In [20]:

```
!pip install beautifulsoup4
```

Requirement already satisfied: beautifulsoup4 in c:\(\text{WprogramdataWanaconda3WlibWsite} \) -packages (4.7.1)

Requirement already satisfied: soupsieve>=1.2 in c:\(\mathbb{W}\)programdata\(\mathbb{W}\)anaconda\(\mathbb{W}\)lib\(\mathbb{W}\)site -packages (from beautifulsoup4) (1.8)

In [21]:

```
!pip install requests
```

Requirement already satisfied: requests in c:\programdata\programd

Requirement already satisfied: certifi>=2017.4.17 in c:\(\mathbb{W}\)programdata\(\mathbb{W}\)anaconda\(3\)\(\mathbb{W}\)lib\(\mathbb{W}\)site-packages (from requests) (2019.3.9)

Requirement already satisfied: urllib3<1.25,>=1.21.1 in c:\programdata\program

Requirement already satisfied: chardet<3.1.0,>=3.0.2 in c:\programdata\program

Requirement already satisfied: idna<2.9,>=2.5 in c:\(\text{WprogramdataWanaconda3WlibWsite} \) -packages (from requests) (2.8)

In [4]:

```
File "<ipython-input-4-f94f6171dbe1>", line 1 id => #
```

SyntaxError: invalid syntax

```
In [6]:
```

```
html = """
<html>
<head></head>
<body>
<div id="result">

<a class="red">go to page1</a>
<a class="blue">go to page2</a>

</div>
</body>
</html>
"""
```

In [8]:

```
from bs4 import BeautifulSoup

dom = BeautifulSoup(html, "|xml")
```

In [13]:

```
print(type(dom.select("a")))
print(dom.select("a.red"))
print(dom.select("a + a"))
<class 'list'>
```

```
<crass Tist >
[<a class="red">go to page1</a>]
[<a class="blue">go to page2</a>]
```

In [22]:

```
import requests

html = download("http://pythonscraping.com/pages/page3.html")
exercise = BeautifulSoup(html.text, "lxml")
```

In [34]:

```
exercise.select("td > img")
```

Out[34]:

```
[<img src="../img/gifts/img1.jpg"/>,
  <img src="../img/gifts/img2.jpg"/>,
  <img src="../img/gifts/img3.jpg"/>,
  <img src="../img/gifts/img4.jpg"/>,
  <img src="../img/gifts/img6.jpg"/>]
```

```
In [23]:
def download(url, params={}, retries=3):
    resp = None
    header = { "user-agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTM
L. like Gecko) Chrome/74.0.3729.108 Safari/537.36"}
    trv:
        resp = requests.get(url, params=params, headers = header)
       resp.raise for status()
    except requests.exceptions.HTTPError as e:
        if 500 <= e.response.status code < 600 and retries > 0:
            print(retries)
            resp = download(url. params. retries - 1)
            print(e.response.status_code, "\n")
            print(e.response.reason, "₩n")
            print(e.request.headers, "\n")
    return resp
In [25]:
html = download("https://www.google.com/search", {"q":"박보영"})
In [28]:
google = BeautifulSoup(html.text, "lxml")
In [30]:
len(google.select(".r > a"))
Out[30]:
In [31]:
[_["href"] for _ in google.select(".r > a")]
Out[31]:
['https://ko.wikipedia.org/wiki/%EB%B0%95%EB%B3%B4%EC%98%81'.
```

https://ko.wikipedia.org/wiki/%FB%B0%95%FB%B3%B4%FC%98%81%FC%9D%98 %FC%9F%91%FD% 92%88_%EB%AA%A9%EB%A1%9D',

'https://namu.wiki/w/%EB%B0%95%EB%B3%B4%EC%98%81',

https://www.msn.com/ko-kr/entertainment/news/%EB%B0%95%EB%B3%B4%EC%98%81-tvn-%E B%93%9C%EB%9D%BC%EB%A7%88-%EC%96%B4%EB%B9%84%EC%8A%A4-%EC%97%AC%EC%A3%BC%EC%9D%B8% EA%B3%B5/ar-BBLZzVh',

'https://twitter.com/hashtag/%EB%B0%95%EB%B3%B4%EC%98%81',

'https://www.hankyung.com/entertainment/article/201905030904H',

'https://news.joins.com/article/22895953',

'https://www.mk.co.kr/star/hot-issues/view/2019/05/293931/']

```
In [40]:
```

naver = BeautifulSoup(html.text, "lxml")

```
html = download("https://search.daum.net/search", {"q":"박보영"})
daum = BeautifulSoup(html.text. "lxml")
In [42]:
[_["href"] for _ in daum.select("div.wrap_tit > a")]
Out [42]:
['http://v.media.daum.net/v/20190507140359069?f=o'.
 'http://v.media.daum.net/v/20190507175153358?f=o',
 'http://v.media.daum.net/v/20190507185002021?f=o',
 'http://v.media.daum.net/v/20190507175700537?f=o'.
  'http://papa0717.tistorv.com/223'.
  'http://dudrms606.tistorv.com/367'.
  'http://adam24eve.tistory.com/858',
  'http://k3k2v.tistorv.com/35'.
  'https://ko.wikipedia.org/wiki/%EB%B0%95%EB%B3%B4%EC%98%81'.
  https://ko.wikipedia.org/wiki/%EB%B0%95%EB%B3%B4%EC%98%81%20%28%EB%B2%95%EC%A1%B
0%EC%9D%B8%29 '.
 http://cafe.daum.net/subdued20club/ReHf/2280152?g=%EB%B0%95%EB%B3%B4%EC%98%81'.
 http://cafe.daum.net/subdued20club/ReHf/2282606?a=%EB%B0%95%EB%B3%B4%EC%98%81'.
 'http://cafe.daum.net/ASMONACOFC/gAVU/1243818?q=%EB%B0%95%EB%B3%B4%EC%98%81',
  'http://cafe.daum.net/ok1221/9Zdf/1524913?g=%EB%B0%95%EB%B3%B4%EC%98%81',
  'http://tip.daum.net/answer/54503481?g=%EB%B0%95%EB%B3%B4%EC%98%81',
  'http://tip.daum.net/answer/54767339?g=%EB%B0%95%EB%B3%B4%EC%98%81'.
  'http://tip.daum.net/answer/55072609?g=%EB%B0%95%EB%B3%B4%EC%98%81'.
  'http://tip.daum.net/answer/54866982?g=%EB%B0%95%EB%B3%B4%EC%98%81'.
  'http://gall.dcinside.com/board/lists/?id=parkboyoung',
  'http://cafe.daum.net/parkbovoungfd'.
  'http://channels.vlive.tv/FCE49/video',
  'http://www.bic.or.kr/'.
  'iavascript:;'.
  'iavascript:;'.
  'iavascript:;'.
  'iavascript:;',
 'iavascript:;']
In [44]:
html = download("https://search.naver.com/search.naver", {"guery":"박보영"})
```

In [51]:

```
print([(_["href"], _.text) for _ in naver.select("dt > a")])
print(len(naver.select("dt > a")))
```

