```
    print("{0} {1} [{2}]".format(">"*link[1], link[0], len(links)))
```

```
 https://blog.naver.com/shkwon1128?Redirect=Log&logNo=221530738046 (https://blog.n
aver.com/shkwon1128?Redirect=Log&logNo=221530738046) [32]
 https://blog.naver.com/js2y86?Redirect=Log&logNo=221530397718 (https://blog.nave
r.com/js2y86?Redirect=Log&logNo=221530397718) [32]
 https://blog.naver.com/dlqlwm14?Redirect=Log&logNo=221529807195 (https://blog.nav
er.com/dlqlwm14?Redirect=Log&logNo=221529807195) [32]
 https://blog.naver.com/lyj0088?Redirect=Log&logNo=221530708253 (https://blog.nave
r.com/lyj0088?Redirect=Log&logNo=221530708253) [33]
 https://komartin.blog.me/221530734469 (https://komartin.blog.me/221530734469) [0]
> https://blog.naver.com (https://blog.naver.com) [0]
> https://blog.naver.com/prologue/PrologueList.nhn?blogId=shkwon1128 (https://blo
g.naver.com/prologue/PrologueList.nhn?blogId=shkwon1128) [78]
> https://blog.naver.com/MyBlog.nhn (https://blog.naver.com/MyBlog.nhn) [0]
> https://blog.naver.com/prologue/PrologueList.nhn?blogId=shkwon1128 (https://blo
g.naver.com/prologue/PrologueList.nhn?blogId=shkwon1128) [78]
> https://blog.naver.com/prologue/PrologueList.nhn?blogId=shkwon1128&skinType=&ski
nId=&from=menu (https://blog.naver.com/prologue/PrologueList.nhn?blogId=shkwon1128
&skinType=&skinId=&from=menu) [78]
> https://blog.naver.com/PostList.nhn?blogId=shkwon1128&skinType=&skinId=&from=men
u (https://blog.naver.com/PostList.nhn?blogId=shkwon1128&skinType=&skinId=&from=me
```

20190509

In [5]:

```
import requests

#url = 'http://example.webscraping.com/places/default/search'
url = "http://example.webscraping.com/places/ajax/search.json"
params = {
    "search_term":"",
    "page_size":10,
    "page":0
}

params["search_term"] = "korea"

html = download(url, params)

html.text
```

Out[5]:

```
'{"records": [{"pretty_link": "<div><a href=WW"/places/default/view/North-Korea-165
WW"><img src=WW"/places/static/images/flags/kp.pngWW" /> North Korea</a></div>", "co
untry": "North Korea", "id": 4153377}, {"pretty_link": "<div><a href=WW"/places/defa
ult/view/South-Korea-211WW"><img src=WW"/places/static/images/flags/kr.pngWW" /> Sou
th Korea</a></div>", "country": "South Korea", "id": 4153423}], "num_pages": 1, "err
or": ""}Wn'
```

In [8]:

```
print(html.json(), "Wn")
print(type(html.json()))
```

```
{'records': [{'pretty_link': '<div><a href="/places/default/view/North-Korea-165"><i
mg src="/places/static/images/flags/kp.png" /> North Korea</a></div>', 'country': 'N
orth Korea', 'id': 4153377}, {'pretty_link': '<div><a href="/places/default/view/Sou
th-Korea-211"><img src="/places/static/images/flags/kr.png" /> South Korea</a></div
>', 'country': 'South Korea', 'id': 4153423}], 'num_pages': 1, 'error': ''} 

<class 'dict'>
```

In [10]:

```
# dom 사용방식
from bs4 import BeautifulSoup

for _ in html.json()["records"]:
    print(_["pretty_link"])
    dom = BeautifulSoup(_["pretty_link"], "lxml")
    print([_["href"] for _ in dom.find_all("a")])
    break
```

```
<div><a href="/places/default/view/North-Korea-165"><img src="/places/static/images/
flags/kp.png" /> North Korea</a></div>
['/places/default/view/North-Korea-165']
```

