

## Chap 3. 항목별 내용 추출 후 다양한 형식의 파일로 저장하기

### 1. 작업 개요

지난 챕터에서는 웹 페이지에서 전체 텍스트 파일을 한꺼번에 다 추출한 것 기억하죠?

이번 시간에는 제목과 내용과 저자 등의 다양한 정보를 각각 따로 추출해서 다양한 형식의 파일로 저장하는 방법을 살펴보겠습니다. 그리고 많은 양의 데이터를 수집하기 위해 페이지를 변경하면서 조회하는 방법도 함께 살펴보겠습니다.

이렇게 하려면 HTML 코드에서 해당 부분이 어떤 태그에 들어 있는지 찾아서 각각 별도로 추출하도록 BeautifulSoup에게 시키면 됩니다.

지난 시간보다 개발자 도구를 더 많이 사용해야 하는데 개발자 도구에서 특정 태그를 찾는 연습이 많이 필요합니다~

그리고 추출된 데이터들을 xls 형식과 csv 형식과 txt 형식의 파일로 저장하는 방법을 전해 드리겠습니다. 실무에서나 데이터 분석 작업시에는 csv 형식이나 xls 형식으로 저장하는 경우도 아주 많기 때문에 이번 시간에 공부할 내용을 꼭 익혀 두세요.

#### \* 학습목표

1. 검색된 결과에서 항목별로 데이터를 추출할 수 있다.
2. 항목별로 추출된 데이터들을 표형태로 만들어서 다양한 형식의 파일로 저장할 수 있다.
3. 수집할 데이터의 양이 많을 경우 페이지를 바꾸면서 작업할 수 있다.

위의 학습 목표를 하나씩 정복해 보자구요~~



```

37 time.sleep(2)
38
39 #Step 7.Beautiful Soup 로 본문 내용만 추출하기
40 from bs4 import BeautifulSoup
41 html_1 = driver.page_source
42 soup_1 = BeautifulSoup(html_1, 'html.parser')
43
44 content_1 = soup_1.find('div','srchResultListW').find_all('li')
45 for i in content_1 :
46     print(i.get_text().replace("\n",""))
47
48 #Step 8. 총 검색 건수를 보여주고 수집할 건수 입력받기
49 import math
50 total_cnt = soup_1.find('div','searchBox pd').find('span','num').get_text()
51 print('키워드 %s (으)로 총 %s 건의 학위논문이 검색되었습니다' %(query_txt,total_cnt))
52 collect_cnt = int(input('이 중에서 몇 건을 수집하시겠습니까?: '))
53 collect_page_cnt = math.ceil(collect_cnt / 10)
54 print('%s 건을 수집하기 위해 %s 페이지를 조회합니다.' %(collect_cnt,collect_page_cnt))
55 print('=' *80)
56
57 #Step 9. 각 항목별로 데이터를 추출하여 리스트에 저장하기
58 no2 = [ ]          #번호 저장
59 title2 = [ ]       #논문제목 저장
60 writer2 = [ ]      #논문저자 저장
61 org2 = [ ]         #소속기관 저장
62 no = 1
63
64 for a in range(1, collect_page_cnt + 1) :
65
66     html_2 = driver.page_source
67     soup_2 = BeautifulSoup(html_2, 'html.parser')
68
69     content_2 = soup_2.find('div','srchResultListW').find_all('li')
70
71     for b in content_2 :
72         #1. 논문제목 있을 경우만
73         try :
74             title = b.find('div','cont').find('p','title').get_text()
75         except :

```

[ 파이썬 능력자 너도 될 수 있어~! - 서진수 저 - ]

```

76         continue
77     else :
78         f = open(ft_name, 'a' , encoding="UTF-8")
79         print('1.번호:',no)
80         no2.append(no)
81         f.write('\n'+ '1.번호:' + str(no))
82
83         print('2.논문제목:',title)
84         title2.append(title)
85         f.write('\n' + '2.논문제목:' + title)
86
87         writer = b.find('span','writer').get_text()
88         print('3.저자:',writer)
89         writer2.append(writer)
90         f.write('\n' + '3.저자:' + writer)
91
92         org = b.find('span','assigned').get_text()
93         print('4.소속기관:' , org)
94         org2.append(org)
95         f.write('\n' + '4.소속기관:' + org + '\n')
96
97         f.close( )
98
99         no += 1
100        print("\n\n")
101
102        if no > collect_cnt :
103            break
104
105        time.sleep(1)          # 페이지 변경 전 1초 대기
106
107        a += 1
108        b = str(a)
109
110    try :
111        driver.find_element(By.LINK_TEXT , '%s' %b).click()
112    except :
113        driver.find_element(By.LINK_TEXT('다음 페이지로')).click()