[('https://help.naver.com/support/alias/search/word/word 1.naver'. '도움말'). ('ht tps://music.naver.com/album/index.nhn?albumld=2500698'. '영화 너의결혼식 ...'). ('https://music.naver.com/album/index.nhn?albumId=565464', '오 나의 귀신님 ...'). ('?where=nexearch&sm=tab etc&guery=%EC%96%B4%EB%B9%84%EC%8A%A4%EA%B3%A0%EC%84%B8%E C%97%B0'. '고세연'). ('?where=nexearch&sm=tab etc&querv=%ED%9E%98%EC%8E%88%EC%97%A C%EC%9E%90%EB%8F%84%EB%B4%89%EC%88%9C%EB%8F%84%EB%B4%89%EC%88%9C'. '도봉순'). ('ht tps://m.post.naver.com/viewer/postView.nhn?volumeNo=19568466&memberNo=36645352&vTv pe=VERTICAL', '배우 박보영, 반할 수밖에 없는 청순 러블리'), ('https://m.post.nave r.com/viewer/postView.nhn?volumeNo=19813069&memberNo=25041664&vType=VERTICAL'. ''어비스' 박보영x안효섭. 케미에 벌써 선덕...복합장르 신기원...'). ('https://m.po st.naver.com/viewer/postView.nhn?volumeNo=17970239&memberNo=23519771&vTvpe=VERTICA L'. '박보영에게 '무장해제' 당한 상대 배우들'). ('http://sports.chosun.com/news/nt vpe.htm?id=201905080100041210002631&servicedate=20190507'. "'어비스' 박보영 · 안효 섭, 현실친구→부활 비주얼 재회 '스펙터클 전개'"), ('http://www.osen.co.kr/article/ G1111136950'. "'어비스' 박보영X안효섭, 추적자 강렬 변신..죽음 본격 추적[Oh!쎈 컷]"), ('http://www.dt.co.kr/contents.html?article_no=2019050702109919040003&ref=n aver', '박보영, 안효섭과 한 지붕 아래 동거 시작[어비스 2화]'), ('http://sports.don ga.com/3/a||/20190507/95412193/1', '[DAY컷] '어비스' 박보영, 안효섭 목욕 염탐? 티격태격 동거 시작'). ('https://cafe.naver.com/heidicafe/1241714'. '어비스 박보영 패션 트렌치코트 예쁘네요~'), ('https://cafe.naver.com/imsanbu/44492596', '박보영 이런 모습도 있네요?'), ('https://cafe.naver.com/dieselmania/26934064'. '드라마 어 비스 제작발표회 박보영'), ('https://cafe.naver.com/momakakao/11285454', '박보영은 언제쯤 결혼을 할까요?'), ('https://cafe.naver.com/linearbeit/43556', '2019.04.29 박보영 쏘피 바디피트 팬사인회 현장'), ('https://namu.wiki/w/%EB%B0%95%EB%B3%B4%EC% 98%81', '박보영 - 나무위키'), ('http://gall.dcinside.com/board/lists/?id=parkboyou ng', '박보영 갤러리 - 디시인사이드'), ('https://movie.daum.net/person/main?person] d=105244', '박보영'), ('https://kin.naver.com/qna/detail.nhn?d1id=3&dirld=3021401& docId=326457379&qb=67CV67007.liB&enc=utf8§ion=kin&rank=1&search_sort=0&spg=1'... '박보영vs한지민'), ('https://kin.naver.com/qna/detail.nhn?d1id=3&dirld=306&docld=3 25917258&gb=67CV67007.JiB&enc=utf8§ion=kin&rank=2&search_sort=0&spg=1'. '제기 박보영 옛날 턱이락 똑같은데 어'), ('https://kin.naver.com/qna/detail.nhn?d1id=3&di rld=31402&docld=322399276&qb=67CV67007JiB&enc=utf8§ion=kin&rank=3&search_sort= 0&spq=1', '박보영 이 사진 고화질 부탁 내공 5'), ('https://kin.naver.com/gna/detai L.nhn?d1id=3&dirld=30104&docld=319183446&gb=67CV67007.JiB&enc=utf8§ion=kin&rank =4&search_sort=0&spq=1', '유라 vs 박보영'), ('https://blog.naver.com/shkwon1128?Re direct=Log&logNo=221530738046', '통해 부활한 박보영과 안효섭의 러블리스릴러 이야 기'). ('https://blog.naver.com/is2v86?Redirect=Log&logNo=221454815826', '박보영 : 오 나의 귀신님 나봉선'). ('https://blog.naver.com/dlglwm14?Redirect=Log&logNo=2215 29807195', '박보영 나이 단발 사진 키 알아보기'), ('https://blog.naver.com/lvi0088? Redirect=Loa&loaNo=221530708253'. '어비스 드라마 인물관계도 등장인물 몇부작 줄거리 박보영 안효섭'). ('https://komartin.blog.me/221530734469'. '어비스 드라마 인물관계 도 & 줄거리 (박보영, 안효섭, 이성재...')]

In [53]:

```
html = download("https://www.naver.com/")
naver2 = BeautifulSoup(html.text, "lxml")
```

```
In [77]:
[ _.text for _ in naver2.select(".ah_roll_area li.ah_item .ah_k")]
Out [77]:
['하나경'.
 '울산대교'
 '강은비'.
 '한서희'.
 '정용국'.
 '맥반석닭갈비'.
 '고양 창릉'.
 '생방송 오늘저녁'
 '삼성바이오로직스'.
 '3기 신도시'
 '이다해'.
 '어비스'.
 '부천 대장동'.
 '강이준'.
 '심재철'.
 '양의지'.
 '초면에 사랑합니다'.
 '규현'.
 '공민규'.
 '한선교']
In [78]:
seed = "http://example.webscraping.com/places/default/index/"
html = download(seed)
dom = BeautifulSoup(html.content, "lxml")
In [79]:
len(dom.select("a"))
Out [79]:
16
In [83]:
links = [_["href"] for _ in dom.select("a")]
# #은 inner link로 아무 동작 하지 않게 하는 것
In [85]:
len(links)
Out[85]:
16
In [86]:
#, / <- 애만 가져올 것, http, https
```

```
In [90]:
```

```
/places/default/user/register?_next=/places/default/index
/places/default/user/login?_next=/places/default/index
/places/default/index
/places/default/search
/places/default/view/Afghanistan=1
/places/default/view/Aland=Islands=2
/places/default/view/Albania=3
/places/default/view/Algeria=4
/places/default/view/American=Samoa=5
/places/default/view/Andorra=6
/places/default/view/Angola=7
/places/default/view/Anguilla=8
/places/default/view/Antarctica=9
/places/default/view/Antigua=and=Barbuda=10
/places/default/index/1
```

In [96]:

```
len(seed), len(unseen) , len(link), len(newLink)
```

Out[96]:

```
(52, 15, 23, 53)
```

In [121]:

```
threshold = 3
def urlExtractor(seed. limit):
   if limit > threshold:
       return []
   html = download(seed)
   dom = BeautifulSoup(html.text, "lxml")
   links = [_["href"] for _ in dom.select("a") if _.has_attr("href")] #핵심
   unseen = list()
   for link in links:
       if len(link) > 1 and link[0] == "/":
           #print(link)
           newLink = requests.compat.urljoin(seed, link)
           if newLink not in unseen:
               unseen.append((newLink, limit+1))
                 print("Skipped2", link)
        elif link.startswith("http"):
           unseen.append((link, limit+1))
         e/se:
             print("Skipped1", link)
     print("{0} [{1}]".format(seed, len(unseen)))
   return unseen
```

In [102]:

```
import time
import random

seed = "http://example.webscraping.com/places/default/index/"

queue = urlExtractor(seed) # => unseen totally
seen = list()

while queue:
    seed = queue.pop(0)
    seen.append(seed)
    time.sleep(random.randint(1,3))
    links = urlExtractor(seed) #links는 unseen 결치면 안되고, seen에 있으면 안됨. 이거 체크
    queue.extend([_ for _ in links if _ not in queue and _ not in seen])
    queue.extend(urlExtractor(seed))
```

```
/places/default/user/register?_next=/places/default/index
/places/default/user/login? next=/places/default/index
/places/default/index
/places/default/search
/places/default/view/Afghanistan-1
/places/default/view/Aland-Islands-2
/places/default/view/Albania-3
/places/default/view/Algeria-4
/places/default/view/American-Samoa-5
/places/default/view/Andorra-6
/places/default/view/Angola-7
/places/default/view/Anguilla-8
/places/default/view/Antarctica-9
/places/default/view/Antigua-and-Barbuda-10
/places/default/index/1
/places/default/user/register
/places/default/user/login
/places/default/index
/places/default/search
/places/default/user/register
/places/default/user/login
/places/default/index
/places/default/search
429
TOO MANY REQUESTS
{'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTM
L. like Gecko) Chrome/74.0.3729.108 Safari/537.36', 'Accept-Encoding': 'gzip, defl
ate', 'Accept': '*/*', 'Connection': 'keep-alive'}
429
TOO MANY REQUESTS
{'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTM
L. like Gecko) Chrome/74.0.3729.108 Safari/537.36'. 'Accept-Encoding': 'gzip. defl
ate', 'Accept': '*/*', 'Connection': 'keep-alive'}
429
TOO MANY REQUESTS
{'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTM
L, like Gecko) Chrome/74.0.3729.108 Safari/537.36', 'Accept-Encoding': 'gzip, defl
ate', 'Accept': '*/*', 'Connection': 'keep-alive'}
429
TOO MANY REQUESTS
{'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTM
L, like Gecko) Chrome/74.0.3729.108 Safari/537.36', 'Accept-Encoding': 'gzip, defl
ate', 'Accept': '*/*', 'Connection': 'keep-alive'}
429
TOO MANY REQUESTS
```

