

# Chapter 10

## Scraping JavaScript

# HTML Source vs. JavaScript Rendered Page

- [http://ipv6.ncnu.org/Course/WebScraping/Exercise/deer\\_or\\_horse.html](http://ipv6.ncnu.org/Course/WebScraping/Exercise/deer_or_horse.html)

```
<body>
```

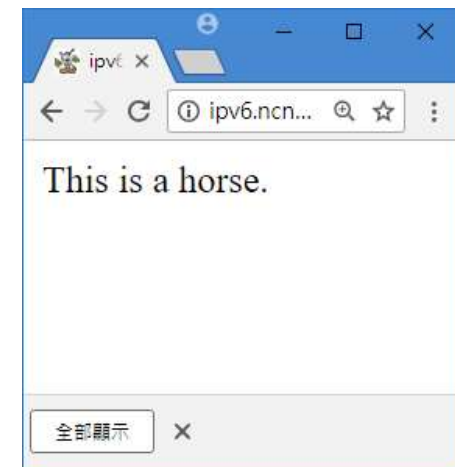
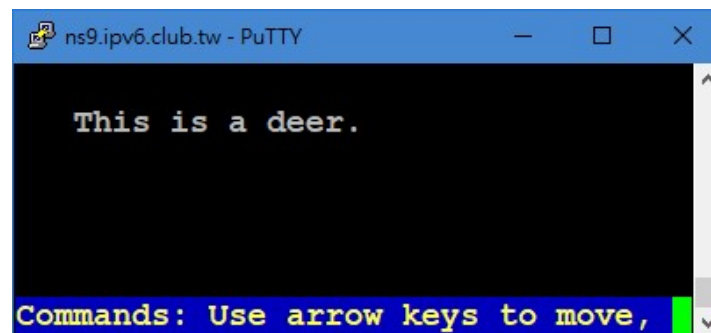
```
This is a <span id="noun">deer</span>.
```

```
</body>
```

```
<script>
```

```
document.getElementById("noun").innerHTML = "horse";
```

```
</script>
```

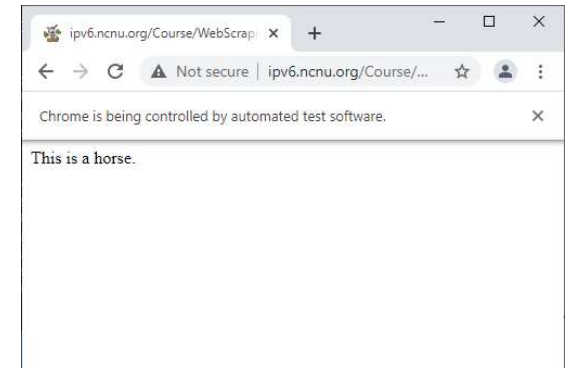




# You need a tool to parse JavaScript beyond HTML!

- `%pip install google-colab-selenium`
- `import google_colab_selenium as gs`
- `chrome = gs.Chrome()`
- `print(chrome.name)`
- `chrome.quit()`

# Retrieve a Webpage



- `import google_colab_selenium as gs`
- `chrome = gs.Chrome()`
- `from selenium.webdriver.common.by import By`
- `url = "http://ipv6.ncnu.org/Course/WebScraping/Exercise/deer_or_horse.html"`
- `chrome.get(url)`
- `print(chrome.title)`
- `print(chrome.find_element(By.ID, 'noun').text)`
- `chrome.quit()`



# Selenium with Form Submission



# seleniumPost.py

```
import google_colab_selenium as gs
chrome = gs.Chrome()
from selenium.webdriver.common.by import By
url = "http://ms15.ipv6.club.tw/~solomon/PHP/hw09-3-form.html"
chrome.get(url)
page1 = chrome.page_source
x = chrome.find_element(by=By.ID, value="x")
y = chrome.find_element(by=By.ID, value="y")
submit_button = chrome.find_element(by=By.ID, value="submit")

x.send_keys("5")
y.send_keys("3")
submit_button.click()

page2 = chrome.page_source
expr = chrome.find_element(by=By.ID, value="expression")
result = chrome.find_element(By.ID, value="result")
print(expr.text, '=', result.text)
print(page1, page2, sep='\n===\n')
```

<https://selenium-python.readthedocs.io/api.html#module-selenium.webdriver.remote.webelement>



You may extract information with BeautifulSoup.

