

예제: Yes24 사이트의 베스트셀러 리스트를 파이썬으로 자동으로 가져오기

1. Yes24 베스트셀러 사이트 접속:

<https://www.yes24.com/Product/Category/BestSeller?categoryNumber=001>

2. 크롬 브라우저 > 오른쪽 상단 세줄 표시되어 있는 꺾임표 클릭 > 도구더보기 메뉴 클릭 > 개발자도구 클릭 > 사이트 화면의 html 코드가 분할면으로 오른쪽에 나타남 > 위 메뉴에 Elements 탭을 클릭되어 있고 사이트 화면의 html 코드를 오른쪽 분할면에 보임 > html-head-body 태그 열고 닫힌 구조 > div 태그와 id 속성, class 속성 존재 > div 태그에 고유한 이름을 지어주었다는 의미 > 거의 처음 보는 태그들
3. 오른쪽 분할화면 위 보면 메뉴의 가장 왼쪽에 화살표 아이콘 (Elements 옆) > 화살표 아이콘 클릭 > 실제 웹사이트 베스트셀러 구역에 마우스 커서를 대고 움직이면 내 마우스가 위치한 그 특정 구역의 태그가 오른쪽 html 화면 코드에 나타남 > 베스트 셀러 1위 구역에 마우스 커서를 대고 클릭하면 오른쪽 html 화면에 베스트셀러 1위에 해당되는 코드를 찾을 수 있음 > 예) unordered list 리스트 태그 아래에 리스트 태그, <li class data-goods-no= '12590468'...>... > 그 아래에 비슷한 태그가 24개 있음 > 2위 해당 태그, 3위해당되는 태그 확인가능 > 은 순서가 없는 리스트 > 마우스를 커서를 상하로 움직이면 오른쪽 화면에 활성화되는 위치가 변함
4. 이 HTML 코드화면의 코드를 간단히 정리하면, html - head - body - ul(id='yesBestList', class= 'sGLI tp_book tp_chkG tp_best tp_list) > 그아래 li(data-goods-no = '....') - li(data-goods-no = '....') - li(data-goods-no = '....') ...
5. body 태그에서 **id 값이 'yesBestList'와 class도 값인 'sGLI tp_book tp_chkG tp_best tp_list'인 태그와 그 아래 베스트셀러 1, 2, 3, ... 관련 리스트 태그 존재함** > 1위 리스트 태그를 한번 더 클릭하면 그 아래 하위 태그인, **class 이름이 'itemUnit'인 <div class='itemUnit'>태그가 존재함** > 다시 클릭하면_그 하위 태그들인 **<div class='item_img'>, <div class='item_info'> div(class='item_btnCol')**로 이루어져 있음
6. 여기에서 두번째인 **class 이름이 'item_info'인 <div> 태그**를 한번 더 클릭해서 들어가면, 그 아래 한 10개정도의 <div> 하위태그가 존재함 > 그 <div>태그위에 마우스 커서를 대면....도서 제목, 저자, 출판사, 증정선물, 가격, 회원리뷰 점수, 배송도착예정 등 정보가 왼쪽에 활성화 됨...이런식으로 하위 태그들이 구조화되어 있음
7. 우리의 작업은 베스트셀러 책 이름 정보만 가져오는 것임
8. 해당되는 하위 코드는 class가 'item_info'인 <div> 태그의 하위 태그 중 **class 이름이 "info_row info_name"인 <div>태그가 존재하고, 그 아래에 class 이름이 "gd_name"인 <a>태그가 존재하고, 책 제목이 텍스트(문자열로) - "초역부처의 말"** > 파이썬으로 추출할 수 있음