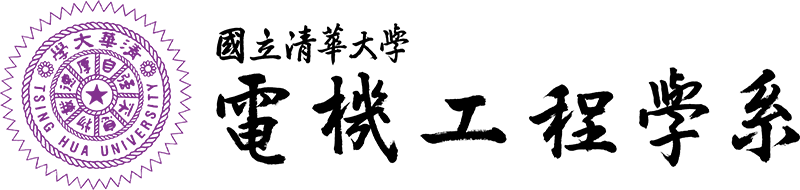
**大學甄試入學工具箱**



|  |  |
| --- | --- |
| 作者 | Jack Cherng <[jfcherng@gmail.com](mailto:jfcherng@gmail.com)> |
| 原始碼 | <https://github.com/jfcherng/CAAC-Toolkit> |

目錄

[一、 環境安裝 2](#_Toc8178583)

[1. 依照作業系統的位元數下載 Python 3.6或更高版本 2](#_Toc8178584)

[2. 安裝 Python 2](#_Toc8178585)

[3. 雙擊 install\_requirements.bat 安裝所需的 Python套件 2](#_Toc8178586)

[二、 第一階段-篩選結果 3](#_Toc8178587)

[1. 修改 do\_crawl.bat 中的參數設定 3](#_Toc8178588)

[2. 抓取網站內容並建立本地資料庫 4](#_Toc8178589)

[3. 修改 do\_lookup.bat 中的參數設定 4](#_Toc8178590)

[4. 從資料庫取出資料 5](#_Toc8178591)

[三、 第二階段-交叉查榜 6](#_Toc8178592)

[1. 將要查詢的校系代碼寫入 department\_ids.txt 中 6](#_Toc8178593)

[2. 雙擊 do\_cross.bat 6](#_Toc8178594)

[3. 等待抓取當年度的網站內容，直到看見 **[Done] It takes XXX seconds.** 後關閉 6](#_Toc8178595)

[4. result\_xxx.xlsx 即為結果，內容以准考證號碼排序。 6](#_Toc8178596)

1. 環境安裝

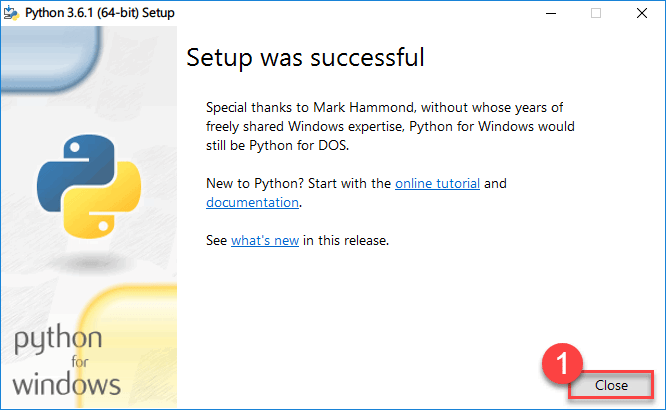
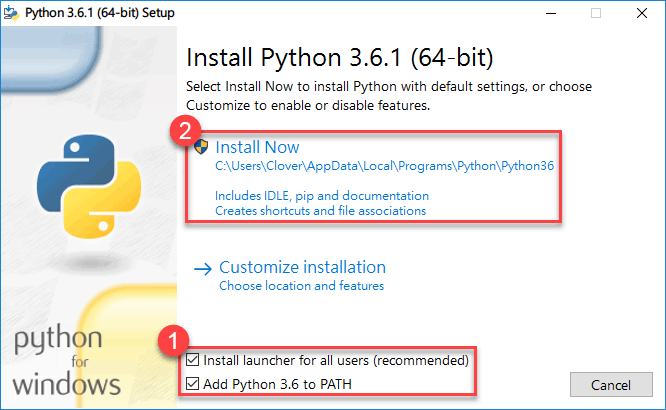
* Python 3

