

⋮ 문제해결을 위한 데이터 분석 및 시각화

■ 데이터 수집을 위한 네이버 API이용하기(2)

한성대학교 노은희 교수

“미래로 향하는 새로운 이정표”



오늘의 학습

학습 내용

- 네이버 지식IN에서 정보를 가져오는 방법을 익힙니다.
- 네이버 뉴스 검색에서 정보를 가져오는 방법을 익힙니다.
- 네이버 영화 검색에서 정보를 가져오는 방법을 익힙니다.

.

네이버 지식IN검색에서 정보 가져오기->pandas로 저장하기

검색질의 :코딩

	Title	Link	Description
0	코딩을 배우면	https://kin.naver.com/qna/detail.naver?d1id=4&...	... 코딩을 배우면서 관련 자격증 따려고 하는데, 해외 국제 대기업들에서 인정받을...
1	엔트리 코딩하는거 있잖아요	https://kin.naver.com/qna/detail.naver?d1id=11...	엔트리 코딩하는거 있잖아요 전자공학기술자란 직업이랑 많이 관계있나요?? 안녕하세요...
2	코딩 추천 교재	https://kin.naver.com/qna/detail.naver?d1id=1&...	코딩 공부를 하려고 하는데 초등학생 코딩어떤 교재가 좋을 까요? 안녕하세요 초등학생...
3	코딩 스크래치로 돈벌수 있나요?	https://kin.naver.com/qna/detail.naver?d1id=1&...	제가 코딩 스크립트 이런거를 배우고있는데 아직 배우는중이라 잘하는게 스크래치 밖에 없...
4	코딩	https://kin.naver.com/qna/detail.naver?d1id=1&...	... 것도 코딩 인가요? 코딩이 뭔지 좀 알려주세요 안녕하세요? 지식파트너 가비아...
...
995	동탄 코딩 학원	https://kin.naver.com/qna/detail.naver?d1id=4&...	프로그래머(웹개발자)가 꿈인 고2 학생입니다 동1 이나 동2 프로그래밍, 코딩 학원...
996	코딩 독학	https://kin.naver.com/qna/detail.naver?d1id=1&...	코딩을 배워보려고 하는데요 자격증 취득이 목표는 아니고 재귀함수를 유연히 접해서 알...
997	코딩의 기준	https://kin.naver.com/qna/detail.naver?d1id=1&...	... 수정도 코딩이라 하나요???? 약간 애매 할 수도 있는데요. 코딩이라 함은...
998	그림의 기본이 선긋기라면 코딩의 기본은?	https://kin.naver.com/qna/detail.naver?d1id=1&...	... 그렇다면 코딩의 기본은 무엇일까요? 그림의 기본이 선긋기라면 코딩의 기본은 ...
999	코딩공부방법 (취미로 시작)	https://kin.naver.com/qna/detail.naver?d1id=1&...	코딩 공부를 취미로 시작해보려고 합니다. 근데 비전공자라 코딩공부방법을 어떻게 시작...