{'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTM L, like Gecko) Chrome/74.0.3729.108 Safari/537.36', 'Accept-Encoding': 'gzip, defl

```
TOO MANY REQUESTS
```

```
{'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTM
L, like Gecko) Chrome/74.0.3729.108 Safari/537.36', 'Accept-Encoding': 'gzip, defl
ate', 'Accept': '*/*', 'Connection': 'keep-alive'}
TOO MANY REQUESTS
{'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTM
L. like Gecko) Chrome/74.0.3729.108 Safari/537.36', 'Accept-Encoding': 'gzip, defl
ate', 'Accept': '*/*', 'Connection': 'keep-alive'}
TOO MANY REQUESTS
{'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTM
L. like Gecko) Chrome/74.0.3729.108 Safari/537.36', 'Accept-Encoding': 'gzip, defl
ate', 'Accept': '*/*', 'Connection': 'keep-alive'}
TOO MANY REQUESTS
{'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTM
L, like Gecko) Chrome/74.0.3729.108 Safari/537.36', 'Accept-Encoding': 'gzip, defl
ate', 'Accept': '*/*', 'Connection': 'keep-alive'}
TOO MANY REQUESTS
{'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTM
L, like Gecko) Chrome/74.0.3729.108 Safari/537.36', 'Accept-Encoding': 'gzip, defl
ate', 'Accept': '*/*', 'Connection': 'keep-alive'}
In [126]:
html = download("https://www.google.com/search", {"q":"박보영"})
google2 = BeautifulSoup(html.content, "lxml")
In [106]:
len(google2.select(".r > a"))
Out[106]:
In [127]:
queue = [(["href"], 0) for _ in google2.select(".r > a")]
```

In [115]:
queue
Out[115]:

[]

In [128]:

```
seen = list()
while queue:
    #(URL, LIMIT)
    link = queue.pop(0)
    seen.append(link[0])
    links = urlExtractor(link[0], link[1])

    queueURL = [_[0] for _ in queue]
    queue.extend([_ for _ in links if _[0] not in queueURL and _ not in seen])
    print("{0} {1} [{2}]".format(">"*link[1], link[0], len(links)))

# for link in queue:
# seen.append(link[0])
# links = urlExtractor(link[0], link[1])
# html = download(link)
# google2 = BeautifulSoup(html.content, "lxml")
# print(link, len([_["href"] for _ in google2.select("a") if _.has_attr("href")]))
```

```
https://ko.wikipedia.org/wiki/%EB%B0%95%EB%B3%B4%EC%98%81 [820]
```

https://ko.wikipedia.org/wiki/%EB%B0%95%EB%B3%B4%EC%98%81%EC%9D%98_%EC%9E%91%ED%92%88 %EB%AA%A9%EB%A1%9D [385]

https://namu.wiki/w/%EB%B0%95%EB%B3%B4%EC%98%81 [1139]

https://www.msn.com/ko-kr/entertainment/news/%EB%B0%95%EB%B3%B4%EC%98%81-tvn-%EB%93%9C%EB%9D%BC%EB%A7%88-%EC%96%B4%EB%B9%84%EC%8A%A4-%EC%97%AC%EC%A3%BC%EC%9D%B8%EA%B3%B5/ar-BBLZzVh [47]

https://twitter.com/hashtag/%EB%B0%95%EB%B3%B4%EC%98%81 [201]

https://www.hankyung.com/entertainment/article/201905030904H [189]

https://www.hankyung.com/entertainment/article/201905077647H [184]

https://news.joins.com/article/22895953 [34]

- > https://ko.wikipedia.org/wiki/%EC%9C%84%ED%82%A4%EB%B0%B1%EA%B3%BC:%EB%8F%99%EC%9D%BC%EC%9D%B4%EC%9D%98%EC%96%B4 %EB%AC%B8%EC%84%9C [200]
- > https://ko.wikipedia.org/wiki/%EB%B0%95%EB%B3%B4%EC%98%81_(%EB%8F%99%EC%9D%8C%E C%9D%B4%EC%9D%98) [51]
- > https://ko.wikipedia.org/wiki/%ED%8C%8C%EC%9D%BC:%EC%98%81%ED%99%94_%27%EB%84%8 8%EC%9D%98_%EA%B2%B0%ED%98%BC%EC%8B%9D%27_%EC%96%B8%EB%A1%A0%EC%8B%9C%EC%82%AC%ED% 9A%8C_%EB%B0%95%EB%B3%B4%EC%98%81.jpg [54]
- > https://ko.wikipedia.org/wiki/%EB%8C%80%ED%95%9C%EB%AF%BC%EA%B5%AD [3400]
- > https://ko.wikipedia.org/wiki/%EC%B6%A9%EC%B2%AD%EB%B6%81%EB%8F%84 [417]
- > https://ko.wikipedia.org/wiki/%EA%B4%B4%EC%82%B0%EA%B5%B0 [380]
- > https://ko.wikipedia.org/wiki/%EC%A6%9D%ED%8F%89%EC%9D%8D [132]
- > https://ko.wikipedia.org/wiki/%EC%A6%9D%ED%8F%89%EA%B5%B0 [247]
- > https://ko.wikipedia.org/wiki/%EB%B0%B0%EC%9A%B0 [249]
- > https://ko.wikipedia.org/wiki/2006%EB%85%84 [1266]
- > https://ko.wikipedia.org/wiki/%EB%8B%A8%EA%B5%AD%EB%8C%80%ED%95%99%EA%B5%90 [162 3]
- > https://ko.wikipedia.org/wiki/%EB%B0%95%EB%B3%B4%EC%98%81%EC%9D%98_%EC%88%98%EC%83%81_%EB%B0%8F_%ED%9B%84%EB%B3%B4_%EB%AA%A9%EB%A1%9D [244]
- > http://www.vlive.tv/channels/FCE49/ [0]
- > https://ko.wikipedia.org/wiki/V_LIVE [129]
- > https://ko.wikipedia.org/wiki/1990%EB%85%84 [1487]
- > https://ko.wikipedia.org/wiki/2%EC%9B%94_12%EC%9D%BC [552]
- > https://ko.wikipedia.org/wiki/%EB%8C%80%ED%95%9C%EB%AF%BC%EA%B5%AD [3400]
- > https://ko.wikipedia.org/wiki/%EB%B0%B0%EC%9A%B0 [249]
- > https://ko.wikipedia.org/w/index.php?title=%EB%B0%95%EB%B3%B4%EC%98%81&action=edit§ion=1 [34]
- > https://ko.wikipedia.org/wiki/%EB%B9%84%EB%B0%80%EC%9D%98_%EA%B5%90%EC%A0%95 [10 9]
- > https://ko.wikipedia.org/wiki/%EB%A7%88%EB%85%80%EC%9C%A0%ED%9D%AC [256]
- https://ko.wikipedia.org/wiki/%EB%8B%AC%EB%A0%A4%EB%9D%BC_%EA%B3%A0%EB%93%B1%EC%96%B4 [131]
- > https://ko.wikipedia.org/wiki/%EC%99%95%EA%B3%BC_%EB%82%98_(%EB%93%9C%EB%9D%BC%EB%A7%88) [544]
- > https://ko.wikipedia.org/wiki/%ED%8F%90%EB%B9%84_%EC%9C%A4%EC%94%A8 [588]
- > https://ko.wikipedia.org/wiki/%EC%A0%95%EA%B8%80%ED%94%BC%EC%89%AC [132]
- > https://ko.wikipedia.org/wiki/%EC%9A%B8%ED%95%99%EA%B5%90_%EC%9D%B4%ED%8B%B0 [13 4]
- > https://ko.wikipedia.org/wiki/%EC%B4%88%EA%B0%90%EA%B0%81_%EC%BB%A4%ED%94%8C [10 3]
- > https://ko.wikipedia.org/wiki/%EA%B3%BC%EC%86%8D%EC%8A%A4%EC%BA%94%EB%93%A4 [16
- > https://ko.wikipedia.org/wiki/%EC%A0%84%EC%A3%BC%EA%B5%AD%EC%A0%9C%EC%98%81%ED%9 9%94%EC%A0%9C [645]
- > https://ko.wikipedia.org/wiki/%EC%98%B4%EB%8B%88%EB%B2%84%EC%8A%A4_%EC%98%81%ED%99%94 [74]
- > https://ko.wikipedia.org/wiki/%EC%8B%9C%EC%84%A0 1318 [143]
- > https://ko.wikipedia.org/wiki/%EB%AF%B8%ED%99%95%EC%9D%B8_%EB%8F%99%EC%98%81%EC%83%81:_%EC%AO%88%EB%8C%80%ED%81%B4%EB%A6%AD%EA%B8%88%EC%A7%80 [92]
- > https://ko.wikipedia.org/wiki/%EB%8A%91%EB%8C%80%EC%86%8C%EB%85%84_(%EC%98%81%E D%99%94) [256]