In [14]:

```
#정규식 사용방식
import re

text = html.json()["records"][0]["pretty_link"]
group = re.findall("<a href=W"(.+)W"><img src=W"(.+)W" /> (.+)</a>", text)
requests.compat.urljoin(url, group[0][0]), W
requests.compat.urljoin(url, group[0][1]), group[0][2]
```

Out[14]:

```
('http://example.webscraping.com/places/default/view/North-Korea-165',
 'http://example.webscraping.com/places/static/images/flags/kp.png',
 'North Korea')
```

In [143]:

```
driver = webdriver.Chrome()
url = 'http://example.webscraping.com/places/default/search'
driver.get(url)

# / -> 자식 ... // -> 자손 . -> 현재 위치
# driver.find_element_by_id("serch_term").clear()
# driver.find_element_by_id("serch_term").send_keys("korea")
# driver.find_element_by_css_selector("#serch_term").send_keys("korea")
driver.find_element_by_xpath("//input[@id='search_term']").send_keys("korea")
driver.find_element_by_id("search").click()
```

In [50]:

```
dom = BeautifulSoup(driver.page_source, "lxml")
```

In [5]:

```
!pip install selenium
from selenium import webdriver
driver = webdriver.Chrome()
```

```
Requirement already satisfied: selenium in c:WusersWjWappdataWlocalWprogramsWpython
Wpython37WlibWsite-packages (3.141.0)
Requirement already satisfied: urllib3 in c:WusersWjWappdataWlocalWprogramsWpythonWp
ython37WlibWsite-packages (from selenium) (1.24.1)

You are using pip version 19.0.3, however version 19.1 is available.
You should consider upgrading via the 'python -m pip install --upgrade pip' command.
```

In [154]:

```
url = 'http://example.webscraping.com/places/default/search'

driver.get(url)
driver.find_element_by_id("search_term").clear()
driver.find_element_by_xpath("//input[@id='search_term']").send_keys("korea")
driver.find_element_by_id("search").click()

#AJAX -> Data요청 -> 브라우저 받아서 -> DOM 업데이트 -> 렌더링
wait.until(lambda x:x.find_element_by_xpath("//div[@id='results']//a"))

for _ in driver.find_elements_by_xpath("//div[@id='results']//a"):
    print(_.tag_name)
    print(_.text)
    print(_.get_attribute("href"))
```

```
a
North Korea
http://example.webscraping.com/places/default/view/North-Korea-165 (http://example.w
ebscraping.com/places/default/view/North-Korea-165)
a
South Korea
http://example.webscraping.com/places/default/view/South-Korea-211 (http://example.w
ebscraping.com/places/default/view/South-Korea-211)
```

In [153]:

```
from selenium.common.exceptions import NoSuchElementException

wait = webdriver.support.ui.WebDriverWait(driver, 10, 0.5, NoSuchElementException)

#wait.until(lambda x:x.find_elements_by_xpath("//div[@id='results']//a"))
```

In [55]:

```
for _ in dom.select("#results a"):
    #print(_)
    print("나라명: ", _.text.strip())
    print("주소: ", _["href"])
```

```
나라명:  North Korea
주소:  /places/default/view/North-Korea-165
나라명:  South Korea
주소:  /places/default/view/South-Korea-211
```

In [120]:

```
for _ in driver.find_elements_by_xpath("//div[@id='results']//a"):
    print(_.tag_name)
    print(_.text)
    print(_.get_attribute("href"))
```

In [56]:

```
%%writefile account.json
{
    "id" : "본인 아이디",
    "pw" : "본인 패스워드"
}
```

```
Writing account.json
```

In [59]:

```
import json

with open("../account.json", "r", encoding="utf-8") as fp:
    account = json.load(fp)
```

