import urllib.request

import urllib.parse

from bs4 import BeautifulSoup

import re

def bind\_url():

default\_url = "https://openapi.naver.com/v1/search/image.xml?"

query = "&query=" + urllib.parse.quote\_plus(str(input("검색어를 입력하세요:")))

*# key = 'Hibrg723Mi3GaN01MmE\_'*

sort = '&sort=sim'

start = '&start=1'

display = '&display=100'

full\_url = default\_url + sort + start + display + query

return full\_url

*# target = '&target=blog'*

def fetch\_contents\_from\_url():

url = bind\_url()

headers = {

'Host': 'openapi.naver.com',

'User-Agent': 'curl/7.43.0',

'Accept': '\*/\*',

'Content-Type': 'application/xml',

'X-Naver-Client-Id': "Hibrg723Mi3GaN01MmE\_",

'X-Naver-Client-Secret': "3Fv7a7WfKW",

}

print(url)

req = urllib.request.Request(url, *headers*=headers)

f = urllib.request.urlopen(req)

html = f.read()

return html

def extract\_text\_in\_tags(*tags*, *tagname*="title"):

txt = []

reg = "[^<" + tagname + ">][^<]+"

print(tags)

for tag in tags:

txt.append(re.search(reg, tag).group())

return txt

def get\_contents\_from\_html():

html = fetch\_contents\_from\_url()

title\_tags = re.findall("<title>[^<]+</title>", html.decode('utf-8'))

link\_tags = re.findall("<link>[^<]+</link>", html.decode('utf-8'))

titles = extract\_text\_in\_tags(title\_tags, *tagname*="title")

links = extract\_text\_in\_tags(link\_tags, *tagname*="link")

f = open("image4.html", "w")

f.write("<html><body>")

print(titles)

for i in range(1, len(titles)):

f.write("<p>" + titles[i] + "</p>")

f.write("<img src=" + links[i] + "</>")

f.write("</body></html>")

f.close()

get\_contents\_from\_html()

이미지 크롤링 <p>태그와 <img src> 소스 코드를 가져올 수 있도록 하는 코드입니다.