

Chap 4. 상세 정보 수집 후 다양한 형식의 파일로 저장하기

1. 작업 개요

지난 챕터까지는 특정 키워드로 검색된 요약 정보를 추출했는데 이번 시간에는 검색된 결과의 상세 정보까지 수집하여 다양한 형식의 파일로 저장하는 방법을 학습합니다.

그림으로 이번 시간에 학습할 내용을 살펴볼까요?

먼저 riss.kr 사이트에 자동 접속한 후 “전염병” 키워드로 자동 검색한 후 “학위논문” 메뉴를 클릭하세요. 검색된 결과에서 아래 그림에서 표시한 정보들을 수집하겠습니다.

검색키워드 **전염병** (검색결과 1,447 건) [검색분석정보 보기](#)

☐ | [내보내기](#) | [내서재담기](#) | [한글로보기](#) **제목** [정확도순](#) [내림차순](#) [10개씩 출력](#) [조회](#)

☐ 1 **코로나(COVID-19) 등 호흡기전염병 병원 내 감염 예방을 위한 표준개발**

황창의 **계명대학교대학원** **2021** **국내박사**

저자 **소속기관** **발표년도** **학위**

예로부터 전염병은 인류의 건강을 위협해 왔다. 2000년대 이래 5차례 유행한 전염병이 발생하였다. 이러한 전염병은 발생되는 간격이 점점 짧아지고 발생 지속시간은 점점 길어지고 있었다. 기후 및 환경의 변화와 갈수록 활발해지는 인류의 사회활동 등으로 인해 전염병이 다시 발생하는 것은 불가피한 상황이다. 이렇게 발생한 5번의 전염병 중 4번은 호흡기전염병이었으며, 1번은 소화기 전염병이었다. 호흡기 감염병은 소화기 감염병과 비교하여 확진자 및 사망자가 훨씬 많았으며 공기, 비말 등을 통해 전파될 수 있기 때문에 위험도가 가장 높았다. 이러한 전염병이 발생한 후, 병원 내 감염에 대한 병원 공간 분야의 예방이 완벽되지 않았기 때문에 병원 내 감염이 발생하고 병원의 의료기능이 일시적으로 상실될 수 있었다. 호흡기전염병 병원 내 감염 문제를 해결하기 위하여, 본 논문은 COVID-19 등 호흡기전염병의 병원 내 감염 ...

이번에 수집할 초록 정보는 위 게시글의 제목을 클릭한 후 상세 정보에서 초록 정보를 수집해야 합니다.

http://www.riss.kr/link?id=T15788444

저자 **황창의** [연구자관계분석](#)

발행사항 **대구: 계명대학교대학원, 2021**

학위논문사항 **학위논문(박사)- 계명대학교대학원: 건축학과 융합건축학과 2021. 2**

발행연도 **2021**

작성언어 **한국어**

주제어 **호흡기전염병; 병원 내 감염; 예방; 표준개발**

발행국(도시) **대구**

형태사항 **x ii, 262; 26 cm**

일반주기명 **지도교수: 이종국**

UCI식별코드 **I804:22003-000000119282**

소장기관 **계명대학교 동산도서관**

673 **714** **44**

상세조회 **다운로드** **내보내기**

[원문보기](#) [음성듣기](#) [인용하기](#) [내보내기](#) [내서재담기](#) [공유하기](#) [오류접수](#)

부가정보 **초록**

국문 초록 (Abstract) [koko i 다국어번역](#)

예로부터 전염병은 인류의 건강을 위협해 왔다. 2000년대 이래 5차례 유행한 전염병이 발생하였다. 이러한 전염병은 발생되는 간격이 점점 짧아지고 발생 지속시간은 점점 길어지고 있었다. ...


이렇게 수집한 데이터는 아래 예시와 같이 화면에 먼저 출력을 한 후 파일로 저장합니다.

[화면 출력 예시]



```
1 번째 정보를 추출하고 있습니다=====
1.번호 : 1
2.제목 : 코로나(COVID-19) 등 호흡기전염병 병원 내 감염 예방을 위한 표준개발
3.작성자 : 황창의
4.소속기관 : 계명대학교대학원
5.발표년도 : 2021
6.학위여부 : 국내박사
7.초록내용 : 예로부터 전염병은 인류의 건강을 위협해 왔다.2000년대 이래 5차례 유행한 전염병이 발생하였다. 이러한 전염병은 발생하는 간
격이 점점 짧아지고 발생 지속시간은 점점 길어지고 있었다. ...
8.논문 URL 주소: http://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cdb09&control_no=aea8ae1ec75c689effe0bdc3ef48d
419&keyword=전염병
```

수집된 데이터들은 아래 그림과 같이 자동으로 "사이트명-수집날짜와시간-카테고리 " 형식으로 폴더를 생성한 후 그 폴더 아래에 csv , xls 형식으로 저장됩니다

[자동 생성 폴더 예시]

 RISS-2022-03-19-10-52-00-학위논문

위 폴더 안에 아래의 형식으로 파일로 저장됩니다.

 RISS-2022-03-19-10-52-00-학위논문.csv
 RISS-2022-03-19-10-52-00-학위논문.xls

[파일 저장 결과 예시]

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	번호	제목	저자	소속(발행)기관	날짜	학위(논문일경우)	초록(논문일경우)	자료URL주소					
2	1	코로나(COVID-19) 등 호흡기전염병 예방을 위한 표준개발	황창의	계명대학교대학원	2021	국내박사	예로부터 전염병은 인류의 건강을 위협해 왔다.2000년대 이래 5차례 유행한 전염병이 발생하였다. 이러한 전염병은 발생하는 간격이 점점 짧아지고 발생 지속시간은 점점 길어지고 있었다. ...	http://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cd					
3	2	전염병에 대한 기록과 견해와 고찰	이성훈	칼빈대학교 신학	2021	국내석사	초록이 없습니다	http://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cd					
4	3	전염병 위험지각이 관광행동의도에 미치는 영향	김혜진	세종대학교 대학	2021	국내박사	관광은 관광객에게 재화	http://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cd					
5	4	국제보건규칙 개정에 따른 우리나라의 대응	김윤호	경북대학교 보건	2007	국내석사	국제보건규칙 개정에 따	http://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cd					
6	5	전염병과 팬데믹 상황에서 미디어의 역할	이재준	칼빈대학교 신학	2021	국내석사	초록이 없습니다	http://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cd					
7	6	전염병으로 인한 사회적 변화로서 제1박연경	부산대학교	2021	국내석사	과거부터 전염병은 정치	http://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cd						
8	7	전염병 모델의 연대기적 분석과 후속	이상원	高麗大學校 大學	2012	국내석사	전염병은 고대 시대부터	http://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cd					
9	8	항공운송을 통해 전파되는 전염병의 특성	최성희	韓國外國語大學	2012	국내석사	Today, airplanes and ai	http://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cd					
10	9	고등학생의 전염병 관련 과학적 소양	강보경	廣南大學校	2020	국내석사	과학적 소양(Scientific	Li http://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cd					
11	10	우리나라 법정전염병 관리실태 및 예	심상숙	경희대학교 대학	1994	국내석사	In modern countries, c	http://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cd					
12	11	국제법상 전염병 통제에 관한 연구	박진아	고려대학교 대학	2012	국내박사	The effective control of	http://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cd					
13	12	학령전 아동의 위한 호흡기전염병 예	김일옥	梨花女子大學校	1999	국내박사	There are many critical	http://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cd					
14	13	조선총독부의 급성전염병 예방 대책	백선래	한양대학교 대학	2021	국내박사	이 논문에서는 일본이	http://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cd					
15	14	Nvivo 12를 활용한 항공산업 전염병	황예린	한국교통대학교	2020	국내석사	항공 교통의 발달로 국	http://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cd					
16	15	코로나 19에 대한 정보과잉인지와 보	최요재	국민대학교 일반	2021	국내석사	본 연구에서는 미디어	http://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cd					
17	16	Timeliness and trade-off between tin	Hyo-Soon	서울대학교 대학	2012	국내박사	서론: 최근 국외여행의	http://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cd					
18	17	전염병 유행기의 비대면 자원봉사활동	박지훈	한양대학교 공공	2021	국내석사	초록이 없습니다	http://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cd					
19	18	階層分析的 意思決定方法에 의한 疫	이민희	건양대학교	2009	국내석사	본 연구에서는 보건 환	http://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cd					
20	19	BSL-3 인수공통전염병 연구시설의 개	박희문	서울산업대학교	2010	국내석사	우리나라 축산업은 국	http://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cd					
21	20	해외유입 전염병 신속대응을 위한 감	박경운	송실대학교 정보	2011	국내석사	초록이 없습니다	http://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cd					