1. 依照作業系統的位元數下載 Python 3.6或更高版本

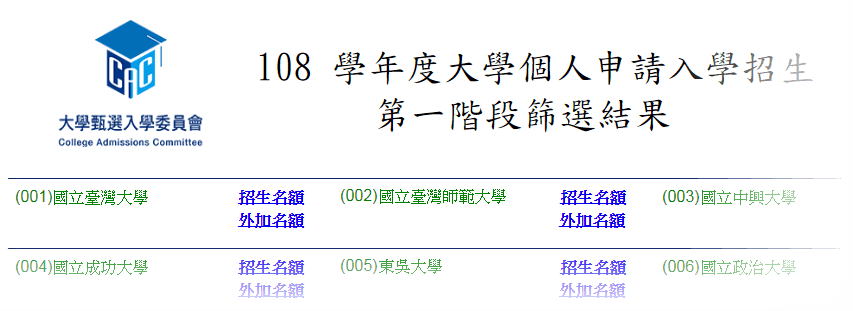
* 32位元：<https://www.python.org/ftp/python/3.7.3/python-3.7.3.exe>
* 64位元：<https://www.python.org/ftp/python/3.7.3/python-3.7.3-amd64.exe>

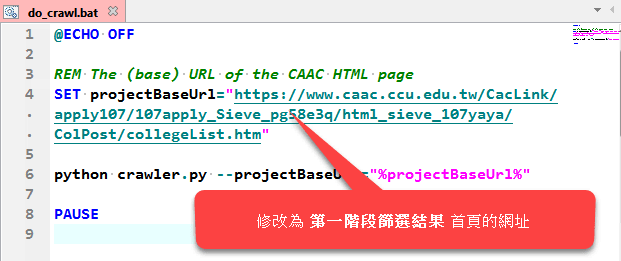
※備註：現代的電腦應該已經大多為 **64位元** 的作業系統了。

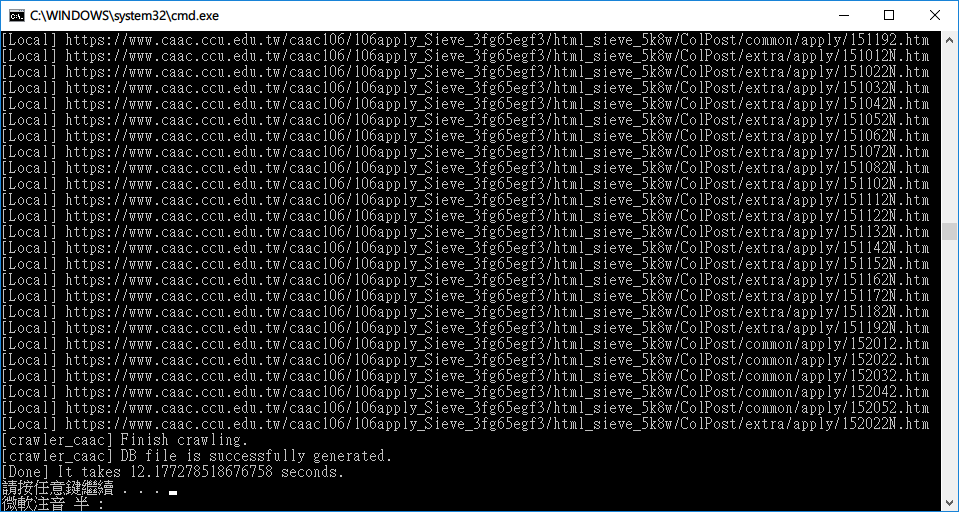
1. 安裝 Python



1. 雙擊 install\_requirements.bat 安裝所需的 Python套件
2. 第一階段-篩選結果
   1. 修改 do\_crawl.bat 中的參數設定
      1. 對 do\_crawl.bat 按 右鍵→編輯
      2. 打開網頁瀏覽器，找到當年度 **第一階段篩選結果** 的網頁

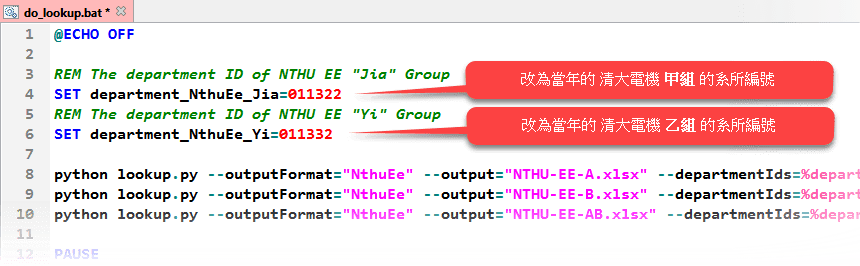


* + 1. 複製該網頁的網址，貼上到 do\_ crawl.bat 中的 **projectBaseUrl**
    2. 儲存並關閉檔案
  1. 抓取網站內容並建立本地資料庫
     1. 雙擊 do\_crawl.bat
     2. 等待抓取當年度的網站內容，直到看見 **[Done] It takes XXX seconds.** 後關閉

