

```
In [ ]: Crawling
        Scraping - content-type;text/html,application/json,/image/*,...
                    => Focused crawling+scraping
        Dynamic HTML (AJAX - XmlHttpRequest/JS/Doc...)
                    => DOM(Req/Resp), Data(Req/Resp => content-type), [웹툰]*
        Cookie/Session

        HTTP
        Client          Server
        Req =====> login, 회원가입, ... => RESTful API(Put, Post)
        Resp <===== HTML Form[method, action / JS]
                    변수=우리정보
        Cookie(*)       Session
```

```
In [1]: from requests import request
        from requests.compat import *
        from bs4 import BeautifulSoup

        headers = {
            'user-agent': 'Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/80.0.3987.163 Safari/537.36'
        }
```

```
In [2]: url = 'https://pythonscraping.com/pages/cookies/login.html'
        resp = request('GET', url)
```

```
In [4]: dom = BeautifulSoup(resp.text, 'html.parser')
```

```
In [7]: form = dom.select_one('form')
```

```
In [8]: form.attrs['action']
```

```
Out[8]: {'method': 'post', 'action': 'welcome.php'}
```

```
In [12]: [tag.attrs['name'] for tag in form.select('[name]')]
```

```
Out[12]: ['username', 'password']
```

```
In [13]: urljoin(resp.url, form.attrs['action'])
```

```
Out[13]: 'https://pythonscraping.com/pages/cookies/welcome.php'
```

```
In [24]: params = {
            'username': 'testtest',
            'password': 'password'
        }
        resp = request(form.attrs['method'],
                        urljoin(resp.url, form.attrs['action']),
                        data = params)
```

```
In [25]: resp.headers
        Client          Server
        id/wrong pw ==> req
                    <== resp (headers cookie X)
        id/pw        ==> req
                    <== resp (headers set-cookie)
```

```
Out[25]: {'Server': 'nginx', 'Date': 'Mon, 17 Jul 2023 01:27:49 GMT', 'Content-Type': 'text/html; charset=UTF-8', 'Content-Length': '123', 'Connection': 'keep-alive', 'X-Powered-By': 'PHP/7.4.33, PleskLin', 'Set-Cookie': 'loggedin=1, username=testest', 'Vary': 'Accept-Encoding', 'Content-Encoding': 'gzip'}
```

```
In [29]: request('GET', urljoin(resp.url, 'profile.php')).text
```

```
Out[29]: 'You\'re not logged into the site!<br>Visit <a href="login.html">the login page</a> to login'
```

```
In [30]: request('GET', urljoin(resp.url, 'profile.php'),  
              cookies={'loggedin': '1', 'username': 'test'}).text
```

```
Out[30]: "Hey test! Looks like you're still logged into the site!"
```

```
In [ ]: 1. FORM(action, method)  
        2. FORM tags[name] = tags[value] ==> 검색창  
        3. FORM.method, FORM.action, GET;params, POST;data  
        4. Cookie 관리(set-cookie 확인, dict관리)
```

```
In [32]: resp.request.url, resp.request.body  
        POST => body(data)
```

```
Out[32]: ('https://pythonscraping.com/pages/cookies/welcome.php',  
          'username=testest&password=password')
```

```
In [33]: from requests.cookies import cookiejar_from_dict
```

```
In [36]: c = cookiejar_from_dict({'loggedin': '1', 'username': 'whoami'})
```

```
In [37]: request('GET', urljoin(resp.url, 'profile.php'),  
              cookies=c).text
```

```
Out[37]: "Hey whoami! Looks like you're still logged into the site!"
```

```
In [41]: for row in resp.headers['set-cookie'].split(','):  
          print(row.strip().split('='))  
  
          ['loggedin', '1']  
          ['username', 'testest']
```

```
In [42]: from requests.sessions import Session  
          session = Session()  
          => set-cookie => cookie jar 관리
```

```
In [43]: resp = session.request(form.attrs['method'],  
                               urljoin(resp.url, form.attrs['action']),  
                               data = params)
```

```
In [46]: session.request('GET', urljoin(resp.url, 'profile.php')).text
```

```
Out[46]: "Hey testest! Looks like you're still logged into the site!"
```

```
In [47]: url = 'https://lms.sunde41.net'  
          resp = request('GET', url)
```

```
In [50]: dom = BeautifulSoup(resp.text, 'html.parser')
```

```
In [53]: form = dom.select_one('form')
```

```
In [61]: params = {}

for tag in form.select('[name]'):
    params[tag.attrs['name']] = tag.attrs['value'] \
        if tag.has_attr('value') else ''
```

```
In [65]: params
```

```
Out[65]: {'next': '/', 'email': '', 'password': '', 'remember': ''}
```

```
In [ ]: %%writefile lms.txt
{
    "id": "value",
    "pw": "pw"
}
```

```
In [63]: import json
# 이걸 내꺼 - 작성 X
with open('lms.txt', 'r', encoding='utf-8') as fp:
    lms = json.load(fp)
```

```
In [66]: params['email'] = lms['id']
params['password'] = lms['pw']
```

```
In [67]: session.cookies.clear()
```

```
In [69]: form.attrs
```

```
Out[69]: {'action': '/auth/login',