- > https://ko.wikipedia.org/wiki/%ED%94%BC%EB%81%93%EB%8A%94_%EC%B2%AD%EC%B6%98 [14 2]
- > https://ko.wikipedia.org/wiki/%EA%B2%BD%EC%84%B1%ED%95%99%EA%B5%90:_%EC%82%AC%EB%9D%BC%EC%A7%84 %EC%86%8C%EB%85%80%EB%93%A4 [108]
- > https://ko.wikipedia.org/wiki/TvN_%EA%B8%88%ED%86%A0%EB%93%9C%EB%9D%BC%EB%A7%88
- > https://ko.wikipedia.org/wiki/%EC%98%A4_%EB%82%98%EC%9D%98_%EA%B7%80%EC%8B%A0%E B%8B%98 [283]
- > https://ko.wikipedia.org/wiki/%EB%8F%8C%EC%97%B0%EB%B3%80%EC%9D%B4_(%EC%98%81%ED%99%94) [116]
- https://ko.wikipedia.org/wiki/%EC%97%B4%EC%A0%95%EA%B0%99%EC%9D%80%EC%86%8C%EB%A6%AC%ED%95%98%EA%B3%A0%EC%9E%88%EB%84%A4 [111]
- > https://ko.wikipedia.org/wiki/JTBC_%EA%B8%88%ED%86%A0%EB%93%9C%EB%9D%BC%EB%A7%88
- > https://ko.wikipedia.org/wiki/%ED%9E%98%EC%8E%88%EC%97%AC%EC%9E%90_%EB%8F%84%EB%B4%89%FC%88%9C [416]
- > https://ko.wikipedia.org/wiki/%EB%84%88%EC%9D%98_%EA%B2%B0%ED%98%BC%EC%8B%9D [5 5]
- > https://ko.wikipedia.org/w/index.php?title=%EB%B0%95%EB%B3%B4%EC%98%81&action=edit§ion=2 [34]
- > https://ko.wikipedia.org/wiki/%EC%B6%A9%EC%B2%AD%EB%B6%81%EB%8F%84 [417]
- > https://ko.wikipedia.org/wiki/%EC%A6%9D%ED%8F%89%EA%B5%B0 [247]
- > https://ko.wikipedia.org/wiki/%EB%8C%80%ED%95%9C%EB%AF%BC%EA%B5%AD_%EC%9C%A1%EA%B5%B0 %ED%8A%B9%EC%88%98%EC%A0%B4%EC%82%AC%EB%A0%B9%EB%B6%80 [363]
- > https://ko.wikipedia.org/wiki/%EC%98%88%EB%B9%84%EC%97%AD [116]
- > https://ko.wikipedia.org/wiki/%EB%8C%80%ED%95%9C%EB%AF%BC%EA%B5%AD_%EC%9C%A1%EA%B5%BO [1078]
- > https://ko.wikipedia.org/wiki/%EC%9B%90%EC%82%AC [52]
- > https://ko.wikipedia.org/wiki/%EC%A6%9D%ED%8F%89%EC%B4%88%EB%93%B1%ED%95%99%EA%B5%90 [368]
- > https://ko.wikipedia.org/wiki/%EC%A6%9D%ED%8F%89%EC%97%AC%EC%9E%90%EC%A4%91%ED%9 5%99%EA%B5%90 [238]
- > https://ko.wikipedia.org/wiki/%EB%8C%80%EC%84%B1%EC%97%AC%EC%9E%90%EC%83%81%EC%97%85%EA%B3%A0%EB%93%B1%ED%95%99%EA%B5%90 [347]
- > https://ko.wikipedia.org/wiki/%EB%8B%A8%EA%B5%AD%EB%8C%80%ED%95%99%EA%B5%90 [162 3]
- > https://ko.wikipedia.org/wiki/%EC%84%9C%EC%9A%B8%EA%B5%AD%EC%AO%9C%EC%B2%AD%EC%86%8C%EB%85%84%EC%98%81%ED%99%94%EC%AO%9C [171]
- > https://ko.wikipedia.org/wiki/%EB%8B%A8%EA%B5%AD%EB%8C%80%ED%95%99%EA%B5%90 [162 3]
- $> https://ko.wikipedia.org/w/index.php?title=\%EB\%B0\%95\%EB\%B3\%B4\%EC\%98\%81\&action=edit\§ion=3\ [34]$
- > https://ko.wikipedia.org/w/index.php?title=%EB%B0%95%EB%B3%B4%EC%98%81&action=edit§ion=4 [34]
- > https://ko.wikipedia.org/wiki/%EB%B9%84%EB%B0%80%EC%9D%98_%EA%B5%90%EC%A0%95 [10 9]
- > https://ko.wikipedia.org/wiki/%EB%A7%88%EB%85%80%EC%9C%A0%ED%9D%AC [256]
- > https://ko.wikipedia.org/wiki/%ED%95%9C%EA%B0%80%EC%9D%B8 [461]
- > https://ko.wikipedia.org/wiki/%EB%8B%AC%EB%A0%A4%EB%9D%BC_%EA%B3%A0%EB%93%B1%EC% 96%B4 [131]
- > https://ko.wikipedia.org/wiki/%EC%99%95%EA%B3%BC_%EB%82%98_(%EB%93%9C%EB%9D%BC%EB%A7%88) [544]
- > https://ko.wikipedia.org/wiki/%ED%8F%90%EB%B9%84 %EC%9C%A4%EC%94%A8 [588]
- > https://ko.wikipedia.org/wiki/%ED%8F%90%EB%B9%84 %EC%9C%A4%EC%94%A8 [588]
- > https://ko.wikipedia.org/wiki/%EC%9C%A0%EC%8A%B9%ED%98%B8 [582]
- > https://ko.wikipedia.org/wiki/%EC%A3%BC%EB%AF%BC%EC%88%98 [125]
- > https://ko.wikipedia.org/wiki/%EC%A0%95%EA%B8%80%ED%94%BC%EC%89%AC [132]
- > https://ko.wikipedia.org/wiki/%EA%B9%80%ED%8F%AC%EC%99%B8%EA%B3%A0 [660]
- > https://ko.wikipedia.org/wiki/%ED%94%BC%EB%B2%84%EB%94%94%EC%83%81 [85]
- > https://ko.wikipedia.org/wiki/2010%EB%85%84 [1866]
- > https://ko.wikipedia.org/wiki/9%EC%9B%94_23%EC%9D%BC [555]