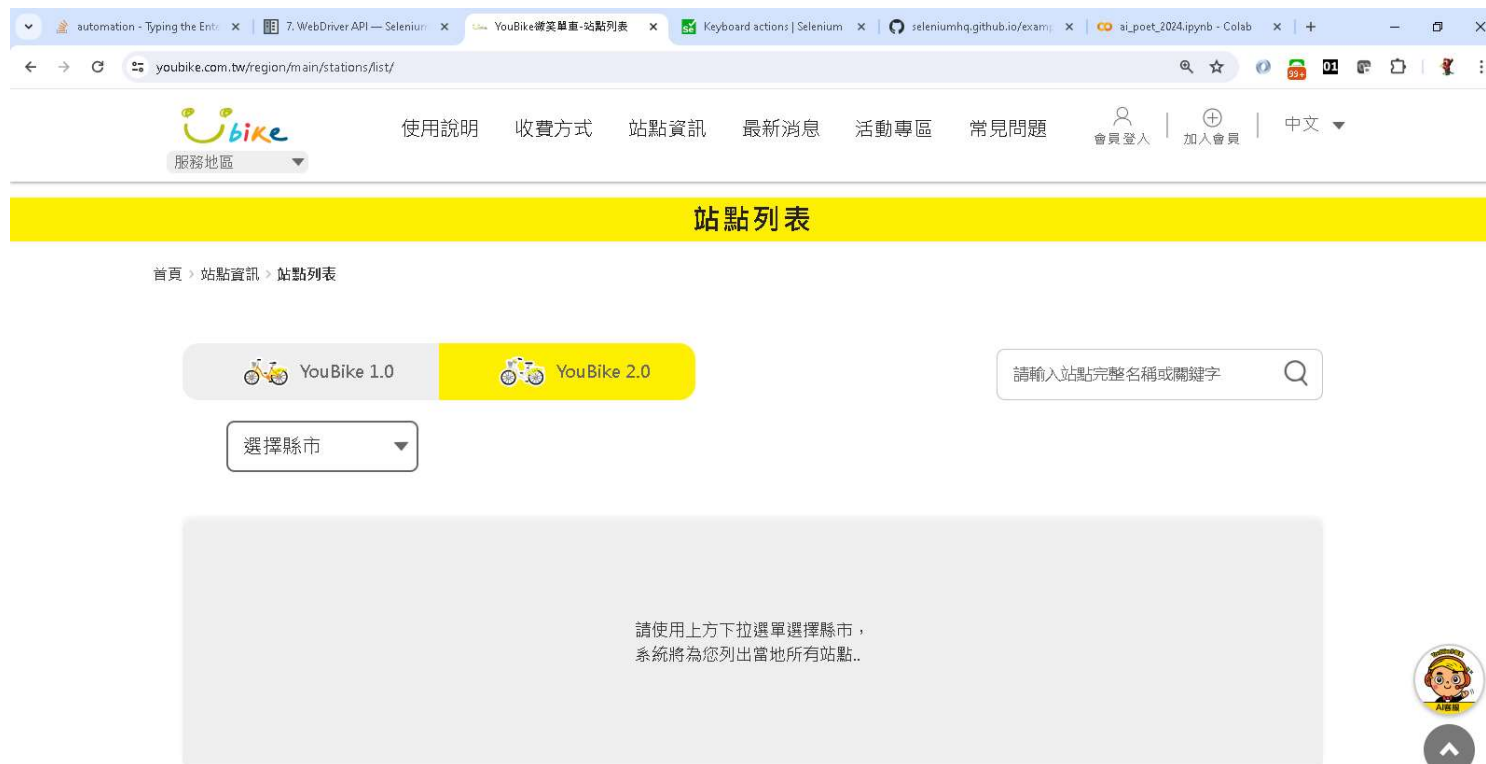
## Exercise: operation.html

- Visit the webpage in last week  
(`http://ms15.ipv6.club.tw/~solomon/PHP/operation.html`)  
which asks users the values of  $x$  and  $y$  and perform an operation of addition, subtraction, multiplication, or division.
- Use Selenium to write a program to test whether any one of the operation is not calculated correctly.
- The program may run as follow:
  - `Operation + is incorrect.`
  - `Operation - is correct.`
  - `Operation * is incorrect.`



