

# Web Scraping with Python

Chapter 10  
폼과 로그인 뚫기  
(by 컴맹)

# Chap10. 폼과 로그인 뚫기

## 10.2 기본적인 폼전송



# Chap10. 폼과 로그인 뚫기

## 10.2 기본적인 폼전송

Get	Post
url 뒤에 정보 전달	header에 정보전달
간단한데이타를 전송할때 유리함.	많은 데이타전송가능 최소한의 보안효과
전달할데이타의 제한(4096바이트)	

← → ↻ pythonscrapping.com/pages/files/form.html

**Tell me your name!**

First name:

Last name:



pythonscrapping.com/pages/files/processing.php

Hello there, Ryan Mitchell!

Network tab showing request details for processing.php:

- Host: pythonscrapping.com
- Origin: https://pythonscrapping.com
- Referer: https://pythonscrapping.com/pages/files/form.html
- Sec-Fetch-Dest: document
- Sec-Fetch-Mode: navigate
- Sec-Fetch-Site: same-origin
- Sec-Fetch-User: ?1
- Upgrade-Insecure-Requests: 1
- User-Agent: Mozilla/5.0 (Windows NT 6.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/68.0.3440.106 Safari/537.36

Form Data:

- firstname: Ryan
- lastname: Mitchell

# Chap10. 폼과 로그인 뚫기

## 10.3 라디오버튼, 체크박스, 기타필드

html form

```
<html>
<head>
</head>
<body>
  <form action = "#" accept-charset="utf-8" name = "person_info" method = "get">
    <fieldset style = "width:150">
      <legend>개인 정보 입력</legend>
      이름 : <input type = "text" name = "name" value="Ryan" /><br><br>
      주민번호 : <input type = "text" name = "security_number" value="123456-1234567"/><br><br>
      아이디 : <input type = "text" name = "id" value="mitchell"/><br><br>
      비밀번호 : <input type = "password" name = "password" value="1234"/><br><br>
      성별 : 남<input type = "radio" name = "gender" checked/>
             여<input type = "radio" name = "gender" /><br><br>
      관심사 : 연애<input type = "checkbox" name = "checkbox1" checked />
             스포츠<input type = "checkbox" name = "checkbox2" />
             IT<input type = "checkbox" name = "checkbox3" /><br><br>
      <input type = "submit" value = "submit"/>
      <input type = "reset" value = "reset"/><br><br>
    </fieldset>
  </form>
</body>
</html>
```

get url

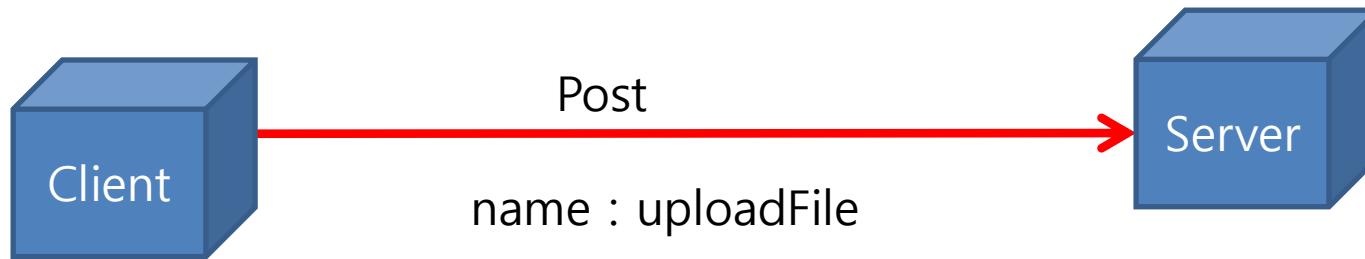
http://domainname.com?name=Ryan&security\_number=123456-1234567&id=mitchell&password=1234&gender=on&checkbox1=on

Python 매개변수객체

{'name':'Ryan', 'security\_number':'123456-1234567','id':'mitchell', 'password':'1234','gender':'on',checkbox1:'on'}