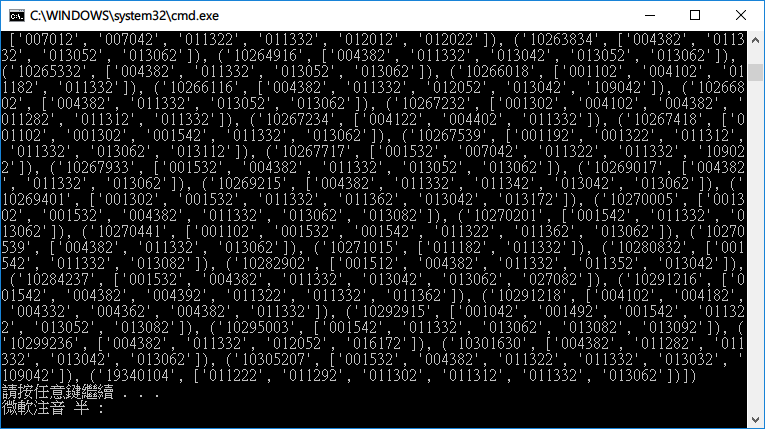


此過程視網路環境的不同，可能耗費5至40分鐘。

* + 1. 如果想要**完全**重新抓取網站內容，可以先刪除 **crawler\_XXX** 資料夾
  1. 修改 do\_lookup.bat 中的參數設定
     1. 對 do\_lookup.bat 按 右鍵→編輯
     2. 修改參數設定
* department\_NthuEe\_Jia 為當年度的 清大電機**甲**組 的系所編號（6位數）
* department\_NthuEe\_Yi 為當年度的 清大電機**乙**組 的系所編號（6位數）

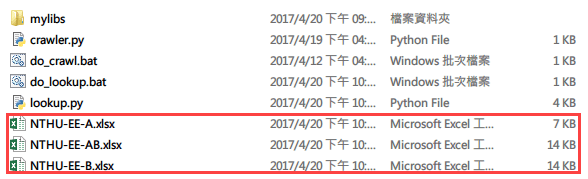


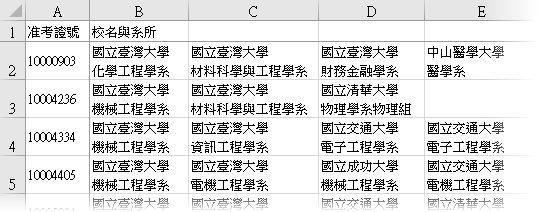
* + 1. 儲存並關閉檔案
  1. 從資料庫取出資料
     1. 雙擊 do\_lookup.bat



如果跑出如上圖一堆數字，那麼表示正確執行，可關閉該視窗。

* + 1. 產生的三個 .xlsx 檔案即為結果
* NTHU-EE-A.xlsx 為報名 清大電機**甲**組 者
* NTHU-EE-B.xlsx 為報名 清大電機**乙**組 者
* NTHU-EE-AB.xlsx 為報名 清大電機**甲**組 **或** 清大電機**乙**組 者





1. 第二階段-交叉查榜

本階段的資料來源為：大學交叉查榜 <https://www.com.tw/cross/>

1. 將要查詢的校系代碼寫入 department\_ids.txt 中

|  |  |
| --- | --- |
|  | 如左圖，一行寫一個校系代碼。  圖中範例為108年清大電機甲、乙組的校系代碼。 |

1. 雙擊 do\_cross.bat
2. 等待抓取當年度的網站內容，直到看見 **[Done] It takes XXX seconds.** 後關閉
3. result\_xxx.xlsx 即為結果，內容以准考證號碼排序。

※備註：因「不明」與「落榜」的在網頁上看起來一樣，因此「不明」實際上可能是「落榜」。也可能在 Excel 裡看到被分發到「不明狀態」的校系，實際上該分發結果（皇冠）是**正確**的，只是該「不明」實際上是「落榜」。這可能與各學校放榜的個資保護方式有關。