          'method': 'POST',
          'name': 'login_user_form',
          'class': ['m-form']}
```

```
In [70]: resp = session.request(form.attrs['method'],
                                urljoin(resp.url, form.attrs['action']),
                                data=params)
```

```
In [76]: dom = BeautifulSoup(session.request('GET', url).text, 'lxml')
```

```
In [79]: dom.select('a[href^="/course"]')
```

```
Out[79]: [<a class="m-menu__link" href="/course/4">
  <i class="m-menu__link-bullet m-menu__link-bullet--line"><span></span></i><span class
="m-menu__link-text">수업 게시판</span></a>,
  <a class="m-menu__link" href="/course/assignment/list/4"><i class="m-menu__link-bullet
m-menu__link-bullet--line"><span></span></i><span class="m-menu__link-text">과제/제출 게
시판</span>
  <span class="m-badge m-badge--orange strong float-right" style="margin-top:10px;">
    new</span>
  </a>,
  <a href="/course/assignment/submit/4/34" style="color:#6a1aff">제출</a>,
  <a href="/course/assignment/submit/4/38" style="color:#6a1aff">제출</a>,
  <a href="/course/assignment/submit/4/39" style="color:#6a1aff">제출</a>,
  <a href="/course/assignment/submit/4/41" style="color:#ff3d3d">제출</a>]
```

```
<a href="/course/assignment/submit/4/43" style="color:#ff3d3d">제출</a>]
```

```
In [85]: resp = request('GET', urljoin(url, '/course/4'),  
                      cookies=session.cookies)
```

```
In [86]: dom = BeautifulSoup(resp.text, 'lxml')
```

```
In [89]: for node in dom.select('a[href^="/static"]'):  
        print(urljoin(resp.url, node.attrs['href']))
```

```
https://lms.sunde41.net/static/uploads/lectures/4/%E1%84%83%E1%85%A6%E1%84%8B%E1%85%B5%E1%84%90%E1%85%A5%E1%84%87%E1%85%A6%E1%84%8B%E1%85%B5%E1%84%89%E1%85%B3.pdf  
https://lms.sunde41.net/static/uploads/lectures/4/%E1%84%86%E1%85%A9%E1%84%83%E1%85%B3%E1%86%AB%20%E1%84%8F%E1%85%A5%E1%86%B7%E1%84%91%E1%85%B2%E1%84%90%E1%85%A5%20%E1%84%80%E1%85%AA%E1%84%92%E1%85%A1%E1%86%A8%E1%84%8C%E1%85%A1%E1%84%80%E1%85%A1%20%E1%84%8B%E1%85%A1%E1%86%AF%E1%84%8B%E1%85%A1%E1%84%8B%E1%85%A3%20%E1%84%92%E1%85%A1%E1%86%AF%20%E1%84%87%E1%85%AE%E1%84%83%E1%85%A9%E1%86%BC%20%E1%84%89%E1%85%A9%E1%84%89%E1%85%AE%E1%84%8C%E1%85%A5%E1%86%B7%E1%84%8B%E1%85%B4%20%E1%84%86%E1%85%A9%E1%84%83%E1%85%B3%E1%86%AB%E1%84%80%E1%85%A5%E1%86%BA.pdf  
https://lms.sunde41.net/static/uploads/lectures/4/0627.html  
https://lms.sunde41.net/static/uploads/lectures/4/0628.html  
https://lms.sunde41.net/static/uploads/lectures/4/Functional%20Programming%20in%20Python.pdf  
https://lms.sunde41.net/static/uploads/lectures/4/Regular%20Expression.pdf  
https://lms.sunde41.net/static/uploads/lectures/4/2021.png  
https://lms.sunde41.net/static/uploads/lectures/4/2018.png  
https://lms.sunde41.net/static/uploads/lectures/4/0629.html  
https://lms.sunde41.net/static/uploads/lectures/4/HTTP.pdf  
https://lms.sunde41.net/static/uploads/lectures/4/0630.html  
https://lms.sunde41.net/static/uploads/lectures/4/0703.html  
https://lms.sunde41.net/static/uploads/lectures/4/Mastering%20Python%20Decorators.pdf.pdf  
https://lms.sunde41.net/static/uploads/lectures/4/%E1%84%80%E1%85%A2%E1%86%A8%E1%84%8E%E1%85%A6%E1%84%8C%E1%85%B5%E1%84%92%E1%85%A3%E1%86%BC.pdf  
https://lms.sunde41.net/static/uploads/lectures/4/DOM.pdf  
https://lms.sunde41.net/static/uploads/lectures/4/0704.html  
https://lms.sunde41.net/static/uploads/lectures/4/OOP.zip  
https://lms.sunde41.net/static/uploads/lectures/4/AOP.pdf  
https://lms.sunde41.net/static/uploads/lectures/4/%EC%9D%8C%EC%84%B1%EC%9D%B8%EC%8B%9D_%EC%9A%94%EC%95%BD_%EA%B9%80%EA%B4%91%ED%98%B8.pdf  
https://lms.sunde41.net/static/uploads/lectures/4/%EC%9E%90%EC%97%B0%EC%96%B4%EC%B2%98%EB%A6%AC%EB%B0%8F%EC%96%B8%EC%96%B4%EB%AA%A8%EB%8D%B8%EB%A7%81_%EA%B9%80%EA%B4%91%ED%98%B8.pdf  
https://lms.sunde41.net/static/uploads/lectures/4/0706.html  
https://lms.sunde41.net/static/uploads/lectures/4/NumPy.pdf  
https://lms.sunde41.net/static/uploads/lectures/4/numpy.ipynb  
https://lms.sunde41.net/static/uploads/lectures/4/0710.html  
https://lms.sunde41.net/static/uploads/lectures/4/data.pdf  
https://lms.sunde41.net/static/uploads/lectures/4/0711.html  
https://lms.sunde41.net/static/uploads/lectures/4/Crawling.pdf  
https://lms.sunde41.net/static/uploads/lectures/4/PageRank.pdf  
https://lms.sunde41.net/static/uploads/lectures/4/data.zip  
https://lms.sunde41.net/static/uploads/lectures/4/Tidy%20Data.pdf  
https://lms.sunde41.net/static/uploads/lectures/4/0713.html  
https://lms.sunde41.net/static/uploads/lectures/4/recommendation.pdf  
https://lms.sunde41.net/static/uploads/lectures/4/recommendation.ipynb  
https://lms.sunde41.net/static/uploads/lectures/4/recommendation2.ipynb  
https://lms.sunde41.net/static/uploads/lectures/4/The%20Split-Apply-Combine%20Strategy%20for%20Data%20Analysis.pdf
```