In [60]:

```
account["id"]
```

Out[60]:

```
'본인 아이디'
```

In [16]:

```
#네이버 로그인
#(JS) 암호화 -> 토큰(Auth) -> 암호화 -> sso -> 세션
```

In [18]:

```
!pip install selenium
```

```
Requirement already satisfied: selenium in c:WprogramdataWanaconda3WlibWsite-package
s (3.141.0)
Requirement already satisfied: urllib3 in c:WprogramdataWanaconda3WlibWsite-packages
(from selenium) (1.24.1)
```

In [20]:

```
from selenium import webdriver
```

In [34]:

```
# driver = webdriver.Chrome()
# driver.implicitly_wait(3)
# driver.get("http://www.cgv.co.kr/reserve/show-times/")
# driver.find_element_by_xpath('//*[@id="contents"]/div[1]/div/div[2]/ul/li[1]/div/ul/li[21]/a').cl
# driver.implicitly_wait(2)
# driver.find_element_by_xpath('//*[@id="slider"]/div/ul/li[4]/div/a/strong').click()
# html = driver.page_source
# soup = BeautifulSoup(html, 'lxml')
# avengers = soup.select('body > div > div.sect-showtimes > ul > li:nth-child(1)')
# browser
# for n in avengers:
#     print(n.text.strip())
```

In [135]:

```
driver = webdriver.Chrome()
url = "https://nid.naver.com/nidlogin.login?url=http%3A%2F%2Fmail.naver.com%2F"
driver.get(url)
#driver.close()
```

In [136]:

```
driver.find_element_by_id("pw").clear()
driver.find_element_by_id("id").clear()
driver.find_element_by_name("id").send_keys("")
driver.find_element_by_css_selector("#pw").send_keys("")
```

In [137]:

```
driver.find_element_by_css_selector("input.btn_global").click()
#driver.find_element_by_xpath('//*[@id="frmNIDLogin"]/fieldset/input').click()
```

In [91]:

```
for _ in driver.find_elements_by_css_selector('strong.mail_title'):
    print(_.text)
```

```
[멜론] 신규 기기(브라우저)에서 로그인 되었습니다.
[신한카드] 결제일별 이용기간 (신용공여기간) 변경안내
[삼성보안포럼] 학생 논문 발표 희망자 초록접수합니다. (~7/5, 금요일까지)
FW: 2019 성공취업! 청년드림잡콘서트와함께
정보통신대학원 클라우드컴퓨팅 SaaS 실습.xls
SaaS 실습 설명 - 보기 권한 공유
서재영님, 예매 확인메일입니다.
서재영님, 예매 취소메일입니다.
(광고) 클릭 한 번으로 CGV 씨네샵 10% 할인받는 법!
The Spirit of Yonsei (섬김과 나눔, 기부와 헌신) - 5월호
서재영님, 온라인입금 내역이 확인되었습니다.
[멜론] 신규 기기(브라우저)에서 로그인 되었습니다.
최근의 IBM 무료 시험판 경험에 대해 알려주십시오 - 1분 설문조사
서재영님, 예매 확인메일입니다.
RE: 안녕하세요, 조교님. AI 이노베이션 스퀘어 자연어처리 수업 결석 일자 관련 메일입니
다.
```

