

反爬：登入授權模擬

- 了解「登入權限機制」的反爬蟲機制
- 「登入權限機制」反爬蟲的因應策略

作業目標

- 找一個需要登入的網站試試看，並說明思考流程（如果不知道要用哪個網站的話，可以試試看 <https://github.com/new> (<https://github.com/new>) 網址，未登入時會被導向登入頁）

In [13]:

```
import requests
from bs4 import BeautifulSoup

r = requests.get('https://github.com/new')
soup = BeautifulSoup(r.text, 'lxml')
soup.title # 當我們直接存取 https://github.com/new 網址時，網頁會被導向登入頁
```

Out[13]:

```
<title>Sign in to GitHub · GitHub</title>
```

In [2]:

```
import requests
from bs4 import BeautifulSoup

cookies = {
    'logged_in': 'yes',
    'user_session': '9-S3QlTWrtHybLYX0gYeGmdqlMdM_rTnfi3jbRLgdvYKJWoQ'
}

r = requests.get('https://github.com/new', cookies=cookies)
soup = BeautifulSoup(r.text, 'lxml')
soup.title # 加上檢查的 cookies，就可以正確取得資料
```

Out[2]:

```
<title>Create a New Repository</title>
```

補充：從瀏覽器找 **cookie** 的方法

Search or jump to...

Pull requests

Issues

Marketplace

Explore

v123582

Repositories

Find a repository...

New

JokerCatz starred init-engineer/init.engineer 1 hour ago

init-engineer/init.engineer

這是一份純靠北工程師的專案，請好好愛護它，謝謝。

PHP 242 1 issue needs help Updated Jan 5

See what launched at GitHub Universe

Say hello to our new GitHub for mobile experience, get more from GitHub Actions and Packages, and learn how we're shaping the future of software, together.

Elements

Console

Sources

Network

Performance

Memory

Application

Security

Audits

Redux

Application

Manifest

Service Workers

Clear storage

Storage

Local Storage

Session Storage

IndexedDB

Web SQL

Cookies

https://github.com

Cache

Cache Storage

Application Cache

Filter

Name	Value	Dom...	Path	Expir...	Size	Http...	Secure	Sam...
_octo	GH1.1.1694594442.1541084337	.gith...	/	2020...	32			
dotcom_user	v123582	.gith...	/	2040...	18	✓	✓	
has_recent_activity	1	githu...	/	2020...	20			
ignored_unsupported_br...	false	githu...	/	Sess...	39			
logged_in	yes	.gith...	/	2040...	12	✓	✓	
tz	Asia%2FTaipei	githu...	/	Sess...	15		✓	
user_session	38VpOGmMB99pzxZcO3NEei9kP-VYr...	githu...	/	2020...	60	✓	✓	

yes

1. 選取開發者工具的 Application 區塊

2. 左邊選取 Storage > Cookies > 目前存取的網站

3. 這個區塊就是目前網站所使用的 Cookies 資訊

In []: