태그 추출
for tag in tag_list:                          # tag_list 내 데이터를 하나씩 읽어오기
    print (tag.text)                          # 텍스트만 출력
```

[다음 바로가기](#)  
[네이버 바로가기](#)

# BeautifulSoup

## ➤ Finding Tags with CSS Classes

```
tag_list = soup.select(".footage_cls")
for tag in tag_list:
    print (tag.text)
```

```
# footage_cls 클래스 추출
# tag_list 내 데이터를 하나씩 읽어오기
# 텍스트만 출력
```

----- 출력결과 -----

2021, ABC Company  
Korea

## ➤ Find tags by CSS ID

```
tag_list = soup.select("#company")
for tag in tag_list:
    print (tag.text)
```

```
# id가 company인 태그 추출
# tag_list 내 데이터를 하나씩 읽어오기
# 텍스트만 출력
```

----- 출력결과 -----

2021, ABC Company

# BeautifulSoup

## ➤ 태그 속성(attribute) 값을 이용해서 태그 찾기

```
tag_list = soup.select("a[href]")  
for tag in tag_list:  
    print(tag.text, tag.attrs["href"])
```

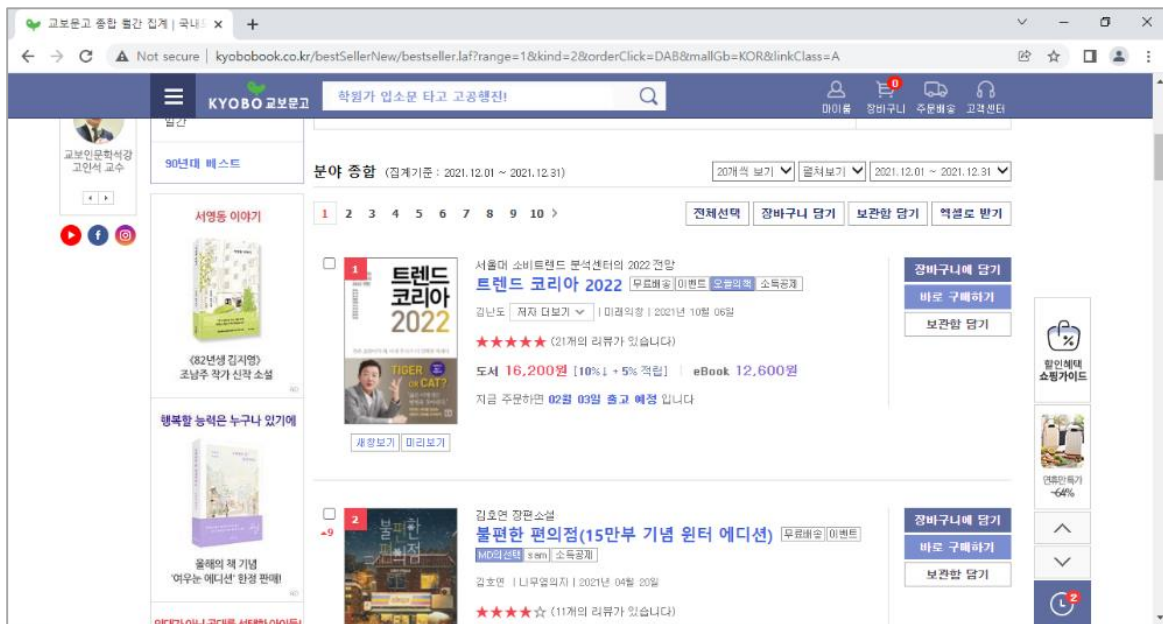
# a[href] 추출  
# tag\_list 내 데이터를 하나씩 읽어오기  
# 텍스트와 url 출력

다음 바로가기 <http://www.daum.net>  