In [92]:

```
dom = BeautifulSoup(driver.page_source, "lxml")
for _ in dom.select("strong.mail_title"):
    print(_.text)
```

```
메일 제목:[멜론] 신규 기기(브라우저)에서 로그인 되었습니다.
메일 제목:[신한카드] 결제일별 이용기간 (신용공여기간) 변경안내
메일 제목:[삼성보안포럼] 학생 논문 발표 희망자 초록접수합니다. (~7/5, 금요일까지)
메일 제목:FW: 2019 성공취업! 청년드림잡콘서트와함께
메일 제목:정보통신대학원 클라우드컴퓨팅 SaaS 실습.xls
메일 제목:SaaS 실습 설명 - 보기 권한 공유
메일 제목:서재영님, 예매 확인메일입니다.
메일 제목:서재영님, 예매 취소메일입니다.
메일 제목:(광고) 클릭 한 번으로 CGV 씨네샵 10% 할인받는 법!
메일 제목:The Spirit of Yonsei (섬김과 나눔, 기부와 헌신) - 5월호
메일 제목:서재영님, 온라인입금 내역이 확인되었습니다.
메일 제목:[멜론] 신규 기기(브라우저)에서 로그인 되었습니다.
메일 제목:최근의 IBM 무료 시험판 경험에 대해 알려주십시오 - 1분 설문조사
메일 제목:서재영님, 예매 확인메일입니다.
메일 제목:RE: 안녕하세요, 조교님. AI 이노베이션 스퀘어 자연어처리 수업 결석 일자 관
련 메일입니다.
```

In [95]:

```
#스팸 메일 탭으로 이동
driver.find_element_by_xpath('//*[@id="5_fol"]/span/a[1]').click()
```

In [97]:

```
for _ in driver.find_elements_by_css_selector('strong.mail_title'):
    print(_.text)
```

```
Microsoft's hottest, most in-demand courses are back!
(광고) [D-1] 데이터사이언스 80기가 곧 시작됩니다.
(광고)[원격지원] 팀뷰어 사용기회~ BIG 프로모션 안내 !!
Present code like never before & Slides turns 6!
Present code like never before & Slides turns 6!
【MasterCard】ナイキシューズ(NIKE)今日限り活動特価6399円
Get ready for Mother's Day
Pre-order Oculus Quest and Oculus Rift S now.
[New] Never leave your IDE
Galaxstore - GALAX RTX SERIES AND HOF GRAPHICS CARDS ARE HERE!
[New] Never leave your IDE
[다나와] 개인정보처리방침 개정 안내
Your IBM subscription is ready to use.
Details on Data Scientist Masters program: Simplilearn
【Getchu.com】NEKO WORK 最新作『LOVE－ラヴキューブ－』予約開始！
```

In [99]:

```
dom = BeautifulSoup(driver.page_source, "lxml")
for _ in dom.select("strong.mail_title"):
    print(_.text)
```

```
메일 제목:Microsoft's hottest, most in-demand courses are back!
메일 제목:(광고) [D-1] 데이터사이언스 80기가 곧 시작됩니다.
메일 제목:(광고)[원격지원] 팀뷰어 사용기회~ BIG 프로모션 안내 !!
메일 제목:Present code like never before & Slides turns 6!
메일 제목:Present code like never before & Slides turns 6!
 메일 제목:【MasterCard】ナイキシューズ(NIKE)今日限り活動特価6399円
메일 제목:Get ready for Mother's Day
메일 제목:Pre-order Oculus Quest and Oculus Rift S now.
메일 제목:[New] Never leave your IDE
메일 제목:Galaxstore - GALAX RTX SERIES AND HOF GRAPHICS CARDS ARE HERE!
메일 제목:[New] Never leave your IDE
메일 제목:[다나와] 개인정보처리방침 개정 안내
메일 제목:Your IBM subscription is ready to use.
메일 제목:Details on Data Scientist Masters program: Simplilearn
메일 제목:【Getchu.com】NEKO WORK 最新作『LOVE－ラヴキューブ－』予約開始！
```

