

1. 웹 크롤링

웹 크롤링이란 웹페이지에서 원하는 정보가 있을 때 이를 데이터로 저장하기 위해 정보를 불러오는 작업을 말합니다.

자세한 사항은 아래에서 확인할 수 있는데, 먼저 중요한 몇 가지를 확인하고 가겠습니다.

우선, 웹 크롤링이 불가능한 페이지가 존재합니다. 엄밀히 말해서 웹 페이지는 특정 기업의 자산입니다. 그들의 데이터라고 표현하는게 맞죠.

어마어마한 비용을 들여서 운영중이고 쌓고 있는 데이터인데, 이것 무분별하게 가져올 수는 없습니다.

이러한 이유로 크롤링이 가능한 지 아닌지는 사이트에 따라 다릅니다. 예를 들면, 네이버와 같은 사이트에서는 크롤링이 가능합니다. 하지만, 넷플릭스와 같은 사이트에서는 (유익미한) 크롤링이 불가능합니다. 즉, 민감한 정보에 대해서는 크롤링을 막아놓았습니다.

또한, 크롤링을 할 때 여러페이지를 크롤링 하는 경우가 있는데, 여기서 여러페이지가 패턴이 없는 여러페이지를 동시에 크롤링 하기에는 쉽지 않습니다.

마지막으로, 우리가 웹 페이지에서 데이터가 필요한 경우, 보통 반복되는 여러페이지에서 데이터를 필요로 하는 경우가 많습니다.

예를 들면, 네이버 뉴스 데이터를 필요로 한다고 했을 때, 보통 뉴스 검색 시에 나오는 1페이지 데이터만을 필요로 하는 경우보다는 특정 검색어에 대해 여러 페이지의 데이터를 필요로 하는 경우가 많습니다.

링크 : [https://search.naver.com/search.naver?](https://search.naver.com/search.naver?where=news&query=%EB%B9%84%ED%8A%B8%EC%BD%94%EC%9D%B8&sm=tab_opt&sort=1&photo=0&field=0&pd=3&ds=2021.06.15&de=2021.09.15&mynews=0&office_type=0&office_section_code=0&news_office_checked=&nso=so:dd,from20210615to20210915&a:all&start=)

[where=news&query=%EB%B9%84%ED%8A%B8%EC%BD%94%EC%9D%B8&sm=tab_opt&sort=1&photo=0&field=0&pd=3&ds=2021.06.15&de=2021.09.15&mynews=0&office_type=0&office_section_code=0&news_office_checked=&nso=so:dd,from20210615to20210915&a:all&start=">](https://search.naver.com/search.naver?where=news&query=%EB%B9%84%ED%8A%B8%EC%BD%94%EC%9D%B8&sm=tab_opt&sort=1&photo=0&field=0&pd=3&ds=2021.06.15&de=2021.09.15&mynews=0&office_type=0&office_section_code=0&news_office_checked=&nso=so:dd,from20210615to20210915&a:all&start=)

그렇기 때문에, URL 분석을 필요로 하는 경우가 많아집니다.

단일 페이지의 데이터만을 필요로 하다면 URL 분석이 의미가 없지만, 여러 페이지의 경우 URL에서 반복되는 요소가 있기 때문에 필요로 하게 됩니다.

예를 보여드리겠습니다.

비트코인으로 검색 후, 기간을 2021.06.15 ~ 2021.09.15로 설정했을 때, URL은 이렇게 됩니다. 1페이지 :

[https://search.naver.com/search.naver?](https://search.naver.com/search.naver?where=news&query=%EB%B9%84%ED%8A%B8%EC%BD%94%EC%9D%B8&sm=tab_opt&sort=1&photo=0&field=0&pd=3&ds=2021.06.15&de=2021.09.15&mynews=0&office_type=0&office_section_code=0&news_office_checked=&nso=so:dd,from20210615to20210915&a:all&start=)

