

# 使用說明書

## CSR REPORT CRAWLER v1.0

公開資訊觀測站 CSR 報告書自動下載工具

更 新 日 期 ： 2 0 2 3 年 9 月 1 0 日

# 目錄

1.	簡介 .....	1
2.	安裝說明 .....	1
3.	主要功能介紹與操作指南.....	4
4.	常見問題 .....	10
5.	聯絡資訊 .....	11

## 1. 簡介

歡迎使用 CSR Report Crawler，「公開資訊觀測站 CSR 報告書自動下載工具」。本說明書旨在提供使用者一個快速而簡易的指南，幫助您在各種 Windows 電腦環境中無縫地下載 CSR 報告書。不再需要繁瑣的手動操作，只需簡單幾步，即可完成下載。

## 2. 安裝說明

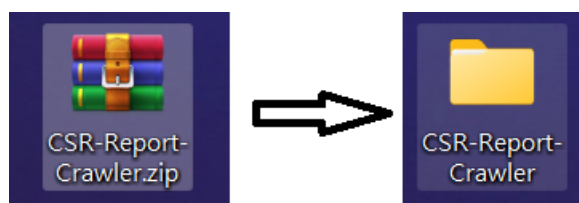


圖 2.1. 下載檔案並解壓縮

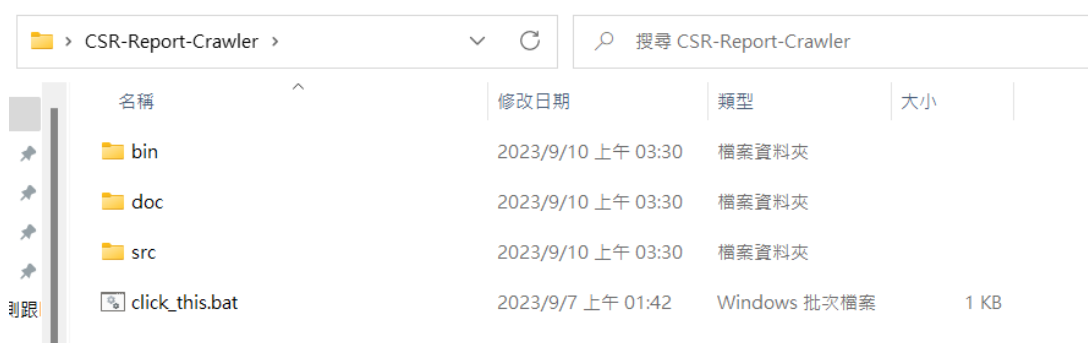


圖 2.2. 下載檔案內容

如圖 2.2 所示，解壓縮後的資料夾中包含三個子資料夾和一個.bat 檔案。其中，doc 資料夾用於存放文件。在開始安裝程式前，請確保 bin 和 src 資料夾的檔案都是完整的。

