

lxml 사용 기초 스크래핑(1)

[1]:

```
1 # pip install lxml 을 윈도우 cmd 창에서 실행한다.
2 # pip install requests
3 # pip install cssselect
```

[2]:

```
1 from typing import get_args
2 import requests
3 from lxml.html import fromstring, tostring
```

[3]:

```
1 def main():
2     """
3     네이버 지식인 스크래핑 메인 함수
4     """
5
6     # 세션 사용
7     session = requests.Session()
8
9     # 스크래핑 대상 URL (GET, POST)
10    response = session.get("https://kin.naver.com/search/list.nhn?que
11
12    # 신문사 링크 리스트 획득
13    urls = scrape_news_list_page(response)
14
15    # 디렉터리 확인
16    # print(urls)
17
18    # 결과 출력
19    for name, url in urls.items():
20        # url 출력
21        print(name, url)
22
23        # 파일 쓰기
24        # 생략
25
```

[4]:

```
1 def scrape_news_list_page(response):
2     # url 리스트 선언
3     urls = {}
4
5     # 태그 정보 문자열 저장
6     #print(response.content)
7     root = fromstring(response.content)
8     #print(root)
9
10    for a in root.xpath('//ul[@class="basic1"]/li/dl/dt/a[@class="_nc
11        # a 구조 확인
12        # print(a)
13
14        # a 문자열 출력
15        # print(tostring(a, pretty_print=True))
16
17        name, url = extract_conents(a)
18        # 딕셔너리 삽입
19        urls[name] = url
20
21    return urls
```

[5]:

```
1 def extract_conents(doc):
2     # 링크 주소
3     link = doc.get("href")
4     name = doc.text_content()
5     return name, link
6
```

[6]:

```
1 # 스크래핑 시작
2 if __name__ == '__main__':
3     main()
```

플라스크에서 파이썬 실행 https://kin.naver.com/qna/detail.naver?d1id=1&dirId=10402&docId=436285489&qb=7YyM7J207I2s&enc=utf8§ion=kin&rank=1&search_sort=0&spq=0 (https://kin.naver.com/qna/detail.naver?d1id=1&dirId=10402&docId=436285489&qb=7YyM7J207I2s&enc=utf8§ion=kin&rank=1&search_sort=0&spq=0)

v파이썬 파이썬 https://kin.naver.com/qna/detail.naver?d1id=1&dirId=10402&docId=433970253&qb=7YyM7J207I2s&enc=utf8§ion=kin&rank=2&search_sort=0&spq=0 (https://kin.naver.com/qna/detail.naver?d1id=1&dirId=10402&docId=433970253&qb=7YyM7J207I2s&enc=utf8§ion=kin&rank=2&search_sort=0&spq=0)

파이썬 코딩 질문 https://kin.naver.com/qna/detail.naver?d1id=1&dirId=10402&docId=433475801&qb=7YyM7J207I2s&enc=utf8§ion=kin&rank=3&search_sort=0&spq=0 (https://kin.naver.com/qna/detail.naver?d1id=1&dirId=10402&docId=433475801&qb=7YyM7J207I2s&enc=utf8§ion=kin&rank=3&search_sort=0&spq=0)

해킹 파이썬, 칼리리눅스 https://kin.naver.com/qna/detail.naver?d1id=1&dirId=11001&docId=436726953&qb=7YyM7J207I2s&enc=utf8§ion=kin&rank=4&search_sort=0&spq=0 (https://kin.naver.com/qna/detail.naver?d1id=1&dirId=11001&docId=436726953&qb=7YyM7J207I2s&enc=utf8§ion=kin&rank=4&search_sort=0&spq=0)

파이썬, 자바? https://kin.naver.com/qna/detail.naver?d1id=4&dirId=40608&docId=435292280&qb=7YyM7J207I2s&enc=utf8§ion=kin&rank=5&search_sort=0&spq=0 (https://kin.naver.com/qna/detail.naver?d1id=4&dirId=40608&docId=435292280&qb=7YyM7J207I2s&enc=utf8§ion=kin&rank=5&search_sort=0&spq=0)

파이썬 식 좀 알려주세요 https://kin.naver.com/qna/detail.naver?d1id=1&dirId=10402&docId=434384411&qb=7YyM7J207I2s&enc=utf8§ion=kin&rank=6&search_sort=0&spq=0 (https://kin.naver.com/qna/detail.naver?d1id=1&dirId=10402&docId=434384411&qb=7YyM7J207I2s&enc=utf8§ion=kin&rank=6&search_sort=0&spq=0)

파이썬 코드 https://kin.naver.com/qna/detail.naver?d1id=1&dirId=10402&docId=435106645&qb=7YyM7J207I2s&enc=utf8§ion=kin&rank=7&search_sort=0&spq=0 (https://kin.naver.com/qna/detail.naver?d1id=1&dirId=10402&docId=435106645&qb=7YyM7J207I2s&enc=utf8§ion=kin&rank=7&search_sort=0&spq=0)

요새 파이썬으로 개발하는 이유 https://kin.naver.com/qna/detail.naver?d1id=1&dirId=10402&docId=424704487&qb=7YyM7J207I2s&enc=utf8§ion=kin&rank=8&search_sort=0&spq=0 (https://kin.naver.com/qna/detail.naver?d1id=1&dirId=10402&docId=424704487&qb=7YyM7J207I2s&enc=utf8§ion=kin&rank=8&search_sort=0&spq=0)

코드업 1602 파이썬 https://kin.naver.com/qna/detail.naver?d1id=1&dirId=10402&docId=433352731&qb=7YyM7J207I2s&enc=utf8§ion=kin&rank=9&search_sort=0&spq=0 (https://kin.naver.com/qna/detail.naver?d1id=1&dirId=10402&docId=433352731&qb=7YyM7J207I2s&enc=utf8§ion=kin&rank=9&search_sort=0&spq=0)

파이썬 질문있습니다! 도와주세요 ㅠㅠ https://kin.naver.com/qna/detail.naver?d1id=1&dirId=10402&docId=433165654&qb=7YyM7J207I2s&enc=utf8§ion=kin&rank=10&search_sort=0&spq=0 (https://kin.naver.com/qna/detail.naver?d1id=1&dirId=10402&docId=433165654&qb=7YyM7J207I2s&enc=utf8§ion=kin&rank=10&search_sort=0&spq=0)

[]:

1	
---	--

[]:

1	
---	--