## Chap10. 폼과 로그인 뚫기

### 10.4 파일과 이미지 전송



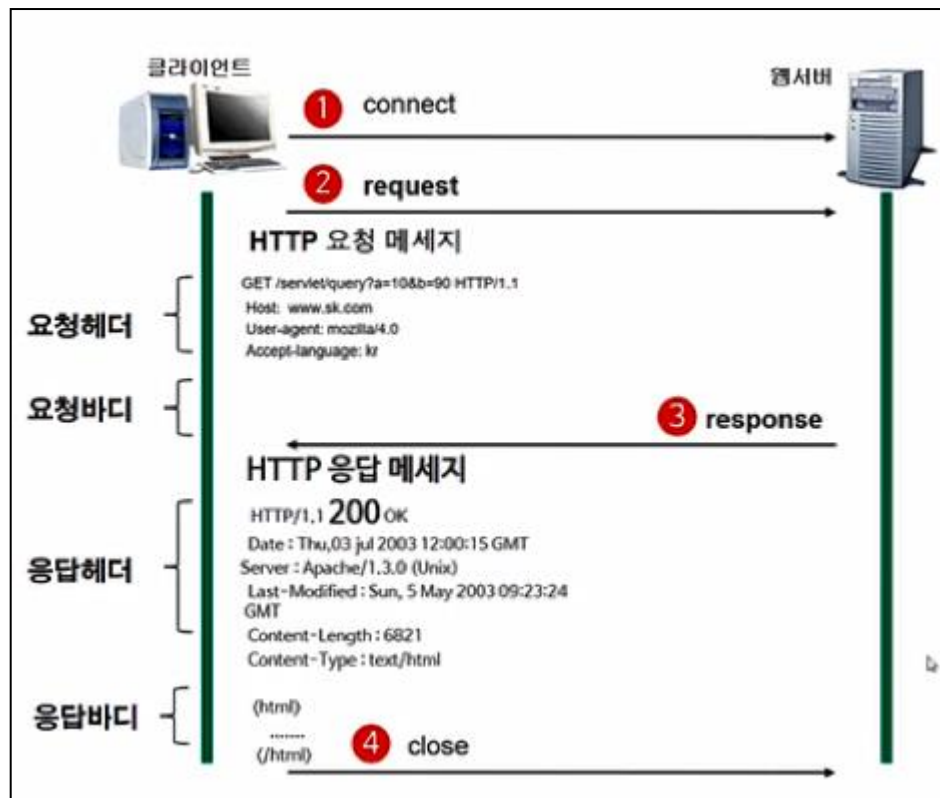
request 객체사용

<https://github.com/psf/requests/blob/master/requests/api.py>

# Chap10. 폼과 로그인 뚫기

## 10.5 로그인과 쿠키처리

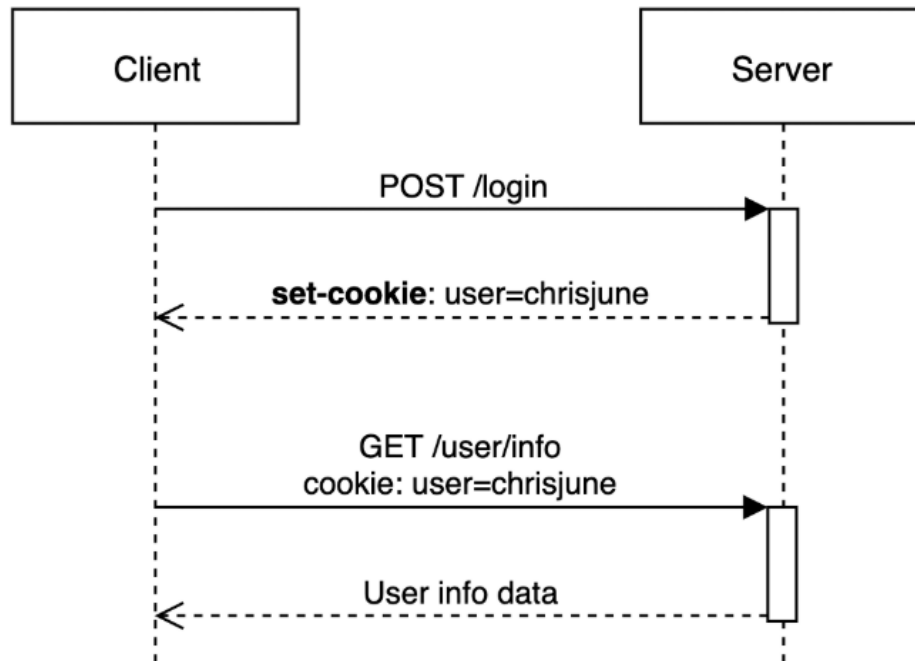
- HTTP
  - 비연결 지향(Connectionless)
  - 무상태(Stateless)