이번 시간에 배우는 기술 중 가장 핵심 기술은 특정 제목이나 링크를 눌러서 나오는 새로운 창에서 정보를 추출하고 이전 페이지로 다시 돌아가는 부분입니다.

그리고 자동으로 조회하는 날짜와 시간으로 폴더를 생성하여 파일을 저장하는 기법도 아주 중요하니까 꼭 열심히 공부해주세요.

[파이썬 능력자 너도 될 수 있어~! - 서진수 저 -]

2. 전체 소스 코드

```

1  # Chap 15.riss.kr 에서 특정 키워드로 논문 / 학술 자료 검색하기
2
3  #Step 1. 필요한 모듈을 로딩합니다
4  from selenium import webdriver
5  from selenium.webdriver.common.by import By
6  from selenium.webdriver.common.keys import Keys
7  from selenium.webdriver.chrome.service import Service
8  import time
9
10 #Step 2. 사용자에게 검색 관련 정보들을 입력 받습니다.
11 print("=" *100)
12 print(" 이 크롤러는 RISS 사이트의 논문 및 학술자료 수집용 웹크롤러입니다.")
13 print("=" *100)
14 query_txt = input('1.수집할 자료의 키워드는 무엇입니까?: ')
15
16 #Step 3. 수집된 데이터를 저장할 파일 이름 입력 받기
17 f_dir = input("2.파일을 저장할 폴더명만 쓰세요(기본값:c:\py_temp\):")
18 if f_dir == "":
19     f_dir="c:\py_temp\"
20
21 #Step 4. 크롬 드라이버 설정 및 웹 페이지 열기
22 s = Service("c:/py_temp/chromedriver.exe")
23 driver = webdriver.Chrome(service=s)
24
25 url = 'http://www.riss.kr/'
26 driver.get(url)
27 driver.maximize_window()
28 time.sleep(2)
29
30 #Step 5. 자동으로 검색어 입력 후 조회하기
31 element = driver.find_element(By.ID,'query')
32 element.send_keys(query_txt)
33 element.send_keys(Keys.ENTER)
34
35 #Step 6. 학위 논문 선택하기
36 driver.find_element(By.LINK_TEXT,'학위논문').click()
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 #Step 8. 총 검색 건수를 보여주고 수집할 건수 입력 받기

45 import math

46 total_cnt = soup_1.find('div','searchBox pd').find('span','num').get_text()

47 print('키워드 %s (으)로 총 %s 건의 학위논문이 검색되었습니다' %(query_txt,total_cnt))

48 cnt = int(input('이 중에서 몇 건을 수집하시겠습니까?: '))

49 page_cnt = math.ceil(cnt / 10)

50 print('%s 건의 데이터를 수집하기 위해 %s 페이지의 게시물을 조회합니다.' %(cnt,page_cnt))

51 print("\n")

52

53 #Step 9. 데이터 수집하기

54 no2=[] # 게시글 번호 컬럼

55 title2=[] # 게시글 제목 컬럼

56 author2=[] # 논문 저자 컬럼

57 company2=[] # 소속 기관 컬럼

58 date2=[] # 게시글 날짜 컬럼

59 suksa2=[] # 국내석사 컬럼

60 contents2=[] # 초록내용

61 full_url2=[] # 논문 원본 URL

62

63 no = 1 # 게시글 번호 초기값

64

65 for a in range(1,page_cnt+1) :

66 print("\n")

67 print("%s 페이지 내용 수집 시작합니다 ======" %a)

68

69 html = driver.page_source

70 soup = BeautifulSoup(html, 'html.parser')

71 content_list = soup.find('div','srchResultListW').find_all('li')

72

73 for i in content_list:

74 # 논문 제목 체크하기

75 try:

76 title=i.find('p','title').get_text().strip()

[파이썬 능력자 너도 될 수 있어~! - 서진수 저 -]

```

77     except :
78         continue
79     else :
80         # 1.게시글 번호
81         print("\n")
82         print("%s 번째 정보를 추출하고 있습니다======" %no)
83         no2.append(no)
84         print("1.번호 : %s" %no)
85
86         # 2. 논문 제목
87         title2.append(title.strip())
88         print("2.제목 : %s" %title.strip())
89
90         # 3. 작성자
91         try :
92             author=i.find('p','etc').find('span','writer').get_text().strip()
93         except :
94             author = '작성자가 없습니다'
95             print("3.작성자 : %s" %author.strip())
96             author2.append(author.strip())
97         else :
98             author2.append(author.strip())
99             print("3.작성자 : %s" %author.strip())
100
101         # 4. 소속기관
102         try :
103             company=i.find('p','etc').find('span','assigned').get_text().strip()
104         except :
105             company='소속 기관이 없습니다'
106             company2.append(company.strip())
107             print("4.소속기관 : %s" %company.strip())
108         else :
109             company2.append(company.strip())
110             print("4.소속기관 : %s" %company.strip())
111
112         # 5. 발표날짜
113         try :
114             date_1 =i.find('p','etc').find_all('span')
115             date_2 = date_1[2].get_text().strip()

```

[파이썬 능력자 너도 될 수 있어~! - 서진수 저 -]

```

116         except :
117             date_2='발표날짜가 없습니다'
118             date2.append(date_2)
119             print("5.발표년도 : %s" %date_2)
120         else :
121             date2.append(date_2)
122             print("5.발표년도 : %s" %date_2)
123
124         # 6.학위여부
125         try :
126             suksa_1 =i.find('p','etc').find_all('span')
127             suksa_2 = suksa_1[3].get_text().strip()
128         except :
129             suksa_2='학위가 없습니다'
130             suksa2.append(suksa_2)
131             print("6.학위여부 : %s" %suksa_2)
131         else :
132             suksa2.append(suksa_2)
133             print("6.학위여부 : %s" %suksa_2)
134
135         # 7.초록 내용-해당 논문의 상세 내역에서 추출할 수 있음.
136         url_1 = i.find('p','title').find('a')['href']
137         full_url = 'http://www.riss.kr'+url_1
138         time.sleep(1)
139         driver.get(full_url)
140
141         html_1 = driver.page_source
142         soup_1 = BeautifulSoup(html_1, 'html.parser')
143         try :
144             cont=soup_1.find("div","text").find('p').get_text().replace("\n","").strip()
145         except :
146             cont='초록이 없습니다'
147             contents2.append(cont)
148             print("7.초록내용 : %s" %cont)
149         else :
150             contents2.append(cont)
151             print("7.초록내용 : %s" %cont)
152
153         time.sleep(1)

```

[파이썬 능력자 너도 될 수 있어~! - 서진수 저 -]