- > https://ko.wikipedia.org/wiki/%EA%B5%AD%EA%B0%80%EC%9D%B8%EA%B6%8C%EC%9C%84%EC%9R%90%FD%9A%RC [230]
- > https://ko.wikipedia.org/wiki/%EC%98%B4%EB%BB%88%EB%B2%84%EC%8A%A4_%EC%98%81%ED% 99%94 [74]
- > https://ko.wikipedia.org/wiki/%EC%8B%9C%EC%84%A0_1318 [143]
- > https://ko.wikipedia.org/wiki/%EC%98%B4%EB%BB%88%EB%B2%84%EC%8A%A4_%EC%98%81%ED% 99%94 [74]
- > https://ko.wikipedia.org/wiki/%EC%A0%84%EC%A3%BC%EA%B5%AD%EC%A0%9C%EC%98%81%ED%9 9%94%EC%A0%9C [645]
- > https://ko.wikipedia.org/wiki/%EC%9A%B8%ED%95%99%EA%B5%90_%EC%9D%B4%ED%8B%B0 [13 4]
- > https://ko.wikipedia.org/wiki/%EA%B9%80%EC%88%98%EB%A1%9C_(%EB%B0%B0%EC%9A%B0)
- > https://ko.wikipedia.org/wiki/%EC%99%95%EA%B3%BC_%EB%82%98_(%EB%93%9C%EB%9D%BC%E B%A7%88) [544]
- > https://ko.wikipedia.org/wiki/%EA%B5%AC%ED%98%9C%EC%84%A0 [759]
- > https://ko.wikipedia.org/wiki/%EC%B5%9C%EA%B0%95%EC%B9%A0%EC%9A%B0 [591]
- > https://ko.wikipedia.org/wiki/%EC%97%90%EB%A6%AD_(%EA%B0%80%EC%88%98) [849]
- > https://ko.wikipedia.org/wiki/%EC%B4%88%EA%B0%90%EA%B0%81_%EC%BB%A4%ED%94%8C [10 3]
- > https://ko.wikipedia.org/wiki/%EC%A7%84%EA%B5%AC [408]
- > https://ko.wikipedia.org/wiki/%EB%AC%B8%ED%99%94%EC%B2%B4%EC%9C%A1%EA%B4%80%EA%B4%91%FB%B6%80 [414]
- > https://ko.wikipedia.org/w/index.php?title=%EB%B0%95%EB%B3%B4%EC%98%81&action=edit§ion=5 [34]
- > https://ko.wikipedia.org/wiki/%EA%B3%BC%EC%86%8D%EC%8A%A4%EC%BA%94%EB%93%A4 [16 4]
- > https://ko.wikipedia.org/wiki/%FC%B0%A8%FD%83%9C%FD%98%84 [1866]
- > https://ko.wikipedia.org/wiki/%EC%A2%85%ED%95%A9%EB%B3%91%EC%9B%90_2 [376]
- > https://ko.wikipedia.org/wiki/%EC%B0%A8%ED%83%9C%ED%98%84 [1866]
- > https://ko.wikipedia.org/wiki/%EC%8A%A4%ED%83%80%EC%9D%98_%EC%97%B0%EC%9D%B8 [28 8]
- > https://ko.wikipedia.org/wiki/%EC%B5%9C%EC%A7%80%EC%9A%B0 [1239]
- > https://ko.wikipedia.org/wiki/%EA%B3%BC%EC%86%8D%EC%8A%A4%EC%BA%94%EB%93%A4 [16 4]
- > https://ko.wikipedia.org/wiki/%EC%B2%AD%EB%A3%A1%EC%98%81%ED%99%94%EC%83%81 [19 1]
- > https://ko.wikipedia.org/wiki/%EB%B0%B1%EC%83%81%EC%98%88%EC%88%A0%EB%8C%80%EC%8 3%81 [3537]
- > https://ko.wikipedia.org/wiki/%ED%95%9C%EA%B5%AD%EC%98%81%ED%99%94%ED%8F%89%EB%A 1%AO%EA%B0%80%ED%98%91%ED%9A%BC%EC%83%81 [804]
- > https://ko.wikipedia.org/wiki/%EB%94%94%EB%A0%89%ED%84%B0%EC%8A%A4_%EC%BB%B7_%EC%8B%9C%EC%83%81%EC%8B%9D [385]
- > https://ko.wikipedia.org/wiki/%EC%B2%AD%EB%A3%A1%EC%98%81%ED%99%94%EC%83%81 [19 1]
- > https://ko.wikipedia.org/wiki/%EA%B3%BC%EC%86%8D%EC%8A%A4%EC%BA%94%EB%93%A4 [16 4]
- > https://ko.wikipedia.org/w/index.php?title=%EB%B0%95%EB%B3%B4%EC%98%81&action=edit§ion=6 [34]
- > https://ko.wikipedia.org/wiki/%EB%AF%B8%ED%99%95%EC%9D%B8_%EB%8F%99%EC%98%81%EC%83%81:_%EC%A0%88%EB%8C%80%ED%81%B4%EB%A6%AD%EA%B8%88%EC%A7%80 [92]
- > https://ko.wikipedia.org/wiki/%EB%AF%B8%ED%99%95%EC%9D%B8_%EB%8F%99%EC%98%81%EC%83%81: %EC%A0%88%EB%8C%80%ED%81%B4%EB%A6%AD%EA%B8%88%EC%A7%80 [92]
- > https://ko.wikipedia.org/wiki/%EB%AF%B8%ED%99%95%EC%9D%B8_%EB%8F%99%EC%98%81%EC%83%81:_%EC%A0%88%EB%8C%80%ED%81%B4%EB%A6%AD%EA%B8%88%EC%A7%80 [92]
- > https://ko.wikipedia.org/wiki/%EA%B9%80%ED%83%9C%EA%B2%BD_(%EC%98%81%ED%99%94_%EA%B0%90%EB%8F%85) [77]
- > https://ko.wikipedia.org/wiki/%EB%8A%91%EB%8C%80%EC%86%8C%EB%85%84_(%EC%98%81%E D%99%94) [256]
- > https://ko.wikipedia.org/wiki/%EC%A1%B0%EC%84%B1%ED%9D%AC_(%EC%98%81%ED%99%94_%EA%B0%90%EB%8F%85) [80]