# Chap10. 폼과 로그인 뚫기

## 10.5 로그인과 쿠키처리

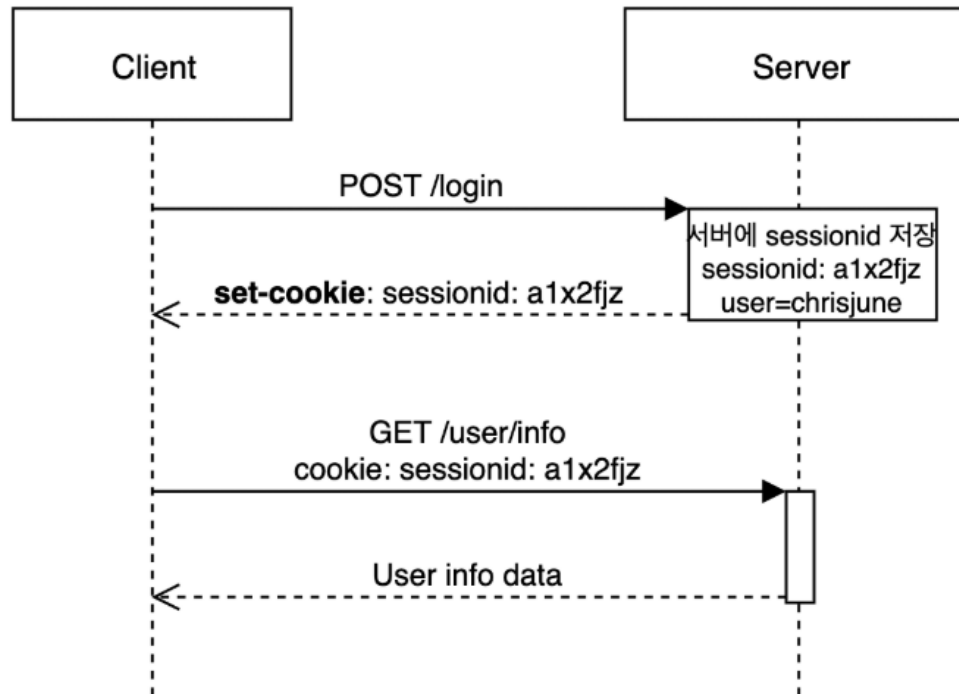
- 쿠키
  - 클라이언트 로컬에 저장되는 Key-Value쌍의 작은 데이터 파일



# Chap10. 폼과 로그인 뚫기

## 10.5 로그인과 쿠키처리

- 세션
  - 서버 메모리에 저장되는 정보





## Chap10. 폼과 로그인 뚫기

### 10.5 로그인과 쿠키처리

쿠키	세션
클라이언트(브라우저)에 저장	서버에 저장
보안에 취약	보안이 더 높음
속도 빠름	속도 느림
일반적인 정보를 저장	보안이 요구되는 자료저장

# Chap10. 폼과 로그인 뚫기

## 10.5 로그인과 쿠키처리

← → ↻ ⓘ 주의 요함 | pythonscrapping.com/pages/cookies/login.html

### Log In Here!

Warning: Your browser must be able to use cookies in order to view our site!

Username (use anything!):

Password (try "password"):

← → ↻ ⓘ 주의 요함 | pythonscrapping.com/pages/cookies/welcome.php


### Welcome to the Website!

Whoops! You logged in wrong. Try again with any username, and the password "password"  
[Log in here](#)

Network tab selected. Filter: All. Timeline shows a request to welcome.php.

Name	Value
loggedin	1
username	abc

← → ↻ ⓘ 주의 요함 | pythonscrapping.com/pages/cookies/profile.php



# Web Scrapping with Python

Network tab selected. Filter: All. Timeline shows a request to profile.php.

Name	Value	D...	P...
loggedin	1	p...	/...
username	abc	p...	/...

# Chap11. 자바스크립트 스크레이핑

## 11.1 자바스크립트

- 객체 기반의 스크립트 프로그래밍 언어
- <https://youtu.be/wcsVjmHrUQg>



# Chap11. 자바스크립트 스크레이핑

## 11.1.1 자바스크립트 라이브러리

- 제이쿼리
  - 오픈 소스 기반의 자바스크립트 라이브러리
- 구글 애널리틱스
  - 웹사이트 방문자의 데이터를 수집해서 분석함으로써 온라인 비즈니스의 성과를 측정하고 개선하는 데 사용하는 웹로그분석 도구
  - <https://analytics.google.com/>
- 구글 지도
  - [cloud.google.com/maps-platform/maps](https://cloud.google.com/maps-platform/maps)

