



教師跨領域研究學社群「智慧觀光服務跨領域研究社群(第三次)」

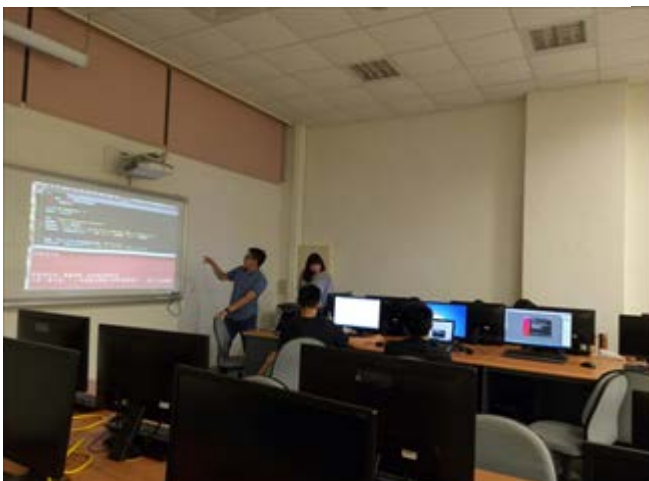
時間：2019-11-11
地點：蘭陽校區 CL324

接續上一堂的課程內容，在同學們了解在瀏覽網頁時，資料是如何傳輸，以什麼樣的形式傳輸後，即可利用 Python 程式語言實作爬蟲，取得網頁上的資料。這次的課程使用 Python 的專案管理器安裝了兩個模塊：BeautifulSoup 以及 requests。

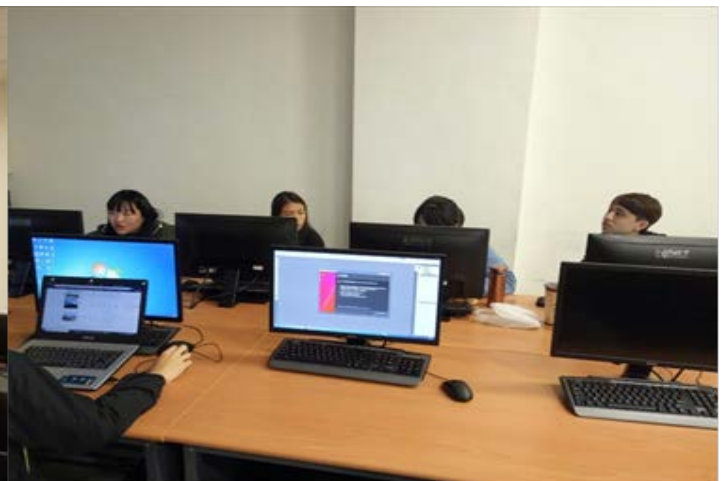
BeautifulSoup 是一個 Python 的函式庫模塊，可以讓同學們撰寫少量的程式碼，就可以快速解析網頁的 HTML 原始碼，並且從 HTML 原始碼解析出同學們「有興趣」，也就是同學們的「目標資料」。這個模組降低了開發的門檻，僅利用不到二十行的程式碼，即可萃取出想要的內容。

requests 也是一個 Python 實用的模塊，用來處理對目標發送請求的所有設定，例如：統一資源定位符、HTTP 方法、傳送資料時所帶的參數.....等等。

課程中運用這兩個 Python 模塊，實作一個 crawler.py 爬蟲應用程式，取得使用者在終端機輸入的關鍵字在 Google 搜尋引擎查詢後的內容，取得在搜尋引擎第一頁出現的所有標題。



講解程式碼



同學們實作爬蟲