

```

83     # 상품 하나 하나 접근
84     for li in boxItems:
85         # 이미지를 링크값을 사용했었인가?
86         # 직접 다운로드 해서 우리 서버에 업로드(ftp) 할것인가?
87         print( '썸네임', li.find_element_by_css_selector('img').get_attribute('src') )
88         print( '링크', li.find_element_by_css_selector('a').get_attribute('onclick') )
89         print( '상품명', li.find_element_by_css_selector('h5.proTit').text )
90         print( '코멘트', li.find_element_by_css_selector('p.proSub').text )
91         print( '가격', li.find_element_by_css_selector('p.proPrice').text )
92         area = ''
93         for info in li.find_elements_by_css_selector('.info-row .proInfo'):
94             print( info.text )
95         print('='*100)
96         # 데이터 모음
97         # li.find_elements_by_css_selector('.info-row .proInfo')[1].text
98         # 데이터가 부족하거나 없을수도 있으므로 직접 인덱스로 표현은 위험성이 있음
99         obj = TourInfo(
100             li.find_element_by_css_selector('h5.proTit').text,
101             li.find_element_by_css_selector('p.proPrice').text,
102             li.find_elements_by_css_selector('.info-row .proInfo')[1].text,
103             li.find_element_by_css_selector('a').get_attribute('onclick'),
104             li.find_element_by_css_selector('img').get_attribute('src')
105         )
106         tour_list.append( obj )
107     except Exception as e1:
108         print( '오류', e1 )
109
110 print( tour_list, len(tour_list) )
111
112 for tour in tour_list:
113     # tour => TourInfo
114     print( type(tour) )
115     # 링크 데이터에서 실데이터 획득
116     # 분해
117     arr = tour.link.split(',')
118     if arr:
119         # 대체
120         link = arr[0].replace('searchModule.OnClickDetail(', '')
121         # 슬라이싱 => 앞에 ' ', 뒤에 ' 제거
122         detail_url = link[1:-1]
123         # 상세 페이지 이동 : URL 값이 완성된 형태인지 확인 (http~)
124         driver.get( detail_url )
125         time.sleep(2)
126         # pip install bs4
127         # 현재 페이지를 BeautifulSoup 의 DOM으로 구성
128         soup = bs( driver.page_source, 'html.parser' )
129         # 현재 상세 정보 페이지에서 스키출 정보 획득
130         data = soup.select('.tip-cover')
131         # print( type(data), len(data), type(data[0].contents) )
132         # 디비 입력 => pip install pymysql
133         # 데이터 sum
134         content_final = ''
135         for c in data[0].contents:
136             content_final += str(c)
137
138         # html 콘텐츠 데이터 전처리 (디비에 입력 가능토록)
139         import re
140         content_final = re.sub(" ", "", content_final)
141         content_final = re.sub(re.compile(r'\\r\\n|\\r|\\n|\\n\\r+'), '', content_final)
142
143         # 콘텐츠 내용에 따라 전처리 => data[0].contents
144         tour_list = [tour.title, tour.price[:-1], tour.area.replace('출발 가능 기간 : ', ''), content_final, keyword ]
145         db.db_insertCrawlingData(tour_list)
146
147 # 게시판에서 데이터를 가져올때
148 # 데이터가 많으면 세션(혹시 로그인을 해서 접근되는 사이트일 경우) 관리
149 # 특정 단위별로 로그아웃 로그인 계속 시도
150 # 특정 게시물리 사라질 경우 => 팝업 발생 (없는 ...) => 팝업 처리 검토
151 # 게시판 스캔시 => 임계점을 모름!!
152 # 게시판 스캔 => 메타 정보 획득 => loop 를 돌려서 일괄적으로 방문 접근 처리
153
154 # searchModule.SetCategoryList(1, '') 스크립트 실행
155 # 16은 임시값, 게시물을 넣어갔을때 현상을 확인차
156 for page in range(1, 2):#16):
157     try:
158         # 자바스크립트 구동하기
159         driver.execute_script("searchModule.SetCategoryList(%s, '')" % page)
160         time.sleep(2)
161         print("%s 페이지 이동" % page)
162         #####
163         # 여러 사이트에서 정보를 수집할 경우 공통 정보 정의 단계 필요
164         # 상품명, 코멘트, 기간1, 기간2, 가격, 평점, 썸네일, 링크(상품상세정보)
165         boxItems = driver.find_elements_by_css_selector('.oTravelBox>.boxList>li')

```

## DbMgr.py

```
1 # 디버 처리, 연결, 해제, 검색어 가져오기, 데이터 삽입
2 # import pymysql as my
3 import cx_Oracle
4 import os
5
6 #오라클 라이브러리 연결 : 오라클 드라이버 등록과 같음
7 # 방법 1 : 환경변수에 등록
8 location = "C:/instantclient_19_8"
9 os.environ["PATH"] = location + ";" + os.environ["PATH"]
10
11 # 해당 데이터베이스와 연결하기 위한 코드 작성
12 # dbURL = "localhost:1521/xe"
13 # dbUser = "c##crawler"
14 # dbPass = "c##crawler"
15 #
16 # conn = cx_Oracle.connect(dbUser, dbPass, dbURL)
17 # 또는
18 # conn = cx_Oracle.connect("c##student/c##student@localhost:1521/xe")
19
20 class DBHelper:
21     """
22     멤버변수 : 커넥션
23     """
24     conn = None
25     """
26     생성자
27     """
28     def __init__(self):
29         self.db_init()
30     """
31     멤버 함수
32     """
33     def db_init(self):
34         self.conn = conn = cx_Oracle.connect("c##crawler/c##crawler@localhost:1521/xe")
35
36     def db_free(self):
37         if self.conn:
38             self.conn.close()
39
40     # 검색 키워드 가져오기 => 웹에서 검색
41     def db_selectKeyword(self):
42         # 커서 오픈
43         # with => 닫기를 처리를 자동으로 처리해준다 => 1/0 많이 사용
44         rows = None
45         with self.conn.cursor() as cursor:
46             sql = "select * from tbl_keyword"
47             cursor.execute(sql)
48             rows = cursor.fetchall()
49             print(rows)
50         return rows
51
52     def db_insertCrawlingData(self, T):
53         with self.conn.cursor() as cursor:
54             sql = "insert into tbl_crawlingdata values( cdata_seq.nextval,:1,:2,:3,:4,:5,default )"
55             cursor.execute(sql, T)
56             self.conn.commit()
57
58     # 단독으로 수행시에만 작동 => 테스트코드를 삼인해서 사용
59 if __name__ == '__main__':
60     db = DBHelper()
61     print( db.db_selectKeyword() )
62     print(db.db_insertCrawlingData(['1','2','3','4','5']))
63     db.db_free()
```

## Tour.py

```
1 # 상품 정보를 담는 클래스
2 class TourInfo:
3     # 멤버변수 (실제 컬럼보다는 작게 세팅했음)
4     title = ''
5     price = ''
6     area = ''
7     link = ''
8     img = ''
9     contents = ''
10     # 생성자
11     def __init__(self, title, price, area, link, img, contents=None):
12         self.title = title
13         self.price = price
14         self.area = area
15         self.link = link
16         self.img = img
17         self.contents = contents
```

## 콘솔창 결과

1 페이지 이동  
썸네임 [http://tourimage.interpark.com/product/tour/00161/A30/280/A3015008\\_7\\_410.jpg](http://tourimage.interpark.com/product/tour/00161/A30/280/A3015008_7_410.jpg)  
링크 `searchModule.OnClickDetail('http://tour.interpark.com/goods/detail/?BaseGoodsCd=A3015008','')`  
상품명 [4개국10일] 로마&런던야간+융프라우\_8대특식  
코멘트 10일 동안의 정통서유럽 여행  
가격 3,240,000 원~  
여행 기간 : 8박10일  
출발 가능 기간 : 2021.09.02~2021.12.30  
평점 9.7  
15개의 상품평

썸네임 [http://tourimage.interpark.com/product/tour/00161/A30/280/A3018067\\_1\\_343.jpg](http://tourimage.interpark.com/product/tour/00161/A30/280/A3018067_1_343.jpg)  
링크 `searchModule.OnClickDetail('http://tour.interpark.com/goods/detail/?BaseGoodsCd=A3018067','')`  
상품명 [스위스/남프랑스/이탈리아 12일] 융프라우VIP패스+프로방스+로마관광2일/바티칸+폴로세움  
코멘트 여유로운 여행\_서유럽  
가격 3,200,000 원~  
여행 기간 : 10박12일  
출발 가능 기간 : 2022.04.05~2022.07.19  
평점 0.0  
0개의 상품평