In [101]:

```
driver.get("https://mail.naver.com")
```

In [102]:

```
driver.get_cookies()
```

Out[102]:


```
[{'domain': '.naver.com',
  'httpOnly': False,
  'name': 'NID_JKL',
  'path': '/',
  'secure': True,
  'value': 'CGOsOIoXUlYz+bQ3cwje0xLUXiUqWbpAY6FW2mMq64o='},
 {'domain': '.naver.com',
  'expiry': 2524640389.396513,
  'httpOnly': False,
  'name': 'NNB',
  'path': '/',
  'secure': False,
  'value': 'RJ6ECESECTKFY'},
 {'domain': 'mail.naver.com',
  'httpOnly': False,
  'name': 'NMUSER',
  'path': '/',
  'secure': False,
  'value': 'urKlKqEwaqbsKoEqKqbr1NOT+6lNBXKsKxE9FqtmKobwKon9KxUmaqgsaqRJaw/wFxRpad/s
yqvs6xRpadUstonstoRzaqROW9e7EoRpadUsawlGW430DVd974lR74lC+4kZ74FTWLm/axgmar05pzk/7xER
brkoWrlvMBil74lR74lC+4kZ74FTWLm/axgmam=='},
 {'domain': '.naver.com',
  'httpOnly': False,
  'name': 'nid_inf',
  'path': '/',
  'secure': False,
  'value': '1820031956'},
 {'domain': '.naver.com',
  'httpOnly': True,
  'name': 'NID_AUT',
  'path': '/',
  'secure': False,
  'value': 'AjxspHLA1ffNS7VIF3wCvg2kVqt09G/P0VjWI5H057XNRa0aZMgSWvVsptr0VnQn'},
 {'domain': '.naver.com',
  'httpOnly': False,
  'name': 'NID_SES',
  'path': '/',
  'secure': False,
  'value': 'AAABi3rwMQhHB674SinxTE6WyzCYteh1De6guPdCFEhsrle2aG6YsJgCobLKVcsPAZCE6DEP
nSlV05QQr518VnAC66ZPpghpRaAt100Yy+d9TiM+U2kKm770770b3yBZ6lGpWu13cD4EgsxKd1aD7xFXXs3T
J3CaqkArNOnn02WMOoZcChPsEl5FHXoKYAELuVgH5biRu4sHf1+E9HYRssrpojGH6YllADgbbXASMrpgukE8
2nWAGKp3l1zbak1CuyZC9rKEQfgnND6jsvwBFoWhnGSutUq2EgG6TLz+FDcCilZrochaFlQ1cxbA0z59vWxJ
uizhZvY+RDBcbSGBlFJpZ7ZxpYv5pwPYjHrb4x0ri4nxre5mQaJZQ7qU1M1icknk8GqB3l2Zf5Gos1FKPw1w
7QqaXKaHhboF2fZ87Me5wMn+aB9rFO6ciOr36UwNm/NNYJtkFyVB83WGdluV1Wm4sXOn/exrq3CMx2Qp5MJX
L3GkSntrtnX6QB95xp4CldSVv+vUjMDsgUsavuDEe5olxql='}]
```


In [103]:

```
driver.get("https://mail.naver.com")
driver.page_source
```

...

In [108]:

```
#selenium에서 일반 dom crawling으로 cookie 넘기기

from requests import Session

session = Session()
```