네이버 바로가기 <http://www.naver.com>  
구글 <http://www.google.com>

# [실습] 서점 베스트셀러 정보 가져오기

## ● Step 01. 수집할 데이터 확인하기

- 교보문고 베스트셀러 목록이 있는 웹 페이지 주소를 확인하고, 도서, 저자, 출판사, 가격 데이터가 어느 위치에 있는지 대략적으로 확인
- 크롬으로 웹 페이지 접속
  - 크롬으로 교보문고 베스트셀러 웹 페이지인 "http://www.kyobobook.co.kr/bestSellerNew/bestseller.laf"에 접속한 후, 왼쪽 메뉴바에서 [월간]을 선택
  - 한 페이지에 종합 월간 베스트셀러 목록이 20개씩 보여지는 것을 확인
  - 정확한 데이터를 가져오기 위해서는 어떤 웹 페이지에 어떤 데이터가 있는지 확인





# [실습] 서점 베스트셀러 정보 가져오기

## ● Step 01. 수집할 데이터 확인하기

### ➤ 데이터 위치 확인

- 도서제목을 누르면 도서별 상세 페이지가 보이는데 도서명, 저자, 가격 등이 일관성 있게 배치되어 있음

트렌드 코리아 2022 | 김난도 | x +

Not secure | kyobobook.co.kr/product/detailViewKor.laf?mallGb=KOR&ejkGb=KOR&barcode=9788959897001

홈 > 국내도서 > 경제/경영 > 경제일반 > 경제전망

하루 10분! 문해력 충전타임

클릭해주세요

이웃권 sam

설연휴 특별기획 sam이웃권 증정!

2022년 일인년 신년기프트카드

교보인문기행 시인 윤동주

트렌드 코리아 2022

서울대 소비트렌드 분석센터의 2022 전망

김난도, 전미영, 최지혜, 이향은, 이준영, 이수진, 서유현, 권정윤, 한다혜, 이혜원 지음

미래의창 | 2021년 10월 06일 출간