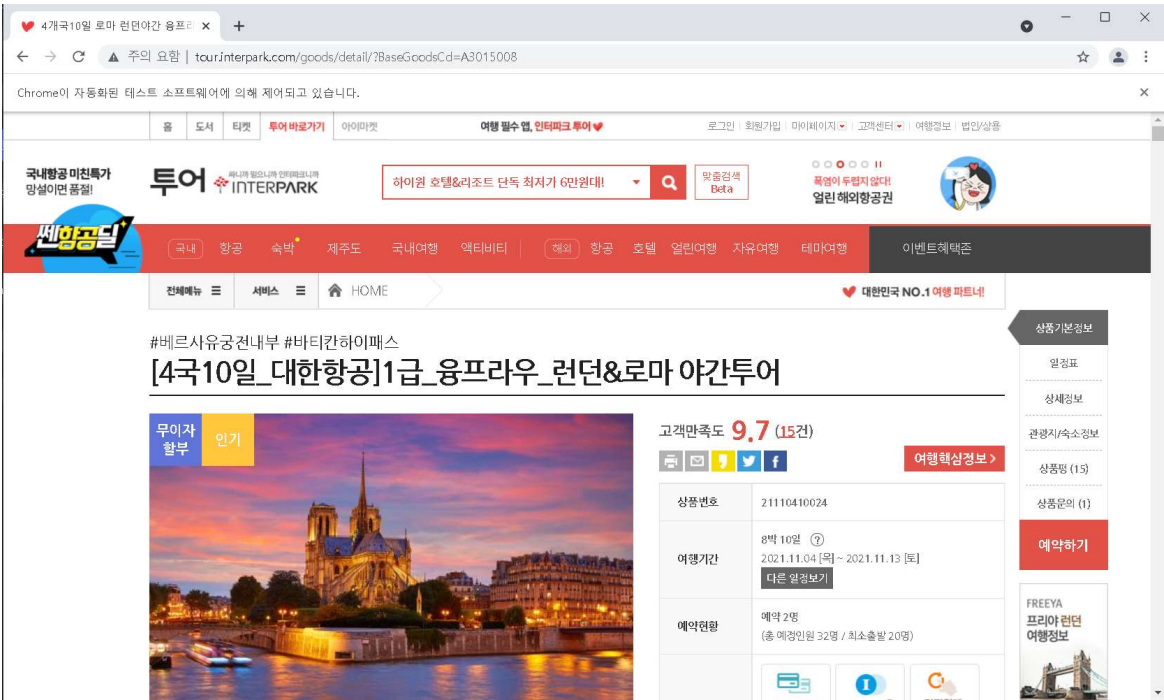
썸네임 [http://tourimage.interpark.com/product/tour/00161/A30/280/A3017982\\_1\\_870.jpg](http://tourimage.interpark.com/product/tour/00161/A30/280/A3017982_1_870.jpg)  
링크 `searchModule.OnClickDetail('http://tour.interpark.com/goods/detail/?BaseGoodsCd=A3017982','')`  
상품명 [2인이상 단독여행\_이탈리아 8일] 로마+피렌체+베네치아★항공제외  
코멘트 가이드 안내 + 자유 일정의 장점만 쏙쏙  
가격 1,600,000 원~  
여행 기간 : 6박7일  
출발 가능 기간 : 2022.12.31~2022.12.31  
평점 0.0  
0개의 상품평

썸네임 [http://tourimage.interpark.com/product/tour/00161/A30/280/A3016977\\_15\\_120.jpg](http://tourimage.interpark.com/product/tour/00161/A30/280/A3016977_15_120.jpg)  
링크 `searchModule.OnClickDetail('http://tour.interpark.com/goods/detail/?BaseGoodsCd=A3016977','')`  
상품명 [서유럽 3국 10일] 융프라우+리기산\_로마야간투어  
코멘트 2대알프스 스위스 길게보기  
가격 2,250,000 원~  
여행 기간 : 8박10일  
출발 가능 기간 : 2021.09.07~2021.12.28  
평점 0.0  
0개의 상품평

썸네임 [http://tourimage.interpark.com/product/tour/00161/A30/280/A3013963\\_23\\_130.jpg](http://tourimage.interpark.com/product/tour/00161/A30/280/A3013963_23_130.jpg)  
링크 `searchModule.OnClickDetail('http://tour.interpark.com/goods/detail/?BaseGoodsCd=A3013963','')`  
상품명 [서유럽 6국 12일] 바티칸하이패스+특식+호텔UP!  
코멘트 ★선착순★ 특별한 혜택이 팡팡!!  
가격 3,175,000 원~  
여행 기간 : 10박12일  
출발 가능 기간 : 2021.09.02~2021.12.30  
평점 8.9  
85개의 상품평