# Chap11. 자바스크립트 스크레이핑

## 11.2 Ajax와 DHTML

- Ajax (synchronous Javascript And Xml)
  - JavaScript를 사용한 비동기 통신, 클라이언트와 서버간에 XML 데이터를 주고받는 기술
- DHTML (Dynamic HTML)
  - 정적 마크업 언어인 HTML과 클라이언트 기반 스크립트 언어(자바스크립트 같은) 그리고 스타일 정의 언어인 CSS를 조합하여 대화형 웹 사이트를 제작하는 기법

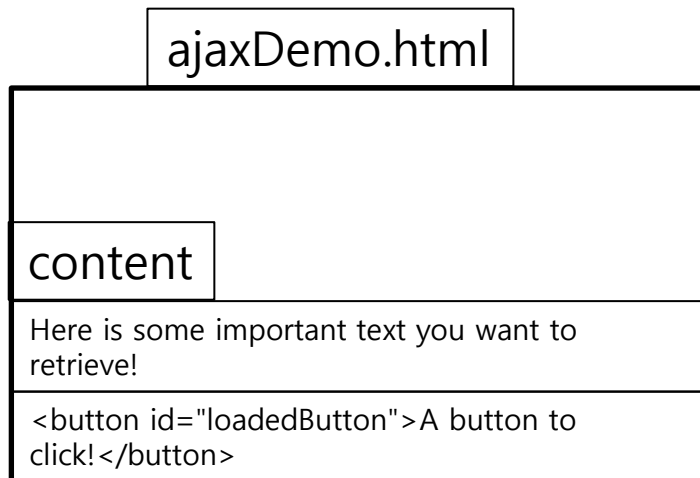
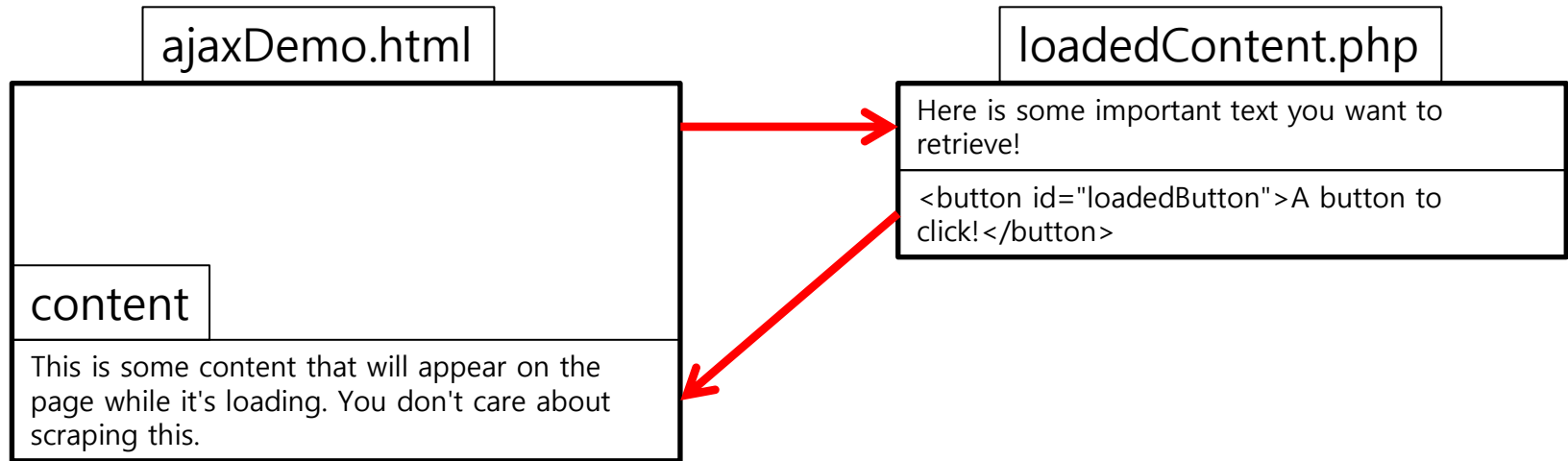
# Chap11. 자바스크립트 스크레이핑

## 11.2.1 셀레니움

- 팬텀JS (<http://phantomjs.org>)
- 셀레니움 설치
  - `conda install selenium`
  - `pip install selenium`
- 크롬 드라이버 다운로드
  - <https://sites.google.com/a/chromium.org/chromedriver/downloads>

# Chap11. 자바스크립트 스크레이핑

## 11.2.1 셀레니움



# Chap11. 자바스크립트 스크레이핑

## 11.3 리다이렉트처리