[where=news&query=%EB%B9%84%ED%8A%B8%EC%BD%94%EC%9D%B8&sm=tab_opt&sort=1&photo=0&field=0&pd=3&ds=2021.06.15&de=2021.09.15&mynews=0&office_type=0&office_section_code=0&news_office_checked=&nso=so:dd,from20210615to20210915&a:all&start=">](https://search.naver.com/search.naver?where=news&query=%EB%B9%84%ED%8A%B8%EC%BD%94%EC%9D%B8&sm=tab_opt&sort=1&photo=0&field=0&pd=3&ds=2021.06.15&de=2021.09.15&mynews=0&office_type=0&office_section_code=0&news_office_checked=&nso=so:dd,from20210615to20210915&a:all&start=)

2페이지 : [https://search.naver.com/search.naver?](https://search.naver.com/search.naver?where=news&sm=tab_pge&query=%EB%B9%84%ED%8A%B8%EC%BD%94%EC%9D%B8&sort=1&photo=0&field=0&pd=3&ds=2021.06.15&de=2021.09.15&mynews=0&office_type=0&office_section_code=0&news_office_checked=&nso=so:dd,from20210615to20210915&a:all&start=)

[where=news&sm=tab_pge&query=%EB%B9%84%ED%8A%B8%EC%BD%94%EC%9D%B8&sort=1&photo=0&field=0&pd=3&ds=2021.06.15&de=2021.09.15&mynews=0&office_type=0&office_section_code=0&news_office_checked=&nso=so:dd,from20210615to20210915&a:all&start=">](https://search.naver.com/search.naver?where=news&sm=tab_pge&query=%EB%B9%84%ED%8A%B8%EC%BD%94%EC%9D%B8&sort=1&photo=0&field=0&pd=3&ds=2021.06.15&de=2021.09.15&mynews=0&office_type=0&office_section_code=0&news_office_checked=&nso=so:dd,from20210615to20210915&a:all&start=)

3페이지 : [https://search.naver.com/search.naver?](https://search.naver.com/search.naver?where=news&sm=tab_pge&query=%EB%B9%84%ED%8A%B8%EC%BD%94%EC%9D%B8&sort=1&photo=0&field=0&pd=3&ds=2021.06.15&de=2021.09.15&mynews=0&office_type=0&office_section_code=0&news_office_checked=&nso=so:dd,from20210615to20210915&a:all&start=)

[where=news&sm=tab_pge&query=%EB%B9%84%ED%8A%B8%EC%BD%94%EC%9D%B8&sort=1&photo=0&field=0&pd=3&ds=2021.06.15&de=2021.09.15&mynews=0&office_type=0&office_section_code=0&news_office_checked=&nso=so:dd,from20210615to20210915&a:all&start=">](https://search.naver.com/search.naver?where=news&sm=tab_pge&query=%EB%B9%84%ED%8A%B8%EC%BD%94%EC%9D%B8&sort=1&photo=0&field=0&pd=3&ds=2021.06.15&de=2021.09.15&mynews=0&office_type=0&office_section_code=0&news_office_checked=&nso=so:dd,from20210615to20210915&a:all&start=)

자, 부분부분 떼어오겠습니다. 참고로, &는 and기호로 연결되었다고 생각하시면 됩니다.

또한, 기업의 홈페이지 주소마다 고유하게 갖고 있는 부분이 있고, 우리에게 의미가 있는 부분이 있는데, 우리에게 의미가 있는 부분을 설명하겠습니다.

(1) <https://search.naver.com/search.naver?where=news> & 부분은 똑같네요 !

(2) `query=%EB%B9%84%ED%8A%B8%EC%BD%94%EC%9D%B8&sm=tab_opt&sort=1&photo=0&field=0&pd=3` 부분도 똑같죠? \n

여기서, "query="이 검색어를 뜻합니다. 즉, query=비트코인 인데, 저 부분이 유니코드로 변환되면서 알 수 없는 언어로 나올 뿐입니다. 고쳐서 검색해볼까요?

(3) `sort=1`도 똑같네요. sort는 관련도순 최신순 오래된순을 나타냅니다. 각각 0,1,2의 값을 가지고 있습니다.

(4) `photo=0`도 똑같네요. 이 부분은 유형 부분을 가리킵니다. 0부터 순서대로 이뤄집니다.

(5) `field=0&pd=3`도 똑같네요 이 부분은 아마도 연관된 기사를 나타낼 때 사용되는 것 같습니다.