```

```

114 print("요청하신 작업이 모두 완료되었습니다")
115
116 # Step 10. 수집된 데이터를 xls와 csv 형태로 저장하기
117 import pandas as pd
118
119 df = pd.DataFrame()
120 df['번호']=no2
121 df['제목']=pd.Series(title2)
122 df['저자']=pd.Series(writer2)
123 df['소속(발행)기관']=pd.Series(org2)
124
125 # xls 형태로 저장하기
126 df.to_excel(fx_name,index=False, encoding="utf-8" , engine='openpyxl')
127
128 # csv 형태로 저장하기
129 df.to_csv(fc_name,index=False, encoding="utf-8-sig")
130 print('요청하신 데이터 수집 작업이 정상적으로 완료되었습니다')
*****
소스코드가 많이 길죠?
다음 단계에서 위 내용을 자세하게 설명해 드리겠습니다 ^^

```

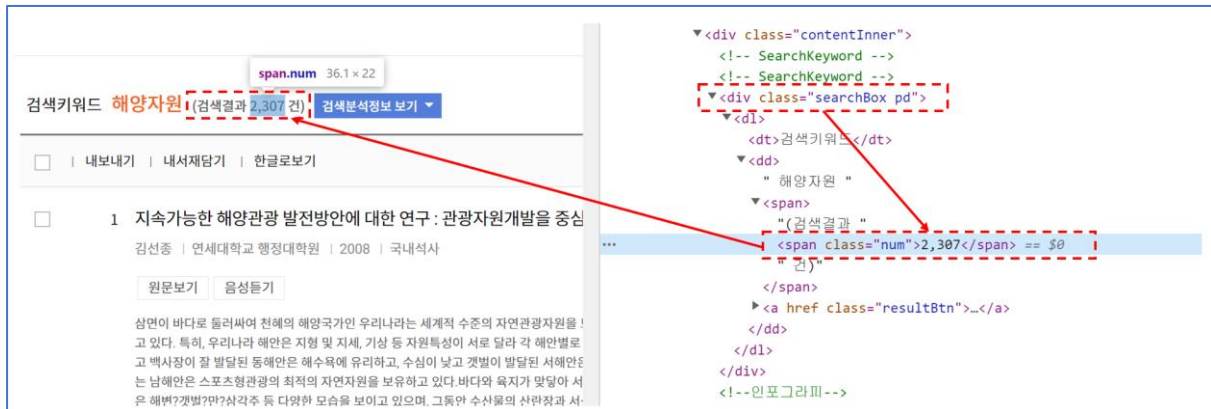
### 3. 소스 코드 설명

위 코드에서 1행에서 Step 7번의 47번 행까지는 앞에서 살펴본 내용이라서 설명은 생략하고 48번 행인 Step 8부터 설명하겠습니다.

먼저 사용자에게 주어진 키워드로 검색 한 결과가 총 몇 건인지를 보여 준 후 수집할 데이터의 건수를 입력 받도록 코드를 작성했습니다. 아래 그림에서 총 건수인 추출할 부분을 표시했습니다.



개발자 도구를 활용해서 위 그림에서 추출할 총 건수가 있는 정보를 확인해보겠습니다.



위 그림을 보면 오른쪽의 개발자 도구 부분에서 'div' 태그에 class 값이 'searchBox pd' 값 아래에 'span' 태그의 class 값이 'num' 인 값을 가져오면 총 검색된 건 수를 가져올 수 있겠죠?

그래서 아래 코드의 50번 행과 같이 코드를 작성했습니다.

그리고 52번 행에서 사용자에게 수집할 자료의 건수를 입력 받은 후 53번 행에서 페이지 번호를 계산했습니다.

#### 48 #Step 8. 총 검색 건수를 보여주고 수집할 건수 입력 받기

```

49 import math
50 total_cnt = soup_1.find('div','searchBox pd').find('span','num').get_text()
51 print('키워드 %s (으)로 총 %s 건의 학위논문이 검색되었습니다' %(query_txt,total_cnt))
52 collect_cnt = int(input('이 중에서 몇 건을 수집하시겠습니까?: '))
53 collect_page_cnt = math.ceil(collect_cnt / 10)
54 print('%s 건을 수집하기 위해 %s 페이지를 조회합니다.' %(collect_cnt,collect_page_cnt))
55 print('=' *80)

```

이제 본격적으로 데이터를 수집해야겠죠?

데이터를 수집한 후 저장을 할 때 csv, xls 형식의 경우는 표 형태로 저장이 되기 때문에 먼저 어떤 컬럼으로 저장할 것인지를 정해야 합니다.

우리가 수집할 데이터와 저장할 결과물 형태는 아래 그림과 같습니다.