9.4 (리뷰 250개) [블로그 리뷰쓰기](#)

국내도서 주간베스트 12위 | 경제/경영 주간베스트 2위

주요 일간지 북섹션 추천도서

정가: 18,000원

판매가: 16,200원 [10%↓ 1,800원 할인]

혜택: [기본적립] 900원 적립 [5% 적립]

[추가적립] 5만원 이상 구매 시 2,000원 추가적립 [안내](#)

[회원혜택] 회원 등급 별, 3만원 이상 구매 시 2~4% 추가적립 [안내](#)

[리뷰적립] 리뷰 작성 시 e교환권 최대 300원 추가적립 [안내](#)

추가혜택: [포인트 안내](#) [도서소독공제 안내](#) [추가혜택 더보기](#)

이 상품의 이벤트 3건 더보기

· 여러분과 함께 만든 산책의 리스트

· 데일리바이트 X 비즈니스군

· 앞서가는 당신을 위한 뉴스레터

할인혜택 쇼핑가이드

연휴인 특가 ~64%

배송비: 무료 [배송비 안내](#)

배송일정: 서울특별시 종로구 세종대로 기준 [지역변경](#)

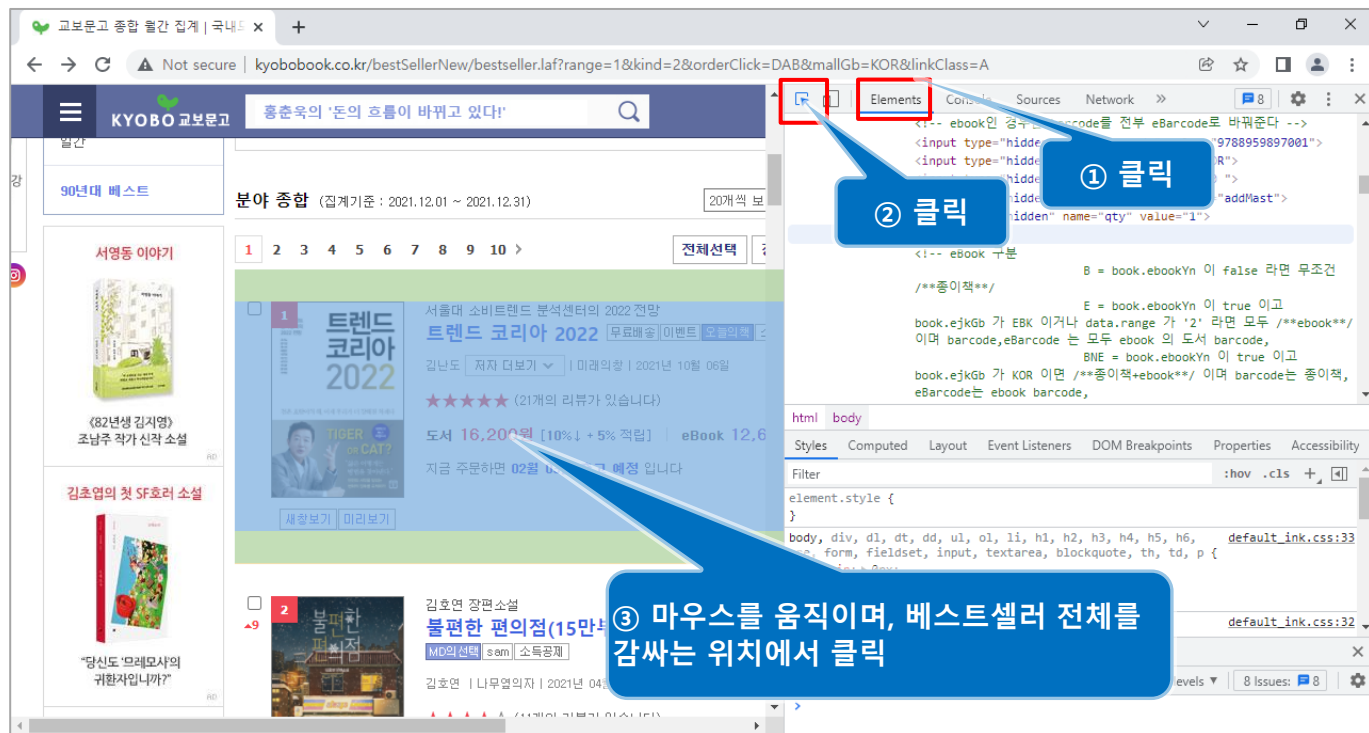
02월 03일 출고 예정 [배송일정 안내](#)

경제경영

# [실습] 서점 베스트셀러 정보 가져오기

## ● Step 02. 웹 구조 파악하기

- 크롬 개발자 도구를 활용해 베스트셀러 목록 페이지와 도서별 상세 페이지 URL 을 확인 후 웹 페이지 접속
- 크롬 개발자 도구 활용하기
  - 베스트셀러 목록 페이지로 돌아와 [F12] 키를 누르면 화면의 우측에 크롬 개발자 도구가 나타남
  - [Elements] 탭에서 Inspector 아이콘을 클릭한 후, 도서 목록에서 마우스를 움직이면서 원하는 도서 전체가 선택될 때 클릭하면 해당 HTML 코드를 확인

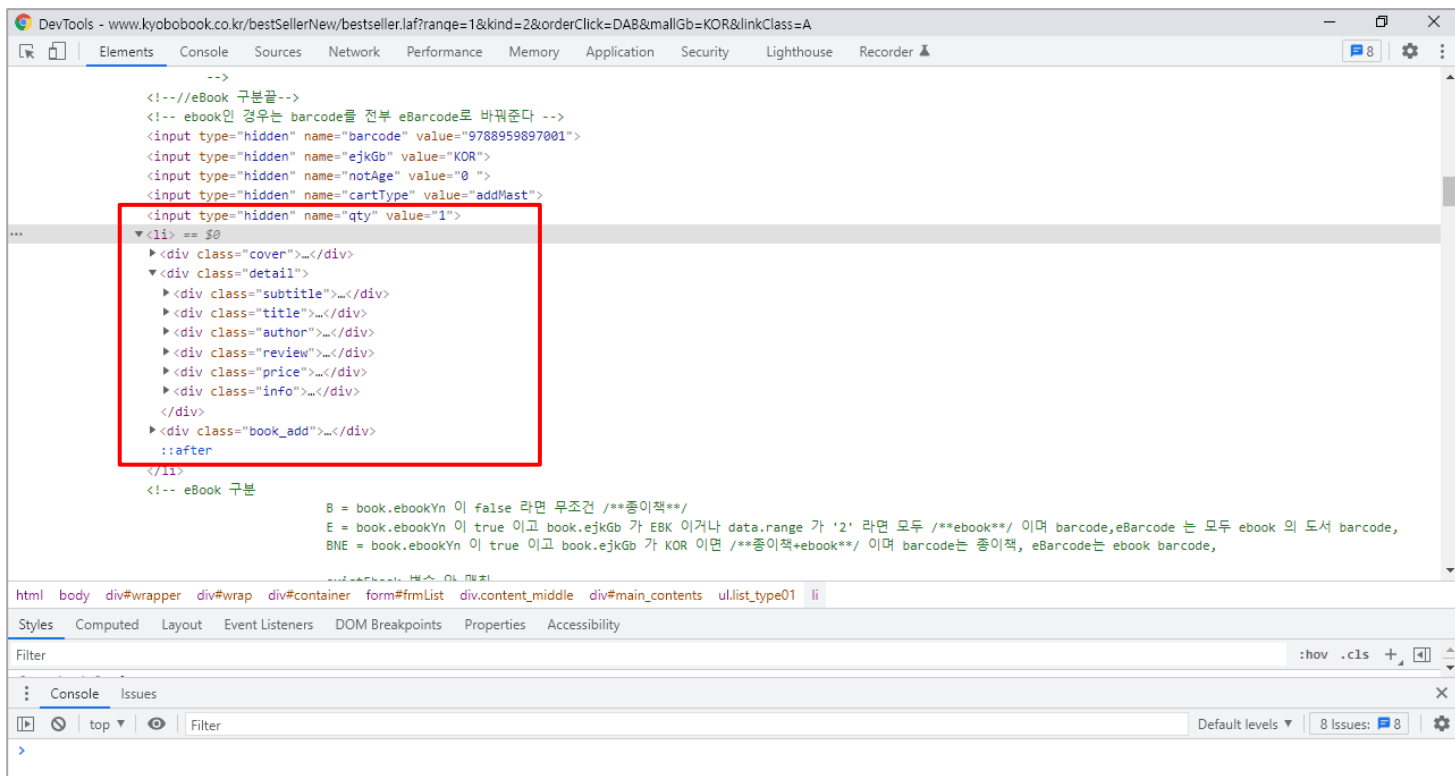


