112年第一學期大數據分析個人期末報告

學習爬蟲之路-Ajax爬蟲實作

**學號:411077017 姓名:石祐安**

透過python撰寫Ajax爬蟲，將Hahow上所有課程的名稱、評分、價格和購買人數等資訊爬回來，並且將資料存放在Excel檔裡，方便進行統計和分析資料。

1. **前言**

在這個資訊爆炸的時代，大量的資訊都以一一呈現在各個網站上。然而，若要了解各個網站上的資訊，一則一則地慢慢尋找實在是太耗時了，不符合效益。爬蟲能應用在多種情境，尤其對於資訊相關的工作而言，能夠迅速地將資料取回進行分析。未來作為一名資訊從業者，肯定會常常需要從網路上取得資料進行分析。為了提升自己的能力，我決定踏上學習爬蟲學習之路，增進自身能力。

1. **相關文獻**

[**https://www.youtube.com/watch?v=xua4Gno7xLo&t=2390s**](https://www.youtube.com/watch?v=xua4Gno7xLo&t=2390s)

**【Python】requests發送http請求｜模擬發文、爬取資訊**

[**https://www.youtube.com/watch?v=1PHp1prsxIM&t=4798s**](https://www.youtube.com/watch?v=1PHp1prsxIM&t=4798s)

【Python爬蟲】2小時初學者課程：一次學會PTT爬蟲、Hahow爬蟲、Yahoo電影爬蟲！

透過影片學習如何實作Ajax爬蟲。

1. **步驟**

透過爬取Hahow平台上所有線上課程的課程資訊練習，來實作Ajax爬蟲

1. 找到Hahow網站上存放課程資訊的API

一張含有 文字, 螢幕擷取畫面, 軟體, 多媒體軟體 的圖片

自動產生的描述

1. 設定好想取得資料的網頁，透過requests發送請求去抓取網頁資料



一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

1. 取得response後，要將資料轉成json形式才可分析資料，透過json形式的資料將每頁中需要的資料以陣列的方式儲存起來，並且印出進度來讓使用者知道程式有正常運作。

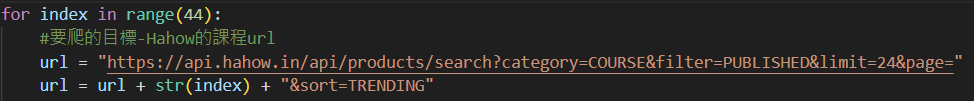
一張含有 文字, 字型, 螢幕擷取畫面, 時鐘 的圖片

自動產生的描述

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

1. 重複上述二到三步驟，將網頁上所有課程頁面執行過一次，爬取所有課程內容



1. 所有資料蝦取完畢後，印出完成提示，並將資料儲存至excel檔裡

一張含有 字型, 文字, 圖形, 印刷術 的圖片

自動產生的描述

一張含有 文字, 螢幕擷取畫面, 字型, 行 的圖片

自動產生的描述

1. **系統建構-資料分析**

完整程式碼截圖

一張含有 文字, 螢幕擷取畫面, 軟體, 陳列 的圖片

自動產生的描述

一張含有 文字, 螢幕擷取畫面, 字型, 行 的圖片

自動產生的描述

運行過程和結果

一張含有 文字, 字型, 螢幕擷取畫面, 圖書 的圖片

自動產生的描述

一張含有 文字, 螢幕擷取畫面, 圖書, 功能表 的圖片

自動產生的描述

產生excel檔

一張含有 文字, 字型, 螢幕擷取畫面, 圖形 的圖片

自動產生的描述

一張含有 文字, 螢幕擷取畫面, 數字, 字型 的圖片

自動產生的描述

1. **結論**

經過這次實作，讓我了解了爬蟲的運作原理，從發出請求，到分析資料，再到最後的呈現資料，這過程看似困難，但只要肯開始學，就會發現其實也蠻簡單的。經過這次的經驗，也讓我更願意去學習新知，進入學習的正循環。

1. **PDF連結**

[**https://docs.google.com/document/d/1wSQ\_zDJfT\_dez49LFy1ZkGUlRiYK7k54/edit?usp=drive\_link&ouid=108116234271939464090&rtpof=true&sd=true**](https://docs.google.com/document/d/1wSQ_zDJfT_dez49LFy1ZkGUlRiYK7k54/edit?usp=drive_link&ouid=108116234271939464090&rtpof=true&sd=true)

1. **影片連結**

[**https://vimeo.com/898755487/6e2b15723e?share=copy**](https://vimeo.com/898755487/6e2b15723e?share=copy)