1000 rows × 3 columns

네이버 지식IN검색에서 정보 가져오기->pandas로 저장하기->엑셀로 저장하기

코딩.xlsx - Excel						
파일 홈 삽입 페이지 레이아웃 수식 데이터 검토 보기 개발 도구 파워 피벗 수행할 작업을 알려 주세요. 공유						
<div> <div>기본</div> <div>페이지 나누기 미리 보기 레이아웃</div> <div>페이지 레이아웃</div> <div>사용자 지정 보기</div> <div>눈금자</div> <div>수식 입력줄</div> <div>눈금선</div> <div>머리글</div> <div>표시</div> <div>확대/축소</div> <div>100%</div> <div>선택 영역 확대/축소</div> <div>확대/축소</div> <div>새 창</div> <div>모두 정렬</div> <div>틀 고정</div> <div>나누기</div> <div>숨기기</div> <div>숨기기 취소</div> <div>창</div> <div>나란히 보기</div> <div>동시 스크롤</div> <div>창 위치 다시 정렬</div> <div>창 전환</div> <div>매크로</div> <div>매크로</div> </div>						
E5						
	A	B	C	D	E	F
1		Title	Link	Description		
2	0	코딩을 배우면	https://kin.naver.com/qna/detail.naver?d1id=4&dirId=40608&docId=418387206&q=7L2U65Sp&enc=utf8&section=kin.qna&rank=1&search_sort=0&spq=0	... 코딩을 배우면서 관련 자격증 따려고 하는데, 해외 국제 대기업들에서 인정받을 수 있는 코딩 관련 자격증이 있을까요? 제 꿈이 개발자는 아니긴 한데, 코딩도 배워두면 취업에 유용할 것 같아서요! 그 쪽 진로가 아니긴 한데 코딩...		
3	1	엔트리 코딩하는거 있잖아요	https://kin.naver.com/qna/detail.naver?d1id=11&dirId=1118&docId=328179617&q=7L2U65Sp&enc=utf8&section=kin.qna&rank=2&search_sort=0&spq=0	엔트리 코딩하는거 있잖아요 전자공학기술자란 직업이랑 많이 관계있나요?? 안녕하세요 메가스터디 IT 아카데미... 코딩이라는 말의 뜻에서부터 알수 있듯이 엔트리카딩은 입문자가 쉽게 이해할 수 있도록 설계된 코딩입니다....		
4	2	코딩 추천 교재	https://kin.naver.com/qna/detail.naver?d1id=1&dirId=104&docId=367291003&q=7L2U65Sp&enc=utf8&section=kin.qna&rank=3&search_sort=0&spq=0	코딩 공부를 하려고 하는데 초등학교 코딩어떤 교재가 좋을 까요? 안녕하세요 초등학교 코딩 교재 찾고계신데 저희가 도와드리겠습니다. 코딩은 사람이 이해하는 언어를 컴퓨터가 이해하는 언어로 바꿔주는것을 말합니다. 꾸준히...		
5	3	코딩 스크래치로 돈 벌수 있나요?	https://kin.naver.com/qna/detail.naver?d1id=1&dirId=1040202&docId=417358318&q=7L2U65Sp&enc=utf8&section=kin.qna&rank=4&search_sort=0&spq=0	제가 코딩 스크립트 이런거를 배우고있는데 아직 배우는중이라 잘하는게 스크래치밖에 없는데 뭐 없을까요?... 스크래치로는 뭐 취직을 한다거나 이런 부분으로는 안되고요 가르치면서 돈버는?그런쪽만 될거예요 코딩을 직장에서...		
			https://kin.naver.com/qna/detail.naver?d1id=1&dirId=1040202&docId=417358318&q=7L2U65Sp&enc=utf8&section=kin.qna&rank=4&search_sort=0&spq=0	... 것도 코딩 인가요? 코딩이 뭔지 좀 알려주세요 안녕하세요? 지식		

네이버 지식IN에서 정보 가져오기

```
1 import os
2 import sys
3 import urllib.request
4 import pandas as pd
5 import json
6 import re
7
8 client_id = "lbzft0g8Ql4mcLqZZPL6"
9 client_secret = "22o8drQ854"
10
11 query = urllib.parse.quote(input("검색질의:")) # 검색어 입력
12 idx = 0
13 display = 100
14 start = 1
15 end = 1000
16 sort = "sim"
17
18 kin_df = pd.DataFrame(columns = ('Title', 'Link', 'Description')) # 컬럼수 3개의 데이터 프레임
19 |
20 for start_index in range(start, end, display):
21
22     url = "https://openapi.naver.com/v1/search/kin?query=" + query +
23         + "&display=" + str(display) +
24         + "&start=" + str(start_index) +
25         + "&sort=" + sort
```

네이버 지식IN검색에서 정보 가져오기

```
27 request = urllib.request.Request(url)
28 request.add_header("X-Naver-Client-Id",client_id)
29 request.add_header("X-Naver-Client-Secret",client_secret)
30 response = urllib.request.urlopen(request)
31
32 rescode = response.getcode()
33 if(rescode==200):
34     response_body = response.read()
35     response_dic = json.loads(response_body.decode('utf-8')) # 딕셔너리로 받아오기
36
37     items = response_dic['items'] # items키에 해당하는 값 가져오기
38
39     for item_index in range(0,len(items)):
40         r_tag = re.compile('<.*?>') # 태그를 제외하기 위한 패턴 지정하기
41
42         title = re.sub(r_tag,'',items[item_index]['title'])
43         link = items[item_index]['link']
44         description = re.sub(r_tag,'',items[item_index]['description'])
45
46         kin_df.loc[idx] = [title,link,description]
47         idx += 1
48
49     else:
50         print("Error Code:" + rescode)
51 kin_df
```