In [110]:

```
for cookie in driver.get_cookies():
    print(cookie["name"], cookie["value"])
    session.cookies.set(cookie["name"], cookie['value'])
```

```
NID_JKL CGOs0loXUlYz+bQ3cwje0xLUXiUqWbpAY6FW2mMq64o=
NNB RJ6ECESECTKFY
NMUSER urKlKqEwaqbsKoEqKqbr1NOT+6lNBXKsKxE9FqtmKobwKon9KxUmaqgsaqRJaw/wFxRpad/syqvs6
xRpadUstonstoRzaqROW9e7EoRpadUsawlGW430DVd974lR74lC+4kZ74FTWLm/axgmar05pzk/7xERbrkoW
rlvMBil74lR74lC+4kZ74FTWLm/axgmam==
nid_inf 1820031956
NID_AUT AjxspHLA1ffNS7VIF3wCvg2kVqt09G/P0VjWl5H057XNRa0aZMgSWvVsptr0VnQn
NID_SES AAABhK3KZJP7fFsX/90kkbNwR4BQP0qlqvJQuq00s+Urt7WzgX8tqSluYZcD+jm6eFgBTBf1Sn2/
955/bgXl8MPdqFoyQXjSvNLrD1+tvylumop+G5KM2kQ2Rnan2CV0cvUEYrRoSPdY0NrLHSliPePqNAV+juHw
hRVv5ug6+TGGGoxfR8do7dca31Yjob0uKqOR0t3j9Qf61Fn6Vc247zCV6pKp5m1h2gCRjMKP/p0+pDq0yLmX
gNhELTt8iSi45bMTlOPPpP5kJ1JA0kNDmSdNUqxeEUnclQXnEh0v9mGphtYYZzLmMruKi5EbLlrQZ4GL50gW
TmYXzWBw6wqWoEkXZe6ol5WemDVBnFKq9lOgHAe++61MDPbuZPo0MgnR8TADc18bMZW5D9tbcTGL1wsF3ALt
SncPbl/5ePnLs3FdMAA8JyH9Ez8BRSkzUlgZ0xFsTppBLZ9aaiKQx0/n2Gskusslyp4un/5bHhLFbu27aXcE
BXfs3zMFLJePYfp7Xv/rwyipG2HN7UrkXxeYmMiPiyM=
```

In [111]:

```
html = session.get("https://mail.naver.com")
html.text
```

...

In [112]:

```
# html = download("https://mail.naver.com")
# html.text
```

In [113]:

```
# html = session.post("")
# html.json
```

In [115]:

```
driver = webdriver.Chrome()
url = "https://logins.daum.net/accounts/signinform.do?url=https%3A%2F%2Fmail.daum.net%2F"
driver.get(url)

driver.find_element_by_id("id").clear()
driver.find_element_by_id("inputPwd").clear()
driver.find_element_by_id("id").send_keys("")
driver.find_element_by_id("inputPwd").send_keys("")
```

In [116]:

```
driver.find_element_by_id("loginBtn").click()
```

In [117]:

```
# driver.find_element_by_id("daumLogo").click()
# driver.find_element_by_xpath('//*[@id="mArticle"]/div[1]/div[2]/ul/li[1]/em/a').click()
```

In [21]:

```
driver = webdriver.Chrome()
driver.close()
```

In [42]:

```
#로그인이 새 창에서 뜨는 식으로 되어 있을 경우
driver = webdriver.Chrome()
driver.get('https://www.kt.com/')

#driver.switch_to.window(driver.window_handles[-1])
#print(driver.window_handles)
#driver.switch_to.window(driver.window_handles[0])
#print(driver.window_handles)
```

In [30]:

```
driver.find_element_by_xpath('//*[@id="cfmClFloating"]/div/div/div[2]/span[1]/a[1]').click()
```

In [38]:

```
print(driver.window_handles)
driver.switch_to.window(driver.window_handles[-1])
driver.find_element_by_id("userId").clear()
driver.find_element_by_id("password1").clear()
driver.find_element_by_id("userId").send_keys("qpwoes")
driver.find_element_by_id("password1").send_keys("qjrfpvp")
```

```
['CDwindow-27782DEC6B88DAAA6490CE4756AC46CE', 'CDwindow-1BE8DEC46EB3CB58FF445EC8E3DF
46CF']
```

In [ ]:

```
driver.switch_to.window(driver.window_handles[0])
```

In [ ]:

```
프로젝트 => 검색엔진(데이터), 문서분류, 스팸분류, 감성분석
for 검색엔진(데이터), 문서분류
다음, 네이트 뉴스
경로, 제목, 요약
=> vkdlfrudfh, ___, sodyd(.text)
파일경로 => 고유번호 + 카테고리(6)

정치 - 몇 번째 뉴스.txt 저장 ... 240개 문서
```

In [43]:

```
driver.close()
#driver.switch_to.window(driver.window_handles[0])
```