검색키워드 **해양자원** (검색결과 2,307 건) 검색분석정보 보기

☐ | 내보내기 | 내서재담기 | 한글로보기

**제목** 정확도순 내림차순 10개씩 출력 조회

☐ 1 **지속가능한 해양관광 발전방안에 대한 연구 : 관광자원개발을 중심으로**

김전중 연세대학교 행정대학원 | 2008 | 국내석사

**저자** **소속기관**

삼면이 바다로 둘러싸여 천혜의 해양국가인 우리나라는 세계적 수준의 자연관광자원을 보유하고 있으며 연안을 중심으로 관광활동에 좋은 지형적 지리적 조건을 갖추고 있다. 특히, 우리나라 해안은 지형 및 지세, 기상 등 자원특성이 서로 달라 각 해안별로 독특한 자연조건을 갖고 있어 해양관광발전 잠재력이 매우 크다. 수질이 깨끗하고 백사장이 잘 발달된 동해안은 해수욕에 유리하고, 수심이 낮고 갯벌이 발달된 서해안은 생태관광의 보고이며, 기후가 온난하고 호수와 같은 잔잔한 내해를 이루고 있는 남해안은 스포츠형관광의 최적의 자연자원을 보유하고 있다. 바다와 육지가 맞닿아 서로 밀접한 영향을 주고받으면서 특수한 자원?환경시스템을 이루는 연안(沿岸)은 해변?갯벌?만?삼각주 등 다양한 모습을 보이고 있으며, 그동안 수산물의 산란장과 서식지로서 항만?임해공단 등 경제활동공간으로서 도시가 발달하고 인구가 집...

A	B	C	D
번호	제목	저자	소속(발행)기관
1	지속가능한 해양관광 발전방안에 대한 연구 : 관광자원개발을 중심으로	김전중	연세대학교 행정대학원
2	2 島嶼觀光資源開發이 海洋觀光産業에 미치는 經濟的 波及效果	김두성	부산대학교
3	3 여수지역 해양생물자원을 활용한 관광지 활성화에 관한 연구	권진구	전남대학교 대학원
4	4 해양관광유형과 관광자원 가치가 주민참여의도에 미치는 영향 : 당진시를 중심으로	이길호	경기대학교 관광전문대학원
5	5 국가관할권 이원지역 해양유전자원의 접근 및 이익공유에 관한 연구	조은영	중앙대학교 대학원
6	6 해양경제계획 이전 권역중접수역에서의 자원개발과 환경보호	이영주	연세대학교 대학원
7	7 海洋觀光資源開發에 관한 研究 : 邊山半島를 中心으로	박명엽	경기대학교 대학원
8	8 해양관광 자원 개발에 대한 연구 : 중국 칭다오와 부산을 중심으로	장영일	경기대학교 대학원
9	9 GIS기반 해양광물자원정보시스템 개발에 관한 연구	김동일	인하대학교 대학원
10	10 국가관할권 이원지역 해양생물자원의 이용과 규율 : UN해양법협약 원칙의 확충	서진웅	성균관대학교
11	11 韓·中의 海洋觀光資源 開發에 관한 연구 : 釜山 靑島를 中心으로	왕엽	동신대학교 대학원
12	12 해양관광자원의 경제적 가치평가 : 중국 대련시 부가장해수욕장을 사례로	류환청	강원대학교
13	13 해양관광자원을 활용한 제주 신양향 리모델링 계획	김예림	서울대학교 환경대학원
14	14 해양관광자원의 비시장가치 추정에 관한 연구	박미옥	동명대학교 대학원

[ 파이썬 능력자 너도 될 수 있어~! - 서진수 저 - ]

위 그림에서처럼 번호 / 제목 / 저자 / 소속(발행기관) 정보를 추출하여 표 형태로 저장해야 하기 때문에 코드에 먼저 위 컬럼을 저장할 리스트를 설정해야 하겠죠?

그래서 아래 코드와 같이 리스트를 먼저 선언을 했습니다.

### 57 #Step 9. 각 항목별로 데이터를 추출하여 리스트에 저장하기

```
58 no2 = [ ]          #번호 저장
59 title2 = [ ]       #논문제목 저장
60 writer2 = [ ]      #논문저자 저장
61 org2 = [ ]         #소속기관 저장
```

이제 각각의 정보를 추출하겠습니다.

먼저 번호는 한 건 추출할 때마다 우리가 연속적인 번호를 줄 예정이라서 62번 행에서 no 값으로 지정을 했습니다.

그리고 riss.kr 사이트의 경우 한 페이지에 검색 결과가 10건씩 출력이 되기 때문에 만약 10건이 넘어가는 게시물을 수집해야 할 경우 페이지 번호를 바꾸면서 수집 작업을 진행해야 합니다.

그래서 아래의 64번 행과 같이 반복문을 사용하여 페이지 번호를 변경할 수 있도록 먼저 구성을 하고 66번 행에서 해당 페이지에서 전체 소스코드를 가져와서 67번 행에서 파싱을 한 후 69번 행에서 게시물이 들어 있는 'li' 태그 값들을 전부 수집합니다.

```
64 for a in range(1, collect_page_cnt + 1) :
65
66     html_2 = driver.page_source
67     soup_2 = BeautifulSoup(html_2, 'html.parser')
68
69     content_2 = soup_2.find('div','srchResultListW').find_all('li')
```

이제 본격적으로 각 항목별 데이터를 수집해 볼까요?