네이버 뉴스검색에서 정보 가져오기

검색질의: 코로나

	Title	Originallink	Link	Description	PubDate
0	[단독] 파우치 "한국, 코로나 지난해보다 훨씬 나아질 것"	https://news.sbs.co.kr/news/endPage.do?news_id...	https://news.naver.com/main/read.naver?mode=LS...	세계보건기구, WHO는 지난 2년 동안 코로나로 전 세계 1,490만 명이 목숨을 ...	Fri, 06 May 2022 18:06:00 +0900
1	코로나 확진 미접종자, 18개월 뒤 1회 접종으로 충분	https://www.donga.com/news/article/all/2022050...	https://news.naver.com/main/read.naver?mode=LS...	신종 코로나바이러스 감염증(코로나19) 확진 판정을 받은 백신 미접종자는 1년 반(...	Fri, 06 May 2022 15:59:00 +0900
2	항저우 아시안게임 연기...시진핑 "끝까지 제로코로나"	https://news.jtbc.joins.com/article/article.as...	https://news.naver.com/main/read.naver?mode=LS...	중국에서 코로나에 걸린 사람들이 최근 빠르게 늘고 있어서 그런 거 아니냐 하는 추측...	Fri, 06 May 2022 19:58:00 +0900
3	WHO "코로나로 2년간 1천490만명 사망... 500명 중 1명꼴"	https://www.yna.co.kr/view/AKR2022050507435100...	https://news.naver.com/main/read.naver?mode=LS...	직접 사망자에 의로 체계 붕괴에 따른 간접적 피해자도 포함 김지연 기자 임은진 특파...	Fri, 06 May 2022 00:24:00 +0900
4	中 코로나19 확산에...9월 항저우 아시안게임 연기	https://news.kmib.co.kr/article/view.asp?arcid...	https://news.naver.com/main/read.naver?mode=LS...	확산하는 코로나19 영향으로 보인다. 항저우는 상하이와 약 200km 떨어진 도시로...	Fri, 06 May 2022 15:24:00 +0900

네이버 뉴스에서 정보 가져오기

```
1 import os
2 import sys
3 import urllib.request
4 import pandas as pd
5 import json
6 import re
7
8 client_id = "lbzft0g8Ql4mcLqZZPL6"
9 client_secret = "22o8drQ854"
10 |
11 query = urllib.parse.quote(input("검색질의:")) # 검색어 입력
12 idx = 0
13 display = 100
14 start = 1
15 end = 1000
16 sort = "sim"
17
18 news_df = pd.DataFrame(columns = ('Title', 'Originallink', 'Link', 'Description', 'PubDate')) # 컬럼수 5개의 데이터
19
20 for start_index in range(start, end, display):
21
22     url = "https://openapi.naver.com/v1/search/news?query=" + query +
23         "&display=" + str(display) +
24         "&start=" + str(start_index) +
25         "&sort=" + sort
```


네이버 뉴스검색에서 정보 가져오기(pandas)

```
26 request = urllib.request.Request(url)
27 request.add_header("X-Naver-Client-Id", client_id)
28 request.add_header("X-Naver-Client-Secret", client_secret)
29 response = urllib.request.urlopen(request)
30
31
32 rescode = response.getcode()
33 if(rescode==200):
34     response_body = response.read()
35     response_dic = json.loads(response_body.decode('utf-8')) # 딕셔너리로 받아오기
36
37     items = response_dic['items'] # items키에 해당하는 값 가져오기
38
39     for item_index in range(0, len(items)):
40         r_tag = re.compile('<.*?>') # 태그를 제외하기 위한 패턴 지정하기
41
42         title = re.sub(r_tag, '', items[item_index]['title'])
43         originallink = items[item_index]['originallink']
44         link = items[item_index]['link']
45         description = re.sub(r_tag, '', items[item_index]['description'])
46         pubDate = re.sub(r_tag, '', items[item_index]['pubDate'])
47
48         news_df.loc[idx] = [title, originallink, link, description, pubDate]
49         idx += 1
50
51     else:
52         print("Error Code:" + rescode)
53 news_df
54 #news_df.to_excel('코로나.xlsx')
```

