requests와 bs4를 이용한 데이터 수집

```
import requests
from bs4 import BeautifulSoup
import json
import os

response = requests.get('http://class.gnu.ac.kr/~jslee/')

htmlObj = response.content
#print(htmlObj)
#-------------------------------------------------
bsObj = BeautifulSoup(htmlObj, 'html.parser')
print(bsObj)     #html과 함께 출력
#print(bsObj.text)     #문서의 내용만 출력



#----- <a> 태그만 추출 -----------------------------
listObj = bsObj.select('body > a')
#print(listObj)



#----- 추출한 정보 딕셔너리에 담기  --------------------
dictObj = {}   #빈 딕셔너리 만들기
for i in listObj:
    dictObj[i.text] = i.get('href')
print(dictObj)
            '''
            for i in dictObj.keys():
                print(i, ":" ,dictObj[i])
            '''

#----- json로 딕셔너리 저장하기  ----------------------------
with open(os.path.join('result.json'), 'w+') as json_file :
      json.dump(dictObj, json_file)
```

Open API - Naver

```
import urllib.request

client_id = "qOt_gqRZWOPjQZfL7dXP"
client_secret = "9yuHTz94mv"
search_text = urllib.parse.quote("빅데이터")
page_start = 1
display = 5


        # 'news', 'blog', 'cafearticle' 'movie' 'encyc'
url = "https://openapi.naver.com/v1/search/news?
        query=%s&start=%s&display=%s"
        %(search_text,page_start, display )
                # + search_text  # json 결과
                #url = search/encyc?query=%s&start=%s&display=%s
                #url = search/news.xml?query=" + search_text # xml 결과

reqObj = urllib.request.Request(url)
print(reqObj)

reqObj.add_header("X-Naver-Client-Id",client_id)
reqObj.add_header("X-Naver-Client-Secret",client_secret)

htmlObj = urllib.request.urlopen(reqObj)
print(htmlObj)

htmlObj_body = htmlObj.read()
print(htmlObj_body.decode('utf-8'))
```