# Exercise: YouBike

# 站點列表 (https://www.youbike.com.tw/region/main/stations/list/)



The screenshot shows the YouBike website's station list page. The browser's address bar displays the URL: [youbike.com.tw/region/main/stations/list/](https://www.youbike.com.tw/region/main/stations/list/). The page features a navigation menu with links for "使用說明", "收費方式", "站點資訊", "最新消息", "活動專區", and "常見問題". There are also links for "會員登入" and "加入會員", and a language selector set to "中文". A yellow banner at the top of the main content area reads "站點列表". Below this, a breadcrumb trail shows "首頁 > 站點資訊 > 站點列表". The main content area includes two tabs for "YouBike 1.0" and "YouBike 2.0", a search box with the placeholder text "請輸入站點完整名稱或關鍵字", and a dropdown menu labeled "選擇縣市". A large grey box contains the instruction: "請使用上方下拉選單選擇縣市，系統將為您列出當地所有站點..". A chatbot icon is visible in the bottom right corner.

# 選擇「高雄市」

The screenshot displays the YouBike 2.0 website interface. At the top, there is a navigation bar with links for '使用說明', '收費方式', '站點資訊', '最新消息', '活動專區', and '常見問題'. On the right side of the navigation bar, there are options for '會員登入', '加入會員', and '中文'. Below the navigation bar, a yellow banner reads '站點列表'. A search bar on the right contains the placeholder text '請輸入站點完整名稱或關鍵字'. A dropdown menu is open on the left, listing various cities and counties: '選擇縣市', '臺北市', '新北市', '桃園市', '新竹縣', '新竹市', '新竹科學園區', '苗栗縣', '臺中市', '嘉義市', '臺南市', '高雄市' (highlighted in blue), and '屏東縣'. Below the dropdown menu, a yellow button with a bicycle icon and the text 'YouBike 2.0' is visible. A large grey box in the center contains the instruction: '請使用上方下拉選單選擇縣市，系統將為您列出當地所有站點..'. In the bottom right corner, there is a circular icon for 'AI 服務' and an upward-pointing arrow.

# 輸入站點名稱並按 ENTER

The screenshot shows the YouBike website interface. At the top, there are navigation links for '使用說明', '收費方式', '站點資訊', '最新消息', '活動專區', and '常見問題'. Below these, there are options for '會員登入' and '加入會員', along with a language selector set to '中文'. The main content area features a search bar with '英雄' entered, highlighted by a red box. Below the search bar, there are filters for 'YouBike 1.0' and 'YouBike 2.0', and a dropdown menu for '高雄市'. A grid of checkboxes lists various districts, all of which are checked. Below the filters, a table displays the search results for '英雄'.

縣市	區域	站點名稱	可借車輛	可停空位
高雄市	苓雅區	高雄國軍英雄館	3	21
高雄市	前金區	高雄女中(英雄路口)	0	30

# Exercise: 晚自習預約管理系統

- 1. Login
  - <http://163.16.245.4/studying/login.aspx>

晚自習管理系統  
高雄女中圖書館  
Kaohsiung Municipal Girls' Senior High School

讀者查詢登入(請擇一輸入並輸入密碼)

