

데싱디바 사이트 한 페이지 이미지 링크 추출

```
In [18]: import requests
from bs4 import BeautifulSoup
import pandas as pd

headers = {'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/58.0.3029.11

#### 1. "데싱디바 > SHOP > 네일 > 전체 페이지"에서 제품별 URL 크롤링 ####
url = 'https://dashingdiva.co.kr/category/%EB%84%A4%EC%9D%BC/82/?page=1'
response = requests.get(url, headers=headers) # 요청

if response.status_code == 200:
    soup = BeautifulSoup(response.text, "lxml")
    # print(soup)

    prdImg = soup.select('div.thumbnail > div > a[href]')
    print(len(prdImg)) # 한 페이지 당 24개 제품

    for img_link in prdImg:
        href = img_link['href'] # 이미지 링크 추출
        print(href) # 주소 앞에 'https://dashingdiva.co.kr' 붙여주기

else:
    print('페이지를 가져올 수 없습니다.')
```

24

```

/product/매직프레스-해피-크리스마스레귤러-라운드/7883/category/82/display/1/
/product/매직프레스-플러스-블랙홀코핀/6874/category/82/display/1/
/product/글레이즈-21-best-네일젤네일-33/7466/category/82/display/1/
/product/매직프레스-클라우드-베어썬-스퀘어/7278/category/82/display/1/
/product/매직프레스-메리-루돌프레귤러-라운드/7881/category/82/display/1/
/product/매직프레스-멜로우-스노우레귤러-라운드/7870/category/82/display/1/
/product/매직프레스-웨딩-링롱-오발/3873/category/82/display/1/
/product/매직프레스-쉬머-퀵팅아몬드/7894/category/82/display/1/
/product/글레이즈-홀리데이-무드/7934/category/82/display/1/
/product/매직프레스-스노우-플라워레귤러-라운드/7868/category/82/display/1/
/product/매직프레스-러브-로지레귤러-라운드/6813/category/82/display/1/
/product/매직프레스-두-낫-디스터브코핀/6080/category/82/display/1/
/product/글레이즈-모브-버건디/7347/category/82/display/1/
/product/매직프레스-워드-러브레귤러-스퀘어/537/category/82/display/1/
/product/매직프레스-오로라-드림레귤러-라운드/6666/category/82/display/1/
/product/매직프레스-크리스탈-드롭레귤러-라운드/7136/category/82/display/1/
/product/매직프레스-플러스-오아시스코핀/6873/category/82/display/1/
/product/매직프레스-체크-워머코핀/7871/category/82/display/1/
/product/글레이즈-애프터글로우/7433/category/82/display/1/
/product/매직프레스-피스-러브-해피니스코핀/6078/category/82/display/1/
/product/매직프레스-누드-글로시아몬드/7927/category/82/display/1/
/product/매직프레스-샤인-플럼코핀/5506/category/82/display/1/
/product/매직프레스-드리미-블랙코핀/7912/category/82/display/1/
/product/매직프레스-웨딩-세레모니레귤러-라운드/6085/category/82/display/1/

```

데싱디바 사이트 전체 페이지 이미지 링크 추출

```

In [3]: import requests
        from bs4 import BeautifulSoup

        headers = {'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/58.0.3029.11'}
        all_product_urls = []      # 전체 제품 링크 저장할 리스트

        ### 1~8 한 페이지씩 방문 ###
        for i in range(1,9):
            url = f'https://dashingdiva.co.kr/category/%EB%84%A4%EC%9D%BC/82/?page={i}'

            ## 이미지 링크 크롤링
            response = requests.get(url, headers=headers)      # request 요청

```

```

if response.status_code == 200:
    soup = BeautifulSoup(response.text, "lxml")
    prdImg = soup.select('div.thumbnail > div > a')
    print(f'{i}번째 페이지 제품 수 :', len(prdImg)) # 페이지 당 제품 수

    # 이미지 링크 추출
    for img_link in prdImg:
        href = img_link['href']
        href = 'https://dashingdiva.co.kr' + href
        all_product_urls.append(href) # 이미지 링크 리스트에 추가

    else:
        print('페이지를 가져올 수 없습니다.')

print()
print(f'총 제품 수는 {len(all_product_urls)}')
print('===== 제품 urls =====')
print(all_product_urls[:3])