```

154
155         # 8.논문 url 주소
156         full_url2.append(full_url)
157         print('8.논문 URL 주소:' , full_url)
158
159         driver.back() # 이전 페이지로 돌아가기
160
161         time.sleep(2)
162
163         no += 1
164
165         if no > cnt :
166             break
167
168     a += 1
169     b = str(a)
170
171     try :
172         driver.find_element(By.LINK_TEXT , '%s' %b).click()
173     except :
174         driver.find_element(By.LINK_TEXT,'다음 페이지로').click()
175
176 print("요청하신 작업이 모두 완료되었습니다")
177
178 # Step 10. 수집된 데이터를 xls와 csv 형태로 저장하기
179 # 현재 날짜와 시간으로 폴더 만들고 파일 이름 설정하기
180 import os
181
182 n = time.localtime()
183 s = '%04d-%02d-%02d-%02d-%02d-%02d' %(n.tm_year, n.tm_mon, n.tm_mday, n.tm_hour, n.tm_min, n.tm_sec)
184
185 os.makedirs(f_dir+'RISS'+ '-' +s+ '-' +'학위논문')
186
187 fc_name = f_dir+'RISS'+ '-' +s+ '-' +'학위논문'+ 'WW'+ 'RISS'+ '-' +s+ '-' +'학위논문'+ '.csv'
188 fx_name = f_dir+'RISS'+ '-' +s+ '-' +'학위논문'+ 'WW'+ 'RISS'+ '-' +s+ '-' +'학위논문'+ '.xls'
189
190 # 데이터 프레임 생성 후 xls , csv 형식으로 저장하기
191 import pandas as pd
192

```

```

193 df = pd.DataFrame()
194 df['번호']=no2
195 df['제목']=pd.Series(title2)
196 df['저자']=pd.Series(author2)
197 df['소속(발행)기관']=pd.Series(company2)
198 df['날짜']=pd.Series(date2)
199 df['학위(논문일경우)']=pd.Series(suksa2)
200 df['초록(논문일경우)']=pd.Series(contents2)
201 df['자료URL주소']=pd.Series(full_url2)
202
203 # xls 형태로 저장하기
204 df.to_excel(fx_name,index=False, encoding="utf-8" , engine='openpyxl')
205
206 # csv 형태로 저장하기
207 df.to_csv(fc_name,index=False, encoding="utf-8-sig")
208
209 print('요청하신 데이터 수집 작업이 정상적으로 완료되었습니다')
*****
소스코드의 내용이 제법 긴데 다음 절에서 중용한 부분을 자세하게 설명하겠습니다.

```


3. 주요 소스 코드 설명

소스 코드의 내용이 길고 앞에서 배웠던 중복된 내용도 많이 있어서 지면 관계상 전체 코드를 다 설명하지 않고 다른 챕터에서 언급되지 않았던 중요한 부분만을 발췌해서 설명하겠습니다.

전체 소스코드는 제가 제공해 드리는 소스코드를 참고해 주세요.

```

53 #Step 9. 데이터 수집하기
54 no2=[]          # 게시글 번호 컬럼
55 title2=[ ]      # 게시글 제목 컬럼
56 author2=[]      # 논문 저자 컬럼
57 company2=[ ]    # 소속 기관 컬럼
58 date2=[ ]       # 게시글 날짜 컬럼
59 suksa2=[ ]      # 국내석사 컬럼
60 contents2=[]    # 초록내용
61 full_url2=[]    # 논문 원본 URL
62
63 no = 1          # 게시글 번호 초기값
64
65 for a in range(1,page_cnt+1) :
66     print("\n\n")
67     print("%s 페이지 내용 수집 시작합니다 ======" %a)
68
69     html = driver.page_source
70     soup = BeautifulSoup(html, 'html.parser')
71     content_list = soup.find('div','srchResultListW').find_all('li')
72
73     for i in content_list:
74         # 논문 제목 체크하기
75         try:
76             title=i.find('div','cont').find('p','title').get_text().strip()
77         except :
78             continue

```

위 코드의 54번 행에서 61번 행까지는 csv, xls 형식으로 저장할 때 필요한 컬럼으로 사용할 리스트를 미리 선언하는 부분입니다. 63번 행은 글 번호를 지정하기 위해 사용할 변수를 선언한 부분입니다.

중요한 부분이 65번 행부터입니다.

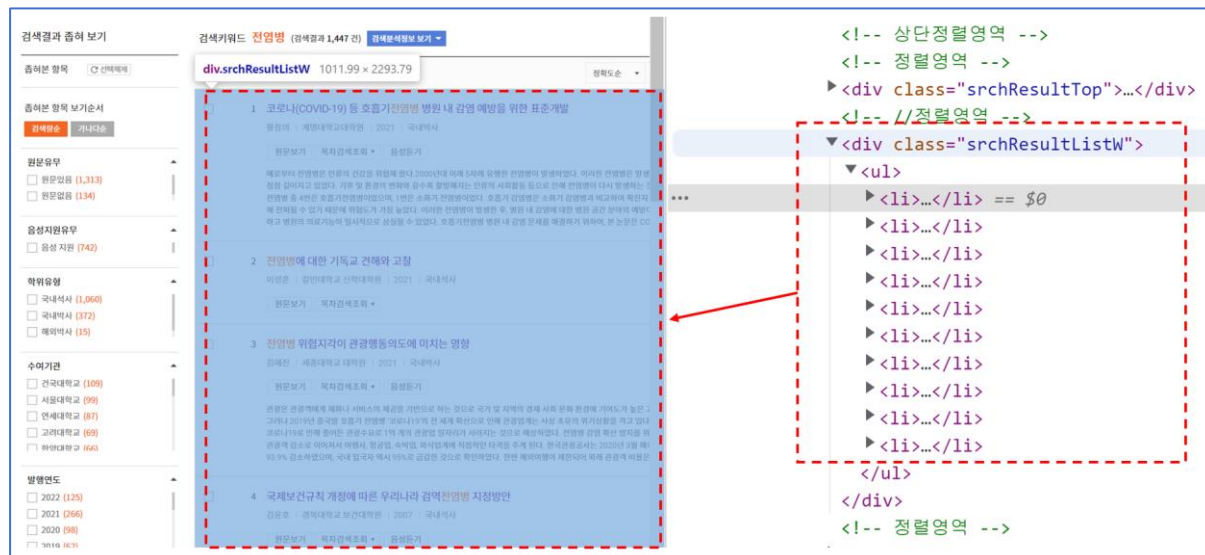
riss.kr 사이트는 한 페이지에 10건의 게시글이 검색되어 출력되므로 만약 사용자가 수집을 요청한 건수가 10건을 넘을 경우 페이지 번호를 계산하여 page_cnt 변수에 넣어 두었습니다.

(페이지 번호 계산 부분은 소스코드의 49번 행을 참고해주세요)

[파이썬 능력자 너도 될 수 있어~! - 서진수 저 -]

69번 행에서 현재 페이지의 전체 소스코드를 모두 가져 온 후 70번 행에서 파싱을 하여 soup 변수에 저장하고 71번 행에서 검색 결과가 저장된 태그 정보를 find() 함수와 find_all() 함수로 추출하였습니다.

아래 그림에서 검색 결과를 저장하고 있는 태그 이름과 클래스 정보를 확인할 수 있습니다.

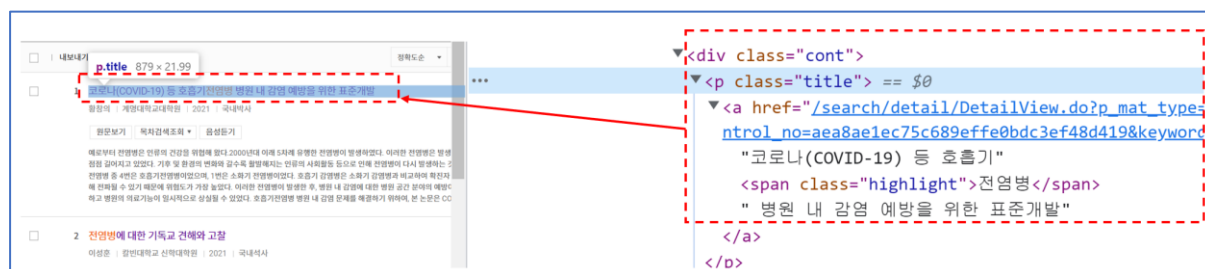


71번 행을 수행하면 content_list 변수에 첫 페이지에 있는 10건의 검색 결과를 저장한 li 태그가 모두 들어가게 됩니다.

그 후 73번 행의 for 반복문을 사용해서 content_list 변수에 있는 첫 번째 li 태그부터 i 변수에 넣고 75번 행 이후부터 나오는 상세 내역들을 추출하는 작업들을 반복적으로 진행하게 됩니다.

이제 각 항목별로 상세 내역을 추출해야 하겠죠?

첫 번째로 제목을 추출하겠습니다.



위 그림을 보면 < div class="cont"> 아래의 < p class="title" > 값 아래에 논문 제목이 있는 것을 확인할 수 있습니다. 그래서 소스코드의 76번 행에서 이 부분의 데이터를 추출하도록 작성한 것을 확인할 수 있습니다.

그런데 제목을 추출할 때 예외처리를 작성한 부분이 보이죠?

이 부분은 riss.kr 사이트에서 논문 제목을 추출할 때 2번째 제목에서 오류가 발생하기 때문에 예외처리를 사용해서 오류가 발생하면 건너 뛰도록 설정을 한 것입니다.

예외처리 부분을 생략하고 제목을 돌려 보면 이 부분을 확인할 수 있습니다.

두 번째로 논문의 저자 정보를 추출하겠습니다.

아래 그림을 보세요.



위 그림을 보면 저자 정보가 <p class="etc"> 아래에 아래에 있는 것을 확인할 수 있습니다.

그래서 소스코드의 90번 행부터 99번 행까지와 같이 작성했습니다.

