



國立清華大學

電機工程學系

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

目錄

| | | |
|----|--|---|
| 一、 | 環境安裝..... | 2 |
| | 1. 依照作業系統的位元數下載 Python 3.6 或更高版本 | 2 |
| | 2. 安裝 Python | 2 |
| | 3. 雙擊 install_requirements.bat 安裝所需的 Python 套件 | 2 |
| 二、 | 第一階段-篩選結果（甄選委員會） | 3 |
| | 1. 修改 do_crawl.bat 中的參數設定 | 3 |
| | 2. 抓取網站內容並建立本地資料庫（只需進行一次） | 4 |
| | 3. 修改 do_lookup.bat 中的參數設定 | 4 |
| | 4. 從資料庫取出資料 | 5 |
| 三、 | 第二階段-交叉查榜 | 6 |
| | 1. 將要查詢的校系代碼寫入 department_ids.txt 中，然後儲存並關閉檔案 | 6 |
| | 2. 雙擊 do_cross.bat | 6 |
| | 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 |
| | 4. result_xxx.xlsx 即為結果 | 6 |
| 四、 | 第二階段-分發結果（甄選委員會） | 7 |
| | 1. 修改 do_crawl.bat 中的參數設定 | 7 |
| | 2. 抓取網站內容並建立本地資料庫（只需進行一次） | 8 |
| | 3. 將要查詢的准考證號寫入 admission_ids.txt 中，然後儲存並關閉檔案 | 8 |
| | 4. 從資料庫取出資料 | 8 |