먼저 논문 제목부터 개발자 도구로 확인해 보겠습니다.

The screenshot shows a web browser window with the search results for '해양자원' (Marine Resources). The search results are displayed in a table-like format. The first result is highlighted, showing the title '지속가능한 해양관광 발전방안에 대한 연구: 관광자원개발을 중심으로' (Research on Sustainable Marine Tourism Development: Centered on Development of Tourism Resources). The HTML source code on the right shows the structure of the search results, with the 'li' tag and its children (span, div, p) highlighted.

위 그림의 오른쪽 부분을 보면 논문 제목은 < li > 태그 아래에 < div class="cont"> 아래에 < p class="title" > 아래에 < a > 태그 안에 있는 것을 확인할 수 있습니다.



```

71     for b in content_2 :
72         #1. 논문제목 있을 경우만
73         try :
74             title = b.find('div','cont').find('p','title').get_text()
75         except :
76             continue
77         else :
78             f = open(ft_name, 'a' , encoding="UTF-8")
79             print('1.번호:',no)
80             no2.append(no)
81             f.write('\n'+ '1.번호:' + str(no))
82
83             print('2.논문제목:',title)
84             title2.append(title)
85             f.write('\n' + '2.논문제목:' + title)

```

riss.kr 사이트의 경우에는 1 페이지에 10건의 게시물이 나오기 때문에 10건의 게시물들이 저장되어 있는 li 태그값들을 전부 contents 변수에 넣은 후 for 반복문을 사용하여 한번에 1개의 게시물의 상세 정보를 추출하기 위해 71번의 for 반복문을 사용하였습니다.

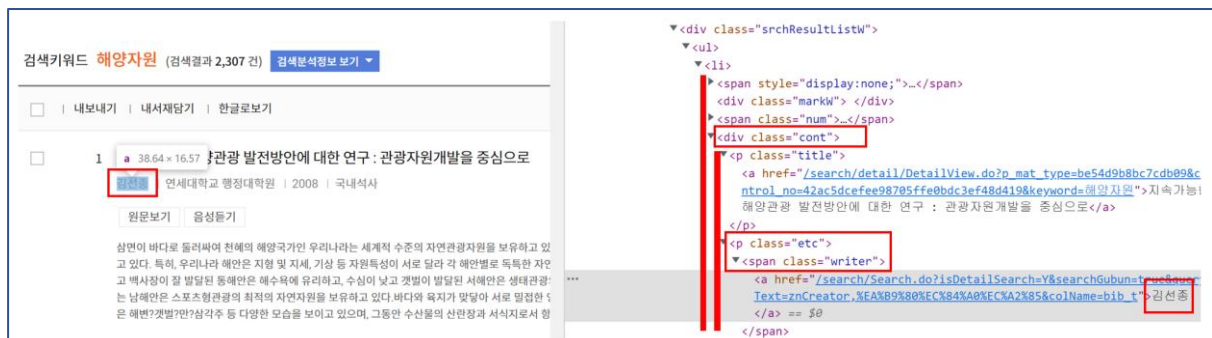
위 코드에서 71번 행의 content\_2 변수에 1 페이지에서 검색된 10건의 검색 결과가 들어 있는데 그 중에서 첫 번째 결과를 b 변수에 넣고 72번 행부터 시작되는 상세 정보를 추출하는 것입니다.

추출하는 첫 번째 데이터로 위 코드의 74번 행에서 논문 제목을 추출합니다.

그리고 78번 행에서 txt 형식의 파일을 open 한 후 80번 행에서 번호 컬럼으로 사용될 리스트에 번호를 추가하고 81번 행에서 텍스트 파일에 번호를 기록합니다.

그리고 84번 행에서 논문제목 컬럼용 리스트에 추출된 논문 제목을 추가한 후 85번 행에서 txt 파일에 논문 제목을 기록합니다.

이제 저자 정보를 확인하겠습니다.



The screenshot displays a web search result for '해양자원' (Marine Resources) on the riss.kr website. The search results show a list of items, with the first item highlighted. The HTML source code on the right shows the structure of the search results, with the title and author information highlighted in red boxes. The author information is '김선중' (Kim Seon-jung).

위 그림의 오른쪽 부분을 보면 < div class="cont"> 아래에 <p class="etc"> 아래에 < span class="writer" > 아래에 <a> 태그에 저자 정보가 있는 것이 확인 됩니다.

그래서 저자 정보를 추출하기 위해서 아래와 같이 코드를 작성했습니다.

```
87         writer = b.find('span','writer').get_text()
88         print('3.저자:',writer)
89         writer2.append(writer)
90         f.write('\n' + '3.저자:' + writer)
```

위 코드의 87번 행에서 저자 정보를 추출 한 후 88번 행에서 화면에 출력하고 89번 행에서 저자 정보용 리스트에 데이터를 추가하고 90번 행에서 txt 파일에 기록했습니다.