```

90 # 3. 작성자
91     try :
92         author=i.find('p','etc').find('span','writer').get_text().strip()
93     except :
94         author = '작성자가 없습니다'
95         print("3.작성자 : %s" %author.strip())
96         author2.append(author.strip())
97     else :
98         author2.append(author.strip())
99         print("3.작성자 : %s" %author.strip())

```

위 코드의 92번 행에서 저자 정보를 추출하는 데 가끔 저자 정보가 없는 경우가 있어서 예외처리를 사용하여 94번 행부터 96번 행까지 지정을 했습니다.

그리고 저자 정보를 추출한 후 98번 행에서 저자 정보를 저장하는 리스트에 추가 한 후 99번 행에서 저자 정보를 화면에 출력하였습니다.

세 번째로 소속 기관 정보를 추출하겠습니다.

먼저 아래 그림을 보세요.



위 그림을 보면 소속 기관 정보는 <p class="etc"> 아래에 아래에 있는 것이 확인됩니다.

그래서 소스 코드의 101번 행에서 110번 행처럼 작성했습니다.

```

101      # 4. 소속기관
102      try :
103          company=i.find('p','etc').find('span','assigned').get_text().strip()
104      except :
105          company='소속 기관이 없습니다'
106          company2.append(company.strip())
107          print("4.소속기관 : %s" %company.strip())
108      else :
109          company2.append(company.strip())
110          print("4.소속기관 : %s" %company.strip())

```

간혹 소속 기관 정보가 없는 곳이 있어서 예외 처리를 사용했고 소속 기관이 없을 경우 105번 행부터 107번 행까지 작업하고 소속 기관 정보를 추출했을 경우 109번 행에서 소속 기관 정보를 저장하는 리스트에 추가한 후 110번 행에서 화면에 출력하였습니다.

네 번째로 발표년도 정보와 학위 정보를 추출하겠습니다.

먼저 아래 그림을 보세요.



위 그림을 보면 발표년도와 학위 정보는 <p class="etc"> 아래의 3번째와 4번째 태그에 있는 것을 알 수 있습니다. 이럴 경우 <p class="etc"> 아래의 모든 태그 값을 find_all() 함수로 다 가져온 후 그 중에서 3번째 값을 추출하면 발표년도가 되고 4번째 값을 추출하면 학위 정보를 추출할 수 있습니다.

그래서 본문 소스코드의 112번 행부터 134번 행까지 소스코드를 작성 했습니다.