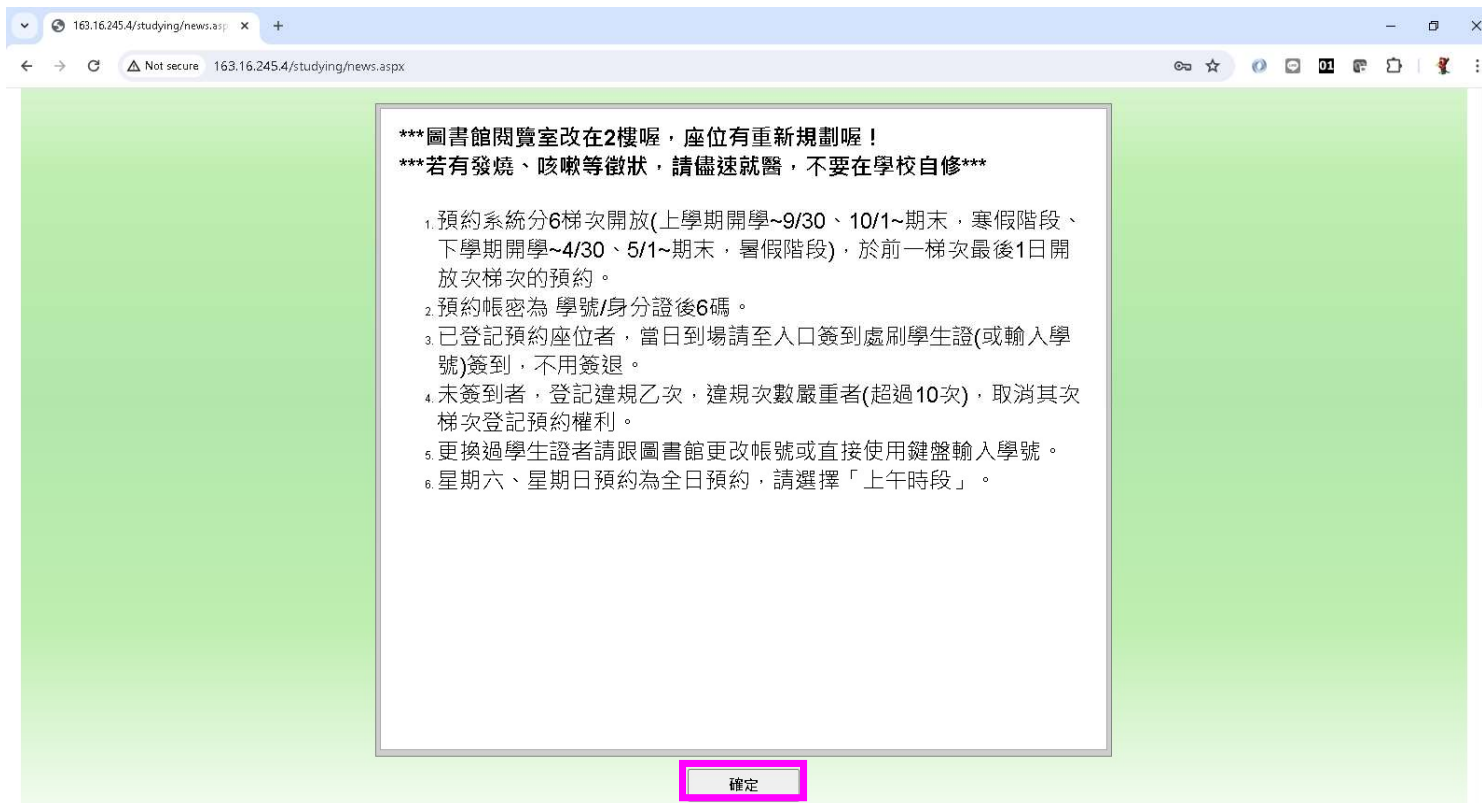
帳號(讀者編號 / 姓名, 教師編號):  密碼:

登入

請輸入讀者帳號密碼

© 1989 - 2024 版權所有，轉載必究  
網站更新日期：2020/09/11

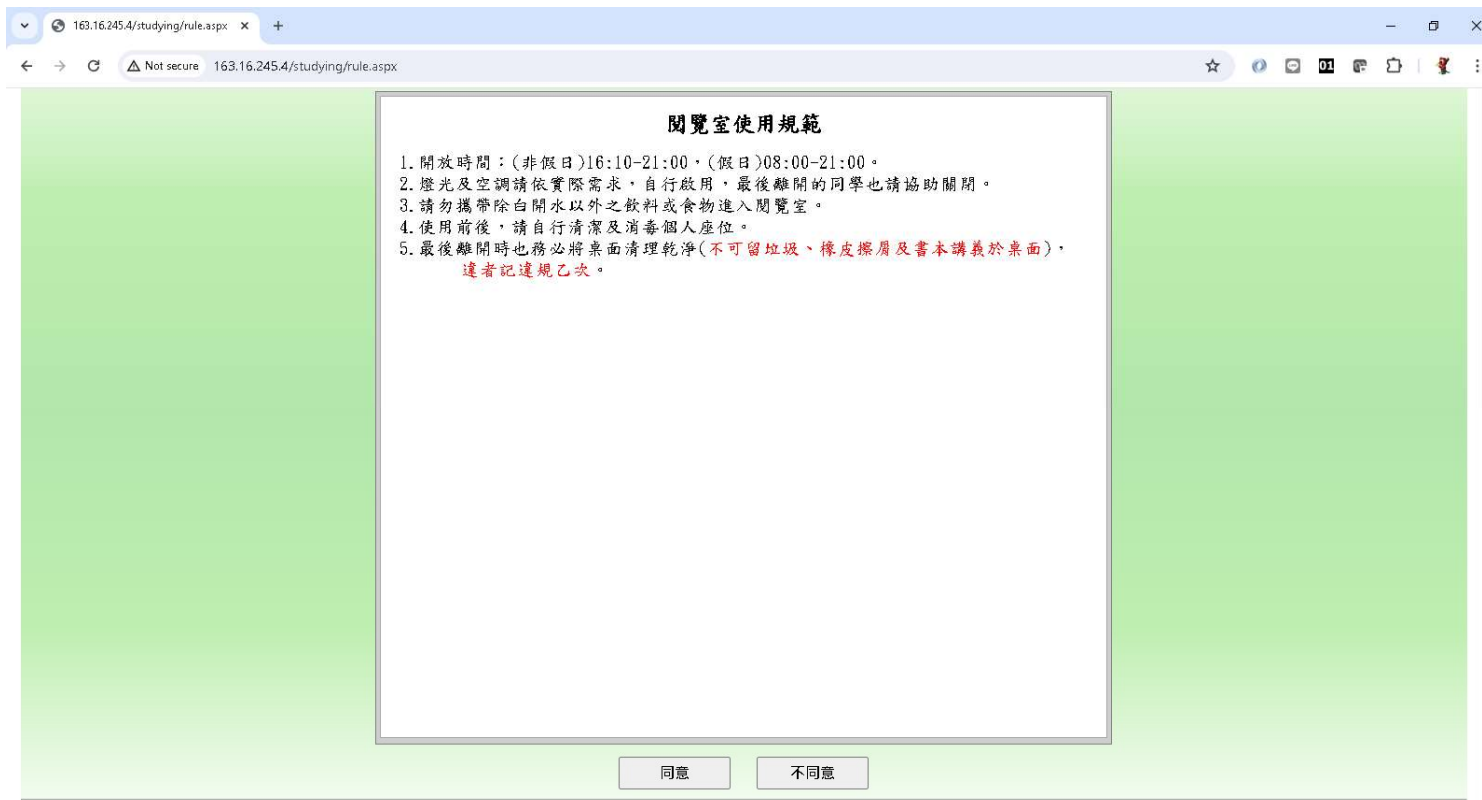
## 2. 強迫每次都要再看一次「最新消息」



### 3. 這裡一定要按「申請」(若按「不申請」就立即登出)



## 4. 必須同意「閱覽室使用規範」(不是「自習室」嗎?)



## 5. 同意之後要再按一次「申請」

 申請  不申請 參加 晚自習, 本人將督促其遵守實施辦法, 並自行負責規劃回家交通及安全事宜。' The footer contains copyright information: '© 1989 - 2024 版權所有, 轉載必究' and '網站更新日期: 2020/09/11'."/>

線上申請

Not secure 163.16.245.4/studying/onlineRegister.aspx

### 晚自習管理系統

高雄女中圖書館  
Kaohsiung Municipal Girls' Senior High School

登出  
巴家寶 您好

座位登記作業 出席狀態查詢 預約查詢 基本資料

敝子弟 畢業校友, 學號: 100001, 姓名: 巴家寶  申請  不申請 參加 晚自習, 本人將督促其遵守實施辦法, 並自行負責規劃回家交通及安全事宜。

© 1989 - 2024 版權所有, 轉載必究  
網站更新日期: 2020/09/11

6. 有人先決定座位再決定哪一天要去自習嗎？

請選取預約方式:

依日期選座位

依座位選日期

# 7. 選日期

自修室：	圖書館2樓閱覽室						
	圖書館2樓閱覽室						
	2024年5月						
	星期日	星期一	星期二	星期三	星期四	星期五	星期六
	28	29	30	1	2	3	4
	5	6	7	8	9	10	11
	12 休館日	13	14	15	16	17 休館日	18 休館日
	19 休館日	20	21	22	23	24	25