이번에는 소속 기관 정보를 살펴보겠습니다.

위 그림의 오른쪽 부분을 보면 소속 기관 정보는 <span class="assigned"> 값 아래에 <a> 태그 아래에 있는 것이 확인됩니다.

그래서 아래와 같이 코드를 작성했습니다.

```
92         org = b.find('span','assigned').get_text()
93         print('4.소속기관:' , org)
94         org2.append(org)
95         f.write('\n' + '4.소속기관:' + org + '\n')
```

위 코드의 92번 행에서 소속 기관 정보를 추출한 후 93번 행에서 화면에 출력하고 94번 행에서 엑셀로 저장할 리스트에 저장한 후 95번 행에서 txt 파일에 기록합니다.

이제 우리가 찾는 정보들은 다 추출했습니다.

데이터를 추출했으니 저장해야 하겠죠?

나머지 코드들이 수집된 데이터를 저장하고 페이지를 변경하는 부분입니다.

```

97         f.close( )
98
99         no += 1
100        print("\n")
101
102        if no > collect_cnt :
103            break
104
105        time.sleep(1)          # 페이지 변경 전 1초 대기
106
107        a += 1
108        b = str(a)
109
110    try :
111        driver.find_element(By.LINK_TEXT ,'%s' %b).click()
112    except :
113        driver.find_element(By.LINK_TEXT('다음 페이지로')).click()

```

위 코드의 97번 행이 txt 파일을 하드 디스크에 기록하라는 의미입니다.

그리고 102 - 103번 행은 사용자가 요청한 건수가 되면 내용이 더 있어도 수집하지 말고 바로 종료하기 위해서 작성한 코드입니다.

107-108번 코드는 페이지 번호를 만드는 부분인데 현재 a 변수에 1이라는 값이 들어 있죠?

전체 코드에서 64번 행을 보시면 됩니다

그래서 1 페이지의 내용을 수집했는데 다음 페이지로 가려면 2 페이지를 클릭해서 바꾸어야 하기 때문에 110 - 113번 행에서 사용할 페이지 번호를 문자형으로 생성하는 부분입니다.

그리고 111번 행이 다음 페이지 번호를 클릭하는 구문인데 riss.kr 사이트의 경우 1-10 페이지까지는 페이지 번호를 클릭하면 정상적으로 페이지 변경이 됩니다. 하지만 만약 11 페이지로 가야 할 경우 11 이라는 번호가 안 나와서 예러가 발생하는 데 이때 11 대신 > 표시를 클릭하도록 예외처리를 사용해서 클릭을 했습니다.

<<	1	2	3	4	5	6	7	8	9	10	>	>>
----	---	---	---	---	---	---	---	---	---	----	---	----

위 코드까지가 페이지를 반복적으로 바꾸면서 사용자가 요청한 데이터를 수집하는 부분입니다  
모든 데이터들의 수집이 끝나면 csv , xls 형태로 저장해야겠죠?

csv 파일과 xls 파일로 저장하는 부분은 아래와 같습니다.

```

116 # Step 10. 수집된 데이터를 xls와 csv 형태로 저장하기
117 import pandas as pd
118
119 df = pd.DataFrame()
120 df['번호']=no2
121 df['제목']=pd.Series(title2)
122 df['저자']=pd.Series(writer2)
123 df['소속(발행)기관']=pd.Series(org2)
124
125 # xls 형태로 저장하기
126 df.to_excel(fx_name,index=False, encoding="utf-8" , engine='openpyxl')
127
128 # csv 형태로 저장하기
129 df.to_csv(fc_name,index=False, encoding="utf-8-sig")
130 print('요청하신 데이터 수집 작업이 정상적으로 완료되었습니다')


```


위 코드의 119번 행에서 DataFrame( ) 함수로 표를 만들겠다고 선언하고 120 - 123번 행에서 표의 각 컬럼에 사용할 제목과 리스트 이름을 지정했습니다.


이때 pd.Series(titles2) 와 같이 Series( ) 함수를 사용했는데 이유는 혹시 해당 리스트의 중간에 빈칸이 있어서 DataFrame 을 생성하라는 의미입니다. Series( ) 함수를 안 쓰면 리스트의 중간에 빈칸이 있을 경우 DataFrame을 만들 수 없다고 에러가 발생합니다.

그리고 126번 행에서 xls 형태로 저장하고 129번 행에서 csv 형태로 저장을 완료했습니다.

위 코드를 실행하면 지정한 경로에 아래와 같이 파일이 저장됨을 확인할 수 있습니다.

 riss.csv

 riss.txt

 riss.xls

위 파일을 열어보면 내용도 저장이 되어 있는 것이 확인됩니다.