# [실습] 서점 베스트셀러 정보 가져오기

## ● Step 02. 웹 구조 파악하기

### ➤ 웹 구조 파악


- html 소스창에서 마우스 커서를 움직여 태그를 분석해보면 도서는 <li> 태그로 구성되어 있음
- 하위 구성요소로 "cover", "detail", "book\_add"라는 클래스를 가지고 있음
- <div class="detail"> 태그를 클릭하면 하위 요소인 title, author, review, price 클래스도 확인할 수 있음

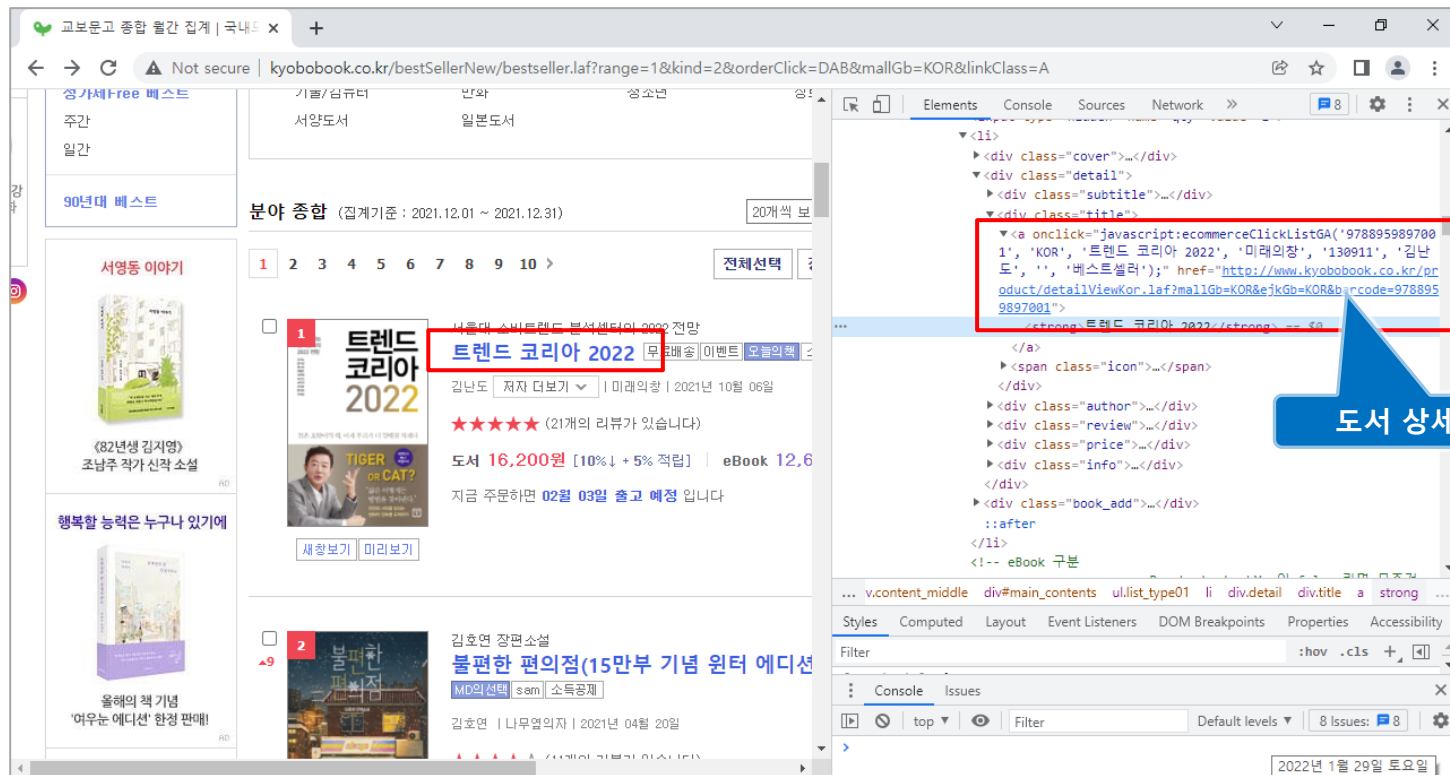


# [실습] 서점 베스트셀러 정보 가져오기

## ● Step 02. 웹 구조 파악하기

### ➤ 연결된 웹 페이지 확인

- Inspector  아이콘이 활성화된 상태에서 베스트셀러 1위 도서의 제목을 클릭해보면 개발자 도구 창에 `<strong> ... </strong>` 태그가 하이라이트되는 것을 확인
- 도서별 상세 페이지 주소가 `<a href>` 태그 안에 있는 것도 확인




The screenshot shows a web browser displaying a list of best-selling books on the Kyobobook website. The first book, '트렌드 코리아 2022', is highlighted with a red box. The developer tools 'Elements' panel is open, showing the HTML structure of the book entry. A red box highlights the 'a' tag containing the book title and its href attribute. A blue callout bubble points to the href attribute with the text '도서 상세페이지 url'.