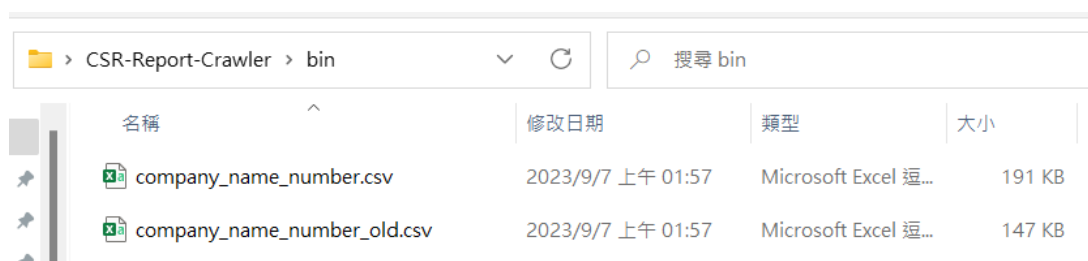


圖 2.3. bin 資料夾須包含 company\_name\_number.csv

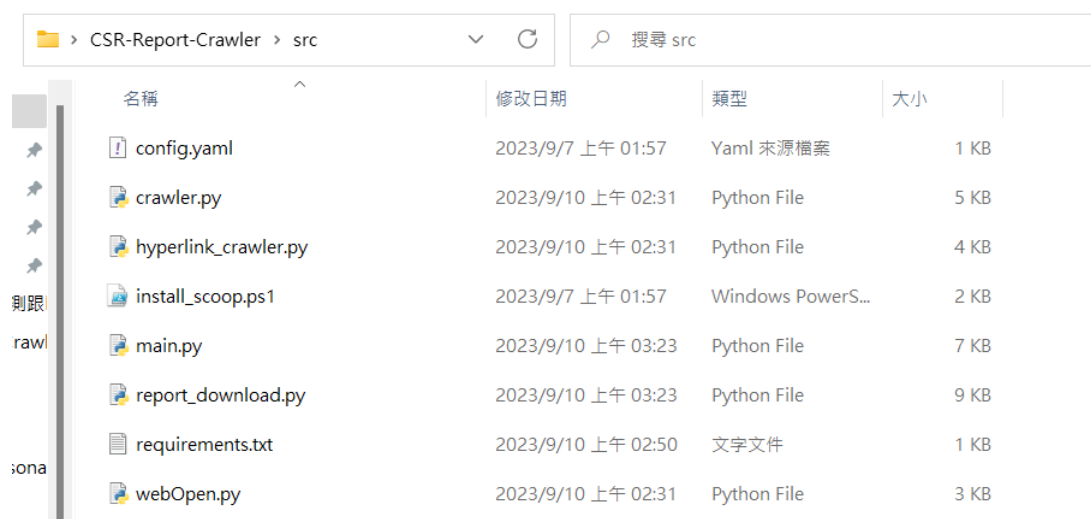


圖 2.4. src 資料夾須包含以上八個檔案

在確認檔案完整無誤後，請返回根目錄並點擊 click\_this.bat，以開始安裝 CSR Report Crawler 程式及其所需的運行環境。

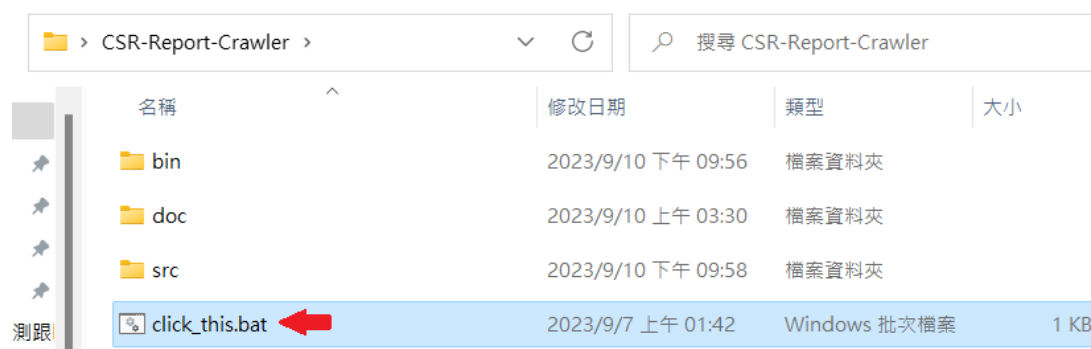


圖 2.5. 雙擊 click\_this.bat



圖 2.6. 按下「是」

請授權此 App 修改裝置設定，以便安裝所需的執行環境。

```
Set-ExecutionPolicy : windows PowerShell updated your execution policy successfully, but the setting is overridden by a
policy defined at a more specific scope. Due to the override, your shell will retain its current effective execution
policy of Bypass. Type "Get-ExecutionPolicy -List" to view your execution policy settings. For more information please
see "Get-Help Set-ExecutionPolicy".
At C:\Users\x0180368905\Desktop\CSR-Report-Crawler\src\install_scoop.ps1:2 char:1
+ Set-ExecutionPolicy -ExecutionPolicy RemoteSigned -Scope CurrentUser ...
+ ~~~~~
+ CategoryInfo          : PermissionDenied: (:) [Set-ExecutionPolicy], SecurityException
+ FullyQualifiedErrorId : ExecutionPolicyOverride,Microsoft.PowerShell.Commands.SetExecutionPolicyCommand
Bat Directory: C:\Users\x0180368905\Desktop\CSR-Report-Crawler
Absolute Requirements Path: C:\Users\x0180368905\Desktop\CSR-Report-Crawler\src\requirements.txt
Absolute Main Path: C:\Users\x0180368905\Desktop\CSR-Report-Crawler\src\main.py
Initializing...
Downloading ...
```

圖 2.7. 載入需要的套件

Windows 的 PowerShell 將會自動開啟，並開始載入程式所需的套件。

```
Set-ExecutionPolicy : windows PowerShell updated your execution policy successfully, but the setting is overridden by a
policy defined at a more specific scope. Due to the override, your shell will retain its current effective execution
policy of Bypass. Type "Get-ExecutionPolicy -List" to view your execution policy settings. For more information please
see "Get-Help Set-ExecutionPolicy".
At C:\Users\x0180368905\Desktop\CSR-Report-Crawler\src\install_scoop.ps1:2 char:1
+ Set-ExecutionPolicy -ExecutionPolicy RemoteSigned -Scope CurrentUser ...
+ ~~~~~
+ CategoryInfo          : PermissionDenied: (:) [Set-ExecutionPolicy], SecurityException
+ FullyQualifiedErrorId : ExecutionPolicyOverride,Microsoft.PowerShell.Commands.SetExecutionPolicyCommand
Bat Directory: C:\Users\x0180368905\Desktop\CSR-Report-Crawler
Absolute Requirements Path: C:\Users\x0180368905\Desktop\CSR-Report-Crawler\src\requirements.txt
Absolute Main Path: C:\Users\x0180368905\Desktop\CSR-Report-Crawler\src\main.py
Initializing...
Downloading ...
Creating shim...
Scoop was installed successfully!
Type 'scoop help' for instructions.
Installing 'dark' (3.11.2) [64bit] from main bucket
dark-3.11.2.zip (3.5 MB) [=====] 100%
Checking hash of dark-3.11.2.zip ... ok.
Extracting dark-3.11.2.zip ... done.
Linking C:\Scoop\apps\dark\current => C:\Scoop\apps\dark\3.11.2
Creating shim for 'dark'.
python-3.11.5-amd64 was installed successfully
Installing 'python' (3.11.5) [64bit] from main bucket
python-3.11.5-amd64.exe (24.7 MB) [=====] 100%
Checking hash of python-3.11.5-amd64.exe ... ok.
Running pre_install script...
Linking C:\Scoop\apps\python\current => C:\Scoop\apps\python\3.11.5
Creating shim for 'python3'.
Creating shim for 'idle'.
Creating shim for 'idle3'.
Persisting Scripts
Persisting Lib\site-packages
Running post_install script...
```