(6) `ds=2021.06.15&de=2021.09.15` 는 무슨 부분인지 아시겠죠?? 아까 검색한 기간에 해당됩니다. 날짜를 바꿔서 검색해볼까요??

(7) `&mynews=0&office_type=0&office_section_code=0&news_office_checked=&nso=so:dd,`이 부분도 동일하네요 !

(8) `from20210615to20210915` 이 부분도 아시겠죠?? 주의할 점은 날짜의 형식이 (6)과 다르다는 것입니다.

(9) `a:all&start=` 이 부분은 페이지 수를 나타냅니다.

1페이지는 1, 2페이지는 11, 3페이지는 21,... 이런식으로 반복됩니다. 왜인지는 저도 모릅니다.

이러한 점을 이용해서 여러페이지를 쉽게 분석할 수 있습니다. 먼저, 단일 페이지 크롤링부터 시작하겠습니다.

우선, 필요한 패키지 및 모듈을 불러옵니다. 보통 크롤링은 BeautifulSoup라는 모듈을 사용하여 분석하는게 쉽습니다.

```
In [1]: from bs4 import BeautifulSoup
```

```

import requests
import pandas as pd
import time
from datetime import datetime

```

In [3]:

```

query = "비트코인" # 검색어
sort = 1 # 최신순
s_date = "2021.06.15" # 시작날짜
e_date = "2021.09.15" # 종료날짜
s_from = "20210615" # 시작날짜
e_to = "20210915" # 종료날짜

mycolumns = ['date', 'press', 'title', 'content', 'link'] # 우리가 분석할 목록을 열로 지정하기 위해 리스트에 열이름을 저장

num = 1 # 1페이지의 값 저장

page = 1

title_list = [] # 제목 모음 리스트
content_list = [] # 내용 모음 리스트
link_list = [] # 링크 모음 리스트
date_list = [] # 날짜 모음 리스트
press_list = [] # 언론사 모음 리스트

url = "https://search.naver.com/search.naver?where=news&sm=tab_pge&query=" + query + "&sort="+str(sort)+"&photo=0&"

response = requests.get(url) # 사이트로부터 Response를 받아옵니다.
html = response.text # BeautifulSoup에 입력인자는 reponse.text 이므로 이를 변수로 지정합니다.

# BeautifulSoup 객체 생성, HTML Parser에 의해 HTML 구조의 웹 페이지를 파악하므로 이러한 문법을 사용합니다.
soup = BeautifulSoup(html, 'html.parser')

li_tmp = soup.find_all('li', {'class' : 'bx'}) # 우리가 필요했던 것은 li 태그의 bx 클래스를 가진 부분입니다. 스테디 때 알려

for i in li_tmp :
    try : # 예외처리 -> 오류가 발생할 시에 멈추지 않고 넘어갈 수 있음
        # 아래는 스테디 참고
        title = i.find('a', {'class' : 'news_tit'}).text # 제목
        content = i.find('div', {'class' : 'news_dsc'}).text # 내용
        link = i.find('a', {'class' : 'news_tit'})['href'] # 링크
        press = i.find('a', {'class' : 'info_press'}).text # 언론사

        date_set = i.findAll('span', {'class' : 'info'})# 날짜
        if date_set :
            # 가끔 17면 1단 이런식으로 된 정보를 제공해주는 기사가 있어서 이런 기사는 리스트의 다음 원소가 date가 됩니다.
            if 'A' in date_set[0].text or '면' in date_set[0].text :
                date = date_set[1].text
            else :
                date = date_set[0].text

        title_list.append(title)
        content_list.append(content)
        link_list.append(link)
        press_list.append(press)
        date_list.append(date)

    except :
        continue

mydata = {
    'date' : date_list,
    'press' : press_list,
    'title' : title_list,
    'content' : content_list,
    'link' : link_list
}

my_df = pd.DataFrame(data = mydata, columns = mycolumns)

```

In [4]:

```
my_df
```

Out[4]:

	date	press	title	content	link
0	3분 전	데일리굿뉴스	코인 거래소 폐업 임박...정부 "자산 증발돼 빨리 옮겨야"	특히 투자한 가상자산이 비트코인이나 이더리움처럼 대표적인 종류가 아닌, 특정 거...	http://goodnews1.com/news/news_view.asp?seq=11...
1	16 분	부산일보	블록체인 자유특구? 실상은 '규제	기존 가상자산거래소의 변형된 형태인 디지털	http://www.busan.com/view/busan/view.php?

