In [1]:
```
import re
import sqlite3
from urllib.request import urlopen
from html import unescape
import pandas as pd
import os
```

In [2]:
```
def fetch(url):
    """
    매개변수로 전달받을 url을 기반으로 웹 페이지를 추출
    웹 페이지의 Content-Type 헤더를 통해 인코딩 형식 확인
    반환값: str 자료형의 HTML
    """
    f = urlopen(url)
    # HTTP 헤더를 기반으로 인코딩 형식 추출
    encoding = f.info().get_content_charset(failobj="utf-8")
    # 추출한 인코딩 형식을 기반으로 문자열 디코딩
    html = f.read().decode(encoding)
    return html
```

In [3]:
```
def scrape(html):
    """
    매개변수 html로 받은 HTML을 기반으로 정규 표현식을 사용해 도서 정보를 추출합니다.
    반환값: 도서(dict) 리스트
    """
    books = []
    # re.findall()을 사용해 도서 하나에 해당하는 HTML을 추출
    for partial_html in re.findall(r'<td class="left"><a.*?</td>', html, re.DOTALL):
        # 도서의 URL을 추출
        url = re.search(r'<a href="(.*?)">', partial_html).group(1)
        url = 'http://www.hanbit.co.kr' + url
        # 태그를 제거해서 도서의 제목 추출
        title = re.sub(r'<.*?>', '', partial_html)
        title = unescape(title)
        books.append(pd.DataFrame({'url': [url], 'title': [title]}))
    return pd.concat(books)
```

In [4]:
```
def save(db_path, books):
    with sqlite3.connect(os.path.join('.', db_path)) as con: # sqlite DB 파일이 존재
        try:
            books.to_sql(name = 'BOOKS_INFO', con = con, index = False, if_exists='a
            #if_exists : {'fail', 'replace', 'append'} default : fail
        except Exception as e:
            print(str(e))

    query = 'SELECT * FROM BOOKS_INFO'
    df = pd.read_sql(query, con = con)
    return df
```

In [5]:
```
html = fetch('http://www.hanbit.co.kr/store/books/full_book_list.html')
```

In [6]:
```
df = scrape(html)
df.reset_index(drop=True, inplace=True)
df2 = save('books.db', df)
df2
```

Out[6]:

	url	title
0	http://www.hanbit.co.kr/store/books/look.php?p...	최신 관리회계
1	http://www.hanbit.co.kr/store/books/look.php?p...	리눅스 입문자를 위한 명령어 사전
2	http://www.hanbit.co.kr/store/books/look.php?p...	파타고니아 이야기
3	http://www.hanbit.co.kr/store/books/look.php?p...	풀스택 서버리스 : 리액트, AWS, 그래프QL을 이용한 최신 애플리케이션 개발
4	http://www.hanbit.co.kr/store/books/look.php?p...	한 권으로 배우는 작고 예쁜 꽃자수
5	http://www.hanbit.co.kr/store/books/look.php?p...	IT CookBook, 처음 만나는 회로이론(2판)
6	http://www.hanbit.co.kr/store/books/look.php?p...	안전필수 시스템 제어 설계
7	http://www.hanbit.co.kr/store/books/look.php?p...	러닝 리액트(2판)
8	http://www.hanbit.co.kr/store/books/look.php?p...	업무에 바로 쓰는 SQL 튜닝
9	http://www.hanbit.co.kr/store/books/look.php?p...	데이터 스토리
10	http://www.hanbit.co.kr/store/books/look.php?p...	상식의 재구성
11	http://www.hanbit.co.kr/store/books/look.php?p...	처음 배우는 네트워크 보안
12	http://www.hanbit.co.kr/store/books/look.php?p...	찾아도 찾아도 끝판왕 1000개 숨은그림찾기 우리 동네
13	http://www.hanbit.co.kr/store/books/look.php?p...	찾아도 찾아도 끝판왕 1000개 숨은그림찾기 숲속 놀이터
14	http://www.hanbit.co.kr/store/books/look.php?p...	IT CookBook, 디지털 콘텐츠 기획(2판)
15	http://www.hanbit.co.kr/store/books/look.php?p...	IT CookBook, C로 배우는 쉬운 자료구조 4판
16	http://www.hanbit.co.kr/store/books/look.php?p...	IT CookBook, 쉽게 배우는 소프트웨어 공학 2판
17	http://www.hanbit.co.kr/store/books/look.php?p...	IT CookBook, 컴퓨터 구조와 원리 3.0
18	http://www.hanbit.co.kr/store/books/look.php?p...	IT CookBook, 최신 기술 동향으로 알아보는 ICT와 4차 산업혁명
19	http://www.hanbit.co.kr/store/books/look.php?p...	초보 판매자가 빅파워셀러로 거듭나는 네이버 스마트스토어 마케팅 시작하기
20	http://www.hanbit.co.kr/store/books/look.php?p...	STEM CookBook, 한 걸음씩 알아가는 선형대수학
21	http://www.hanbit.co.kr/store/books/look.php?p...	STEM CookBook, 해석학 첫걸음
22	http://www.hanbit.co.kr/store/books/look.php?p...	IT CookBook, 난생처음 파이썬 프로그래밍