썸네임 [http://tourimage.interpark.com/product/tour/00161/A30/280/A3014604\\_8\\_990.jpg](http://tourimage.interpark.com/product/tour/00161/A30/280/A3014604_8_990.jpg)  
링크 `searchModule.OnClickDetail('http://tour.interpark.com/goods/detail/?BaseGoodsCd=A3014604','')`  
상품명 [스위스/이태리9일] 1급UP! 리기정상식+몽트뢰/프라이부르크  
코멘트 국적기 직항/루체른/몽트뢰/프라이부르크  
가격 2,840,000 원~  
여행 기간 : 7박9일  
출발 가능 기간 : 2021.09.02~2021.12.30  
평점 9.0  
82개의 상품평

[<Tour.TourInfo object at 0x00000202430847C0>, <Tour.TourInfo object at 0x0000020243084340>, <Tour.TourInfo object at 0x00000202428CBE80>, <Tour.TourInfo object at 0x00000202425C1340>, <Tour.TourInfo object at 0x00000202428B9430>, <Tour.TourInfo object at 0x00000202428B95B0>, <Tour.TourInfo object at 0x0000020243084430>, <Tour.TourInfo object at 0x00000202428B9670>, <Tour.TourInfo object at 0x00000202428B9730>, <Tour.TourInfo object at 0x00000202428B9850>] 10  
<class 'Tour.TourInfo'>