```

112         # 5. 발표날짜
113         try :
114             date_1 =i.find('p','etc').find_all('span')
115             date_2 = date_1[2].get_text().strip()
116         except :
117             date_2='발표날짜가 없습니다'
118             date2.append(date_2)
119             print("5.발표년도 : %s" %date_2)
120         else :
121             date2.append(date_2)
122             print("5.발표년도 : %s" %date_2)
123
124         # 6.학위여부
125         try :
126             suksa_1 =i.find('p','etc').find_all('span')
127             suksa_2 = suksa_1[3].get_text().strip()
128         except :
129             suksa_2='학위가 없습니다'
130             suksa2.append(suksa_2)
131             print("6.학위여부 : %s" %suksa_2)
132         else :
133             suksa2.append(suksa_2)
134             print("6.학위여부 : %s" %suksa_2)

```

위 코드의 114번 행에서 <p class="etc"> 의 모든 태그 값을 다 가져와서 date_1 변수에 할당하고 115번 행에서 date_1[2] 로 3번째 태그 값을 추출합니다.
간혹 발표 날짜가 없는 경우가 있어서 예외 처리를 사용했습니다.

그리고 126번행과 127번 행도 동일한 원리로 학위 정보를 추출하였습니다.
간혹 학위 정보가 없는 경우가 있어서 예외 처리를 사용했습니다.

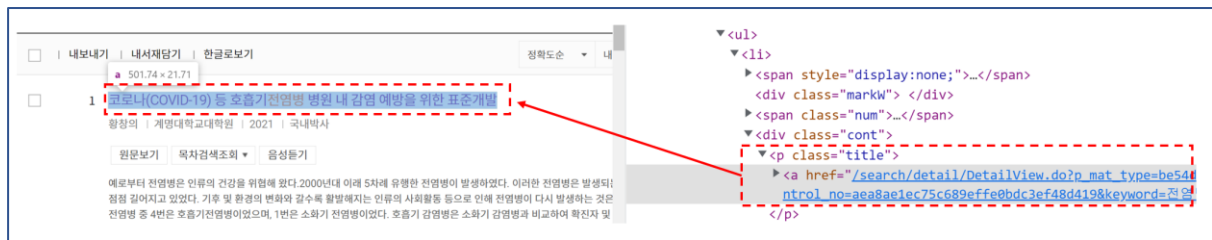
이제 아주 중요한 논문의 초록을 추출하는 부분을 살펴보겠습니다.

논문 초록은 요약 정보에서 추출할 수도 있지만 논문의 상세 페이지로 들어가서 초록을 추출하는 방법을 연습하겠습니다.

논문의 상세 페이지로 가는 방법은 해당 논문의 제목을 클릭하는 방법이 있고 해당 논문의 URL 주소를 찾아서 driver.get() 명령으로 접속하는 방법이 있습니다.

여기서는 실무에서 많이 사용되는 URL 정보를 추출하는 방법을 연습을 하기 위해 후자의 방법을 사용하겠습니다.

특정 URL 주소를 찾기 위해 아래 그림을 보세요.



위 그림에서처럼 논문의 제목에서 가 있는 것을 확인할 수 있습니다. 그런데 URL 부분을 보면 "/search/detail....." 과 같이 맨 앞에 도메인 주소가 빠져 있는 것을 확인할 수 있습니다. 즉 <http://www.riss.kr> 주소가 생략되어 있는 것인데 실제 사용할 때는 도메인 주소를 넣어서 전체 URL 주소를 완성해야 사용할 수 있습니다. 그래서 아래와 같이 소스코드에서 작업을 추가로 진행하여 URL 주소를 완성하였습니다.

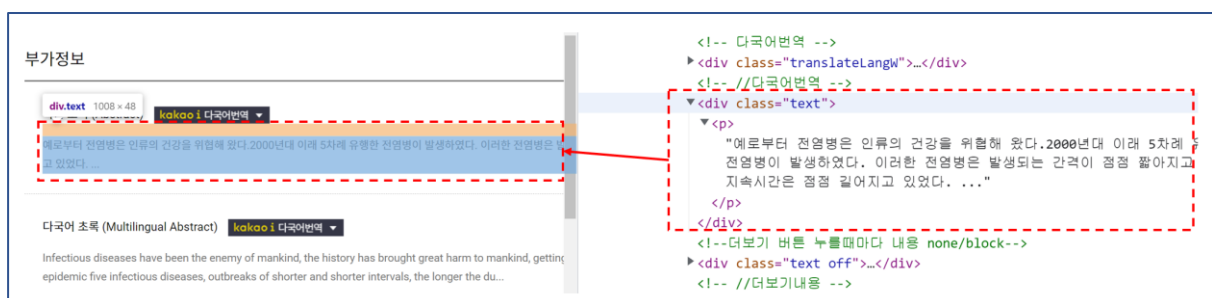
```

136      # 7.초록 내용-해당 논문의 상세 내역에서 추출할 수 있음.
137      url_1 = i.find('p','title').find('a')['href']
138      full_url = 'http://www.riss.kr'+url_1
139      time.sleep(1)
140      driver.get(full_url)

```

위 코드에서 중요한 부분은 137번 행인데 특정 태그의 특정 속성값을 추출하기 위해서는 find('태그이름')['속성이름'] 의 형태를 사용해야 합니다. 즉 URL 주소를 추출하려면 find('a')['href'] 라고 하면 'a' 태그에 있는 'href' 속성값을 추출할 수 있다는 의미입니다. 그 후 138번행에서 도메인 주소까지 합쳐서 full_url 변수에 완성된 URL 주소를 할당하고 140번 행에서 해당 URL에 접속하여 논문의 상세 페이지에 접속을 하도록 코드를 작성했습니다.

이제 상세 페이지에 접속했으니 상세 페이지의 소스코드를 다시 가져와서 파싱 후 논문의 초록 부분이 저장되어 있는 태그를 추출하면 초록 정보를 가져올 수 있습니다.



위 그림을 보면 초록 정보는 <div class="text"> 아래에 <p> 태그에 저장되어 있는 것이 확인됩니다. 그래서 소스코드의 142번행부터 158번 행까지와 같이 작성하였습니다.

```

142         html_1 = driver.page_source
143         soup_1 = BeautifulSoup(html_1, 'html.parser')
144         try :
145             cont=soup_1.find("div","text").find('p').get_text().replace("\n","").strip()
146         except :
147             cont='초록이 없습니다'
148             contents2.append(cont)
149             print("7.초록내용 : %s" %cont)
150         else :
151             contents2.append(cont)
152             print("7.초록내용 : %s" %cont)
153
154         time.sleep(1)
155
156         # 8.논문 url 주소
157         full_url2.append(full_url)
158         print('8.논문 URL 주소:' , full_url)
159
160         driver.back() # 이전 페이지로 돌아가기
161
162         time.sleep(2)
163

```

간혹 초록이 없는 논문이 있어서 144번 행부터 예외 처리를 사용했습니다.

위 코드에서 중요한 부분은 논문 제목에서 URL 정보를 추출하여 새로운 페이지에 접속했기 때문에 142번행-143번행과 같이 새로운 페이지의 소스코드를 가져와서 새로 파싱 한 후 원하는 정보를 추출해야 한다는 부분입니다.

필요한 정보를 다 추출했으니 다시 목록이 있는 페이지로 돌아가야 합니다.

키보드에서 백스페이스 키를 치거나 웹 브라우저에서 이전 페이지로 가기 버튼을 클릭하면 되는데 파이썬 코드에서는 160번 행의 driver.back() 함수를 이용하면 이전 페이지로 다시 돌아갈 수 있습니다.

riss.kr 사이트의 경우는 새로운 페이지를 열어도 현재 창에서 열리기 때문에 이전 목록 페이지로 돌아갈 때는 driver.back() 명령을 사용해야 하는데 어떤 페이지들은 새로운 페이지가 새로운 창으로 열리는 경우도 많이 있습니다. 이렇게 새로운 창에서 열릴 경우는 driver.close() 명령으로 이전 창을 닫아 주는 경우도 많이 있는 데 이런 경우의 예제가 나올 때 이 부분은 다시 설명하겠습니다.

이제 폴더를 새로 생성한 후 저장하는 부분의 소스 코드를 살펴보겠습니다.