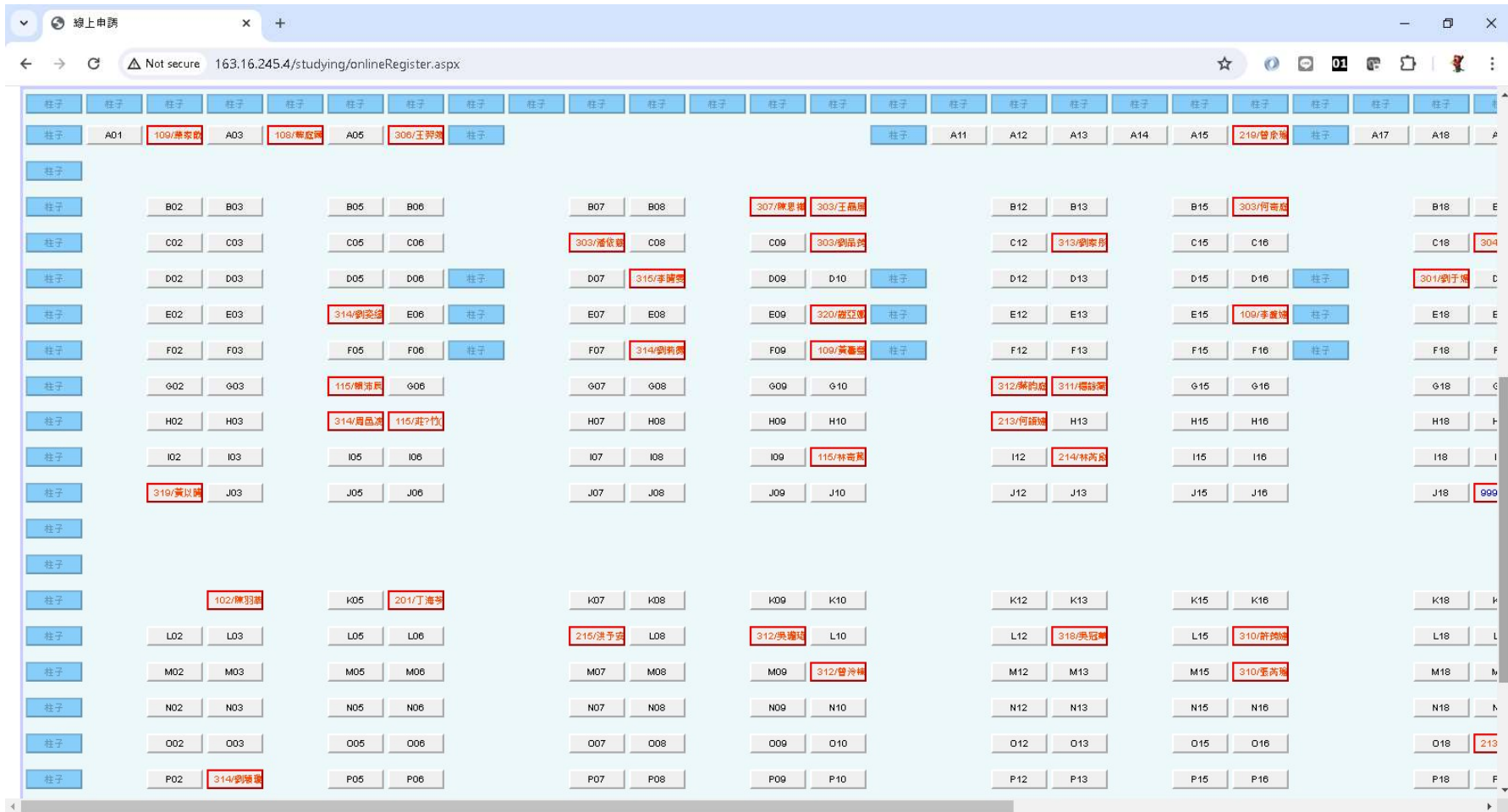
  

選擇自修日期：	26	27	28	29	30	31	1
	2	3	4	5	6	7	8

## 8. 選時段 (雖然只有一個時段可選)

自修室：		圖書館2樓開覽室						
選擇自修日期：	< 2024年5月 >							
	星期日	星期一	星期二	星期三	星期四	星期五	星期六	
	28	29	30	1	2	3	4	
	5	6	7	8	9	10	11	
	12	13	14	15	16	17 休館日	18 休館日	
	19 休館日	20	21	22	23	24	25	
	26	27	28	29	30	31	1	
2	3	4	5	6	7	8		
		選擇自修時段： 2024/05/29,星期三 <input type="radio"/> 16:10-21:00						

# 9. 在平面圖上選一個座位



# 10. 確認申請

## 自修室座位申請確認

學號： 100001

班級： 畢業校友

姓名： 巴家寶

選取自修座位： J19

選取自修日期： 2024/05/29,星期三

選取自修時段： 16:10-21:00

申請此座位

取消申請此座位

返回