https://lms.sunde41.net/static/uploads/ckeditor/2023-06-29\_17-48-32\_%EB%8D%B0%EC%B2%AD%EC%BA%A0%20%EC%B6%9C%EC%84%9D%EC%9D%B8%EC%A0%95%20%EA%B8%B0%EC%A4%80.pdf

```
In [90]: resp = request('GET', 'https://lms.sunde41.net/static/uploads/lectures/4/%E1%84%83%E1%85
root path = /
uri
public path = /uploads/, /static/
```

```
In [91]: resp.headers
```

```
Out[91]: {'Server': 'nginx/1.18.0 (Ubuntu)', 'Date': 'Mon, 17 Jul 2023 02:00:03 GMT', 'Content-Type': 'application/pdf', 'Content-Length': '8931190', 'Last-Modified': 'Thu, 29 Jun 2023 01:02:52 GMT', 'Connection': 'keep-alive', 'ETag': '"649cd83c-884776"', 'Accept-Ranges': 'bytes'}
```

```
In [94]: c = {}
for k,v in session.cookies.get_dict().items():
    c[k] = v
```

```
In [97]: request('GET', url, cookies=c).text
```

```
Out[97]: '<!DOCTYPE html>\n<html lang="ko">\n<head>\n    <title>대시보드 : 고려대학교 지능정보SW아카데미</title>\n    <meta charset="UTF-8">\n    <meta http-equiv="X-UA-Compatible" content="IE=edge,chrome=1">\n    <meta name="viewport" content="width=device-width, initial-scale=1, maximum-scale=1, shrink-to-fit=no">\n    <meta name="robots" content="noindex,no follow">\n    <link href="/static/vendors/bootstrap.css?v=1.4" rel="stylesheet">\n    <link href="/static/vendors/main.css?v=1.8" rel="stylesheet">\n    <link href="/static/vendors/icon.css?v=1.0" rel="stylesheet">\n    <link href="/static/vendors/form.css?v=1.0" rel="stylesheet">\n    <link href="/static/vendors/style.css?v=2.10" rel="stylesheet">\n    <link href="/static/vendors/fullcalendar.min.css?v=1.2" rel="stylesheet">\n<style>\n#att
endace {\n\tdisplay: none;\n}\n.fc-unthemed .fc-event .fc-title {\n    font-size: 11p
x;\n    font-weight: 500;\n}\n.fc-unthemed .fc-event, .fc-unthemed .fc-event-dot {\n
border: 1px solid #ebedf2;\n}\n    .fc .fc-bg-event {\n        opacity: 0.8;\n    ba
ckground: red !important;\n    }\n    .fc .fc-bg-event .fc-event-title {\n        colo
r: #FFF !important;\n        font-style: normal;\n        font-weight: bolder;\n    }\n.
message {\n    cursor: pointer;\n}\n\n.modal table {\n\twidth: 100%;\n\t\tmargin-bottom:
1rem;\n}\n\n.modal table, \n.modal th,\n.modal td {\n    border: 1px solid #ebedf2;\n borde
r-collapse: collapse;\n    padding: 0.75rem;\n}\n</style>\n    <link rel="shortcut icon" h
ref="/static/favicon/favicon.ico" type="image/x-icon">\n    <link rel="icon" href="/stat
ic/favicon/favicon.ico" type="image/x-icon">\n</head>\n<body class="m-page--fluid m-cont
ent--skin-light2 m-header--fixed m-header--fixed-mobile m-aside-left--enabled m-aside-le
ft--skin-dark m-aside-left--fixed m-aside-left--offcanvas m-footer--push m-aside--offcan
vas-default">\n<div class="m-grid m-grid--hor m-grid--root m-page">\n    <header id="m_h
eader" class="m-grid__item m-header " m-minimize-offset="200" m-minimize-mobile-offset
="200">\n        <div class="m-container m-container--fluid m-container--full-height">\n
<div class="m-stack m-stack--ver m-stack--desktop">\n            <div class="m-stack
__item m-brand m-brand--skin-dark ">\n                <div class="m-stack m-stack--
ver m-stack--general">\n                    <div class="m-stack__item m-stack__item-
-middle m-brand__logo">\n                        \n                    </div>\n                <div class="m-stack__item m-stack__item--middle m-brand__tools">\n\t\t\t
<a href="javascript:;" id="m_aside_left_minimize_toggle" class="m-brand__icon m-brand__toggl
er m-brand__togglер--left m--visible-desktop-inline-block m-brand__togglер--active">\n
<span></span>\n                    </a>\n                <a href="ja
vascript:;" id="m_aside_left_offcanvas_toggle"\n                    class="m-
brand__icon m-brand__togglер m-brand__togglер--left m--visible-tablet-and-mobile-inline-
block">\n                        <span></span>\n                    </a>
\n                <a id="m_aside_header_topbar_mobile_toggle" href="javascri
pt:;"\n                    class="m-brand__icon m--visible-tablet-and-mobile-
```