- > https://ko.wikipedia.org/wiki/%EB%8A%91%EB%8C%80%EC%86%8C%EB%85%84_(%EC%98%81%E D%99%94) [256]
- > https://ko.wikipedia.org/wiki/%ED%95%9C%EA%B5%AD%EA%B0%A4%EB%9F%BD [62]
- > https://ko.wikipedia.org/wiki/%EB%8A%91%EB%8C%80%EC%86%8C%EB%85%84_(%EC%98%81%ED%99%94) [256]
- > https://ko.wikipedia.org/wiki/%ED%94%BC%EC%96%B4%EC%84%A0_%EC%A0%8A%EC%9D%80_%EC%98%81%ED%99%94_%ED%8E%98%EC%8A%A4%ED%8B%B0%EB%B2%8C [183]
- > https://ko.wikipedia.org/wiki/%EC%A1%B0%EC%84%B1%ED%9D%AC_(%EC%98%81%ED%99%94_%EA%B0%90%EB%8F%85) [80]
- > https://ko.wikipedia.org/wiki/%ED%95%9C%EA%B5%AD%EC%98%81%ED%99%94%EC%95%84%EC%B9%B4%EB%8D%B0%EB%AF%B8 [102]
- > https://ko.wikipedia.org/wiki/%ED%94%BC%EB%81%93%EB%8A%94_%EC%B2%AD%EC%B6%98 [14 2]
- > https://ko.wikipedia.org/wiki/%ED%94%BC%EB%81%93%EB%8A%94_%EC%B2%AD%EC%B6%98 [14 2]
- > https://ko.wikipedia.org/wiki/%EC%9D%B4%EC%A2%85%EC%84%9D_(%EB%B0%B0%EC%9A%B0) [1202]
- > https://ko.wikipedia.org/w/index.php?title=%EB%B0%95%EB%B3%B4%EC%98%81&action=edit§ion=7 [34]
- > https://ko.wikipedia.org/wiki/%EC%9D%B4%ED%95%B4%EC%98%81_(%EC%98%81%ED%99%94_%EA%B0%90%EB%8F%85) [212]
- > https://ko.wikipedia.org/wiki/%EA%B2%BD%EC%84%B1%ED%95%99%EA%B5%90:_%EC%82%AC%EB%9D%BC%EC%A7%84_%EC%86%8C%EB%85%80%EB%93%A4 [108]
- > https://ko.wikipedia.org/wiki/%EC%9D%B4%ED%95%B4%EC%98%81_(%EC%98%81%ED%99%94_%EA%B0%90%EB%8F%85) [212]
- > https://ko.wikipedia.org/wiki/%EA%B2%BD%EC%84%B1%ED%95%99%EA%B5%90:_%EC%82%AC%EB%9D%BC%EC%A7%84_%EC%86%8C%EB%85%80%EB%93%A4 [108]
- > https://ko.wikipedia.org/wiki/%EA%B2%BD%EC%84%B1%ED%95%99%EA%B5%90:_%EC%82%AC%EB%9D%BC%EC%A7%84_%EC%86%8C%EB%85%80%EB%93%A4 [108]
- > https://ko.wikipedia.org/wiki/%EC%98%A4_%EB%82%98%EC%9D%98_%EA%B7%80%EC%8B%A0%E B%8B%98 [283]
- > https://ko.wikipedia.org/wiki/%EC%9D%91%EB%8B%B5%ED%95%98%EB%9D%BC1994 [752]
- > https://ko.wikipedia.org/wiki/%EB%AF%B8%EC%83%9D_(%EB%93%9C%EB%9D%BC%EB%A7%88)
- > https://ko.wikipedia.org/wiki/%EC%98%A4_%EB%82%98%EC%9D%98_%EA%B7%80%EC%8B%A0%E B%8B%98 [283]
- > https://ko.wikipedia.org/wiki/%EC%98%A4_%EB%82%98%EC%9D%98_%EA%B7%80%EC%8B%A0%E B%8B%98 [283]
- > https://ko.wikipedia.org/wiki/%EB%8F%8C%EC%97%B0%EB%B3%80%EC%9D%B4_(%EC%98%81%ED%99%94) [116]
- > https://ko.wikipedia.org/wiki/%EC%9D%B4%EA%B4%91%EC%88%98_(%EB%B0%B0%EC%9A%B0) [806]
- > https://ko.wikipedia.org/wiki/%EA%B5%AD%EA%B0%80%EC%A0%95%EB%B3%B4%EC%9B%90_%EC%97%AC%EB%A1%A0 %EC%A1%B0%EC%9E%91 %EC%82%AC%EA%B1%B4 [2100]
- > https://ko.wikipedia.org/wiki/CJ_E%26M [643]
- > https://ko.wikipedia.org/wiki/%ED%86%A0%EB%A1%A0%ED%86%A0_%EA%B5%AD%EC%A0%9C_%EC%98%81%ED%99%94%EC%A0%9C [207]
- > https://ko.wikipedia.org/wiki/%EC%97%B4%EC%A0%95%EA%B0%99%EC%9D%80%EC%86%8C%EB%A6%AC%ED%95%98%EA%B3%A0%EC%9E%88%EB%84%A4[111]
- > https://ko.wikipedia.org/wiki/%EC%97%B4%EC%A0%95,_%EA%B0%99%EC%9D%80_%EC%86%8C%EB%A6%AC_%ED%95%98%EA%B3%A0_%EC%9E%88%EB%84%A4 [88]
- > https://ko.wikipedia.org/wiki/%ED%95%9C%EA%B5%AD%EA%B0%A4%EB%9F%BD [62]
- > https://ko.wikipedia.org/w/index.php?title=%EB%B0%95%EB%B3%B4%EC%98%81&action=edit§ion=8 [34]
- > https://ko.wikipedia.org/wiki/%ED%9E%98%EC%8E%88%EC%97%AC%EC%9E%90_%EB%8F%84%EB%B4%89%EC%88%9C [416]
- > https://ko.wikipedia.org/wiki/%EC%A7%80%EC%88%98_(%EB%B0%B0%EC%9A%B0) [191]
- > https://ko.wikipedia.org/wiki/%EB%B0%95%ED%98%95%EC%8B%9D [834]
- > https://ko.wikipedia.org/wiki/%ED%9E%98%EC%8E%88%EC%97%AC%EC%9E%90_%EB%8F%84%EB%B4%89%EC%88%9C [416]
- > https://ko.wikipedia.org/wiki/%EB%AC%B4%EC%9E%90%EC%8B%9D_%EC%83%81%ED%8C%94%EC%

9E%90 [446]

- > https://ko.wikipedia.org/w/index.php?title=%EB%B0%95%EB%B3%B4%EC%98%81&action=edit§ion=9 [34]
- > https://ko.wikipedia.org/wiki/%EC%86%8C%EC%85%9C_%EB%84%A4%ED%8A%B8%EC%9B%8C%ED%81%AC %EC%84%9C%EB%B9%84%EC%8A%A4 [360]
- > https://ko.wikipedia.org/wiki/%EA%B9%80%ED%95%B4%EC%88%99 [709]
- > https://ko.wikipedia.org/wiki/%EC%B2%AD%EB%A3%A1%EC%98%81%ED%99%94%EC%83%81 [19 1]
- > https://ko.wikipedia.org/wiki/%EA%B3%BC%EC%86%8D%EC%8A%A4%EC%BA%94%EB%93%A4 [16 a1
- > https://ko.wikipedia.org/wiki/%EA%B3%BC%EC%86%8D%EC%8A%A4%EC%BA%94%EB%93%A4 [16 4]
- > https://ko.wikipedia.org/wiki/%EB%8A%91%EB%8C%80%EC%86%8C%EB%85%84 [55]
- > https://ko.wikipedia.org/wiki/%EB%8F%8C%EC%97%B0%EB%B3%80%EC%9D%B4_(%EC%98%81%E D%99%94) [116]
- > https://ko.wikipedia.org/wiki/%EC%9B%90%EB%B9%88 [1272]
- > https://ko.wikipedia.org/wiki/%EB%B0%95%ED%9A%A8%EC%8B%A0 [657]
- > https://ko.wikipedia.org/wiki/%EC%9B%90%EB%B9%88 [1272]
- > https://ko.wikipedia.org/wiki/%ED%94%84%EB%A1%9C%ED%8F%AC%EC%A6%88_(%EB%93%9C%EB%9D%BC%EB%A7%88) [125]
- > https://ko.wikipedia.org/wiki/%EB%B0%95%ED%9A%A8%EC%8B%A0 [657]
- > https://ko.wikipedia.org/w/index.php?title=%EB%B0%95%EB%B3%B4%EC%98%81&action=edit§ion=10 [34]
- > https://ko.wikipedia.org/w/index.php?title=%EB%B0%95%EB%B3%B4%EC%98%81&action=edit§ion=11 [34]
- > https://ko.wikipedia.org/wiki/%EB%B0%95%EB%B3%B4%EC%98%81%EC%9D%98_%EC%88%98%EC%83%81_%EB%B0%8F_%ED%9B%84%EB%B3%B4_%EB%AA%A9%EB%A1%9D [244]
- > https://ko.wikipedia.org/w/index.php?title=%EB%B0%95%EB%B3%B4%EC%98%81&action=edit§ion=12 [34]
- > http://entertain.naver.com/read?oid=117&aid=0000107683 [221]
- > http://media.daum.net/entertain/culture/newsview?newsid=20091221201110428 [5]
- > http://www.sisunnews.co.kr/news/articleView.html?idxno=30180 [175]
- > http://entertain.naver.com/read?oid=427&aid=0000000608 [220]
- > http://tydailv.asiae.co.kr/read.php3?aid=1439969448963995002 [85]
- > http://media.daum.net/entertain/drama/newsview?newsid=20160603133010138 [5]
- > http://entertain.naver.com/read?oid=311&aid=0000706247 [225]
- > http://entertain.naver.com/read?oid=465&aid=0000002867 [228]
- > http://entertain.naver.com/read?oid=020&aid=0002215077 [235]
- > http://entertain.naver.com/read?oid=076&aid=0002069547 [229]
- > http://media.daum.net/entertain/enews/newsview?newsid=20080206094708959 [5]
- > http://www.ikoreadaily.co.kr/news/articleView.html?idxno=214444 [155]
- > http://maxmedia.maxmovie.com/126136 [47]
- > http://maxmedia.maxmovie.com/126137 [47]
- > http://entertain.naver.com/read?oid=312&aid=0000024101 [229]
- > https://web.archive.org/web/20170216134707/http://www.ceci.co.kr/article/article_detail.do?idx=1464 [54]