一、環境安裝

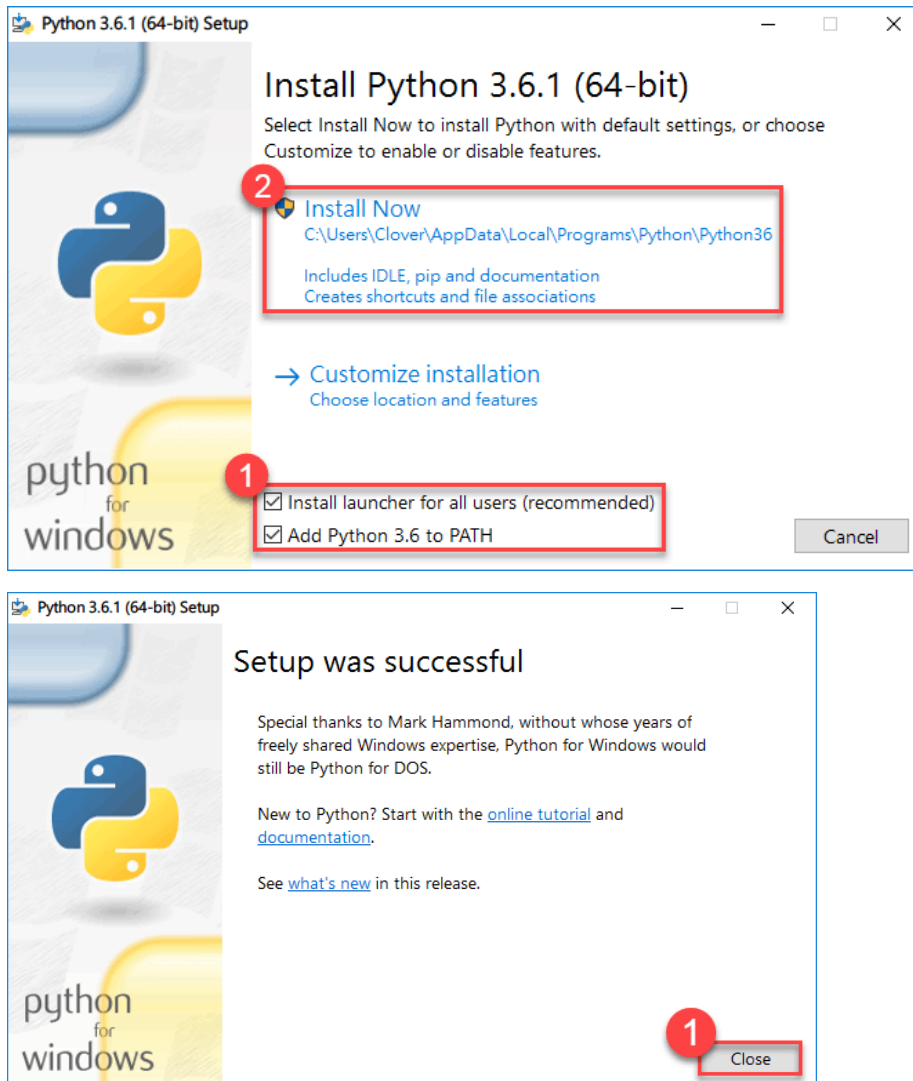
● Python 3

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。

2. 安裝 Python



3. 雙擊 install_requirements.bat 安裝所需的 Python 套件

二、 第一階段-篩選結果（甄選委員會）

1. 修改 do_crawl.bat 中的參數設定
 - i. 對 do_crawl.bat 按 右鍵→編輯
 - ii. 打開網頁瀏覽器，找到當年度 第一階段篩選結果 的網頁



108 學年度大學個人申請入學招生 第一階段篩選結果

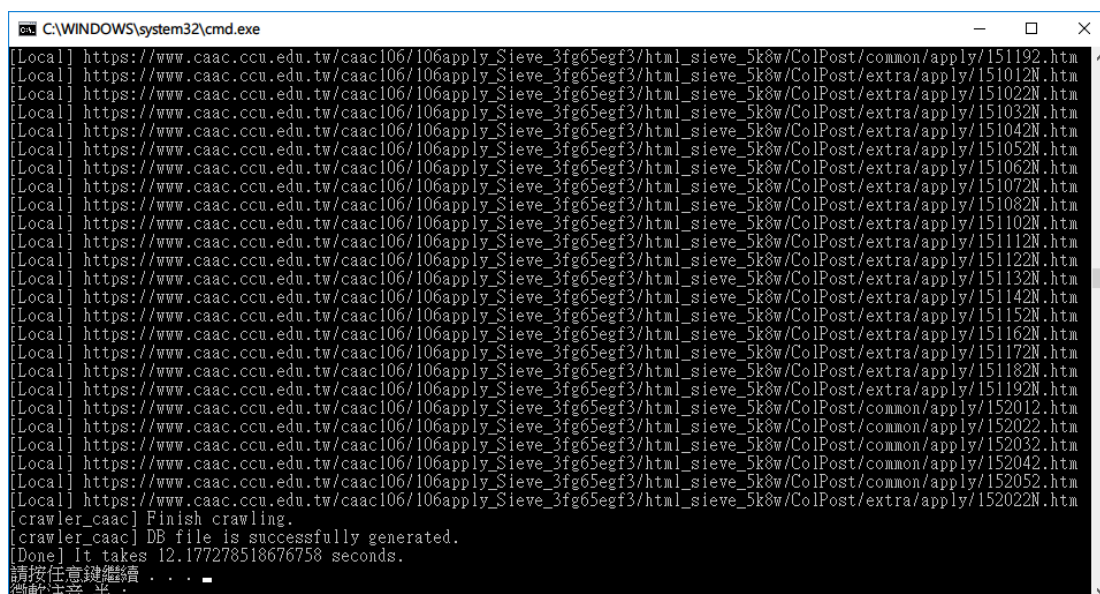
| | | | | |
|-------------|--------------|---------------|--------------|-------------|
| (001)國立臺灣大學 | 招生名額 外加名額 | (002)國立臺灣師範大學 | 招生名額 外加名額 | (003)國立中興大學 |
| (004)國立成功大學 | 招生名額 外加名額 | (005)東吳大學 | 招生名額 外加名額 | (006)國立政治大學 |

- iii. 複製該網頁的網址，貼上到 do_crawl.bat 中的 **projectBaseUrl**

- iv. 儲存並關閉檔案

2. 抓取網站內容並建立本地資料庫（只需進行一次）

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



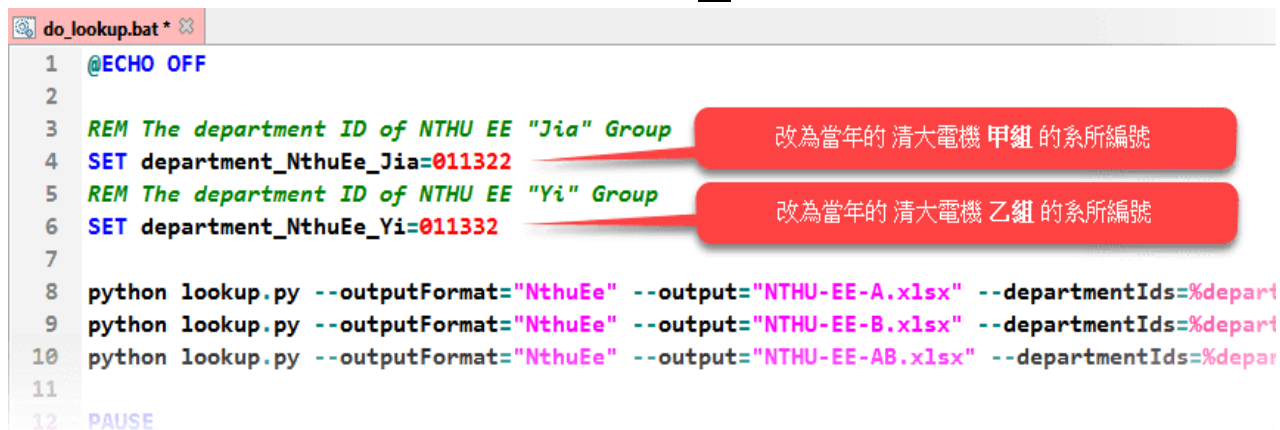
```
C:\WINDOWS\system32\cmd.exe
[Local] https://www.caac.ccu.edu.tw/caac106/106apply_Sieve_3fg65egf3/html_sieve_5k8w/ColPost/common/apply/151192.htm
[Local] https://www.caac.ccu.edu.tw/caac106/106apply_Sieve_3fg65egf3/html_sieve_5k8w/ColPost/extra/apply/151012N.htm
[Local] https://www.caac.ccu.edu.tw/caac106/106apply_Sieve_3fg65egf3/html_sieve_5k8w/ColPost/extra/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 2 至 5 分鐘。

- 如果想要完全重新抓取網站內容，可以先刪除 **`data/crawler XXX`** 資料夾

3. 修改 `do_lookup.bat` 中的參數設定

- 對 `do_lookup.bat` 按 右鍵→編輯
- 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）



```
do_lookup.bat *
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.xlsx" --departmentIds=%department_NthuEe_Jia%
9 python lookup.py --outputFormat="NthuEe" --output="NTHU-EE-B.xlsx" --departmentIds=%department_NthuEe_Yi%
10 python lookup.py --outputFormat="NthuEe" --output="NTHU-EE-AB.xlsx" --departmentIds=%department_NthuEe_Jia%
11
12 PAUSE
```

改為當年的 清大電機 甲組 的系所編號

改為當年的 清大電機 乙組 的系所編號

- 儲存並關閉檔案

4. 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 .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

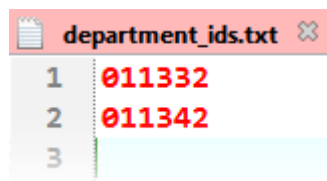
| | | | |
|-----------------|---------------------|----------------------|-------|
| mylibs | 2017/4/20 下午 09:... | 檔案資料夾 | |
| crawler.py | 2017/4/19 下午 04:... | Python File | 1 KB |
| do_crawl.bat | 2017/4/12 下午 04:... | Windows 批次檔案 | 1 KB |
| do_lookup.bat | 2017/4/20 下午 10:... | Windows 批次檔案 | 1 KB |
| lookup.py | 2017/4/20 下午 10:... | Python File | 4 KB |
| NTHU-EE-A.xlsx | 2017/4/20 下午 10:... | Microsoft Excel 工... | 7 KB |
| NTHU-EE-AB.xlsx | 2017/4/20 下午 10:... | Microsoft Excel 工... | 14 KB |
| NTHU-EE-B.xlsx | 2017/4/20 下午 10:... | Microsoft Excel 工... | 14 KB |

| | A | B | C | D | E |
|---|----------|------------------|---------------------|-------------------|------------------|
| 1 | 准考證號 | 校名與系所 | | | |
| 2 | 10000903 | 國立臺灣大學 化學工程學系 | 國立臺灣大學 材料科學與工程學系 | 國立臺灣大學 財務金融學系 | 中山醫學大學 醫學系 |
| 3 | 10004236 | 國立臺灣大學 機械工程學系 | 國立臺灣大學 材料科學與工程學系 | 國立清華大學 物理學系物理組 | |
| 4 | 10004334 | 國立臺灣大學 機械工程學系 | 國立臺灣大學 資訊工程學系 | 國立交通大學 電子工程學系 | 國立交通大學 電子工程學系 |
| 5 | 10004405 | 國立臺灣大學 機械工程學系 | 國立臺灣大學 電機工程學系 | 國立成功大學 機械工程學系 | 國立交通大學 電機工程學系 |
| | | 國立臺灣大學 | 國立臺灣大學 | 國立交通大學 | 國立清華大學 |

三、 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，然後儲存並關閉檔案



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。

2. 雙擊 do_cross.bat
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
4. result_xxx.xlsx 即為結果

其內容以准考證號碼排序，若一名考生出現在多個校系代碼的結果中，在輸出的檔案裡只會出現一次。

※備註：由於學校放榜時可能會做個資保護，因此網路上的資料可能不齊全，某些人的放榜狀態會是「不明」（可能是落榜）。

| | A | B | C | D | E | F | G | H | I |
|---|----------|------|----------------------|-----------------------|------|----------------------|-----|-----------------------|------|
| 1 | 准考證號 | 考生姓名 | 分發結果 | 校系名稱 | 榜單狀態 | | | | |
| 2 | 10001832 | | 國立交通大學 電機工程學系(乙組) | 國立臺灣大學 工商管理學系科技管理組 | 正 | 國立交通大學 電子工程學系(乙組) | 備75 | 國立交通大學 電機工程學系(乙組) | 備103 |
| 3 | 10004829 | | 國立清華大學 電機工程學系乙組 | 國立成功大學 電機工程學系 | 備62 | 國立中央大學 電機工程學系 | 正12 | 國立臺灣科技大學 資訊工程系 | 不明 |
| 4 | 10007501 | | 國立臺灣大學 機械工程學系 | 國立臺灣大學 機械工程學系 | 備6 | 國立臺灣大學 生物產業機電工程學系 | 不明 | 國立交通大學 資訊工程學系資訊工程組 | 備1 |

四、 第二階段-分發結果（甄選委員會）

※備註：通常，只有當 第二階段-交叉查榜 無法使用時（像是來源網站已經倒閉）才需要使用此方式。此方式為從「甄選委員會」的官方網站中抓取資料，因此僅能查詢「統一分發結果」。

1. 修改 do_crawl.bat 中的參數設定
 - i. 對 do_crawl.bat 按 右鍵→編輯
 - ii. 打開網頁瀏覽器，找到當年度 統一分發結果 的網頁



108 學年度大學個人申請入學招生 統一分發結果

| | | | | | |
|-------------|--------------|---------------|--------------|-------------|----------|
| (001)國立臺灣大學 | 招生名額 外加名額 | (002)國立臺灣師範大學 | 招生名額 外加名額 | (003)國立中興大學 | 招生 外加 |
| (004)國立成功大學 | 招生名額 外加名額 | (005)東吳大學 | 招生名額 外加名額 | (006)國立政治大學 | 招生 外加 |
| (007)高雄醫學大學 | 招生名額 | (008)中原大學 | 招生名額 | (009)東海大學 | 招生 外加 |

- iii. 複製該網頁的網址，貼上到 do_crawl.bat 中的 projectBaseUrl

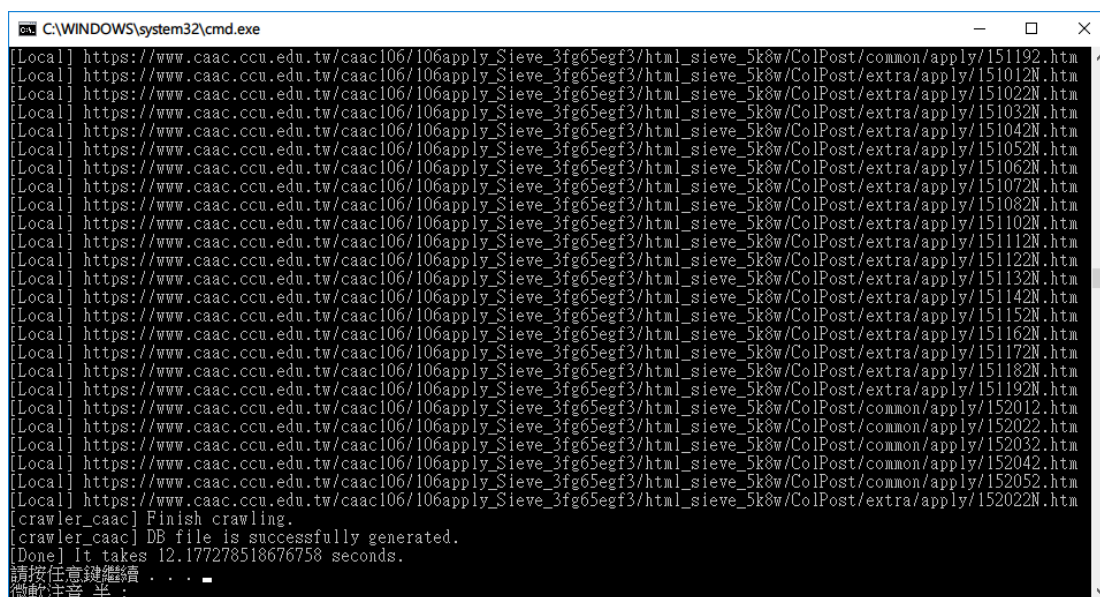
```
do_crawl.bat
1 @ECHO OFF
2
3 REM The (base) URL of the CAAC HTML page
4 SET projectBaseUrl="https://www.cac.edu.tw/CacLink/
5   apply108/108apply_kg_Entrance_fq6ryh7/
6   html_entrance_kk108nonp/result_html/result_apply/
7   collegeList.htm"
8
9 python crawler.py
10
11 PAUSE
```

修改為 統一分發結果 首頁的網址

- iv. 儲存並關閉檔案

2. 抓取網站內容並建立本地資料庫（只需進行一次）

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

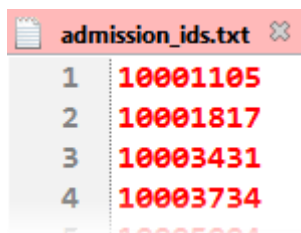