[illegible]

[illegible]

[illegible]



0,\r\n"bbc": 0,\r\n"dt": 1,\r\n"dw": 0,\r\n"dc": 0,\r\n"fi": { }\r\n},\r\n"01D9A7BFBF6C53C000000030": {\r\n"id": 2,\r\n"td": false,\r\n"sh": false,\r\n"st": 0,\r\n"sc": 0,\r\n"si": false,\r\n"bt": 0,\r\n"bi": false,\r\n"cl": 0,\r\n"bc": false,\r\n"lt": 0,\r\n"lw": 0,\r\n"lc": 0,\r\n"rt": 0,\r\n"rw": 0,\r\n"rc": 0,\r\n"tt": 0,\r\n"tw": 0,\r\n"tc": 0,\r\n"bbt": 0,\r\n"bbw": 0,\r\n"bbc": 0,\r\n"dt": 1,\r\n"dw": 0,\r\n"dc": 0,\r\n"fi": {\r\n"wb": {\r\n"fc": 4294967295,\r\n"hc": 4278190080,\r\n"al": 0,\r\n"hs": -1\r\n}\r\n}\r\n},\r\n"cp": {\r\n"01D9A7BFBF6C53C000000031": {\r\n"id": 0,\r\n"he": 1100,\r\n"tc": 0,\r\n"sc": 4294967295,\r\n"uf": false,\r\n"uk": false,\r\n"sm": 0,\r\n"bf": "01D9A7BFBF6C53C000000030",\r\n"f1": "함초롬바탕",\r\n"t1": 1,\r\n"f2": "함초롬바탕",\r\n"t2": 1,\r\n"f3": "한양신명조",\r\n"t3": 2,\r\n"f4": "함초롬바탕",\r\n"t4": 1,\r\n"f5": "함초롬바탕",\r\n"t5": 1,\r\n"f6": "함초롬바탕",\r\n"t6": 1,\r\n"f7": "함초롬바탕",\r\n"t7": 1,\r\n"r1": 100,\r\n"r2": 100,\r\n"r3": 100,\r\n"r4": 100,\r\n"r5": 100,\r\n"r6": 100,\r\n"r7": 100,\r\n"s1": 0,\r\n"s2": 0,\r\n"s3": 0,\r\n"s4": 0,\r\n"s5": 0,\r\n"s6": 0,\r\n"s7": 0,\r\n"e1": 100,\r\n"e2": 100,\r\n"e3": 100,\r\n"e4": 100,\r\n"e5": 100,\r\n"e6": 100,\r\n"e7": 100,\r\n"o1": 0,\r\n"o2": 0,\r\n"o3": 0,\r\n"o4": 0,\r\n"o5": 0,\r\n"o6": 0,\r\n"o7": 0,\r\n"it": false,\r\n"bo": false,\r\n"ut": 0,\r\n"us": 1,\r\n"uc": 0,\r\n"st": false,\r\n"ss": 1,\r\n"so": 0,\r\n"ot": 0,\r\n"ht": 0,\r\n"hc": 0,\r\n"hx": 0,\r\n"hy": 0,\r\n"em": false,\r\n"en": false,\r\n"su": false,\r\n"sb": false\r\n},\r\n"01D9A7BFBF6C53C000000032": {\r\n"id": 1,\r\n"he": 1200,\r\n"tc": 0,\r\n"sc": 4294967295,\r\n"uf": false,\r\n"uk": false,\r\n"sm": 0,\r\n"bf": "01D9A7BFBF6C53C000000030",\r\n"f1": "함초롬바탕",\r\n"t1": 1,\r\n"f2": "함초롬바탕",\r\n"t2": 1,\r\n"f3": "한양신명조",\r\n"t3": 2,\r\n"f4": "함초롬바탕",\r\n"t4": 1,\r\n"f5": "함초롬바탕",\r\n"t5": 1,\r\n"f6": "함초롬바탕",\r\n"t6": 1,\r\n"f7": "함초롬바탕",\r\n"t7": 1,\r\n"r1": 100,\r\n"r2": 100,\r\n"r3": 100,\r\n"r4": 100,\r\n"r5": 100,\r\n"r6": 100,\r\n"r7": 100,\r\n"s1": 0,\r\n"s2": 0,\r\n"s3": 0,\r\n"s4": 0,\r\n"s5": 0,\r\n"s6": 0,\r\n"s7": 0,\r\n"e1": 100,\r\n"e2": 100,\r\n"e3": 100,\r\n"e4": 100,\r\n"e5": 100,\r\n"e6": 100,\r\n"e7": 100,\r\n"o1": 0,\r\n"o2": 0,\r\n"o3": 0,\r\n"o4": 0,\r\n"o5": 0,\r\n"o6": 0,\r\n"o7": 0,\r\n"it": false,\r\n"bo": false,\r\n"ut": 0,\r\n"us": 1,\r\n"uc": 0,\r\n"st": false,\r\n"ss": 1,\r\n"so": 0,\r\n"ot": 0,\r\n"ht": 0,\r\n"hc": 0,\r\n"hx": 0,\r\n"hy": 0,\r\n"em": false,\r\n"en": false,\r\n"su": false,\r\n"sb": false\r\n},\r\n"01D9A7BFBF6C53C000000033": {\r\n"id": 