```

179 # Step 10. 수집된 데이터를 xls와 csv 형태로 저장하기
180 # 현재 날짜와 시간으로 폴더 만들고 파일 이름 설정하기
181 import os
182
183 n = time.localtime()
184 s = '%04d-%02d-%02d-%02d-%02d-%02d' %(n.tm_year, n.tm_mon, n.tm_mday, n.tm_hour, n.tm_min, n.tm_sec)
185
186 os.makedirs(f_dir+'RISS'+ '-' +s+ '-' +'학위논문')
187
188 fc_name = f_dir+'RISS'+ '-' +s+ '-' +'학위논문'+ 'WW'+ 'RISS'+ '-' +s+ '-' +'학위논문'+ '.csv'
189 fx_name = f_dir+'RISS'+ '-' +s+ '-' +'학위논문'+ 'WW'+ 'RISS'+ '-' +s+ '-' +'학위논문'+ '.xls'

```

일반적으로 웹 크롤링을 진행할 때는 1회만 하는 경우보다는 주기적으로 반복하는 경우가 아주 많습니다. 이렇게 주기적으로 웹 크롤링을 진행할 경우 이전에 저장해 두었던 파일에다가 추가를 하거나 잘 못되면 이전 파일을 덮어쓰는 경우도 생길 수 있습니다.

그래서 저의 경우는 웹 크롤링을 진행할 때 크롤링을 하는 날짜와 시간을 가져와서 폴더와 파일 이름으로 사용함으로써 파일을 겹쳐 쓰지 않도록 설정을 합니다.

위 코드의 183번 행은 현재 시스템의 날짜와 시간을 가져오는 코드입니다.

그리고 184번행은 183번 행에서 가져온 날짜와 시간을 사용하기 편리한 형태로 재정의 하는 코드입니다.

그리고 186번 행에서 날짜와 시간과 사이트 이름과 카테고리 이름을 합쳐서 폴더 이름으로 생성하고 188번행과 189번 행에서 파일 이름으로도 활용하도록 지정하였습니다.

이렇게 설정하면 웹 크롤링을 할 때마다 파일 이름을 번거롭게 변경할 필요없이 자동으로 날짜와 시간을 사용하여 중복되지 않도록 폴더를 생성하고 파일명을 지정하여 저장하게 됩니다.

아주 편리한 기능이죠?

여기까지 이번 챕터에서 아주 중요한 부분들을 설명했습니다.

소스코드에 있는 다른 부분들은 이전 차시에서 반복적으로 설명했던 내용들이니까 다 이해하시죠?

열심히 연습해서 꼭 여러분의 실력으로 만드세요~

3. 연습 문제로 실력 굳히기

1.아래 그림과 같이 서울시 응답소 페이지에서 “시장에게 바란다” 메뉴에 있는 게시물 58건을 수집하는 웹 크롤러를 만들어 보세요

[서울시 응답소 URL] - <http://eungdapso.seoul.go.kr/>



위 그림에서 빨간색 박스의 **시장에게 바란다** 안에 있는 **바로가기** 버튼을 누르세요.

번호	제목	날짜	조회수
887	강남구에 비닐봉투·습기·쓰레기 등,	2023-08-10	13
888	서울시청과 관련된 일련된 가짜 뉴스,	2023-08-08	18
889	행정부에서 전자정부의 새로운 내용,	2023-08-08	11
890	서울시내 가동중인 차량 일련된 차량,	2023-08-08	15
891	고려대학교를 위한 일련된 차량,	2023-08-08	13
892	한국대학교를 위한 일련된 차량,	2023-08-08	18
893	서울시내 가동중인 차량 일련된 차량,	2023-08-08	17
894	서울시는 교육이치수위원회 일련된 차량,	2023-08-08	18
895	강제적인 비닐봉투를 수거하는 차량,	2023-08-08	18
896	서울시내 가동중인 차량 일련된 차량,	2023-08-08	18

위 표와 같은 게시물이 보이죠? 민원인의 개인 정보 보호를 위해 위 그림을 일부러 모자이크 처리를 했습니다. 위 표에서 각 게시물의 제목을 누르면 해당 게시물의 상세 내역으로 들어가는데 상세 내용 중에서 제목과 공개일과 상담내용을 수집하되 총 58건의 게시물의 상담내역을 수집하여 txt , csv , xls 형식으로 저장하세요.

The screenshot shows a web interface with a list of posts. Red callouts highlight the following fields:

- 제목** (Title): Located at the top of the post list.
- 공개일** (Publication Date): Located below the title.
- 상담내용** (Consultation Content): Located in the main content area of the post.

The post content includes a title, a date, and a detailed description of the issue, followed by a list of related items and a conclusion.

[크롤러 실행 결과 예시 화면]

연습문제: 서울시 응답소 게시판 크롤링하기

1. 크롤링 할 건수는 몇건입니까?: 58
크롤링 할 총 페이지 번호: 6
2. 결과 파일을 저장할 폴더명만 쓰세요(예:c:\wpy_temp\):

1 페이지 내용 수집 시작합니다 =====

===== 1번째 게시물의 상세내역을 추출합니다=====

1.제목: 강남구에 버림받은 땅 ?? 일원 1동 현대화 시설 요청

2.공개일: 2022-03-10

3.상담내역: 수고하십니다. 강남구에 버림받은 땅이 있습니다. 바로 일원1동입니다. 일원1동 수서1단지 옆 자원회수시설에는 쓰레기처리장, 하수종말처리장, 열병합발전소 등 3대 기피시설이 집중되어 있습니다. 특히 쓰레기처리장은 서울시 8개구로 부터 수집된 쓰레기를 처리 중인데, 일원1동 주민들은 90년대 중반 이후 기피시설 수용에 따른 혜택도 전혀 체감하지 못하고 있습니다. 동일한 시설을 운영 중인 마포구 당인리 발전소나 하남 유니온시티 한번 가보시기 바랍니다. 두 곳은 데마공원으로 지역 명소화가 된 반면, 강남자원회수시설은 그에 비하면 낡은 90년대 공장 이미지입니다. 똑같은 세금내는 강남구민입니다. 왜 일원1동만 피해를 입어야하는지 모르겠습니다. 당인리 발전소, 유니온시티 지하화 및 상부 공원화한 것과 똑같이 강남자원회수시설 현대화를 조속히 추진해주시기 바랍니다. 시설 현대화가 어렵다면 즉각적인 시설 폐쇄 요청드립니다.