네이버 뉴스검색에서 정보 가져오기:엑셀로 저장하기

```
27 request = urllib.request.Request(url)
28 request.add_header("X-Naver-Client-Id",client_id)
29 request.add_header("X-Naver-Client-Secret",client_secret)
30 response = urllib.request.urlopen(request)
31
32 rescode = response.getcode()
33 if(rescode==200):
34     response_body = response.read()
35     response_dic = json.loads(response_body.decode('utf-8')) # 딕셔너리로 받아오기
36
37     items = response_dic['items'] # items키에 해당하는 값 가져오기
38
39     for item_index in range(0,len(items)):
40         r_tag = re.compile('<.*?>') # 태그를 제외하기 위한 패턴 지정하기
41
42         title = re.sub(r_tag,'',items[item_index]['title'])
43         link = items[item_index]['link']
44         description = re.sub(r_tag,'',items[item_index]['description'])
45
46         kin_df.loc[idx] = [title,link,description]
47         idx += 1
48
49     else:
50         print("Error Code:" + rescode)
51 #kin_df
52 kin_df.to_excel('코딩.xlsx')
```

네이버 영화검색에서 정보 가져오기

검색질의:알라딘

	Title	Link	Image	Subtitle	PubDate	Director	Actor	UserRating
0	알라딘	https://movie.naver.com/movie/bi/mi/basic.nhn?...	https://ssl.pstatic.net/imgmovie/mdi/mit110/16...	Aladdin	2019	가이 리치	메나 마수드 윌 스미스 나오미 스콧	9.38
1	알라딘-램프의 신	https://movie.naver.com/movie/bi/mi/basic.nhn?...	https://ssl.pstatic.net/imgmovie/mdi/mit110/18...	Adventures of Aladdin	2019	글렌 캠펔		0.00
2	알라딘 2	https://movie.naver.com/movie/bi/mi/basic.nhn?...	https://ssl.pstatic.net/imgmovie/mdi/mit110/18...	Alad'2	2018	리오넬 스테케티	케브 아담스 자멜 드 부즈바네 사 가이드 에릭 주도트 람지 베디아	4.06
3	애덤 그린의 알라딘	https://movie.naver.com/movie/bi/mi/basic.nhn?...		Adam Green's Aladdin	2016	애덤 그린	애덤 그린	10.00

네이버 영화에서 정보 가져오기

```
1 import os
2 import sys
3 import urllib.request
4 import pandas as pd
5 import json
6 import re
7
8 client_id = "lbzft0g8Ql4mcLqZZPL6"
9 client_secret = "22o8drQ854"
10
11 query = urllib.parse.quote(input("검색질의:")) # 검색어 입력
12 idx = 0
13 display = 100
14 start = 1
15 end = 1000
16 sort = "sim"
17
18 movie_df = pd.DataFrame(columns = ('Title', 'Link', 'Image', 'Subtitle', 'PubDate', 'Director', 'Actor', 'UserRating')) # 컬럼수 8개의 데이터
19
20 for start_index in range(start, end, display):
21
22     url = "https://openapi.naver.com/v1/search/movie?query=" + query +
23           + "&display=" + str(display) +
24           + "&start=" + str(start_index) +
25           + "&sort=" + sort
```


네이버 영화검색에서 정보 가져오기

```
--
27 request = urllib.request.Request(url)
28 request.add_header("X-Naver-Client-Id",client_id)
29 request.add_header("X-Naver-Client-Secret",client_secret)
30 response = urllib.request.urlopen(request)
31
32 rescode = response.getcode()
33 if(rescode==200):
34     response_body = response.read()
35     response_dic = json.loads(response_body.decode('utf-8')) # 딕셔너리로 받아오기
36
37     items = response_dic['items'] # items키에 해당하는 값 가져오기
38
39     for item_index in range(0,len(items)):
40         r_tag = re.compile('<.*?>') # 태그를 제외하기 위한 패턴 지정하기
41         title = re.sub(r_tag, '', items[item_index]['title'])
42         link = items[item_index]['link']
43         image = items[item_index]['image']
44         subtitle = items[item_index]['subtitle']
45         pubDate = items[item_index]['pubDate']
46         director = items[item_index]['director']
47         actor = items[item_index]['actor']
48         userRating = items[item_index]['userRating']
49
50         movie_df.loc[idx] = [title,link,image,subtitle,pubDate,director,actor,userRating]
51         idx += 1
52     else:
53         print("Error Code:" + rescode)
54 movie_df
55 #movie_df.to_excel('알라딘 영화.xlsx')
```