2,\r\n"he": 1000,\r\n"tc": 0,\r\n"sc": 4294967295,\r\n"uf": false,\r\n"uk": false,\r\n"sm": 0,\r\n"bf": "01D9A7BFBF6C53C000000030",\r\n"f1": "한양신명조",\r\n"t1": 2,\r\n"f2": "한양신명조",\r\n"t2": 2,\r\n"f3": "한양신명조",\r\n"t3": 2,\r\n"f4": "한양신명조",\r\n"t4": 2,\r\n"f5": "한양신명조",\r\n"t5": 2,\r\n"f6": "한양신명조",\r\n"t6": 2,\r\n"f7": "명조",\r\n"t7": 2,\r\n"r1": 100,\r\n"r2": 100,\r\n"r3": 100,\r\n"r4": 100,\r\n"r5": 100,\r\n"r6": 100,\r\n"r7": 100,\r\n"s1": 0,\r\n"s2": 0,\r\n"s3": 0,\r\n"s4": 0,\r\n"s5": 0,\r\n"s6": 0,\r\n"s7": 0,\r\n"e1": 100,\r\n"e2": 100,\r\n"e3": 100,\r\n"e4": 100,\r\n"e5": 100,\r\n"e6": 100,\r\n"e7": 100,\r\n"o1": 0,\r\n"o2": 0,\r\n"o3": 0,\r\n"o4": 0,\r\n"o5": 0,\r\n"o6": 0,\r\n"o7": 0,\r\n"it": false,\r\n"bo": false,\r\n"ut": 0,\r\n"us": 1,\r\n"uc": 0,\r\n"st": false,\r\n"ss": 1,\r\n"so": 0,\r\n"ot": 0,\r\n"ht": 0,\r\n"hc": 0,\r\n"hx": 0,\r\n"hy": 0,\r\n"em": false,\r\n"en": false,\r\n"su": false,\r\n"sb": false\r\n}\r\n},\r\n"tp": {\r\n"01D9A7BFBF6C53C000000034": {\r\n"id": 0,\r\n"al": false,\r\n"ar": false,\r\n"tp": [ ]\r\n}\r\n},\r\n"nu": { },\r\n"bu": { },\r\n"pp": {\r\n"01D9A7BFBF6C53C000000035": {\r\n"id": 0,\r\n"ah": 0,\r\n"av": 0,\r\n"ht": 0,\r\n"hi": "",\r\n"h1": 0,\r\n"tp": "01D9A7BFBF6C53C000000034",\r\n"kb": 0,\r\n"kn": true,\r\n"ko": false,\r\n"kk": false,\r\n"kl": false,\r\n"kp": false,\r\n"kw": 0,\r\n"co": 0,\r\n"f1": false,\r\n"st": false,\r\n"sl": false,\r\n"ae": false,\r\n"aa": false,\r\n"mi": -4880,\r\n"ml": 0,\r\n"mr": 0,\r\n"mp": 0,\r\n"mn": 0,\r\n"lt": 0,\r\n"lv": 140,\r\n"bf": "01D9A7BFBF6C53C000000030",\r\n"bl": 0,\r\n"br": 0,\r\n"bt": 0,\r\n"bb": 0,\r\n"bc": false,\r\n"bi": false\r\n}\r\n},\r\n"01D9A7BFBF6C53C000000036": {\r\n"id": 1,\r\n"ah": 0,\r\n"av": 0,\r\n"ht": 0,\r\n"hi": "",\r\n"h1": 0,\r\n"tp": "01D9A7BFBF6C53C000000034",\r\n"kb": 0,\r\n"kn": true,\r\n"ko": false,\r\n"kk": false,\r\n"kl": false,\r\n"kp": false,\r\n"kw": 0,\r\n"co": 0,\r\n"f1": false,\r\n"st": false,\r\n"sl": false,\r\n"ae": false,\r\n"aa": false,\r\n"mi": 0,\r\n"ml": 0,\r\n"mr": 0,\r\n"mp": 0,\r\n"mn": 0,\r\n"lt": 0,\r\n"lv": 160,\r\n"bf": "01D9A7BFBF6C53C000000030",\r\n"bl": 0,\r\n"br": 0,\r\n"bt": 0,\r\n"bb": 0,\r\n"bc": false,\r\n"bi": false\r\n}\r\n},\r\n"st": {\r\n"01D9A7BFBF6C53C000000037": {\r\n"id": 0,\r\n"ty": 0,\r\n"na": "바탕글",\r\n"en": "Normal",\r\n"pp": "01D9A7BFBF6C53C000000036",\r\n"cp": "01D9A7BFBF6C53C000000033",\r\n"ns": "01D9A7BFBF6C53C000000037",\r\n"li": 1042,\r\n"lf": false\r\n}\r\n},\r\n"mp": { },\r\n"ro": {\r\n"hp": "01D9A7BFBF6C53C00000000