圖 2.8. 等待環境安裝

當您第一次開啟程式時，由於需要進行安裝，所以開啟的時間較長。但之後每次開啟，程式只會快速檢查更新後立即啟動。

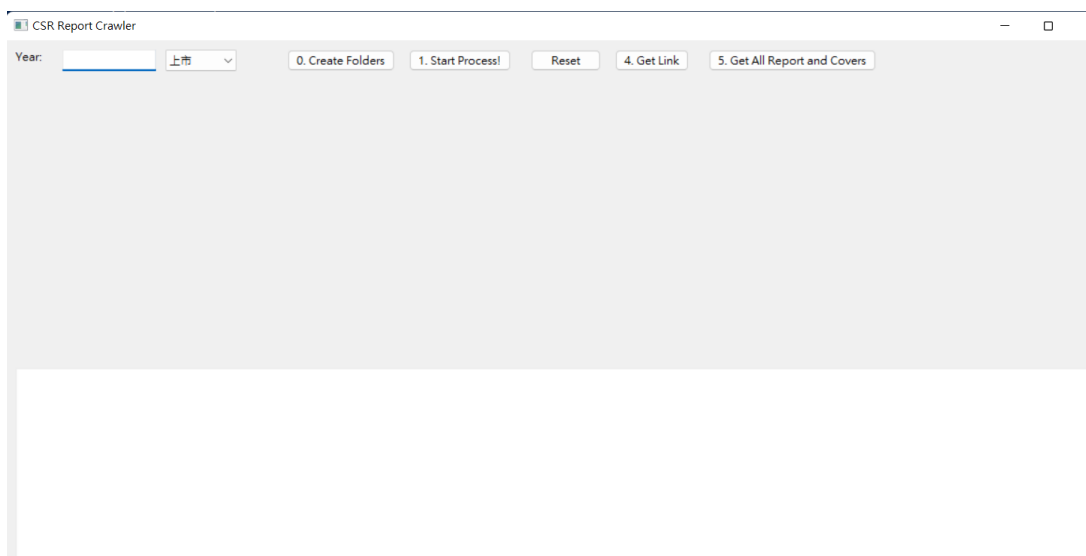


圖 2.9. 安裝完畢後，CSR Report Crawler 程式視窗會自動跳出

### 3. 主要功能介紹與操作指南

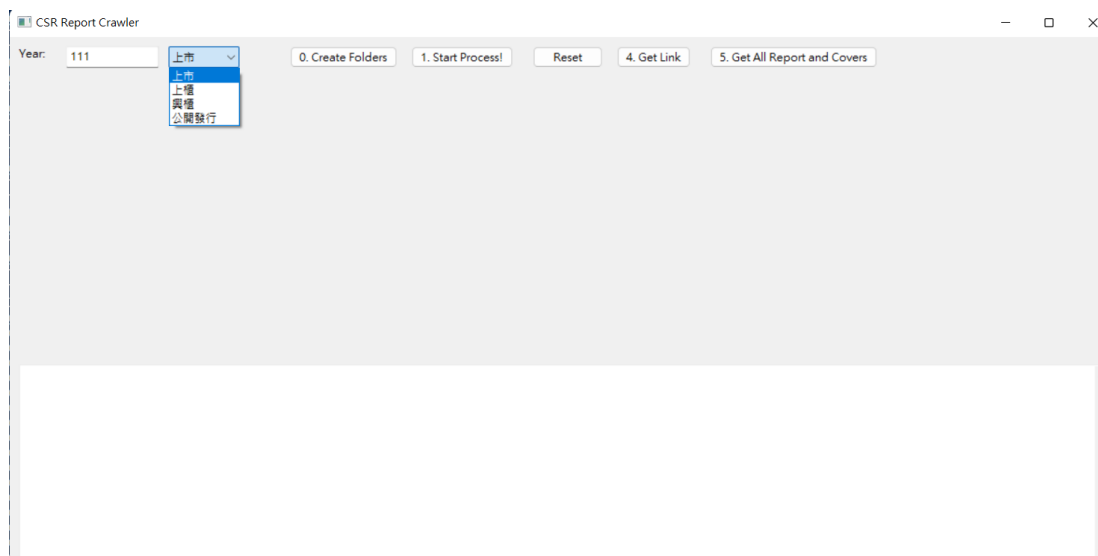


圖 3.1. 輸入條件，點擊“0. Create Folders”按鈕以創建儲存資料夾

在"Year"輸入框中，您應該輸入公司統整資料的民國年份。例如，通常民國 112 年出版的 CSR Report 主要是整理民國 111 年的資料，因此您應在此輸入框中填入「111」。右側的下拉選單允許您選擇想要下載的公司報告書的類型。

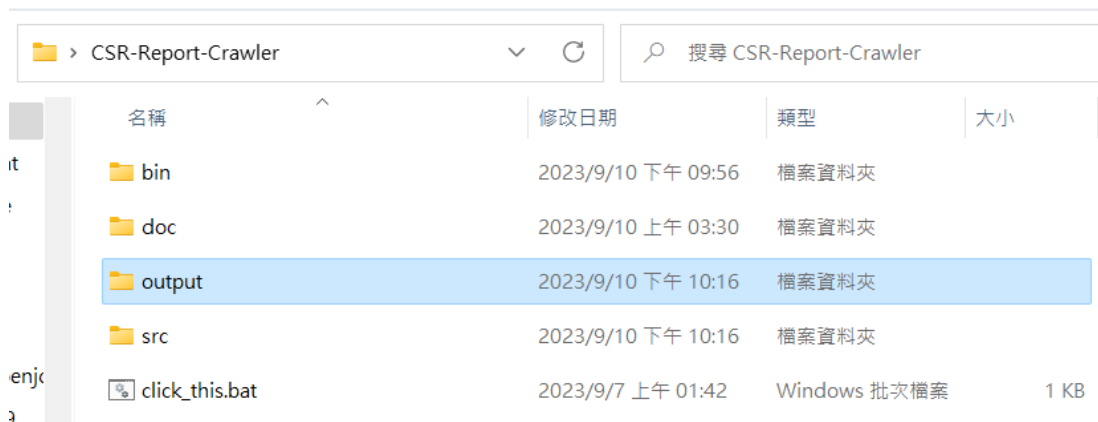


圖 3.2. 根目錄會出現一個 output 資料夾



圖 3.3. 根據創建時間和欲下載的公司類型來命名資料夾

資料夾之名字會根據此規則命名：YYMMDD\_hhmm\_公司種類。

如圖 3.3 所示，「230910\_2218\_上市」代表於西元 20「23」年的「09」月「10」日，「22」點「18」分時，創建的「上市」公司資料夾。



圖 3.4. 當中會包含三個子資料夾

「jpg」資料夾用於存放報告書的封面；「pdf」資料夾將存放已下載的報告書；而「table」資料夾將儲存從公開資訊觀測站爬取的整張表格。此表格會合併報告書的下載連結、公司的完整名稱，以及公司的統編資訊，然後保存成一個單一檔案。

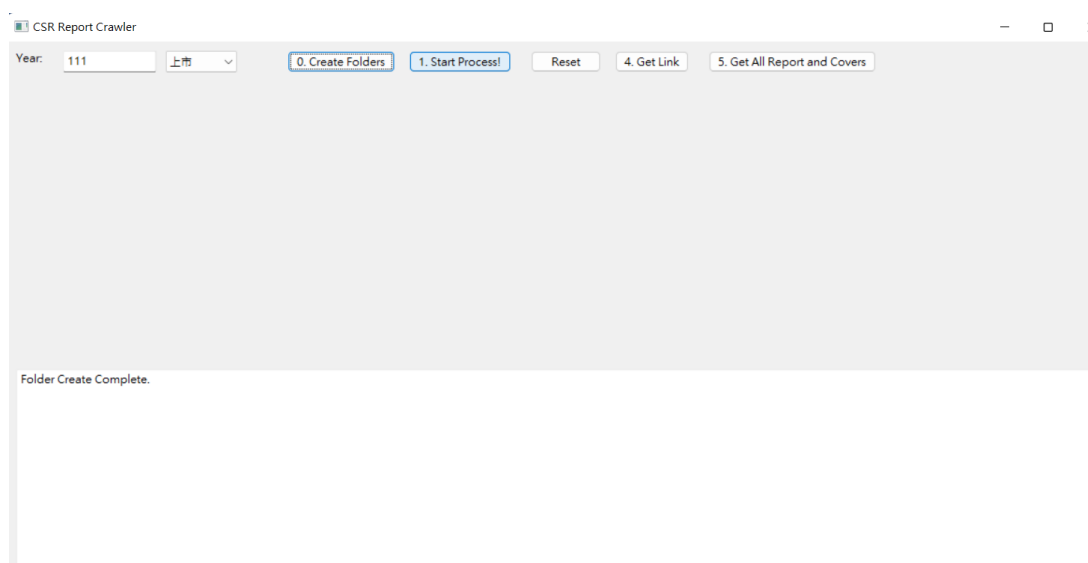


圖 3.5. 資料夾創建完會出現” Folders Create Complete”