#### 4. 연습 문제로 실력 굳히기

1. 사용자에게 RISS 사이트의 모든 카테고리를 다 보여준 후 사용자에게 “국내학술논문” 카테고리를 선택했다고 가정하여 수집할 장르를 2번으로 입력 후 저장할 파일의 이름을 입력 받은 후 “국내학술논문” 카테고리의 다양한 정보들을 수집하여 csv, xls 형식으로 저장하도록 코드를 작성하세요. 수집할 항목은 아래의 예시를 참고하세요.

##### [사용자에게 보여줄 입력 화면 예시 화면]

```
=====
이 크롤러는 RISS 사이트의 논문 및 학술자료 수집용 웹크롤러입니다.
=====
1. 수집할 자료의 키워드는 무엇입니까?: 해양자원

2. 위 키워드로 아래의 장르 중 어떤 장르의 정보를 수집할까요?

1. 학위논문      2. 국내학술논문      3. 해외학술논문      4. 학술지
5. 단행본        6. 공개강의          7. 연구보고서

위 장르 중 수집할 장르의 번호를 입력하세요: 2

3. 결과를 저장할 csv형식의 파일명을 쓰세요(예: c:\Wpy_tempWriss_2.csv): c:\Wpy_tempWriss_2.csv
4. 결과를 저장할 xls형식의 파일명을 쓰세요(예: c:\Wpy_tempWriss_2.xls): c:\Wpy_tempWriss_2.xls
```

위 키워드로 검색하여 총 검색 건수를 아래와 같이 보여준 후 실제 수집할 건수를 입력 받아서 진행하세요.

```
해양자원 키워드로 국내학술논문 부분에서 검색된 자료의 건수는 총 3742 건 입니다
이 중에서 크롤링 할 건수는 몇건입니까?: 25
해양자원 키워드로 국내학술논문을 검색하여 총 3742 건 중 25건의 정보를 수집하겠습니다
```

##### [수집할 항목 예시]

(아래 그림처럼 번호 / 제목 / 작성자 / 소속기관 / 발표년도 / 논문집/자료집 / 논문 URL정보 를 추출하여 저장하세요)

```
1 페이지 내용 수집 시작합니다 =====
```

```
1 번째 정보를 추출하고 있습니다=====
```

```
1.번호 : 1
2.제목 : 발표논문 / 해양생물자원으로서 해조류 : 생물활성물질의 정제와 분자적 응용
3.작성자 : 홍용기(Yong Ki Hong)
4.소속기관 : 한국조류학회
5.발표년도 : 2000
6.논문집/자료집 : 국제심포지움 일정 및 발표논문집 - 21세기, 해양환경과 해양생물자원의 전망
7.논문 URL 주소: http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&control_no=508c47ca1906d5ecffe0bdc3ef48d419&keyword=해양자원
```

( 중간 내역은 생략합니다 )

25 번째 정보를 추출하고 있습니다=====

1. 번호 : 25
2. 제목 : 동아시아태평양 주요국가의 해양관리시스템 분석 : 해양질서관리와 해양자원관리를 중심으로
3. 작성자 : 주종광
4. 소속기관 : 한국해양경찰학회
5. 발표년도 : 2017
6. 논문집/자료집 : 한국해양경찰학회 학술대회
7. 논문 URL 주소 : [http://www.riss.kr/search/detail/DetailView.do?p\\_mat\\_type=1a0202e37d52c72d&control\\_no=437e1724024ae34f4884a65323211ff0&keyword=해양자원](http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&control_no=437e1724024ae34f4884a65323211ff0&keyword=해양자원)