위 그림과 같이 총 검색 건수와 파일을 저장할 경로만 입력 받은 후 저장될 파일 및 폴더 이름은 아래 그림처럼 조회하는 시점의 날짜와 시간으로 자동 설정되도록 코드를 작성하세요.


총 소요시간은 105.3 초 입니다


파일 저장 완료: txt 파일명 : c:\wpy_temp\2022-03-20-20-11-08-서울시응답소\2022-03-20-20-11-08-서울시응답소.txt

파일 저장 완료: csv 파일명 : c:\wpy_temp\2022-03-20-20-11-08-서울시응답소\2022-03-20-20-11-08-서울시응답소.csv

파일 저장 완료: xls 파일명 : c:\wpy_temp\2022-03-20-20-11-08-서울시응답소\2022-03-20-20-11-08-서울시응답소.xls

 2022-03-20-20-11-08-서울시응답소.csv

 2022-03-20-20-11-08-서울시응답소.txt

 2022-03-20-20-11-08-서울시응답소.xls

[txt 형식으로 저장된 파일 예시]

*2022-03-20-20-11-08-서울시응답소.txt - Windows 메모장

파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)

1.제목:강남구에 버림받은 땅 ?? 일원 1동 현대화 시설 요청

2.공개일:2022-03-10

3.상담내역:수고하십니다.강남구에 버림받은 땅이 있습니다. 바로 일원1동입니다.일원1동 수서1단

1.제목:[전기차] 미성년 장애인 가구 추가 보조금 적용件

2.공개일:2022-03-04

3.상담내역:시장님 안녕하세요.미성년 장애아동의 부모로서 이번 서울시 대기환경과의 '전기차 추

1.제목:영업용 택시 전기차 구매 보조금 신청관련

2.공개일:2022-03-04

3.상담내역:오세훈시장님, 시정 돌보시느라 노고가 많으십니다.최근 국내외 할것없이 전기차 시대

Ln 13, Col 1 100% Windows (CRLF) UTF-8

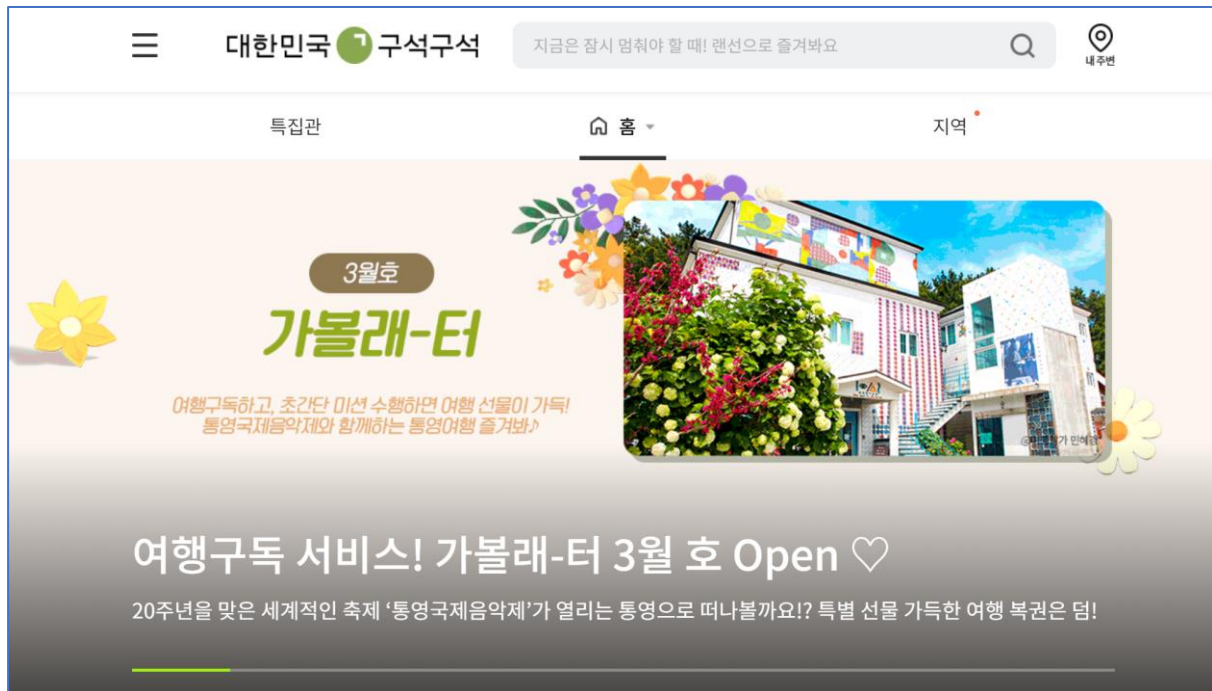
(총 58건이지만 이하 내용은 생략했습니다)

2. 대한민국 구석구석 사이트에서 “제주도여행”으로 검색한 후 검색 결과 수집하기

이번 연습문제는 한국관광공사에서 운영하는 대한민국 구석구석 사이트에서 “제주도여행” 키워드로 검색한 결과 123건을 수집하여 csv, xls, txt 형식으로 저장하는 문제입니다.

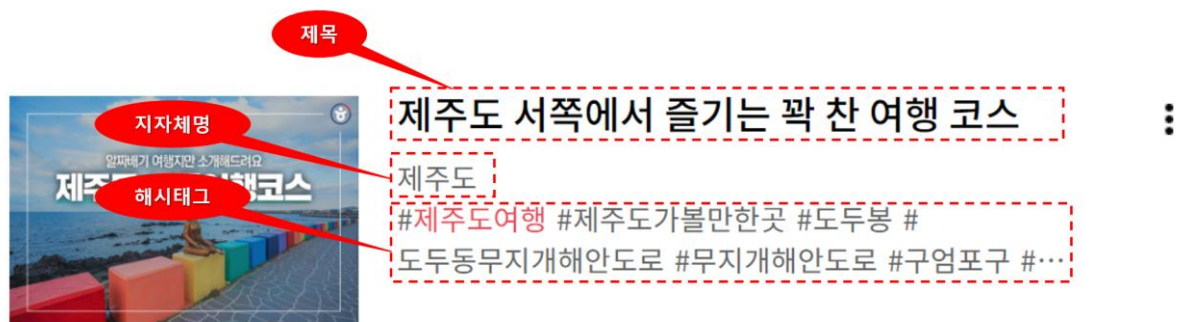
앞에서 공부했던 내용을 바탕으로 저장할 폴더와 파일명에 크롤링을 진행하는 날짜와 시간을 넣어서 생성하세요.

[대한민국 구석구석 URL : <https://korean.visitkorea.or.kr>]



(참고: 위 랜딩페이지는 수시로 업데이트 되어서 다른 화면이 보일 수 있습니다)

위 페이지의 상단에 검색창에 “제주도여행”을 키워드로 입력하여 검색한 후 나온 결과에서 아래 항목들을 수집하여 csv, xls 형식의 파일로 저장하세요.



위 그림에서 3가지 항목과 제목을 클릭하면 상세 페이지로 들어가는데 상세 페이지에 있는 텍스트를 수집하세요.



볼 것도, 즐길 것도 너~무 많은 제주도! 시간이 한정되어 있다면, 알짜배기 여행지만 쓱쓱 골라 떠나야겠죠?

오늘은 제주도 서쪽을 따라 떠나는 짝 찬 여행 코스를 소개해드립니다.

본문 텍스트
추출하기

| 도두봉, 도두동무지개해안도로

[웹 크롤러 실행 화면 예시]

연습문제: 이 크롤러는 대한민국 구석구석 사이트 정보 수집용 웹크롤러입니다.

1. 정보를 수집할 키워드는 무엇입니까?: 제주도여행
2. 몇 건의 정보를 수집할까요? : 123
총 123 건의 정보를 수집하기 위해 13 페이지까지 이동할 예정입니다
3. 파일을 저장할 폴더명만 쓰세요(기본값:c:\wpy_temp\):

위 그림과 같이 키워드와 수집할 건수, 파일을 저장할 폴더명만 입력 받습니다.

소스 코드에 파일이 저장될 기본 폴더를 c:\wpy_temp\로 지정하도록 코딩하세요.

그리고 크롤링하는 현재 날짜와 시간으로 위 폴더 안에 하위 폴더를 만들고 파일까지 생성하세요.

[수집 항목 예시]

현재 1 페이지에 있는 정보를 수집합니다

1 번째 정보를 추출하고 있습니다=====

1.번호 : 1
2.제목 : 제주도 서쪽에서 즐기는 꼭 찬 여행 코스
3.지자체이름: 제주도
4.해시태그: #제주도여행 #제주도가볼만한곳 #도두봉 #도두동무지개해안도로 #무지개해안도로 #구영포구 #영장해안길 #드라이브코스 #오설록 #오설록티뮤지엄 #제주도
5.본문내용:
볼 것도, 즐길 것도 너무 많은 제주도! 시간이 한정되어 있다면, 알짜배기 여행지만 쓱쓱 골라 떠나야겠죠?오늘은 제주도 서쪽을 따라 떠나는 꼭 찬 여행 코스를 소개해드립니다.
첫 번째 여행지는 바로 도두봉. 공항과 가까우며 탁 트인 풍경을 볼 수 있는 곳입니다. 도두봉 옆, 최근 인기가 많은 무지개해안도로도 함께 소개합니다. 우선 도두봉은 도두동 해안가에 있습니다.
해발이 63m 정도로 높지 않고, 도두봉 정상부분에는 키세스 초콜릿 모양의 나무 터널이 있는데요. SNS 인생샷 명소로 사랑받는 장소입니다.올라가는 등산로는 계단으로 이루어져 있습니다. 대략 10분 정도 소요되며, 정상에서 바다와 한라산을 볼 수 있어 가볍게 올라 풍경을 조망하기 좋습니다.
내려 온 뒤 바로 옆에 있는 무지개 해안도로를 꼭 둘러 보는 것을 추천해요! 1km 남짓한 해안도로를 배경으로 팔두초파남보 무지개색으로 칠해진 방호벽을 볼 수 있습니다. 약 15분에서 20분 사이의 거리로 바다를 보며 산책하고 사진을 찍기에 제격입니다.