	전	정글'... 기업 자금줄 막혀 아우성	자산거래소는 비트코인, 이더리움 등 코...	code=...
2	17 분 전	코인데스크 코리아	[저녁브리핑] 코인 가격과 상폐 그 리고 NFT 아트	[글래스노드] "시장 변동성에도 장기보유자 지 표는 역사상 최고치" 이번 주간 ... http://www.coindesk.com/news/articleView...
3	42 분 전	SBS Biz	홍남기 "다주택자 양도세 완화-가상 자산 과세연기 계획없다"	굉장히 불확실하다는 것이 일반적인 평가"라고 말했습니다. 또 내년으로 예정된 비... https://biz.sbs.co.kr/article_hub/20000031156?...
4	45 분 전	세계일보 론사 선정	"'상폐빔' 노리다 눈물 쏟 뺨니다"... 가상화폐 거래소 폐업 가시화	15일 포블게이트도 원화마켓을 일시 중단하고 비트코인을 기초자산으로 코인 간 거... http://www.segye.com/content/html/2021/09/15/2...
5	46 분 전	중앙일보	대마 재배해 '다크웹'서 판 일당...범 죄단체조직죄 첫 적용	아울러 검찰은 김씨 등이 재배 중이던 대마 약 300주(약 30kg)를 압수하고,... https://www.joongang.co.kr/article/25007521
6	47 분 전	데일리안	[코인뉴스] 비트코인·이더리움 거래 량 급감...하락장 징조?	비트코인과 이더리움의 거래량이 하루만에 10% 이상 빠지면서 투자자들의 불안감이... https://www.dailian.co.kr/news/view/1033802/?s...
7	48 분 전	비즈니스포 스트	비트코인 5630만 원대 공방, 가상화 페 '상승' 테마기업 추가 '혼조'	비트코인은 1BTC(비트코인 단위)당 5635만2 천 원에 거래돼 24시간 전보다... https://www.businesspost.co.kr/BP?command=arti...
8	50 분 전	팍스넷뉴스	[블록캐치] 밀크코인, CU포인트와 연동 외	후오비코리아, 추석 한가위 맞아 비트코인 에어 드럼 롤렛 이벤트 진행 가상자산 거... https://www.paxnetnews.com/articles/78613
9	56 분 전	경북도민일 보	비트코인으로 마약 사고판 일당 무 더기 검거	대구경찰청 마약범죄수사대는 15일 SNS(사회 관계망서비스)에서 비트코인을 이용해... http://www.hidomin.com/news/articleView.html?i...

다음은 여러 페이지를 분석하는 방법입니다. 마지막에 본 페이지수만 계속 변경해주면 됩니다.

```
In [8]: query = "비트코인"
sort = 1
s_date = "2021.06.15"
e_date = "2021.09.15"
s_from = "20210615"
e_to = "20210915"

mycolumns = ['date', 'press', 'title', 'content', 'link']

page = 1 # 페이지의 시작 값은 1

title_list = [] # 제목 모음 리스트
content_list = [] # 내용 모음 리스트
link_list = [] # 링크 모음 리스트
date_list = [] # 날짜 모음 리스트
press_list = [] # 언론사 모음 리스트

while page <= 91 : # 몇 페이지만 해볼게요
    #url을 정의합니다.
    url = "https://search.naver.com/search.naver?where=news&sm=tab_pge&query=" + query + "&sort="+str(sort)+"photo"

    response = requests.get(url)
    html = response.text

    soup = BeautifulSoup(html, 'html.parser')
    li_tmp = soup.find_all('li', {'class' : 'bx'})

    for i in li_tmp :
        try :
            title = i.find('a', {'class' : 'news_tit'}).text # 제목
            content = i.find('div', {'class' : 'news_dsc'}).text # 내용
            link = i.find('a', {'class' : 'news_tit'})['href'] # 링크
            press = i.find('a', {'class' : 'info_press'}).text # 언론사

            date_set = i.findAll('span', {'class' : 'info'})# 날짜
            if date_set :
                if 'A' in date_set[0].text or '면' in date_set[0].text :
                    date = date_set[1].text
                else :
                    date = date_set[0].text

            title_list.append(title)
            content_list.append(content)
            link_list.append(link)
            press_list.append(press)
            date_list.append(date)

        except :
            continue

    page += 10 # 10씩 더해지는거 보셨죠??
```

```
mydata = {
    'date' : date_list,
    'press' : press_list,
    'title' : title_list,
    'content' : content_list,
    'link' : link_list
}

my_df = pd.DataFrame(data = mydata, columns = mycolumns)
```