	url	title
23	http://www.hanbit.co.kr/store/books/look.php?p...	IT CookBook, 정보 보안 개론(4판)
24	http://www.hanbit.co.kr/store/books/look.php?p...	세상에서 제일 친절한 엑셀(개정판)
25	http://www.hanbit.co.kr/store/books/look.php?p...	게임세대 내 아이와 소통하는 법
26	http://www.hanbit.co.kr/store/books/look.php?p...	고개를 끄덕이는 것만으로도 위로가 되니까
27	http://www.hanbit.co.kr/store/books/look.php?p...	IT CookBook, 익스플로링 아두이노(2판)
28	http://www.hanbit.co.kr/store/books/look.php?p...	IT CookBook, 난생처음 인공지능 입문
29	http://www.hanbit.co.kr/store/books/look.php?p...	제대로 작성하는 논문 : 시작부터 마무리까지
30	http://www.hanbit.co.kr/store/books/look.php?p...	회사에서 바로 통하는 실무 엑셀 함수&수식 - 모든 버 전용
31	http://www.hanbit.co.kr/store/books/look.php?p...	지리의 쓸모
32	http://www.hanbit.co.kr/store/books/look.php?p...	수학이 외계어처럼 들리는 이공계생을 위한 제로 수학
33	http://www.hanbit.co.kr/store/books/look.php?p...	개발자에서 아키텍트로
34	http://www.hanbit.co.kr/store/books/look.php?p...	머신러닝을 활용한 웹 최적화
35	http://www.hanbit.co.kr/store/books/look.php?p...	STEM CookBook, 이공계생을 위한 확률과 통계(2판)
36	http://www.hanbit.co.kr/store/books/look.php?p...	STEM CookBook, 기초 선형대수학(2판)
37	http://www.hanbit.co.kr/store/books/look.php?p...	파이썬으로 살펴보는 아키텍처 패턴
38	http://www.hanbit.co.kr/store/books/look.php?p...	린 AI
39	http://www.hanbit.co.kr/store/books/look.php?p...	파이토치로 배우는 자연어 처리
40	http://www.hanbit.co.kr/store/books/look.php?p...	NGINX 쿡북
41	http://www.hanbit.co.kr/store/books/look.php?p...	리얼 국내여행 [2021~2022년 최신판]
42	http://www.hanbit.co.kr/store/books/look.php?p...	리얼 제주 [2021~2022년 최신판]
43	http://www.hanbit.co.kr/store/books/look.php?p...	지금 당장 회계공부 시작하라(전면개정판)
44	http://www.hanbit.co.kr/store/books/look.php?p...	IT CookBook, 문제해결을 위한 컴퓨팅 사고와 파이썬
45	http://www.hanbit.co.kr/store/books/look.php?p...	보고서 발표 실무 강의 - 잘 쓰고 제대로 전달하는 보고 의 기술

	url	title
46	http://www.hanbit.co.kr/store/books/look.php?p...	소문난 명강의 : 오준석의 플러터 생존코딩(개정판)
47	http://www.hanbit.co.kr/store/books/look.php?p...	만화로 배우는 서양사 중세 3
48	http://www.hanbit.co.kr/store/books/look.php?p...	IT CookBook, 컴퓨터 활용과 실습 2019
49	http://www.hanbit.co.kr/store/books/look.php?p...	테슬라 웨이

In [ ]: