

各校系招生名額及外加名額第一階段篩選通過名單

作者	ifcherng@gmail.com
原始碼	https://github.com/ifcherng/CAAC-Toolkit

目錄

一、	先決條件.....	2
	1. 依照作業系統的位元數下載 Python 3.6.1 （或更高版本）	2
	2. 安裝 Python	2
二、	使用方式.....	3
	1. 抓取網站內容並建立本地資料庫.....	3
	2. 修改 do_lookup.bat 中的參數設定	3
	3. 從資料庫取出資料.....	4

一、 先決條件

● 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>

檢視電腦的基本資訊

Windows 版本

Windows 10 專業版

© 2016 Microsoft Corporation. 著作權所有，並保留一切權利。



系統

處理器: Intel(R) Xeon(R) CPU E3-1230 V2 @ 3.30GHz 3.30 GHz
已安裝記憶體 (RAM) 4.00 GB
系統類型: **64 位元作業系統** x64 型處理器
手寫筆與觸控: 此顯示器不提供手寫筆或觸控式輸入功能。

電腦名稱、網域及工作群組設定

電腦名稱: VM-Win10
完整電腦名稱: VM-Win10
電腦描述:
工作群組: WORKGROUP

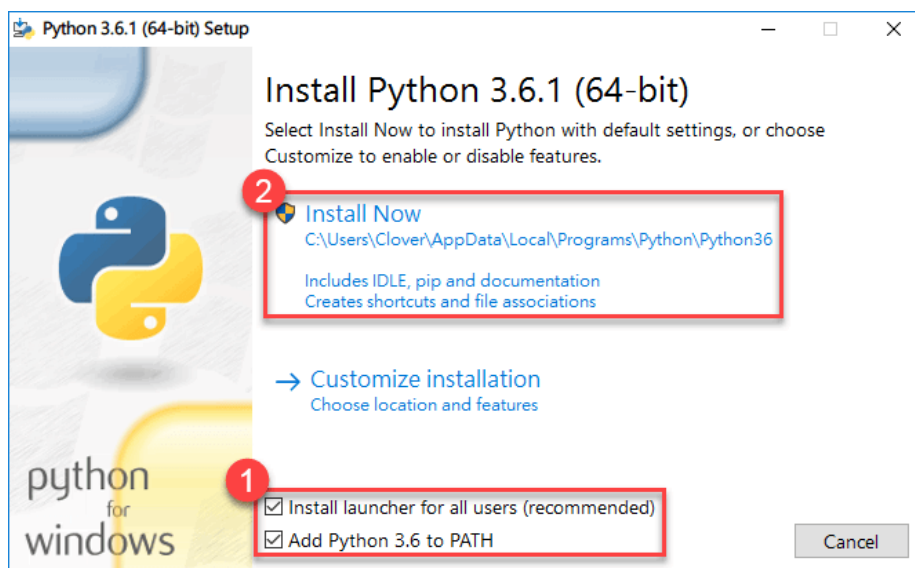
變更設定

Windows 啟用

Windows 已啟用 [閱讀 Microsoft 軟體授權條款](#)

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。

2. 安裝 Python





3. 雙擊 install_requirements.bat 安裝依賴套件

二、 使用方式

1. 抓取網站內容並建立本地資料庫

- i. 雙擊 do_crawl.bat
- ii. 等待抓取當年度的網站內容，直到看見 **[Done] It takes XXX seconds.** 後關閉

```
C:\WINDOWS\system32\cmd.exe
[Local] https://www.caac.ccu.edu.tw/caac106/106apply_Sieve_3fg65egf3/html_sieve_5k8w/ColPost/common/apply/151192.htm
[Local] https://www.caac.ccu.edu.tw/caac106/106apply_Sieve_3fg65egf3/html_sieve_5k8w/ColPost/extra/apply/151012N.htm
[Local] https://www.caac.ccu.edu.tw/caac106/106apply_Sieve_3fg65egf3/html_sieve_5k8w/ColPost/extra/apply/151022N.htm
[Local] https://www.caac.ccu.edu.tw/caac106/106apply_Sieve_3fg65egf3/html_sieve_5k8w/ColPost/extra/apply/151032N.htm
[Local] https://www.caac.ccu.edu.tw/caac106/106apply_Sieve_3fg65egf3/html_sieve_5k8w/ColPost/extra/apply/151042N.htm
[Local] https://www.caac.ccu.edu.tw/caac106/106apply_Sieve_3fg65egf3/html_sieve_5k8w/ColPost/extra/apply/151052N.htm
[Local] https://www.caac.ccu.edu.tw/caac106/106apply_Sieve_3fg65egf3/html_sieve_5k8w/ColPost/extra/apply/151062N.htm
[Local] https://www.caac.ccu.edu.tw/caac106/106apply_Sieve_3fg65egf3/html_sieve_5k8w/ColPost/extra/apply/151072N.htm
[Local] https://www.caac.ccu.edu.tw/caac106/106apply_Sieve_3fg65egf3/html_sieve_5k8w/ColPost/extra/apply/151082N.htm
[Local] https://www.caac.ccu.edu.tw/caac106/106apply_Sieve_3fg65egf3/html_sieve_5k8w/ColPost/extra/apply/151102N.htm
[Local] https://www.caac.ccu.edu.tw/caac106/106apply_Sieve_3fg65egf3/html_sieve_5k8w/ColPost/extra/apply/151112N.htm
[Local] https://www.caac.ccu.edu.tw/caac106/106apply_Sieve_3fg65egf3/html_sieve_5k8w/ColPost/extra/apply/151122N.htm
[Local] https://www.caac.ccu.edu.tw/caac106/106apply_Sieve_3fg65egf3/html_sieve_5k8w/ColPost/extra/apply/151132N.htm
[Local] https://www.caac.ccu.edu.tw/caac106/106apply_Sieve_3fg65egf3/html_sieve_5k8w/ColPost/extra/apply/151142N.htm
[Local] https://www.caac.ccu.edu.tw/caac106/106apply_Sieve_3fg65egf3/html_sieve_5k8w/ColPost/extra/apply/151152N.htm
[Local] https://www.caac.ccu.edu.tw/caac106/106apply_Sieve_3fg65egf3/html_sieve_5k8w/ColPost/extra/apply/151162N.htm
[Local] https://www.caac.ccu.edu.tw/caac106/106apply_Sieve_3fg65egf3/html_sieve_5k8w/ColPost/extra/apply/151172N.htm
[Local] https://www.caac.ccu.edu.tw/caac106/106apply_Sieve_3fg65egf3/html_sieve_5k8w/ColPost/extra/apply/151182N.htm
[Local] https://www.caac.ccu.edu.tw/caac106/106apply_Sieve_3fg65egf3/html_sieve_5k8w/ColPost/extra/apply/151192N.htm
[Local] https://www.caac.ccu.edu.tw/caac106/106apply_Sieve_3fg65egf3/html_sieve_5k8w/ColPost/common/apply/152012.htm
[Local] https://www.caac.ccu.edu.tw/caac106/106apply_Sieve_3fg65egf3/html_sieve_5k8w/ColPost/common/apply/152022.htm
[Local] https://www.caac.ccu.edu.tw/caac106/106apply_Sieve_3fg65egf3/html_sieve_5k8w/ColPost/common/apply/152032.htm
[Local] https://www.caac.ccu.edu.tw/caac106/106apply_Sieve_3fg65egf3/html_sieve_5k8w/ColPost/common/apply/152042.htm
[Local] https://www.caac.ccu.edu.tw/caac106/106apply_Sieve_3fg65egf3/html_sieve_5k8w/ColPost/common/apply/152052.htm
[Local] https://www.caac.ccu.edu.tw/caac106/106apply_Sieve_3fg65egf3/html_sieve_5k8w/ColPost/extra/apply/152022N.htm
[crawler_caac] Finish crawling.
[crawler_caac] DB file is successfully generated.
[Done] It takes 12.177278518676758 seconds.
請按任意鍵繼續 . . .
微軟注音 半：
```

此過程視網路環境的不同，可能耗費 20 至 40 分鐘。

- iii. 如果想要完全重新抓取網站內容，可以先刪除 **crawler XXX** 資料夾
2. 修改 do_lookup.bat 中的參數設定
 - i. 對 do_lookup.bat 按 右鍵→編輯
 - ii. 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 @ECHO OFF
2
3 REM The department ID of NTHU EE "Jia" Group
4 SET department_NthuEe_Jia=011322
5 REM The department ID of NTHU EE "Yi" Group
6 SET department_NthuEe_Yi=011332
7
8 python lookup.py --outputFormat="NthuEe" --output="NTHU-EE-A.csv" --departmentIds=%department_NthuEe_Jia%
9 python lookup.py --outputFormat="NthuEe" --output="NTHU-EE-B.csv" --departmentIds=%department_NthuEe_Yi%
10 python lookup.py --outputFormat="NthuEe" --output="NTHU-EE-AB.csv" --departmentIds=%department_NthuEe_Jia%,%
    department_NthuEe_Yi%
11
12 PAUSE
13

```

修改此為當年度 "清大電機 甲組" 的系所編號

修改此為當年度 "清大電機 乙組" 的系所編號

- iii. 儲存並關閉檔案
3. 從資料庫取出資料
 - i. 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。

- ii. 產生的三個 .csv 檔案即為結果

.git	2017/4/6 下午 12:21	檔案資料夾	
crawler_106	2017/4/6 下午 01:54	檔案資料夾	
libs	2017/4/6 下午 12:07	檔案資料夾	
mylibs	2017/4/6 下午 12:08	檔案資料夾	
.gitignore	2017/4/6 下午 12:07	文字文件	1 KB
crawler.py	2017/4/6 下午 12:07	Python File	1 KB
do_crawl.bat	2017/4/6 下午 12:07	Windows 批次檔案	1 KB
do_lookup.bat	2017/4/6 下午 12:07	Windows 批次檔案	1 KB
lookup.py	2017/4/6 下午 12:07	Python File	4 KB
NTHU-EE-A.csv	2017/4/6 下午 02:20	Microsoft Excel 逗點分隔值檔案	8 KB
NTHU-EE-AB.csv	2017/4/6 下午 02:20	Microsoft Excel 逗點分隔值檔案	38 KB
NTHU-EE-B.csv	2017/4/6 下午 02:20	Microsoft Excel 逗點分隔值檔案	35 KB
使用說明.doc	2017/4/6 下午 01:47	Microsoft Word 97 - 2003 文件	96 KB

- 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