d to establish a new connection: [WinError 10060] 연결된 구성원으로부터 응답이 없어 연결하지 못했거나, 호스트로부터 응답이 없어 연결이 끊어졌습니다'))

In [112]:

```
requests.compat.urlparse(queue[0])
```

Out[112]:

ParseResult(scheme='https', netloc='ko.wikipedia.org', path='/wiki/%EB%B0%95%EB%B 3%B4%EC%98%81', params='', query='', fragment='')

In [142]:

```
tofind = "https://blog.naver.com/js2y86?Redirect=Log&logNo=221530397718" requests.compat.urlparse(tofind)[1]
```

Out[142]:

'blog.naver.com'

In [144]:

whos					
------	--	--	--	--	--

Variable	Type	Data/Info	
BeautifulSoup	type	<pre><class 'bs4.beautifulsoup'=""></class></pre>	
a	list	n=247	
daum	BeautifulSoup	html Wn <html cl<="">제>WnWn<!--</td--></html>	
html>₩n	0	at the arms of the section of the section of	
dom	BeautifulSoup	[if HTML5] [endif]-< f]>\n\n	
tml>₩n			
download	function	<pre><function 0x0000027afa6ced08="" at="" download=""></function></pre>	
exercise	BeautifulSoup	<html>Wn<head>Wn<style>Wn<>/div>Wn</body>Wn</h</td></tr><tr><td>tml>₩n</td><td>0</td><td></td></tr><tr><td>google</td><td>BeautifulSoup</td><td><!DOCTYPE html>\mathbb{\text{Mn<html} it<> {}</style></head></html>	
	D 1:1 10	41000TVD5 11 15W 41 1 114 51[] 4/ 1 1 5 4/1	
google2	BeautifulSoup	html Wn <html it<=""> {}</html>	
		n food.	
html	Response	<response [200]=""></response>	
link	tuple	n=2	
links	list	n=54	
naver	BeautifulSoup	html Wn <html la<=""> </html>	
	D 1:1 10	41000TV05 11 15W 41 1 1 4 5 1 15W 4/1 15W 4/1	
naver2	BeautifulSoup	html Wn <html la<="">ript>WnWn</html>	
tml>\mathfrak{Wn}			
newLink	str	http://example.webscrapin<>om/places/default/i	
ndex/1		00500	
queue	list	n=29506	
queueURL	list	n=29453	
random	module	<pre><module 'c:<="" 'random'="" from="">aconda3\\\Ib\\\Int\\Int\\Int\\Int\\Int\\Int\\In</module></pre>	
m.py'>		Conduit a contract from 100 NMM-contract MMM in it	
requests	module	<pre><module '<="" 'requests'="" from="">WWrequestsWWinit_</module></pre>	
py'>	-4-	http://www.lo.wahaaaai.c	
seed	str	http://example.webscrapin<>com/places/default/	
search	1:-4	100	
seen	list	n=182 3	
threshold	int		
time tofind	module str	<pre><module 'time'="" (built-in)=""> https://blog.naver.com/js<>ct=Log&logNo=221530</module></pre>	
397718	511	111 tp5+//b10g.11ave1.0011/j54>0t-L0g&10gN0=221530	
	liet	n=15	
unseen	list	<pre>r=15 <function 0x0000027afd4f3d08="" at="" urlextractor=""></function></pre>	
urlExtractor	function	Tunction utlextractor at 0x000002/AFD4F3D08>	

In [163]:

```
html = download("https://search.naver.com/search.naver", {"query":"박보영"}, {"where":"post"})
dom = BeautifulSoup(html.text, "Ixml")
#print(naver2)
queue = [(_["href"], 0) for _ in dom.select("dt > a.sh_blog_title")]
#seed = "https://blog.naver.com/js2y86/221530397718"
threshold = 3
def urlExtractor(seed. limit):
   if limit > threshold:
        return []
   html = download(seed)
   dom = BeautifulSoup(html.text, "lxml")
   iframe = dom.select one("iframe")
   if not iframe or not iframe.has_attr("src"):
       return []
   html = download(requests.compat.urljoin(seed, iframe["src"]))
   dom = BeautifulSoup(html.text, "lxml")
   links = [_["href"] for _ in dom.select("a") if _.has_attr("href")] #핵심
   unseen = list()
   for link in links:
       if len(link) > 1 and link[0] == "/":
           #print(link)
           newLink = requests.compat.urljoin(seed, link)
           if newLink not in unseen:
               unseen.append((newLink, limit+1))
                 print("Skipped2", link)
       elif link.startswith("http"):
           unseen.append((link, limit+1))
         e/se:
             print("Skipped1", link)
     print("{0} [{1}]".format(seed, len(unseen)))
   return unseen
seen = list()
for link in queue:
   #(URL. LIMIT)
   link = queue.pop(0)
   seen.append(link[0])
   links = urlExtractor(link[0], link[1])
   queueURL = [_[0] for _ in queue]
   queue.extend([_ for _ in links if _[0] not in queueURL and _ not in seen and requests.compa
t.urlparse(_[0])[1] == "blog.naver.com"])
```