資料夾建立完畢後，請點擊“1. Start Process”按鈕。系統會根據條件啟動程式。



圖 3.6. Chrome 受到自動測試軟體控制

當 Chrome 視窗自動開啟，它會開始根據您輸入的條件搜尋資料。接著，系統將依照您的設定從公開資訊觀測站爬取公司的資訊。在 Chrome 視窗自動開啟的過程中，請儘量避免讓鼠標停留在視窗上，因為這可能會導致部分網頁元件被觸發，從而影響網頁的結構。

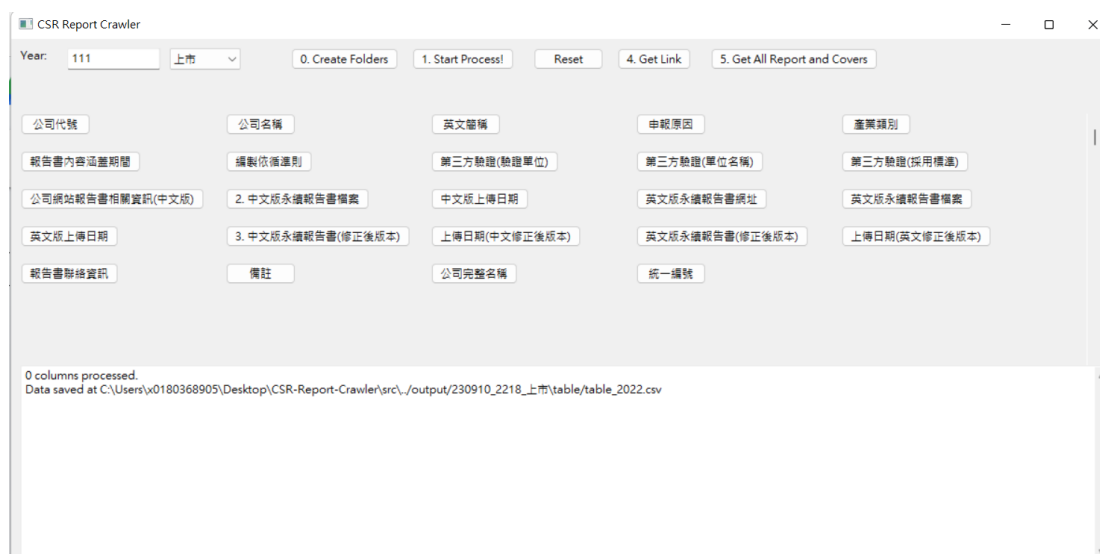


圖 3.7. 出現對應公開資訊觀測站欄位的選項按鈕



當 Chrome 藉由自動測試控制軟體完成資料爬取後，它會自動關閉。爬取的資料會暫時儲存在“output”資料夾內的“table”子資料夾中，且其儲存路徑將會顯示在下方的文字窗口。請您暫時不要開啟這個表格檔，因為後續我們會合併報告書的連結並覆蓋此檔案。