## [저장된 결과 화면 예시] – csv / xls 모두 동일한 내용입니다

	A	B	C	D	E	F	G	H	I	J	K
1	번호	제목	저자	소속(발행)기관	날짜	학위(논문일 경우)	초록(논문일 경우)	자료URL주소			
2	1	발표논문 / 해양생물자원으로 활용하기(Yong Ki Hong)	한국조류학회		2000	국제심포지움 일정 및 발표논문집 - 21세기, 해양환경과 해양생물		<a href="http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원">http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원</a>			
3	2	수거된 해양폐기물 자원화 기 걸상인(Sang-In Keel), 한국해양환경-에너지학회			2002	한국해양환경-에너지학회지		<a href="http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원">http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원</a>			
4	3	수거된 해양폐기물 자원화 기 걸상인(Sang-In Keel), 한국해양환경-에너지학회			2002	한국해양환경-에너지학회지		<a href="http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원">http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원</a>			
5	4	해양수산생명자원의 국내 접:최석문,박수진	해양환경안전학회		2019	해양환경안전학회 학술발표대회 논문집		<a href="http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원">http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원</a>			
6	5	동아시아태평양 주요국가의 해양자원관리	한국해양경찰학회		2017	한국해양경찰학회 학술대회		<a href="http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원">http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원</a>			
7	6	21세기 신성장동력원 해양유:박수진	한국해양수산개발원		2009	해양국토21		<a href="http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원">http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원</a>			
8	7	해양관광 활성화를 위한 해양 하경회(Ha, Kyoung-H)	한국해양관광학회		2018	해양관광학연구		<a href="http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원">http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원</a>			
9	8	중국 해양자원개발의 해양강:주현희(Hyun-Hee Ju)	한국해양비즈니스학회		2014	해양비즈니스		<a href="http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원">http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원</a>			
10	9	우리나라 관할해역 내 해양자 박수진(Su Jin Park)	한국해양환경-에너지학회		2014	한국해양환경공학회 학술대회논문집		<a href="http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원">http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원</a>			
11	10	해양수산생명자원의 국내 접:최석문,박수진	한국해양환경-에너지학회		2017	한국해양환경공학회 학술대회논문집		<a href="http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원">http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원</a>			
12	11	발표논문 / 해양생물자원으로 활용하기(Yong Ki Hong)	한국조류학회		2000	국제심포지움 일정 및 발표논문집 - 21세기, 해양환경과 해양생물		<a href="http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원">http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원</a>			
13	12	수거된 해양폐기물 자원화 기 걸상인(Sang-In Keel), 한국해양환경-에너지학회			2002	한국해양환경-에너지학회지		<a href="http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원">http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원</a>			
14	13	수거된 해양폐기물 자원화 기 걸상인(Sang-In Keel), 한국해양환경-에너지학회			2002	한국해양환경-에너지학회지		<a href="http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원">http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원</a>			
15	14	해양수산생명자원의 국내 접:최석문,박수진	해양환경안전학회		2019	해양환경안전학회 학술발표대회 논문집		<a href="http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원">http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원</a>			
16	15	동아시아태평양 주요국가의 해양자원관리	한국해양경찰학회		2017	한국해양경찰학회 학술대회		<a href="http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원">http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원</a>			
17	16	21세기 신성장동력원 해양유:박수진	한국해양수산개발원		2009	해양국토21		<a href="http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원">http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원</a>			
18	17	해양관광 활성화를 위한 해양 하경회(Ha, Kyoung-H)	한국해양관광학회		2018	해양관광학연구		<a href="http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원">http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원</a>			
19	18	중국 해양자원개발의 해양강:주현희(Hyun-Hee Ju)	한국해양비즈니스학회		2014	해양비즈니스		<a href="http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원">http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원</a>			
20	19	우리나라 관할해역 내 해양자 박수진(Su Jin Park)	한국해양환경-에너지학회		2014	한국해양환경공학회 학술대회논문집		<a href="http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원">http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&amp;control_no=437e1724024ae34f4884a65323211ff0&amp;keyword=해양자원</a>			

2. 네이버 ( <https://www.naver.com> )에서 아래 예시와 같이 사용자에게 검색어와 저장할 파일들의 경로를 입력 받은 후 해당 키워드로 검색하여 나오는 결과에서 **블로그**만 선택한 후 15건의 데이터를 아래 예시와 같이 xls, csv 형식으로 저장하세요.

(편의상 검색어는 "서진수 빅데이터"로 해보세요 ^^)

연습문제 : 네이버 블로그-저장할 내용을 목록으로 만들어서 xls , csv 형식으로 저장하기

- 1.크롤링할 키워드는 무엇입니까?: 서진수 빅데이터
- 2.수집할 데이터는 몇 건입니까?: 15
- 3.csv 형태로 저장할 경로와 파일명을 입력하세요(예:c:\wpy\_temp\naver.csv): c:\wpy\_temp\naver.csv
- 4.xls 형태로 저장할 경로와 파일명을 입력하세요(예:c:\wpy\_temp\naver.xls): c:\wpy\_temp\naver.xls

[ 네이버 검색 결과 ]


**서진수 빅데이터**


통합 **VIEW** 이미지 지식iN 인플루언서 동영상 쇼핑 뉴스 어학사전 지도 ..

• 전체 • **블로그** • 카페


 어쩐지 오늘은 | 2015.10.31.

**마소캠퍼스 빅데이터 분석 오픈 세미나!(서진수 대표님) 151030**

진행한 빅데이터 분석 오픈 세미나 강연 후기입니다 진행은 서진수 대표님이... 하락 중인 데이터를 보고 어떤 인사이트를 얻을 수 있는지, 차례 차례 알려주셨습니다...




 카루의 프리랜서 라이프 | 2017.03.07.

**빅데이터 전문가 서진수 대표의 트렌드 분석 강의**

빅데이터 전문가 서진수 대표님. 빅데이터 강의는 분야를 막론하고 누구든... 미래 빅데이터를 추출하는 것은 기술적으로 그리 어렵지 않다. 오히려 데이터를 어떻게...




 한국정보인재개발원-빅데이터,파이썬,R활용,6시그마 | 2019.06.11.