```

```

1번째 페이지 제품 수 : 24
2번째 페이지 제품 수 : 24
3번째 페이지 제품 수 : 24
4번째 페이지 제품 수 : 24
5번째 페이지 제품 수 : 24
6번째 페이지 제품 수 : 24
7번째 페이지 제품 수 : 24
8번째 페이지 제품 수 : 12

```

총 제품 수는 180

===== 제품 urls =====

```

['https://dashingdiva.co.kr/product/매직프레스-놀이농pick-해피-크리스마스레귤러-라운드/7883/category/82/display/1/', 'https://dashingdiva.co.kr/product/글레이즈-21-best-네일젤네일-33/7466/category/82/display/1/', 'https://dashingdiva.co.kr/product/매직프레스-플러스-블랙홀코핀/6874/category/82/display/1/']

```

한 상품 정보 뽑기

```

In [60]: url = 'https://dashingdiva.co.kr/product/%EB%A7%A4%EC%A7%81%ED%94%84%EB%A0%88%EC%8A%A4-%ED%95%B4%ED%94%BC-%ED%81%AC%EB%A6%AC%E

response = requests.get(url, headers=headers) # request 요청

```

```

if response.status_code == 200:
    soup = BeautifulSoup(response.text, "lxml")

    # 종류, 제품명 추출
    product_info = soup.select('div.headingArea > h2')
    if product_info:      # 리스트에 요소가 있는지 확인
        product_info_txt = product_info[0].get_text() # 첫 번째 요소의 텍스트 추출
        product_type = product_info_txt.split(' - ')[0]
        product_name = product_info_txt.split(' - ')[1]
        print('종류 :', product_type)
        print('제품명 :', product_name)

    # 상세설명 추출
    summary = soup.select('span.summary')[0].get_text()
    print('상세 설명 :', summary)

    # 키워드 추출
    keywords = soup.select('div.headingArea > div.simple')[0].get_text()
    print('키워드 :', keywords)

    # 상품코드 추출
    product_code = soup.select('div.infoArea tr:nth-child(7) > td > span')[0].get_text()
    print('상품코드 :', product_code)

    # 상품 가격 추출
    product_price = soup.select('div.price-box > ul > li.prd-price')[0].get_text()
    print('상품 가격 :', product_price)

else:
    print('페이지를 가져올 수 없습니다.')

```

종류 : 매직프레스

제품명 : 해피 크리스마스(레귤러 라운드)

상세 설명 : 산타와 루돌프 디자인으로 크리스마스 분위기 물씬 나는 네일

키워드 : 놀이농PICK, 23.11 출시, 크리스마스, 3D엠보, 루돌프, 산타

상품코드 : MDR3W029RR

상품 가격 : 10,800원

전체 상품 정보 뽑기

```
In [4]: import requests
        from bs4 import BeautifulSoup

        product_types = []
        product_names = []
        summary_lst = []
        keywords_lst = []
        product_codes = []
        product_prices = []

        for idx, url in enumerate(all_product_urls):
            response = requests.get(url, headers=headers)    # request 요청
            if idx % 50 == 0:
                print(f'===== {idx}번째 크롤링 시작 =====')

            if response.status_code == 200:
                soup = BeautifulSoup(response.text, "lxml")

                # 종류, 제품명 추출
                product_info = soup.select('div.headingArea > h2')
                if product_info:    # 리스트에 요소가 있는지 확인
                    product_info_txt = product_info[0].get_text()    # 첫 번째 요소의 텍스트 추출
                    product_type = product_info_txt.split('-')[0]
                    product_name = product_info_txt.split('-')[1]
                    product_types.append(product_type.strip())
                    product_names.append(product_name.strip())
                else:
                    product_types.append(None)
                    product_names.append(None)

                # 상세설명 추출
                summary = soup.select('span.summary')
                if summary:
                    summary = summary[0].get_text()
                    summary_lst.append(summary)
                else:
                    summary_lst.append(None)
```

```

# 키워드 추출
keywords = soup.select('div.headingArea > div.simple')
keywords = keywords[0].get_text()
keywords_lst.append(keywords)

# 상품 코드 추출
if len(keywords)==0:
    product_code = soup.select('div.infoArea tr:nth-child(6) > td > span')
    if product_code:
        product_code = product_code[0].get_text()
        product_codes.append(product_code)
    else:
        product_codes.append(None)
else:
    product_code = soup.select('div.infoArea tr:nth-child(7) > td > span')
    if product_code:
        product_code = product_code[0].get_text()
        product_codes.append(product_code)
    else:
        product_codes.append(None)

# 상품 가격 추출
product_price = soup.select('div.price-box > ul > li.prd-price')
if product_price:
    product_price = product_price[0].get_text()
    product_prices.append(product_price[:-1])
else:
    product_prices.append(None)

else:
    print('페이지를 가져올 수 없습니다.')

print('- 크롤링 완료 -')