In [9]: my_df

Out[9]:

	date	press	title	content	link
0	7분 전	데일리굿 뉴스	코인 거래소 폐업 임박...정부 "자산 증발돼 빨리 옮겨야"	특히 투자한 가상자산이 비트코인이나 이더리움처럼 대표적인 종류가 아닌, 특정 거...	http://goodnews1.com/news/news_view.asp?seq=11...
1	20분 전	부산일보	블록체인 자유특구? 실상은 '규제 정글'... 기업 자금줄 막혀 아우성	기존 가상자산거래소의 변형된 형태인 디지털자산거래소는 비트코인, 이더리움 등 코...	http://www.busan.com/view/busan/view.php?code=...
2	21분 전	코인데스크코리아	[저녁브리핑] 코인 가격과 상폐 그리고 NFT 아트	[글래스노드] "시장 변동성에도 장기보유자 지표는 역사상 최고치" 이번 주간 ...	http://www.coindeskorea.com/news/articleView....
3	46분 전	SBS Biz	홍남기 "다주택자 양도세 완화·가상 자산 과세연기 계획없다"	굉장히 불확실하다는 것이 일반적인 평가"라고 말했습니다. 또 내년으로 예정된 비...	https://biz.sbs.co.kr/article_hub/20000031156?...
4	49분 전	세계일보 언론사 선정	"'상폐빔' 노리다 눈물 쏟 뺨니다"...가상화폐 거래소 폐업 가시화	15일 포블게이트도 원화마켓을 일시 중단하고 비트코인을 기초자산으로 코인 간 거...	http://www.segye.com/content/html/2021/09/15/2...
...
95	4시간 전	EBN	증권사 리서치센터, 블록체인·암호화폐 '삼매경'	다만 당장은 비트코인, 이더리움과 같은 암호화폐 시세를 분석하는 대신 블록체인 ...	https://www.ebn.co.kr/news/view/1500782/?sc=Naver
96	5시간 전	스포츠서울	업비트 10개월 만에 고객센터 재개장	한편 비트코인은 4% 이상 상승하며 5600만원대에 거래되고 있다. 15일 국내...	http://www.sportsseoul.com/news/read/1063540?r...
97	5시간 전	서울경제	[점심 브리핑]주요 암호화폐 상승세...美 SEC 위원장 "암호화폐 거래소 규제해...	15일 오후 1시 19분 빙썬 기준 국내 비트코인(BTC) 가격은 전날 같은 시...	https://www.sedaily.com/NewsView/22RH43GWM5
98	5시간 전	코인데스크코리아	[JK(장칼)의 코인 투자 전망] "4만7300달러가 비트코인 상승세 좌우한다"	출처=트레이딩뷰 횡보구간 이탈 시도하는 비트코인...4만7300달러 지지하면 상...	http://www.coindeskorea.com/news/articleView....
99	5시간 전	코인리더스	빅데이터로 본 최근 '암호화폐 절대강자' 비트코인(BTC) VS 이더리움(ETH) ...	9월 7일(현지 시각), 엘살바도르가 세계 최초로 시가총액 1위 가상자산인 비트...	http://coinreaders.com/12423

100 rows × 5 columns

여기서부터는 팁입니다. 크롤링보다는 코딩에 대한 팁입니다. 문제는 항상 발생하는 법입니다. 이를 어떻게 해결해야 할까가 결국 여러분의 시간을 결정적으로 오래쓰느냐 적게쓰느냐를 결정하게 됩니다.