```
<a onclick="javascript:ecommerceClickListGA('9788959897001', 'KOR', '트렌드 코리아 2022', '미래의창', '130911', '김난도', '', '베스트셀러');" href="http://www.kyobobook.co.kr/product/detailViewKor.laf?mallGb=KOR&e1kGb=KOR&barcode=9788959897001">트렌드 코리아 2022</a>
```

# [실습] 서점 베스트셀러 정보 가져오기

## ● Step 03. 데이터 태그 파악

- 도서별 상세 페이지에서 도서, 저자, 도서 가격을 가져온다고 할 때 데이터가 어떤 클래스 안에 있는지, 어떤 태그로 구성되어 있는지 확인해야 함
- 데이터별 태그 확인
  - 도서 제목을 클릭해 도서별 상세 페이지로 들어감
  - 도서별 상세 페이지에서 Inspector 아이콘 을 클릭한 후, 도서의 제목을 클릭하면 해당 위치의 HTML을 확인
  - <div class="box\_detail\_point"> <h1 class="title"> <strong> 도서 </strong> 태그로 구성되어 있음



# [실습] 서점 베스트셀러 정보 가져오기

## ● Step 04. 코딩 및 실행하기

- 교보문고의 베스트셀러 웹 페이지에서 도서 20개의 상세 웹 페이지 주소를 추출한 후, 도서별 상세정보인 도서, 저자, 가격 데이터를 크롤링
- 패키지 불러오기