※ 도두봉 -위치: 제주 제주시 도두일동 산1 ※ 도두동무지개해안도로 -위치: 제주 제주시 도두일동 1734
구영포구에 있는 영장해안길은 구영리~고내리까지 이어지는 약 4.8km 거리의 걷기 좋은 길입니다. 애월읍 구영포구 - 돌영전 - 남도리센터 - 신영도대불 - 남두연대 - 다락센터 - 고내 포구로 이어지는 코스로 구성되어 있습니다. 구영포구에서 시작되는 영장해안길은 오른쪽으로 에메랄드빛 제주 바다가 펼쳐져 있는데요. 화려한 암석으로 이루어진 해안길을 걸으며 멋진 풍경을 감상할 수 있습니다.
해안 길 입구에 있는 돌영전은 400년 이상 된 영전이며, 일부 복원해서 관광코스로 활용 중입니다. 구영리 돌영전은 당시 돌 위에 흙으로 작은 독을 만들어 바윗돌 위에 바닷물을 가두고 자연 건조하는 방식으로 생선했다고 합니다.마치 거북이 등처럼 보이기도 하죠?바닷길을 따라 약 1.5 00여 평 규모였다고 합니다. 현대에 들어 육지에서 쓴 소금을 들여오며 소금밭의 기능은 잃게 되었다고 해요.
영장해안길은 드라이브코스로도 유명하며, 일출명소로도 잘 알려져 있습니다. 맑은 날 제주도의 자연을 느끼기에 더없이 좋은 곳으로, 걸어서 약 2시간 정도 소요됩니다.

(지면 관계상 중간 내용은 생략합니다)

[xls , csv 파일 저장 예시]

폴더 경로와 이름

내 PC > SYSTEM (C:) > py_temp > 대한민국구석구석-2022-03-20-22-11-21-제주도여행

이름

파일 이름

대한민국구석구석-2022-03-20-22-11-21-제주도여행.csv

대한민국구석구석-2022-03-20-22-11-21-제주도여행.xls

번호	제목	지자체명	본문내용	해시태그
1	제주도 서쪽에서 즐기는 꼭 찬 여행 코스	제주도	볼 것도, 즐길 것도 너무 많은 제주도! 시간이 한정되어 있다	#제주도여행 #제주도가볼만한곳 #도두봉 #도두동무지개해안도로
2	제주도 서귀포 가볼 만한 곳, 초록이 가득한 제주도 서귀포시	서귀포시	첫 번째 여행지는 바로 도두봉. 공항과 가까우며 탁 트인 풍경	#제주도여행 #제주도가볼만한곳 #돌영아리오티를습지 #서귀포시 #원영
3	제주도 '한정' 여행지, 제주도의 이색 공간을 찾자	제주도	해발이 63m 정도로 높지 않고, 도두봉 정상부분에는 키세스	#제주도여행 #제주도가볼만한곳 #김영갑갤러리 #김영갑갤러리두모악
4	효리내민박 촬영지부터 매화꽃 명소까지 ♥제주도	제주도	내려 온 뒤 바로 옆에 있는 무지개 해안도로를 꼭 둘러 보는	#제주3월여행코스 #제주도3월여행 #효리내민
5	한국관광공사 추천 2월 걷기여행길, 이른 봄 제주	제주도	※ 도두봉 -위치: 제주 제주시 도두일동 산1 ※ 도두	#걷기여행길 #2월걷기여행길 #제주도 #제주도걷기길 #트레킹코스
6	김장열의 회귀 철학을 건축으로 표현한 미술 제주	제주시	구영포구에 있는 영장해안길은 구영리~고내리까지 이어지는	#2021년4월추천가볼만한곳 #4월추천가볼만한곳 #추천가볼만한곳
7	[국내 트레킹 추천] 제주 올레길 걷기 여행, 제주	제주도	해안 길 입구에 있는 돌영전은 400년 이상 된 영전이며, 일부	#제주올레길 #트레킹코스 #트레킹 #제주도가볼만한곳 #제주도여행
8	늦기 전에 떠나야 할 제주 숨은 동백 명소	제주도	영장해안길은 드라이브코스으로도 유명하며, 일출명소로도 잘	#겨울여행 #동백꽃 #제주도 #제주도여행코스 #제주도가볼만한곳
9	알록달록 가을빛 제주	제주도	※ 구영포구 영장해안길 -위치: 제주 제주시 애월읍 구영리	#장애민주천코스 #장애민주천여행지 #공공누리 #무장애여행지 #
10	[제주] 여행자를 위한 안성맞춤 숙소, 제주 아 제주도 서귀포시	서귀포시	'제주도'하면 떠오르는 '녹차'제주도에서는 녹차를 주제로 한	#여행Q레이터 #여행자의방 #한국관광품질인증 #추천숙박 #아트
11	제주도 세계자연유산 탐방 1 코스	제주도 제주시	제주산 녹차는 다원에서 유기농 재배 방식으로 화확비료를 사	#당일코스 #1박2일 #2박3일 #제주도 #가볼만한곳
12	제주도 국가지질공원	제주도 제주시	산노루에서는 최상급의 재료를 사용한 말차 라떼를 꼭 마셔보	#제주도여행 #지질공원 #성산일출봉 #한라산국립공원 #자연환경
13	유채꽃으로 시작하는 봄, 인생사진 찍으러 제주	제주도	※ 산노루 녹차카페-위치: 제주 제주시 한경면 낙원로 32-윤영	#한국관광품질인증 #여행자의방 #추천숙박 #공공누리 #다세
14	[랜선가이드] 소설가 김영하가 들려주는 여행 전국	제주도	제주 오설록 서광 자발과 맞닿아 있는 오설록 티 뮤지엄은 20	#랜선가이드 #김영하작가 #여행의이유 #당신에게 권하는 여행의
15	성산일출봉 [유네스코 세계자연유산]	제주도 서귀포시	'주사'와 제주이마'기 미디어아트를 감상하고, 내부로 들어가	#성산일출봉 #가족여행 #연인과함께 #친구와함께 #아이와함께 #
16	천지연폭포 (제주도 국가지질공원)	제주도 서귀포시	지하의 발효차 숙성과 투어를 마지막으로 티를라스를 마무리	#천지연폭포 #가족여행 #연인과함께 #친구와함께 #아이와함께 #
17	[제주] 이동하기 좋은 위치의 숙소를 찾는다면 제주	제주시	※ 오설록 티 뮤지엄 -위치: 제주 서귀포시	#한국관광품질인증 #여행자의방 #추천숙박 #제주도R호텔
18	가을에 만나는 초록빛 꽃지, 제주 청수마을 제주	제주시	※ 해당 원고에 언급된 업체로부터 어떠한 대가도 받지 않고	#2022년2월추천가볼만한곳 #2월추천가볼만한곳 #추천가볼만한곳
19	서귀포문화재여행	제주도 서귀포시	→ 거리 두기 상황에 따라 운영 여부 및 시간이 변동될 수 있	#축제 #문화 #서귀포 #제주도 #제주도여행코스 #제주도가볼만한
20	우도, 성산 그리고 제주올레, 제주 동북부를	제주도 서귀포시	※ 우리 모두 건강한 일상으로 돌아가기 위해 방역수칙을 준	#장애민주천코스 #서귀포가볼만한곳 #접지코지 #하도해변 #세화

(총 123건이 수집되었지만 지면 관계상 이하 내용은 생략합니다)

이번장의 내용도 열심히 공부해서 꼭 실력으로 만드세요~

[파이썬 능력자 너도 될 수 있어~! - 서진수 저 -]