네이버의 경우 400페이지까지밖에 크롤링이 되지 않습니다. 링크 : https://search.naver.com/search.naver?where=news&sm=tab_pge&query=%EB%B9%84%ED%8A%B8%EC%BD%94%EC%9D%B8&sort=1&photo=0&field=0&pd=3&ds=2021.06.15&

그런데, 보시면 아시겠지만 400페이지는 고작 20일 남짓되는 기간밖에 되지 않습니다. 심지어, 검색량이 폭발하는 기사의 경우 더 적은 기간이 될 수 있습니다. 그런데, 여러분이 만약 10년간의 뉴스 기록을 가져오고 싶으시다면 어떻게 하시겠습니까?? 일일이 손으로 해가는 작업이 있습니다. 그런데, 너무 비효율적이고, 하면서 화가 많이 날 수 있습니다.

아래는 제가 이를 저번 크롤링을 하면서 해결한 방법입니다. 조금 번잡하지만, 결과물은 가져올 수 있었고, 시간을 많이 아낄 수 있었습니다.

참고로, 아래의 코드를 돌리면 오래 반복이 돌아가므로 시간은 좀 오래걸릴 수 있습니다.

In [12]:

```
query = "비트코인"
sort = 1
s_date = "2021.06.15"
e_date = "2021.09.15"
s_from = "20210615"
e_to = "20210915"

mycolumns = ['date', 'press', 'title', 'content', 'link']

num = 1 # 먼저 첫 400페이지는 그냥 긁어와야겠쥬

while True : # 어떤 조건을 만족하기 전까지 반복을 끝내지 않습니다.
    page = 1
```

```

if num >= 2 : # 400페이지를 긁었다면, 그 이후부터는 아래의 코드를 진행합니다.
    year = int(date_list[-1][0:4] + date_list[-1][5:7] + date_list[-1][8:10]) #YYYYMMDD형태
    # 먼저, 마지막 400페이지에 있는 날짜를 가져옵니다. 이 날짜를 시작으로 다시 검색할 것입니다.

    if year <= 20210614 : # 우리가 원하는 날짜 전의 날짜가 나오면, break해서 종료해줍니다.
        break

    s_date, s_from = '2021.06.15', '20210615' # 시작날짜는 동일하게 원했던 날짜로 해주고
    e_date, e_to = date_list[-1][:-1], str(year) # 종료날짜는 방금 전 시도까지 했던 마지막페이지의 날짜로 해줍니다.

title_list = [] # 제목 모음 리스트
content_list = [] # 내용 모음 리스트
link_list = [] # 링크 모음 리스트
date_list = [] # 날짜 모음 리스트
press_list = [] # 언론사 모음 리스트

while page <= 3991 : # 400페이지까지 돌아줍니다.
    url = "https://search.naver.com/search.naver?where=news&sm=tab_pge&query=" + query + "&sort="+str(sort)+"
    response = requests.get(url)
    html = response.text

    soup = BeautifulSoup(html,'html.parser')
    li_tmp = soup.find_all('li',{'class' : 'bx'})

    for i in li_tmp :
        try :
            title = i.find('a',{'class' : 'news_tit'}).text #제목
            content = i.find('div',{'class' : 'news_dsc'}).text #내용
            link = i.find('a',{'class' : 'news_tit'})['href'] #링크
            press = i.find('a',{'class' : 'info_press'}).text #언론사

            date_set = i.findAll('span',{'class' : 'info'})#날짜
            if date_set :
                if 'A' in date_set[0].text or '면' in date_set[0].text :
                    date = date_set[1].text
                else :
                    date = date_set[0].text

            title_list.append(title)
            content_list.append(content)
            link_list.append(link)
            press_list.append(press)
            date_list.append(date)

        except :
            continue

    page += 10

mydata = {
    'date' : date_list,
    'press' : press_list,
    'title' : title_list,
    'content' : content_list,
    'link' : link_list
}

my_df = pd.DataFrame(data = mydata, columns = mycolumns)

num += 1 # 첫 시도가 지났음을 알려주기 위해 1에서 숫자를 더합니다.

if num == 3 : # 이 예제는 너무 오래 돌아가지 않기 위해 2번만 할게요..
    break
time.sleep(3) # 가끔 웹 페이지 충돌 오류가 생길 수 있어서 3~5초간의 텀을 주면 좋습니다.