```

```

===== 0번째 크롤링 시작 =====
===== 50번째 크롤링 시작 =====
===== 100번째 크롤링 시작 =====
===== 150번째 크롤링 시작 =====
- 크롤링 완료 -

```

데이터프레임으로 저장, to_csv

```
In [5]: import pandas as pd
products_info = pd.DataFrame({'product_type':product_types, 'product_name':product_names,
                             'summary':summary_lst, 'keywords':keywords_lst,
                             'product_code':product_codes, 'product_price':product_prices})

products_info.to_csv('dashingdiva.csv', index=False)
```

```
In [6]: import pandas as pd
df = pd.read_csv('dashingdiva.csv')
df.info()
# summary - 실제로 없음
# keyword - 실제로 없음
# product_code - [2+1]상품은 제품코드 미포함
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 180 entries, 0 to 179
Data columns (total 6 columns):
#   Column          Non-Null Count  Dtype
---  -
0   product_type    180 non-null   object
1   product_name    180 non-null   object
2   summary         177 non-null   object
3   keywords        148 non-null   object
4   product_code    177 non-null   object
5   product_price   180 non-null   object
dtypes: object(6)
memory usage: 8.6+ KB
```

```
In [9]: df.head()
```

Out[9]:

	product_type	product_name	summary	keywords	product_code	product_price
0	매직프레스	[놀이농pick] 해피 크리스마스(레귤러 라운드)	산타와 루돌프 디자인으로 크리스마스 분위기 물씬 나는 네일	놀이농PICK,23.11 출시,크리스마스,3D엠보,루돌프,산타	MDR3W029RR	9,700원
1	글레이즈	[2+1] BEST 네일(젤네일) ~33%	NaN	33% 할인,가을 신상 2+1,신상 네일 2개 구매 시 1종 증정	NaN	29,600원
2	매직프레스 플러스	블랙홀(코핀)	블랙 마그네틱에 폴스톤을 더해 어느 각도에서 봐도 영롱한 프리미엄 코핀 네일	NaN	MOL2W047PC	9,900원
3	매직프레스	메리 루돌프(레귤러 라운드)	루돌프와 트리 드로잉에 썸 프렌치를 매치한 크리스마스 네일	23.11 출시,크리스마스,3D엠보,루돌프,트리	MDR3W019RR	9,700원
4	매직프레스	멜로우 스노우(레귤러 라운드)	다양한 겨울 오브제를 품은 귀여운 토끼 눈사람 네일	눈사람,캐릭터,3D엠보,체크	MDR3W028RR	10,800원

In [8]: `df[df.product_code.isna()]`

Out[8]:

	product_type	product_name	summary	keywords	product_code	product_price
1	글레이즈	[2+1] BEST 네일(젤네일) ~33%	NaN	33% 할인,가을 신상 2+1,신상 네일 2개 구매 시 1종 증정	NaN	29,600원
7	매직프레스	[2+1] 겨울 네일 팁	NaN	33% 할인,겨울 신상 2+1,신상 네일 2개 구매 시 1종 증정	NaN	21,600원
27	글레이즈	[2+1] 겨울 젤네일 ~33%	NaN	33% 할인,겨울신상 2+1,신상 네일 2개 구매 시 1종 증정	NaN	29,600원