

更多的資料爬蟲工具介紹 - PyQuery/grab



簡報閱讀



範例與作業



問題討論



學習心得(完成)

更多的資料爬蟲工具介紹 - PyQuery/grab

**Day 09** 靜態網頁資料爬蟲

更多的資料爬蟲工具介紹 - PyQuery/grab



 python


出題教練：張維元

cupoy

本日知識點目標



本日知識點目標

- 利用 Grab 套件的存取 HTML 資源
- 利用 PyQuery 套件的解析 HTML 格式

Grab

應。」

像 request 一樣，Grab 初始化的時候也需要傳入 URL，接著就可以用 body 取回內容

```
1 from grab import Grab
2 g = Grab()
3 resp = g.go('https://google.com')
  resp.body
```

參考資料：grablab.org

PyQuery

「除了 BeautifulSoup 之外，如果你比較熟悉網頁運作的話，其實也可以嘗試看看 PyQuery。PyQuery 是一個利用 CSS Selector 選取資料的解析工具。」

像 BeautifulSoup 一樣，PyQuery 初始化的時候也需要傳入 HTML 字串來初始化一個操作對象，接著就可以用 doc 的方式搭配 CSS 選擇器存取資料。

```
from pyquery import PyQuery as pq
doc = pq("<html><h1>Hello World</h1></html>")
h1 = doc('h1')
print(type(h1), h1.text())
```

參考資料：pythonhosted.org

重要知識點複習





- 利用 Grab 套件的存取 HTML 資源
- 利用 PyQuery 套件的解析 HTML 格式

解題時間



Sample Code & 作業
開始解題



更多的資料爬蟲工具介紹 -
PyQuery/grab



本日知識點目標



Grab



PyQuery



重要知識點複習



解題時間



下一步：閱讀範例與完成作業