```

이렇게 되면 날짜가 겹치는 부분이 생깁니다. 그래서 이를 위해 내용이 겹치는 부분을 drop_duplicate 함수를 써서 없애버리면 됩니다.

```

In [13]: my_df = my_df.drop_duplicates("content")
my_df # 결과물

```

```

Out[13]:

```

	date	press	title	content	link
0	2021.08.23.	청년일보	뉴욕증시, 상승출발...FDA, 화 이자 백신 정식 승인	비트코인 가격은 5월 이후 처음으로 5 만 달러를 웃돌았다. 코인베이스와 페 이팔 ...	http://www.youthdaily.co.kr/news/article.html?... 이팔 ...
1	2021.08.23.	코인리더스	외신 "바이낸스, 정부기관 투자 유치 검토...싱가포르 유력"	또 코인데스크는 브라질 현지 매체 르 탈 도 비트코인(Portal do Bitco...	http://coinreaders.com/12133
2	2021.08.23.	여성신문언 론사 선정	여성·소수자 없이 메타버스 혁 신 없다	대형 포털 사이트 검색창에 '메타버 스'를 입력하면 '메타버스 관련주', '메 타버...	http://www.womennews.co.kr/news/articleView.ht...
3	2021.08.23.	한국경제TV	코로나가 살린 미 증시...백신주 ·코인관련주 상승세 [뉴욕증시 나우]	우선 최근의 또다른 뉴스가 될 수 있 는, 비트코인 상승세와 연관해서 생각 할 수 ...	http://www.wowtv.co.kr/NewsCenter/News/Read?ar...
4	2021.08.23.	톱스타뉴스	미국 뉴욕증시, 잭슨홀 회의 앞 두고 상승 출발...S&P·나스닥	비트코인 가격은 5월 이후 처음으로 5 만 달러를 웃돌았다. 코인베이스와 페	http://www.topstarnews.net/news/articleView.ht...

↑다우↑				이팔 ...	
...
3994	2021.07.28.	부산일보	“약관개정·환불·해지, 멋대로” 가상화폐거래소 불공정약관 철폐	26일 오후 서울 용산구 코인원 고객센터 모니터에 비트코인 등 가상화폐 가격이 ...	http://www.busan.com/view/busan/view.php?code=...
3995	2021.07.28.	아시아타임즈	아마존발 악재 탈출 비트코인 상승세...리플도 10%대 이상 급등	비트코인 상승 4600만원대 거래 알트코인 리플 10% 이상 급등 아마존의 비트...	https://www.asiatime.co.kr/article/20210728500354
3996	2021.07.28.	한국금융신문	[Money Plus] MZ세대, 자신이 좋아하는 것에서 투자 힌트를 찾아라	이들의 재테크는 기존 세대와 달리 유튜브나 책을 통해 치열하게 공부하며, 부동산...	http://www.fntimes.com/html/view.php?ud=202107...
3997	2021.07.28.	파이낸셜뉴스언론사 선정	민형배 "암호화폐는 '디지털 자산'.. '디지털자산법' 대표발의"	현재 '특정 금융거래정보의 보고 및 이용 등에 관한 법률'은 비트코인 등 암호화...	http://www.fnnews.com/news/202107281708303920
3998	2021.07.28.	디지털데일리	[DD 퇴근길] 애플 '아이패드'에 '5세대, 언제 나올까?	일례로 가상화폐 시세 대표격으로 꼽히는 비트코인은 27일에만 해도 한순간 4만달...	http://www.ddaily.co.kr/news/article.html?no=2...

3732 rows × 5 columns

1. Class

class를 간단하게 설명해드리겠습니다.

