

수집할 정보 정의

여러 사이트에서 정보를 수집할 경우 공통 정보 정의 단계 필요

상품명, 코멘트, 기간1, 기간2, 가격, 평점, 썸네일, 링크(상품상세정보)



이 예제는 1개 사이트에서만 정보를 수집하므로 넘어감

수집 대상 정의

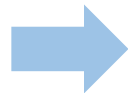
```
boxItems = driver.find_elements_by_css_selector('.oTravelBox>.boxList>li')
```

➡ boxList가 밑에서도 계속 나오므로 확실하게 하기 위해 더 상위 단계(oTravelBox)에서 끊음

너무 많아서 range(1, 2)로 축소해 한 바퀴만 돌려봄

간단하게 상품명만 뽑아 보기(이게 되면 나머지도 다 되는 거!)

```
for li in boxItems:  
    print('상품명', li.find_element_by_css_selector('h5.proTit').text)
```



상품 하나하나 접근하기 위해 for문 돌림

나머지도 이어서 해주기 (코멘트, 가격, 기간1, 기간2, 평점)

```
print('코멘트', li.find_element_by_css_selector('.proSub').text)  
print('가격', li.find_element_by_css_selector('.proPrice').text)
```

➡ 'proSub'와 'proPrice'가 하나씩만 존재하므로 이렇게 해도 됨

```
for info in li.find_elements_by_css_selector('.info-row .proInfo'):  
    print(info.text)
```

➡ 'proInfo'가 여러 개이므로 순서대로 출력해주는 것

➡ '.info-row .proInfo'와 같이 한 칸 띄면 '자손' 개념

나머지도 이어서 해주기 (썸네일)
get_attribute() : 요소의 속성값 획득

```
print('썸네일', li.find_element_by_css_selector('img').get_attribute('src'))
```

➡ 'img'가 뒤에 icon에서 한번 더 나오지만,
첫번째로 나오는 썸네일 img만 사용할 것이기 때문에 element로 해도 됨

➡ 'img'의 'src' 속성이 실제 이미지를 나타냄

추가로 생각해보기

- 이미지를 링크값을 사용할것인가? : 의존적. 파일이 지워지면 사용x
- 직접 다운로드해서 우리 서버에 업로드(ftp)할 것인가? 저작권 등의 문제

나머지도 이어서 해주기 (링크(URL))

```
print('링크', li.find_element_by_css_selector('a').get_attribute('onclick'))
```

- ➡ 'a' 태그도 뒤에 또 나오지만,
첫번째로 나오는 'a' 태그만 사용할 것이기 때문에 element로 해도 됨
- ➡ 'a' 태그의 'onclick'이 URL의 페이지로 tab을 열어주는 속성

추가로 생각해보기

- 스크립트를 구동할 것인가? : Tab이 새로 열림
- URL만 잘라내서 링크를 태울 것인가? : 현재 Tab에서 이동. Tab을 하나만 사용

=> resource 사용의 차이

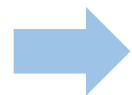
상품 정보 저장 클래스 생성

Tour.py

```
# 상품 정보를 담는 클래스
class TourInfo:
    # 멤버변수 정의 (실제 컬럼보다는 작게 세팅했음: 9 -> 5)
    title = ''
    price = ''
    area = ''
    link = ''
    img = ''
    # 생성자
    def __init__(self, title, price, area, link, img):
        # 멤버변수 초기화
        self.title = title
        self.price = price
        self.area = area
        self.link = link
        self.img = img
```

생성한 모듈 import

```
from Tour import TourInfo ✓
```



생성한 모듈과 클래스 import

```
# 상품 정보를 담은 리스트(Tourinfo 리스트 - Tourinfo 객체들을 모아서 담은 리스트)  
tour_list = [] ✓
```

※ 주의

- chromedriver.exe가 있는 driver_path와 생성한 모듈이 같은 디렉터리 안에 있어야 함!

데이터 모아서 저장

```
obj = TourInfo(  
    li.find_element_by_css_selector('h5.proTit').text,  
    li.find_element_by_css_selector('.proPrice').text,  
    li.find_elements_by_css_selector('.info-row .proInfo')[1].text,  
    li.find_element_by_css_selector('a').get_attribute('onclick'),  
    li.find_element_by_css_selector('img').get_attribute('src')  
)  
tour_list.append(obj)
```



단계(하나의 li)가 끝날때마다 데이터 모음

마무리

```
print(tour_list, len(tour_list))
```

- ➡ 리스트의 element 개수 출력해보기
- ➡ 여기까지 성공적으로 실행되면, range를 (1, 16)으로 잡고 해보기