```
from requests  
from bs4 import BeautifulSoup as bs
```

# [실습] 서점 베스트셀러 정보 가져오기

## ● Step 04. 코딩 및 실행하기

### ➤ 교보문고의 월간 베스트셀러 웹 페이지 가져오기

- requests.get() 함수로 교보문고 월간 베스트셀러 웹 페이지 리소스를 가져와 html 변수에 저장
- BeautifulSoup를 이용해 html 변수에 들어있는 데이터를 파싱
- 파싱한 정보를 담고 있는 book을 출력하면 교보문고 베스트셀러 웹 페이지의 HTML 소스 코드들을 볼 수 있음

```
res = requests.get("http://www.kyobobook.co.kr/bestSellerNew/bestseller.laf?range = ₩
1&kind=2&orderClick = DAB&mallGb=KOR&linkClass=A")
html = res.content                                #웹 리소스 가져오기
book = bs(html, "html.parser")                    #html 파싱
print(book)
```

### 출력결과

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">

<html lang="ko" xml:lang="ko" xmlns="http://www.w3.org/1999/xhtml">
<!-- s:html:head -->
<head>
<title>교보문고 종합 주간 집계 | 국내도서 | 베스트셀러 - 교보문고</title>
<!--MS의 최신 웹브라우저인 edge 브라우저 호환을 위해 넣어줌-->
<meta content="IE=edge" http-equiv="X-UA-Compatible"/>
<meta content="text/html; charset=utf-8" http-equiv="Content-Type">
<meta content="nocache" http-equiv="Pragma"/>
<meta content="0" http-equiv="Expires"/>
<meta content="no-cache" http-equiv="Cache-Control"/>
<link href="http://image.kyobobook.co.kr/newimages/apps/b2c/kyobo.ico" rel="shortcut icon"/>
<link href="http://image.kyobobook.co.kr/ink/css/default_ink.css" rel="stylesheet" type="text/css"/>
```