Class에서 가장 중요한 점은, init 함수와 self, 그리고 Class 내의 함수를 불러오는 방법입니다.

init이란, 클래스를 처음 선언했을 때 자동적으로 초기화되는 것을 말합니다. 즉, init함수에 선언해 놓은 모든 사항들은 여러분이 Class 객체를 선언하게되면 자동적으로 할당이됩니다.

self란, Class 내에서 Class의 소속 객체를 말합니다. 즉, Class의 객체라고 생각하시면 될 것 같습니다. 이는 self.~ 로 표현하며, Class 밖에서는 쓰일 수 없습니다.

아래 예에서 self.name이 의미하는 것은 Class가 가진 name이라는 변수라는 뜻으로 self.name = "정우"는 Class가 가진 name이라는 변수에 "정우"라는 이름을 할당한다 라는 뜻입니다. 참고로, 이는 멤버변수라고도 부르는데, 클래스에 속하는 변수라는 의미로 받아들이면 되겠습니다. 또한, 클래스 객체에 .을 이용해서 이 변수를 외부에서 불러올 수도 있습니다.

참고로, self는 외부에서 클래스에 있는 함수 및 변수를 부를 때는 입력하지 않고, 클래스 내에서만 입력을 하는데, 이는 Class내에서만 통용되는 객체이므로 따로 받지 않게끔 파이썬 내부에서 코드처리가 되어 있어서 그런 것으로 보입니다. 즉, 입력하지 않아도 알아서 self부분은 그 클래스의 일부분으로 인식되어 입력할 필요가 없는 것입니다.

클래스에 있는 함수를 불러오기 위해서는 .을 사용합니다. (아래 참고) 본격적으로 예를 들겠습니다.

In [14]:

```
class LinearAlgebra() :
    def __init__(self) :
        self.name = "정우"
        self.date = "오늘"

    def print_sentence(self, by_what) :
        print(self.name+ "는 " + self.date , by_what + "에 의해 패키지를 한다")
```

In [15]:

```
la = LinearAlgebra()

la.print_sentence("예빈이") # 클래스 내의 함수 외부에서 사용
print(la.name) # 멤버변수 외부에서 부르기
```

정우는 오늘 예빈이에 의해 패키지를 한다
정우

마지막으로, 상속에 관해서 설명하겠습니다.

상속이란, 자식이 부모의 DNA를 물려받는 것처럼, 부모 클래스로부터 성질을 물려받는 것이라 생각하시면 됩니다.

부모 클래스를 어떤 클래스를 선언할 때 괄호안에 넣어주면 됩니다.

super()는 부모클래스를 가리킵니다.

In [23]:

```
class psatLinearAlgebra(LinearAlgebra) : # LinearAlgebra가 부모 클래스
    def __init__(self, team) : # init 함수에 self 이외의 인자가 들어가면 클래스를 선언할 때 이를 괄호안에 입력해줘야 하고, i
        super().__init__()
        self.team = team
```

```
def print_sentence(self, by_what) :  
    print(self.team+"의 "+ self.name + "는 " + self.date , by_what + "에 의해 패키지를 한다")
```

In [24]:

```
pla = psatLinearAlgebra() # 예러가 뜹니다. init함수에 team이 있는데, 괄호안에 아무것도 없으므로
```

```
-----  
TypeError                                 Traceback (most recent call last)  
<ipython-input-24-bbe1c382be96> in <module>  
----> 1 pla = psatLinearAlgebra() # 예러가 뜹니다. init함수에 team이 있는데, 괄호안에 아무것도 없으므로  
  
TypeError: __init__() missing 1 required positional argument: 'team'
```

In [25]:

```
pla = psatLinearAlgebra("선형대수학팀")  
print(pla.team)  
pla.print_sentence("예빈")
```

선형대수학팀
선형대수학팀의 정우는 오늘 예빈에 의해 패키지를 한다

사실 이것 외에도 클래스에는 다양한 성질이 있는데, 위의 기본적인 내용에서 확장되는 것이라고 생각하시고 공부하시면 됩니다.

이런 기본적인 것만 알고 계셔도, 이후에 클래스를 다루는데에 큰 문제가 없을 것이라 생각합니다. 이번주도 수고하셨습니다!

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js