[illegible]

[illegible]

[illegible]

[illegible]

```
In [99]: resp = request('POST', urljoin(url, '/auth/login'), data=params)
```

```
In [109]: url = 'https://nid.naver.com/nidlogin.login?mode=form'
          resp = request('GET', url)
```

```
In [104]: session.cookies.clear()
```

```
In [114]... c = {}
for line in r.splitlines():
    if len(line) > 1:
#         print(line.split('\t')[2])
```

```
c[line.split('\t')[0]] = line.split('\t')[1]
```

```
In [119]... url = 'https://mail.naver.com'
request('GET', url, cookies=c).text
```

```
Out[119]: '<!doctype html><html><head lang="ko"><meta charset="UTF-8"/><meta name="viewport" cont
ent="width=device-width,initial-scale=1,maximum-scale=1,minimum-scale=1,viewport-fit=co
ver"/><title>메일</title><link rel="shortcut icon" href="/favicon.ico" type="image/x-ico
n"/><link rel="apple-touch-icon-precomposed" sizes="72x72" href="https://ssl.pstatic.ne
t/static/pwe/naver_mail/hdpi_72X72.png"><link rel="apple-touch-icon-precomposed" sizes
="96x96" href="https://ssl.pstatic.net/static/pwe/naver_mail/xhdpi_96x96.png"><link rel
="apple-touch-icon-precomposed" sizes="144x144" href="https://ssl.pstatic.net/static/pw
e/naver_mail/xxhdpi_144x144.png"><link rel="apple-touch-icon-precomposed" sizes="192x19
2" href="https://ssl.pstatic.net/static/pwe/naver_mail/xxxhdpi_192x192.png"><script>win
dow.nhn = {};\n      window.gnb_option = {\n          gnb_service: "mail",\n          gnb_t
emplate: "gnb_utf8",\n          gnb_logout: "https://www.naver.com",\n          gnb_one_nav
er: 1,\n          gnb_item_hide_option: 8,\n          gnb_one_flat: 1,\n          gnb_one_pw
e: 1\n      };\n</script><script src="https://ssl.pstatic.net/tveta/libs/glad/prod/gfp-co
re.js"></script><script>window.gladsdk = window.gladsdk || { cmd: [] };</script><script
async src="/builds/v2/libs/jindo.js"></script><script src="/builds/v2/libs/es6-shim.j
s"></script><script src="/builds/v2/libs/lcslog_v0.14.0.js"></script><script async src
="/builds/v2/libs/site-translator_v3.0.6.js"></script><script async src="/builds/v2/lib
s/ncik_v0.8.1.js"></script><script id="postData">try {\n      var __ONEWEB_POST_DATA__
= (POST_DATA);\n      } catch (e) {\n      if (typeof __ONEWEB_POST_DATA__ === '\nun
defined\') {\n      var postDataEl = document.querySelector('#postData');\n
postDataEl.parentNode.removeChild(postDataEl);\n      }\n      }</script><link rel="s
tylesheet" href="/builds/v2/css/index.3e723f4.css"/><script defer="defer" src="/builds/
v2/js/runtime.3e723f4.js"></script><script defer="defer" src="/builds/v2/js/vendor.3e72
3f4.js"></script><script defer="defer" src="/builds/v2/js/app.3e723f4.js"></script></he
ad><body><div id="u_skip"><a href="#content">본문으로 바로가기</a></div><div id="root"></
div><div id="layer"></div><div id="toast"></div></body></html>'
```

```
In [121]... for k,v in c.items():
            session.cookies.set(k,v)
```

```
In [124]... import re
