**大學甄試入學工具箱**

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

目錄

[一、 環境安裝 2](#_Toc480397730)

[1. 依照作業系統的位元數下載 Python 3.6.1 （或更高版本） 2](#_Toc480397731)

[2. 安裝 Python 2](#_Toc480397732)

[二、 第一階段-篩選結果 3](#_Toc480397733)

[1. 抓取網站內容並建立本地資料庫 3](#_Toc480397734)

[2. 修改 do\_lookup.bat 中的參數設定 3](#_Toc480397735)

[3. 從資料庫取出資料 3](#_Toc480397736)

[三、 第二階段-交叉查榜 5](#_Toc480397737)

[1. 先執行過 第一階段-篩選結果 5](#_Toc480397738)

[2. 將要查詢的准考證號碼寫入 admission\_ids.txt 中 5](#_Toc480397739)

[3. 雙擊 do\_cross.bat 5](#_Toc480397740)

[4. 等待抓取當年度的網站內容，直到看見 [Done] It takes XXX seconds. 後關閉 5](#_Toc480397741)

[5. result.csv 即為結果，准考證的順序將與 admission\_ids.txt 中的順序相同。 5](#_Toc480397742)

1. 環境安裝

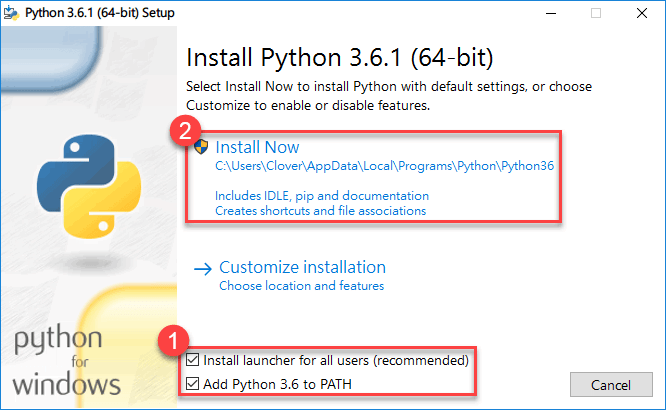
* Python 3

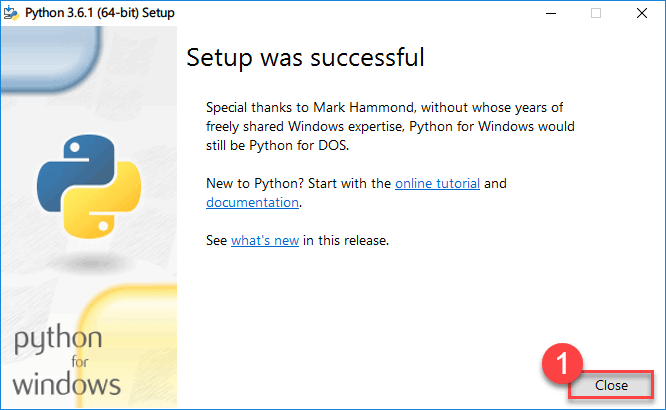
1. 依照作業系統的位元數下載 Python 3.6.1 （或更高版本）

* 32位元：<https://www.python.org/ftp/python/3.6.1/python-3.6.1.exe>
* 64位元：<https://www.python.org/ftp/python/3.6.1/python-3.6.1-amd64.exe>

