

urllib 사용법 및 기본 스크래핑

```
In [1]: import urllib.request as req
```

```
In [23]: # 파일 URL
img_url = "http://post.phinf.naver.net/20160621_169/1466482468068ImSHj_JPEG/I f7GeIb0PZuYwI -GI3xU7ENRr I f I . jpg"
html_url = "http://google.com"
```

```
In [31]: # 다운받을 경로
save_path1 = "C:WMMyexamWtest1.jpg"
save_path2 = "C:WMMyexamWindex.html"
```

```
In [32]: # 예외 처리
try:
    file1, header1 = req.urlretrieve(img_url, save_path1)
    file2, header2 = req.urlretrieve(html_url, save_path2)
except Exception as e:
    print("Download failed.")
    print(e)
else:
    # Header 정보 출력
    print(header1)
    print(header2)
```

Date: Fri, 13 Aug 2021 13:20:13 GMT
Accept-Ranges: bytes
Cache-Control: max-age=15552000
Content-Length: 797094
Content-Type: image/jpeg
Last-Modified: Fri, 13 Aug 2021 13:20:12 GMT
p3p: CP="ALL CURa ADMa DEVa TAId OUR BUS IND PHY ONL UNI PUR FIN COM NAV INT DEM CNT STA POL HEA PRE LOC OTC"
Age: 0
Expires: Wed, 09 Feb 2022 13:20:12 GMT
Connection: close

Date: Fri, 13 Aug 2021 13:20:13 GMT
Expires: -1
Cache-Control: private, max-age=0
Content-Type: text/html; charset=ISO-8859-1
P3P: CP="This is not a P3P policy! See g.co/p3phelp for more info."
Server: gws
X-XSS-Protection: 0
X-Frame-Options: SAMEORIGIN
Set-Cookie: 1P_JAR=2021-08-13-13; expires=Sun, 12-Sep-2021 13:20:13 GMT; path=/; domain=.google.com; Secure
Set-Cookie: NID=221=xrBv6I6ioDZ53wUJ5XohjiAR6lVEuPgYNYFY4iI3gz8mCv1Zb2bJn2Tv16VLq9rrtViC6rN7dRcpfG23VWUJArKa2XPyk09CeRreiwdX-RKa
sLmIXr5eHRtc_pEmARntLpE3VM9AIBnNmHe0lv86AW_Ma02tjn_02YxQk4rrl6l; expires=Sat, 12-Feb-2022 13:20:13 GMT; path=/; domain=.google.c
om; HttpOnly
Accept-Ranges: none
Vary: Accept-Encoding
Connection: close
Transfer-Encoding: chunked

```
In [33]: # 다운로드 파일 정보
print("Filename1 {}".format(file1))
print("Filename2 {}".format(file2))
print()
```

Filename1 C:WMMyexamWtest1.jpg
Filename2 C:WMMyexamWindex.html

```
In [34]: # 성공
print("Download Succeed.")
```

Download Succeed.

```
In [ ]:
```

```
In [ ]:
```