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

- https://blog.naver.com/shkwon1128?Redirect=Log&logNo=221530738046 [32]
- https://blog.naver.com/js2y86?Redirect=Log&logNo=221530397718 [32]
- https://blog.naver.com/dlqlwm14?Redirect=Log&logNo=221529807195 [32]
- https://blog.naver.com/lyj0088?Redirect=Log&logNo=221530708253 [33]
- https://komartin.blog.me/221530734469 [0]
- > https://blog.naver.com [0]
- > https://blog.naver.com/prologue/PrologueList.nhn?blogId=shkwon1128 [78]
- > https://blog.naver.com/MyBlog.nhn [0]
- > https://blog.naver.com/prologue/PrologueList.nhn?blogId=shkwon1128 [78]
- > 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 [27]
- > https://blog.naver.com/library/LibraryView.nhn?blogId=shkwon1128&skinType=&skinId=&from=menu [39]
- > https://blog.naver.com/PostList.nhn?blogId=shkwon1128&categoryNo=37&from=postList[30]
- > https://blog.naver.com/shkwon1128 [78]
- > https://blog.naver.com/shkwon1128 [78]
- > https://blog.naver.com/PostList.nhn?blogId=shkwon1128&categoryNo=37&from=postList&parentCategoryNo=0 [30]
- > https://blog.naver.com/PostList.nhn?blogId=shkwon1128&categoryNo=37&from=postList&parentCategoryNo=0 [30]
- > https://blog.naver.com/library/LibraryView.nhn?blogId=shkwon1128 [39]
- > https://blog.naver.com/prologue/PrologueList.nhn?blogId=js2y86 [0]
- > https://blog.naver.com/prologue/PrologueList.nhn?blogId=js2y86&skinType=&skinId=&from=menu [0]
- > https://blog.naver.com/PostList.nhn?blogId=js2y86&categoryNo=1&skinType=&skinId=&from=menu [0]
- > https://blog.naver.com/PostList.nhn?blogId=js2y86&categoryNo=8&skinType=&skinId=&from=menu [0]
- > https://blog.naver.com/PostList.nhn?blogId=js2y86&categoryNo=27&skinType=&skinId =&from=menu [0]
- > https://blog.naver.com/PostList.nhn?blogId=js2y86&categoryNo=71&skinType=&skinId =&from=menu [0]
- > https://blog.naver.com/PostList.nhn?blogId=js2y86&categoryNo=62&skinType=&skinId =&from=menu [0]
- > https://blog.naver.com/memo/MemologPostList.nhn?blogId=js2y86&Redirect=MemoList&skinType=&skinId=&from=menu [0]
- > https://blog.naver.com/BlogPrivateTagCloud.nhn?blogId=js2y86&Redirect=Dtagcloud&skinType=&skinId=&from=menu [0]
- > https://blog.naver.com/prologue/PrologueList.nhn?blogId=is2v86&categorvNo=1 [0]
- > https://blog.naver.com/PostList.nhn?blogId=js2y86&categoryNo=19&from=postList[0]
- > https://blog.naver.com/js2y86 [37]
- > https://blog.naver.com/is2v86 [37]
- > https://blog.naver.com/PostList.nhn?blogId=js2y86&categoryNo=19&from=postList&parentCategoryNo=0 [0]
- > https://blog.naver.com/PostList.nhn?blogId=js2y86&categoryNo=19&from=postList&parentCategoryNo=0 [0]
- > https://blog.naver.com/PostList.nhn?blogId=dlqlwm14 [0]
- > https://blog.naver.com/prologue/PrologueList.nhn?blogId=dlqlwm14&skinType=&skinId=&from=menu [0]
- > https://blog.naver.com/PostList.nhn?blogId=dIqIwm14&skinType=&skinId=&from=menu [0]
- > https://blog.naver.com/PostList.nhn?blogId=dIqIwm14&categoryNo=20&skinType=&skinId=&from=menu [0]
- > https://blog.naver.com/PostList.nhn?blogId=dlqlwm14&categoryNo=1&parentCategoryNo=1&skinType=&skinId=&from=menu [0]
- > https://blog.naver.com/PostList.nhn?blogId=dIqIwm14&categoryNo=27&parentCategoryNo=27&skinType=&skinId=&from=menu [0]

- > https://blog.naver.com/PostList.nhn?blogId=dIqIwm14&categoryNo=10&parentCategoryNo=10&skinType=&skinId=&from=menu [0]
- > https://blog.naver.com/mapview/MapView.nhn?blogId=dlqIwm14&skinType=&skinId=&from=menu [17]
- > https://blog.naver.com/PostView.nhn?blogId=dIqIwm14&IogNo=220543334886&categoryNo=20&parentCategoryNo=-1&viewDate=¤tPage=&postListTopCurrentPage=&isAfterWrite=true [0]
- > https://blog.naver.com/PostView.nhn?blogId=dIqIwm14&IogNo=220428316411&categoryNo=20&parentCategoryNo=-1&viewDate=¤tPage=&postListTopCurrentPage=&isAfterWrite=true [0]
- > https://blog.naver.com/PostView.nhn?blogId=dIqIwm14&IogNo=220274186328&categoryNo=20&parentCategoryNo=-1&viewDate=¤tPage=&postListTopCurrentPage=&isAfterWrite=true [0]
- > https://blog.naver.com/PostList.nhn?blogId=dIqIwm14&categoryNo=24&from=postList[0]
- > https://blog.naver.com/dlqlwm14 [30]
- > https://blog.naver.com/dlqlwm14 [30]
- > https://blog.naver.com/PostList.nhn?blogId=dIqIwm14&categoryNo=24&from=postList&parentCategoryNo=0 [0]
- > https://blog.naver.com/PostList.nhn?blogId=dIqIwm14&categoryNo=24&from=postList&parentCategoryNo=0 [0]
- > https://blog.naver.com/prologue/PrologueList.nhn?blogId=Iyj0088 [0]
- > https://blog.naver.com/prologue/PrologueList.nhn?blogId=Iyj0088&skinType=&skinId =&from=menu [0]
- > https://blog.naver.com/PostList.nhn?blogId=Iyj0088&skinType=&skinId=&from=menu[0]
- > https://blog.naver.com/PostList.nhn?blogId=Iyj0088&categoryNo=205&parentCategoryNo=205&skinType=&skinId=&from=menu [0]
- > https://blog.naver.com/PostList.nhn?blogId=Iyj0088&categoryNo=126&skinType=&skinId=&from=menu [0]
- > https://blog.naver.com/mapview/MapView.nhn?blogId=Iyj0088&skinType=&skinId=&from =menu [19]
- > https://blog.naver.com/library/LibraryView.nhn?blogId=Iyj0088&skinType=&skinId=&from=menu [29]
- > https://blog.naver.com/memo/MemologPostList.nhn?blogId=Iyj0088&Redirect=MemoList &skinType=&skinId=&from=menu [0]
- > https://blog.naver.com/BlogPrivateTagCloud.nhn?blogId=Iyj0088&Redirect=Dtagcloud &skinType=&skinId=&from=menu [0]
- > https://blog.naver.com/PostView.nhn?blogId=Iyj0088&logNo=60195993223&categoryNo=81&parentCategoryNo=-1&viewDate=¤tPage=&postListTopCurrentPage=&isAfterWrite=true [0]
- > https://blog.naver.com/PostView.nhn?blogId=Iyj0088&logNo=60185939163&categoryNo=146&parentCategoryNo=-1&viewDate=¤tPage=&postListTopCurrentPage=&isAfterWrite=true [0]
- > https://blog.naver.com/PostList.nhn?blogId=Iyj0088&categoryNo=137&from=postList[0]
- > https://blog.naver.com/lyj0088 [123]
- > https://blog.naver.com/lyj0088 [123]
- > https://blog.naver.com/PostList.nhn?blogId=Iyj0088&categoryNo=137&from=postList&parentCategoryNo=0 [0]
- > https://blog.naver.com/PostList.nhn?blogId=Iyj0088&categoryNo=137&from=postList&parentCategoryNo=0 [0]
- > https://blog.naver.com/prologue/PrologueList.nhn?blogId=Ivi0088 [0]
- >> https://blog.naver.com/PostView.nhn?blogId=shkwon1128&logNo=221531564221&redire ct=Dlog [29]
- >> https://blog.naver.com/PostView.nhn?blogId=shkwon1128&logNo=221531564221&redirect=blog[29]
- >> https://blog.naver.com/PostView.nhn?blogId=shkwon1128&logNo=221531344187&redirect=Dlog [29]
- >> https://blog.naver.com/PostView.nhn?blogId=shkwon1128&logNo=221531344187&redire ct=Dlog [29]

- >> https://blog.naver.com/PostView.nhn?blogId=shkwon1128&logNo=221524608145&redire ct=Dlog [30]
- >> https://blog.naver.com/PostView.nhn?blogId=shkwon1128&logNo=221524608145&redire ct=Dlog [30]
- >> https://blog.naver.com/PostView.nhn?blogId=shkwon1128&logNo=221519971769&redire ct=Dlog [29]
- >> https://blog.naver.com/PostView.nhn?blogId=shkwon1128&logNo=221519971769&redire ct=Dlog [29]
- >> https://blog.naver.com/PostView.nhn?blogId=shkwon1128&logNo=221516141442&redire ct=Dlog [33]
- >> https://blog.naver.com/PostView.nhn?blogId=shkwon1128&logNo=221516141442&redire ct=Dlog [33]