# [실습] 서점 베스트셀러 정보 가져오기

## ● Step 04. 코딩 및 실행하기

### ➤ 도서 상세 웹 페이지 주소 추출하기

```
book_list = []  
for book_detail in book.select("div.detail")  
    book_urls = book_detail.select_one("div.title > a").attrs["href"]  
book_list.append(book_urls)  
print(book_list)
```

#book\_list 리스트 변수 만들기  
#book에서 class=detail 태그 가져오기  
#상세주소를 book\_urls에 저장  
#book\_list 변수에 주소를 하나씩 추가

### 출력결과

```
['http://www.kyobobook.co.kr/product/detailViewKor.laf?mallGb=KOR&ejkGb=KOR&barcode=9791196661984', 'http://www.kyobobook.co.kr/product/detailViewKor.laf?mallGb=KOR&ejkGb=KOR&barcode=9791161571188', 'http://www.kyobobook.co.kr/product/detailViewKor.laf?mallGb=KOR&ejkGb=KOR&barcode=9788991622869', 'http://www.kyobobook.co.kr/product/detailViewKor.laf?mallGb=KOR&ejkGb=KOR&barcode=9788959897001', 'http://www.kyobobook.co.kr/product/detailViewKor.laf?mallGb=KOR&ejkGb=KOR&barcode=9791165344252', 'http://www.kyobobook.co.kr/product/detailViewKor.laf?mallGb=KOR&ejkGb=KOR&barcode=9791165216603', 'http://www.kyobobook.co.kr/product/detailViewKor.laf?mallGb=KOR&ejkGb=KOR&barcode=9791191056556', 'http://www.kyobobook.co.kr/product/detailViewKor.laf?mallGb=KOR&ejkGb=KOR&barcode=9791190030922', 'http://www.kyobobook.co.kr/product/detailViewKor.laf?mallGb=KOR&ejkGb=KOR&barcode=9791130677774', 'http://www.kyobobook.co.kr/product/detailViewKor.laf?mallGb=KOR&ejkGb=KOR&barcode=9791130677774', 'http://www.kyobobook.co.kr/product/detailViewKor.laf?mallGb=KOR&ejkGb=KOR&barcode=9791130677774']
```



# 숙제

<http://www.yes24.com/24/Category/BestSeller>

예스24에서 베스트셀러 추출해서 순위, 도서명, 가격 출력하기

```
PS C:\Users\saint\SynologyDrive\Drive\Document\01.주요문서\01.수업\웹프로그래밍\강의자료> python .\crawl.py
1 [도서] 트렌드 코리아 2023
2 [도서] 아버지의 해방일지
3 [도서] 이은경 쌤의 초등어휘일력 365
4 [도서] 불편한 편의점 2
5 [도서] 역행자
6 [도서] 하얼빈
7 [도서] 조국의 법고전 산책
8 [도서] 불편한 편의점
9 [도서] 양강 형제 엽집
10 [도서] 우리, 편하게 말해요
11 [도서] 에그박사 8
12 [도서] 세상에서 가장 쉬운 본질육아
13 [도서] 나의 월급 독립 프로젝트 (리마스터 에디션)
14 [도서] 마흔에 읽는 니체
15 [도서] 데일 카네기 인간관계론
16 [도서] 파친코 2
17 [도서] 스틱!
18 [도서] 작은 땅의 야수들
19 [도서] 파친코 1
20 [도서] 원생 THE ONE THING
21 [도서] 인생의 역사
22 [도서] 이 책은 돈 버는 법에 관한 이야기
23 [도서] 흔한남매 이상한 나라의 고전 읽기 2
24 [도서] 갯마을 차차차 포도에세이
25 [도서] 세상의 마지막 기차역
26 [도서] 자이언트 임팩트
27 [도서] 2022 큰별쌤 최태성의 별별한국사 한국사능력검정시험 심화(1, 2, 3급) 상
28 [도서] ETS 토익 정기시험 기출문제집 1000 Vol.1.3 READING 리딩
29 [도서] 잘될 수밖에 없는 너에게
30 [도서] 다산, 어른의 하루
31 [도서] 그릿 Grit : 100세 기념 리커버 에디션
32 [도서] 부자 아빠 가난한 아빠 20주년 특별 기념판
33 [도서] 올리버쌤의 미국식 아이 영어 습관 365
34 [도서] 2022 큰별쌤 최태성의 별별한국사 한국사능력검정시험 심화(1, 2, 3급) 하
35 [도서] ETS 토익 정기시험 기출문제집 1000 Vol.1.3 LISTENING 리스닝
36 [도서] 아파트 투자는 사이클이다
37 [도서] 마법천자문 55
38 [도서] 정재승의 인간탐구보고서 10
39 [도서] 클루지
40 [도서] 인생의 허무를 어떻게 할 것인가
```