re.search(r'<title>(.)</title>', session.request('GET', url).text)
```

```
Out[124]: <re.Match object; span=(179, 196), match='<title>메일</title>'
```

```
In [ ]: 받은메일함
```

```
In [125]... session.get('https://mail.naver.com/v2/folders/0').text
```

```
Out[125]: '<!doctype html><html><head lang="ko"><meta charset="UTF-8"/><meta name="viewport" cont
ent="width=device-width,initial-scale=1,maximum-scale=1,minimum-scale=1,viewport-fit=co
ver"/><title>메일</title><link rel="shortcut icon" href="/favicon.ico" type="image/x-ico
n"/><link rel="apple-touch-icon-precomposed" sizes="72x72" href="https://ssl.pstatic.ne
t/static/pwe/naver_mail/hdpi_72X72.png"><link rel="apple-touch-icon-precomposed" sizes
="96x96" href="https://ssl.pstatic.net/static/pwe/naver_mail/xhdpi_96x96.png"><link rel
="apple-touch-icon-precomposed" sizes="144x144" href="https://ssl.pstatic.net/static/pw
e/naver_mail/xxhdpi_144x144.png"><link rel="apple-touch-icon-precomposed" sizes="192x19
2" href="https://ssl.pstatic.net/static/pwe/naver_mail/xxxhdpi_192x192.png"><script>win
dow.nhn = {};\n      window.gnb_option = {\n          gnb_service: "mail",\n          gnb_t
emplate: "gnb_utf8",\n          gnb_logout: "https://www.naver.com",\n          gnb_one_nav
er: 1,\n          gnb_item_hide_option: 8,\n          gnb_one_flat: 1,\n          gnb_one_pw
e: 1\n      };\n</script><script src="https://ssl.pstatic.net/tveta/libs/glad/prod/gfp-co
re.js"></script><script>window.gladsdk = window.gladsdk || { cmd: [] };</script><script
```

```

async src="/builds/v2/libs/jindo.js"></script><script src="/builds/v2/libs/es6-shim.js"></script><script src="/builds/v2/libs/lcslog_v0.14.0.js"></script><script async src="/builds/v2/libs/site-translator_v3.0.6.js"></script><script async src="/builds/v2/libs/nclk_v0.8.1.js"></script><script id="postData">try {\n      var __ONEWEB_POST_DATA__ = (POST_DATA);\n    } catch (e) {\n      if (typeof __ONEWEB_POST_DATA__ === \undefined\') {\n        var postDataEl = document.querySelector(\#postData\');\n        postDataEl.parentNode.removeChild(postDataEl);\n      }\n    }</script><link rel="stylesheet" href="/builds/v2/css/index.3e723f4.css"/><script defer="defer" src="/builds/v2/js/runtime.3e723f4.js"></script><script defer="defer" src="/builds/v2/js/vendor.3e723f4.js"></script><script defer="defer" src="/builds/v2/js/app.3e723f4.js"></script></head><body><div id="u_skip"><a href="#content">본문으로 바로가기</a></div><div id="root"></div><div id="layer"></div><div id="toast"></div></body></html>'

```

```

In [126]... url = 'https://mail.naver.com/json/list'
           params = {
               'foldersSN':'0',
               'page':'1',
               'viewMode':'time',
               'previewMode':'1',
               'sortField':'1',
               'sortType':'0',
               'u':'gon0121'
           } # params['u'] <= 본인 아이디
           resp = session.request('POST', url, params=params)

```

```

In [132]... len(resp.json()['mailData'])