```
C:\WINDOWS\system32\cmd.exe
[Local] https://www.caac.ccu.edu.tw/caac106/106apply_Sieve_3fg65egf3/html_sieve_5k8w/ColPost/common/apply/151192.htm
[Local] https://www.caac.ccu.edu.tw/caac106/106apply_Sieve_3fg65egf3/html_sieve_5k8w/ColPost/extra/apply/151012N.htm
[Local] https://www.caac.ccu.edu.tw/caac106/106apply_Sieve_3fg65egf3/html_sieve_5k8w/ColPost/extra/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 2 至 5 分鐘。

- 如果想要完全重新抓取網站內容，可以先刪除 **`data/crawler XXX`** 資料夾

3. 將要查詢的准考證號寫入 `admission_ids.txt` 中，然後儲存並關閉檔案



如左圖，一行寫一個准考證號。

圖中範例為 108 年清大電機甲、乙組考生的准考證號。

4. 從資料庫取出資料

- 雙擊 `do_lookup.bat`
- `result_xxx.xlsx` 即為結果

其內容以准考證號碼排序，若一名考生出現在多個校系代碼的結果中，在輸出的檔案裡只會出現一次。

| | A | B |
|---|----------|----------------------|
| 1 | 准考證號 | 分發結果 |
| 2 | 10001105 | 國立清華大學 材料科學工程學系 |
| 3 | 10001817 | 國立交通大學 電機工程學系(乙組) |
| 4 | 10003431 | |
| 5 | 10003734 | 國立臺灣大學 化學工程學系 |
| 6 | 10005004 | 國立交通大學 資訊工程學系(甲組) |