**[빅데이터 전문가] 서진수 강사님을 소개합니다.**

서진수 강사님께서서는 빅데이터나 데이터베이스, 컴퓨터 프로그래밍 관련 분야 책임 R가기, R라뷰 등 14권이상 집필하였습니다. 뿐만 아니라, 대학교나 대학원, 그리고...




 1-1 | 2021.03.24.

**서진수 대표님의 강의를 듣고**

공동체 수업을 듣고 - 서진수 대표님 22019020 오늘은 빅데이터 관련 회사를... 뉴스는 선생아의 울음소리에 대한 데이터를 모아서 울음의 원인을 판단할 수 있다는...



[ 파이썬 능력자 너도 될 수 있어~! - 서진수 저 - ]

위 출력 결과에서 아래의 내용을 추출하여 파일로 저장하세요.

1.번호: 1  
 2.제목: 마소캠퍼스 빅데이터 분석 오픈 세미나!(서진수 대표님) 151030  
 3.요약내용: 진행한 빅데이터 분석 오픈 세미나 강연 후기입니다. 진행은 서진수 대표님이... 하라중인 데이터를 보고 어떤 인사이트를 얻을 수 있는지, 차례 차례 알려주셨습니다...  
 4.작성일자: 2015. 10.31.  
 5.블로그 닉네임: 어쩐지 오늘은

1.번호: 2  
 2.제목: 빅데이터 전문가 서진수 대표의 트렌드 분석 강의  
 3.요약내용: 빅데이터 전문가 서진수 대표님. 빅데이터 강의는 분야를 막론하고 누구든... 미래 빅데이터를 추출하는 것은 기술적으로 그리 어렵지 않다. 오히려 데이터를 어떻게...  
 4.작성일자: 2017. 03. 07.  
 5.블로그 닉네임: 카루의 프리랜서 라이프

1.번호: 3  
 2.제목: [빅데이터 전문가] 서진수 강사님을 소개합니다.  
 3.요약내용: 서진수 강사님께서서는 빅데이터나 데이터베이스, 컴퓨터 프로그래밍 관련 분야 책임 R짜기, R라뷰 등 14권이상 집필하였습니다. 뿐만 아니라, 대학교나 대학원, 그리고 KBS...  
 4.작성일자: 2019. 06. 11.  
 5.블로그 닉네임: 한국정보인재개발원-빅데이터,파이썬,R활용,6시그마

( 지면 관계상 이하 내용은 생략합니다 )

#### [ csv / xls 형식으로 저장된 파일 예시 ]

	A	B	C	D	E
1	번호	제목	내용	작성일자	블로그닉네임
2	1	마소캠퍼스 빅데이터 분석 오픈 세미나!(서진수 대표님)	진행한 빅데이터 분석 오픈 세미나	2015.10.31.	어쩐지 오늘은
3	2	빅데이터 전문가 서진수 대표의 트렌드 분석	빅데이터 전문가 서진수 대표님	2017.03.07.	카루의 프리랜서 라이프
4	3	[빅데이터 전문가] 서진수 강사님을 소개합니다	서진수 강사님께서서는 빅데이터나 데이터베이스, 컴퓨터 프로그래밍 관련 분야 책임 R짜기, R라뷰 등 14권이상 집필하였습니다. 뿐만 아니라, 대학교나 대학원, 그리고 KBS...	2019.06.11.	한국정보인재개발원-빅데이터,파이썬,R활용,6시그마
5	4	서진수 대표님의 강의를 듣고	공동체 수업을 듣고 - 서진수 대	2021.03.24.	1-1
6	5	소셜포럼] 서진수 대표, 변화의 흐름을 이끄는 창의와 융합의 신(新) 물결 '4차		2017.08.24.	위클리피플 [명사와의 만남] _신지식인 소셜포럼
7	6	방법, 빅데이터 관리가 가장 기본 <(주)컨시어 된 빅데이터 기술 활용법	알린다	2017.09.25.	월간 인터뷰 (Interview)
8	7	빅데이터 활용으로 4차산업시대를 준비하자_스물데이터만으로도 충분히 고		2017.02.28.	위드앤교육컨설팅_퍼실리테이션 전문교육
9	8	빅데이터 전문가 서진수 강사님의 빅데이터	서진수 강사님의 빅데이터 관련	2017.02.27.	한결다운교육연구소
10	9	빅데이터 분석, 웹크롤링 강남 아이티윌과 서	오늘은 강남 아이티윌과 서진수	2017.08.08.	우수훈련기관 아이티윌
11	10	서진수 강사님의 빅데이터-R을 활용한 텍스트	서진수 강사님은 빅데이터 전문	2017.07.25.	IT취업 나도 할 수 있다.

이번 챕터의 내용도 열심히 공부하셔서 꼭 실력으로 만드세요!!



데이터쟁이 서진수가 여러분을 응원합니다!!

[ 파이썬 능력자 너도 될 수 있어~! - 서진수 저 - ]