```

Out[132]: 15

```

In [133]... resp = request('POST', url, params=params, cookies=c)
           len(resp.json()['mailData'])
           스팸 메일함 불러오기

```

Out[133]: 15

```

In [134]... resp = session.request('GET', 'https://cafe.naver.com')

```

```

In [136]... resp.text
           카페 목록 불러오기

```

```

Out[136]: '<!DOCTYPE html><html lang=ko><head><meta charset=utf-8><meta http-equiv=Content-Type content="text/html; charset=utf-8"><meta http-equiv=X-UA-Compatible content="IE=edge"><meta name=robots content="noindex, nofollow"><meta name=viewport content="width=device-width,initial-scale=1"><link rel=icon href=https://ca-fe.pstatic.net/web-section/favicon.ico><title>네이버 카페</title><link href=https://ca-fe.pstatic.net/web-section/css/chunk-vendors.d97ccb6b.css rel=stylesheet><link href=https://ca-fe.pstatic.net/web-section/css/app.64391c97.css rel=stylesheet></head><body><noscript><strong>We're sorry but web-pc doesn't work properly without JavaScript enabled. Please enable it to continue.</strong></noscript><div id=app></div><script src=https://ca-fe.pstatic.net/web-section/js/chunk-vendors.6bebdfe4.js></script><script src=https://ca-fe.pstatic.net/web-section/js/app.0ae076ff.js></script></body></html>'

```

```

In [143]... url = 'https://www.ypbooks.co.kr/kor_index.yp'
           resp = request('GET', url)
           dom = BeautifulSoup(resp.text, 'lxml')

```

```

In [167]... form = dom.select('form')[-1]
           params = {node.attrs['name']:node.attrs['value']}

```

```
        for node in form.select('[name]')
    params['query'] = '뉴진스'
```

In [168]... `form.attrs`

Out[168]: {'name': 'searchFrm', 'method': 'post', 'action': '/search.yp'}

```
In [193]... params = {
    'query': '뉴진스',
    'collection': 'ALL',
    'searchfield': 'ALL',
    'sortField': 'DATE',
    'notSoldOut': 'N',
    'catesearch': 'false',
    'viewStyle': 'list',
    'pageNum': '1'
}
resp = request('POST', 'https://www.ypbooks.co.kr/ypbooks/search/requestAjaxSearchTab.js
              data=params)
```

```
In [ ]: HTTP(Req/Resp)
        RE,DOM,CSSOM
        DHTML/AJAX(XHR)
        Session/Cookie(*)
        => Crawling + Scraping
           뉴스, 이미지, 메일, 소셜미디어, ...
```

In [201]... `from datetime import datetime`

In [204]... `datetime.now().timestamp()`

Out[204]: 1689566353.727519

In [ ]: 1689565992

In [205]... `datetime.fromtimestamp(1689565992)`

Out[205]: datetime.datetime(2023, 7, 17, 12, 53, 12)

```
In [236]... t = int(datetime.now().timestamp())
pw = 'toinstagram'
f'#PWD_INSTAGRAM_BROWSER:0:{t}:{pw}'
```

Out[236]: '#PWD\_INSTAGRAM\_BROWSER:0:1689567174:toinstagram'

```
In [207]... url = 'https://www.instagram.com/api/v1/web/accounts/login/ajax/'
params = {
    'enc_password': '#PWD_INSTAGRAM_BROWSER:10:1689566089:AVxQAGTw2LChRHhfhxod0/Y8FxG6+'
    'optIntoOneTap': 'false',
    'queryParams': '{}',
    'trustedDeviceRecords': '{}',
    'username': 'test@test.com'
}
resp = request('POST', url, data=params)
resp.text
```

Out[207]: '{"message":"CSRF token missing or incorrect","status":"fail"}'



```
In [ ]: Oauth, CSRF
API - 내가(인증)
API(인증)
```

```
In [208]... url = 'https://www.instagram.com'
resp = request('GET', url)
```

```
In [232]... csrf = re.search(r'\\\"csrf_token\\\":\\\"([^\"]+)\"', resp.text).group(1)
```

```
In [237]... url = 'https://www.instagram.com/api/v1/web/accounts/login/ajax/'
params = {
    'enc_password': '#PWD_INSTAGRAM_BROWSER:10:1689566089:AVxQAGTw2LChRHhfhxod0/Y8FxG6+'
    'optIntoOneTap': 'false',
    'queryParams': '{}',
    'trustedDeviceRecords': '{}',
    'username': '본인아이디'
}
t = int(datetime.now().timestamp())
pw = '비밀번호'
params['enc_password'] = f'#PWD_INSTAGRAM_BROWSER:0:{t}:{pw}'
resp = request('POST', url, data=params,
               headers={'x-csrf-token': csrf})
resp.text
```

```
Out[237]: '{"user":true,"userId":"49814451599","authenticated":true,"oneTapPrompt":true,"status":"ok"}'
```