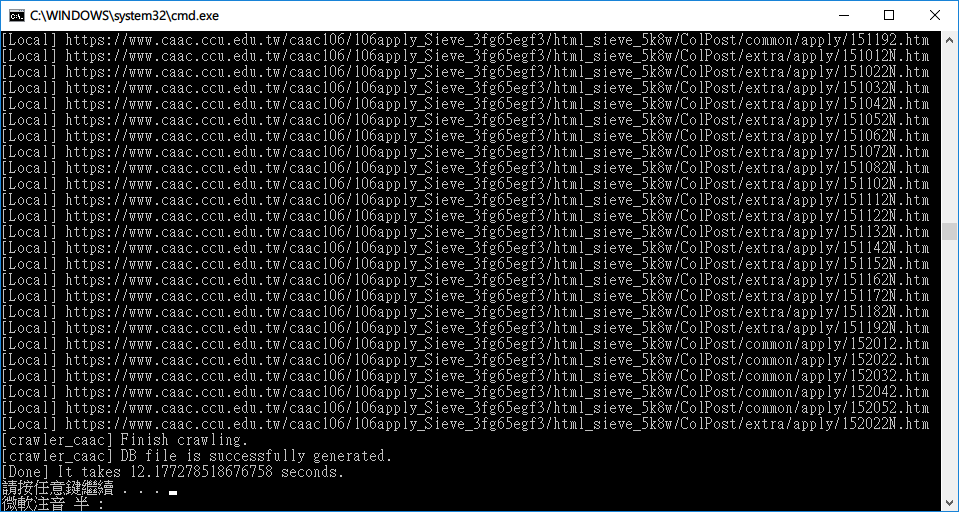
如果不確定作業系統的位元數，你總是可以安裝 **32位元** 的版本。

1. 安裝 Python



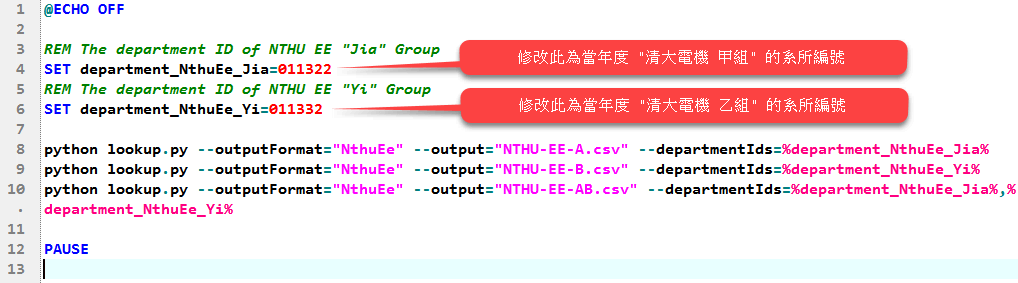


1. 雙擊 install\_requirements.bat 安裝所需的 Python套件
2. 第一階段-篩選結果
   1. 抓取網站內容並建立本地資料庫
      1. 雙擊 do\_crawl.bat
      2. 等待抓取當年度的網站內容，直到看見 **[Done] It takes XXX seconds.** 後關閉

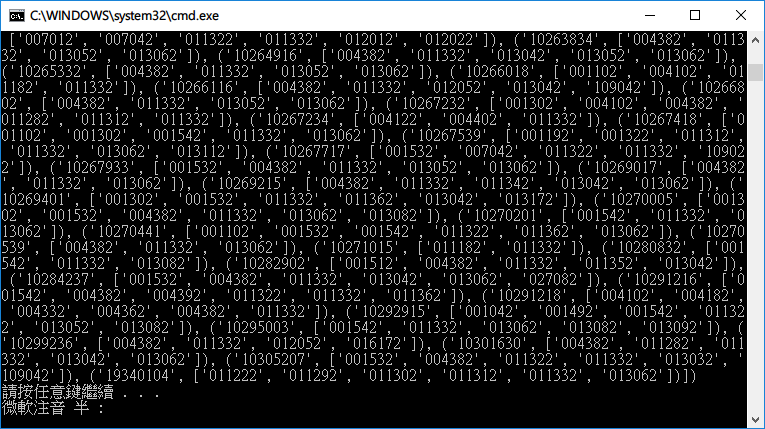


此過程視網路環境的不同，可能耗費20至40分鐘。

* + 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 .csv 檔案即為結果



* NTHU-EE-A.csv 為報名 清大電機**甲**組 者
* NTHU-EE-B.csv 為報名 清大電機**乙**組 者
* NTHU-EE-AB.csv 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：<http://freshman.tw>

1. 先執行過 **第一階段-篩選結果**

如果以前已經做過 **第一階段-篩選結果** 可以跳過此步驟。

※ 主要是 data/crawler\_xxx/sqlite3.db 必須已經產生

1. 將要查詢的准考證號碼寫入 admission\_ids.txt 中

|  |  |
| --- | --- |
| C:\Users\Clover\Desktop\asdasd.png | 如左圖，一行一個准考證號。  可以直接從 Excel 那邊選取整列後直接複製過來。 |

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.csv 即為結果，准考證的順序將與 admission\_ids.txt 中的順序相同。

若 admission\_ids.txt 中有重複的准考證號，此處也會有重複以保證資料筆數相同。