DB결과창

1	1	[4개국10일] 로마+런던야간+윙프라우_8대특석	3,240,000	원	2021.09.02-2021.12.30	인터파크	예약	<div class="default-section special" id="divGoodsSpecialInfo"><h3>인터파크
2	2	[스위스/남프랑스/이탈리아 12일] 윙프라우VIP패스+프로방스+로마관광2일/바티칸+플로렌스	3,200,000	원	2022.04.05-2022.07.19	인터파크	예약	<div class="default-section special" id="divGoodsSpecialInfo"><h3>인터파크
3	3	[2인이상 단체여행_이탈리아 9일] 로마+피렌체+베네치아★항공제외	1,600,000	원	2022.12.31-2022.12.31	인터파크	예약	<div class="default-section special" id="divGoodsSpecialInfo"><h3>인터파크
4	4	[자유형 3국 10일] 윙프라우+리가산_로마야간투어	2,250,000	원	2021.09.07-2021.12.28	인터파크	예약	인터파크
5	5	[4개국10일] 로마+런던야간+윙프라우_8대특석	3,240,000	원	2021.09.02-2021.12.30	인터파크	예약	<div class="default-section special" id="divGoodsSpecialInfo"><h3>인터파크
6	6	[프랑스/스위스/이탈리아 9일] 윙프라우/바티칸하이패스	3,040,000	원	2021.09.04-2021.12.26	인터파크	예약	<div class="default-section special" id="divGoodsSpecialInfo"><h3>인터파크
7	7	[4개국10일] 로마+런던야간+윙프라우_8대특석	3,240,000	원	2021.09.02-2021.12.30	인터파크	예약	<div class="default-section special" id="divGoodsSpecialInfo"><h3>인터파크
8	8	[스위스/남프랑스/이탈리아 12일] 윙프라우VIP패스+프로방스+로마관광2일/바티칸+플로렌스	3,200,000	원	2022.04.05-2022.07.19	인터파크	예약	<div class="default-section special" id="divGoodsSpecialInfo"><h3>인터파크
9	9	[2인이상 단체여행_이탈리아 9일] 로마+피렌체+베네치아★항공제외	1,600,000	원	2022.12.31-2022.12.31	인터파크	예약	<div class="default-section special" id="divGoodsSpecialInfo"><h3>인터파크
10	10	[자유형 3국 10일] 윙프라우+리가산_로마야간투어	2,250,000	원	2021.09.07-2021.12.28	인터파크	예약	인터파크
11	11	[자유형 10일] 1금호텔, 관광의 도시 치비타, 윙프라우, 런던, 로마야간	3,100,000	원	2021.09.02-2021.12.30	인터파크	예약	인터파크
12	12	[스위스/이탈리아 9일] 체로마트+윙프라우+몬호수+사프하우젠	2,990,000	원	2021.09.02-2021.12.30	인터파크	예약	<div class="default-section special" id="divGoodsSpecialInfo"><h3>인터파크
13	13	[자유형 6국 12일] 워킹/베르사유/3대박물관	3,025,000	원	2021.09.02-2021.12.30	인터파크	예약	<div class="default-section special" id="divGoodsSpecialInfo"><h3>인터파크
14	14	[자유형 6국 12일] 바티칸하이패스+특식+호텔UP!	3,175,000	원	2021.09.02-2021.12.30	인터파크	예약	<div class="default-section special" id="divGoodsSpecialInfo"><h3>인터파크
15	15	[스위스/이탈리아 9일] 1금UP! 리기장상식+몽테비/프라이부르크	2,840,000	원	2021.09.02-2021.12.30	인터파크	예약	<div class="default-section special" id="divGoodsSpecialInfo"><h3>인터파크
16	16	[4개국10일] 로마+런던야간+윙프라우_8대특석	3,240,000	원	2021.09.02-2021.12.30	인터파크	예약	<div class="default-section special" id="divGoodsSpecialInfo"><h3>인터파크
17	17	[스위스/남프랑스/이탈리아 12일] 윙프라우VIP패스+프로방스+로마관광2일/바티칸+플로렌스	3,200,000	원	2022.04.05-2022.07.19	인터파크	예약	<div class="default-section special" id="divGoodsSpecialInfo"><h3>인터파크
18	18	[2인이상 단체여행_이탈리아 9일] 로마+피렌체+베네치아★항공제외	1,600,000	원	2022.12.31-2022.12.31	인터파크	예약	<div class="default-section special" id="divGoodsSpecialInfo"><h3>인터파크
19	19	[자유형 3국 10일] 윙프라우+리가산_로마야간투어	2,250,000	원	2021.09.07-2021.12.28	인터파크	예약	인터파크
20	20	[자유형 10일] 1금호텔, 관광의 도시 치비타, 윙프라우, 런던, 로마야간	3,100,000	원	2021.09.02-2021.12.30	인터파크	예약	인터파크
21	21	[프랑스/스위스/이탈리아 9일] 윙프라우/바티칸하이패스	3,040,000	원	2021.09.04-2021.12.26	인터파크	예약	<div class="default-section special" id="divGoodsSpecialInfo"><h3>인터파크
22	22	[스위스/이탈리아 9일] 체로마트+윙프라우+몬호수+사프하우젠	2,990,000	원	2021.09.02-2021.12.30	인터파크	예약	<div class="default-section special" id="divGoodsSpecialInfo"><h3>인터파크
23	23	[자유형 6국 12일] 워킹/베르사유/3대박물관	3,025,000	원	2021.09.02-2021.12.30	인터파크	예약	<div class="default-section special" id="divGoodsSpecialInfo"><h3>인터파크
24	24	[스위스/이탈리아 9일] 체로마트+윙프라우+몬호수+사프하우젠	3,175,000	원	2021.09.02-2021.12.30	인터파크	예약	<div class="default-section special" id="divGoodsSpecialInfo"><h3>인터파크
25	25	[4개국10일] 로마+런던야간+윙프라우_8대특석	3,240,000	원	2021.09.02-2021.12.30	인터파크	예약	<div class="default-section special" id="divGoodsSpecialInfo"><h3>인터파크
26	26	[스위스/남프랑스/이탈리아 12일] 윙프라우VIP패스+프로방스+로마관광2일/바티칸+플로렌스	3,200,000	원	2022.04.05-2022.07.19	인터파크	예약	<div class="default-section special" id="divGoodsSpecialInfo"><h3>인터파크
27	27	[2인이상 단체여행_이탈리아 9일] 로마+피렌체+베네치아★항공제외	1,600,000	원	2022.12.31-2022.12.31	인터파크	예약	<div class="default-section special" id="divGoodsSpecialInfo"><h3>인터파크
28	28	[자유형 3국 10일] 윙프라우+리가산_로마야간투어	2,250,000	원	2021.09.07-2021.12.28	인터파크	예약	인터파크
29	29	[자유형 10일] 1금호텔, 관광의 도시 치비타, 윙프라우, 런던, 로마야간	3,100,000	원	2021.09.02-2021.12.30	인터파크	예약	인터파크
30	30	[4개국10일] 로마+런던야간+윙프라우_8대특석	3,240,000	원	2021.09.02-2021.12.30	인터파크	예약	<div class="default-section special" id="divGoodsSpecialInfo"><h3>인터파크
31	31	[프랑스/남프랑스/이탈리아 12일] 윙프라우VIP패스+프로방스+로마관광2일/바티칸+플로렌스	3,200,000	원	2022.04.05-2022.07.19	인터파크	예약	<div class="default-section special" id="divGoodsSpecialInfo"><h3>인터파크
32	32	[프랑스/스위스/이탈리아 9일] 윙프라우/바티칸하이패스	3,040,000	원	2021.09.04-2021.12.26	인터파크	예약	<div class="default-section special" id="divGoodsSpecialInfo"><h3>인터파크