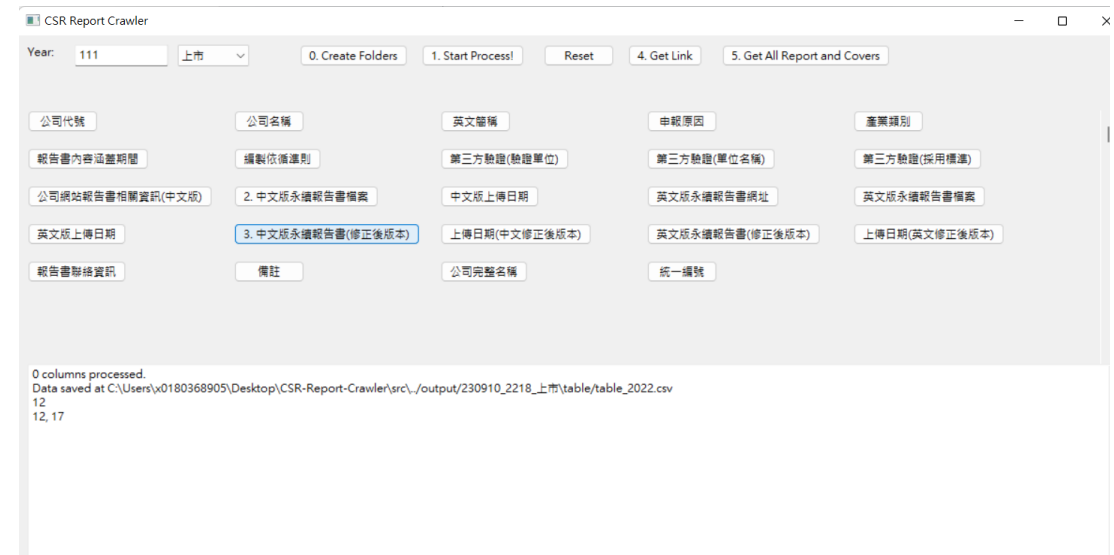


圖 3.8. 根據提示，依次按下 2 號、3 號及 4 號按鈕

在已爬取的資料中，您需要下載中文版的永續報告書位於「中文版永續報告書檔案」和「中文版永續報告（修正後版本）」這兩個欄位中。按提示順序點擊這兩按鈕，使系統能提取超連結。如超連結位置改變，點選相應欄位按鈕。最後再點按鈕“4. Get Link”使系統取得報告連結。



圖 3.9. Chrome 視窗再度出現

當提取報告書連結時，Chrome 視窗將重新開啟。請稍候幾秒，系統將自動載入報告書的連結，完成後視窗將自動關閉。

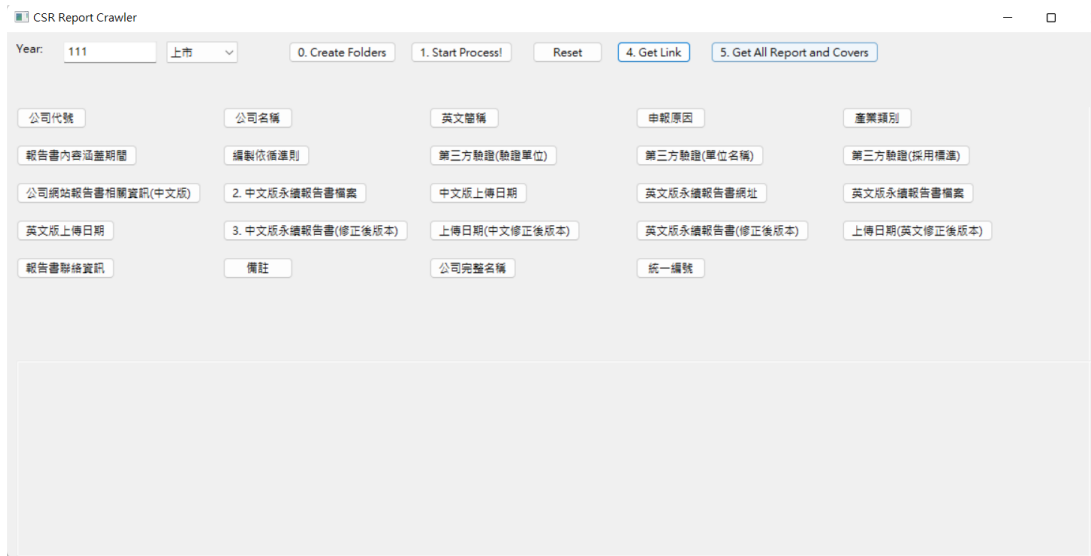


圖 3.10. 當連結下載完畢後，下方的文字框可能會短暫消失，這是正常的

當連結下載完成後，直接點擊“5. Get All Reports and Covers”按鈕，系統便會開始下載報告書及其封面。

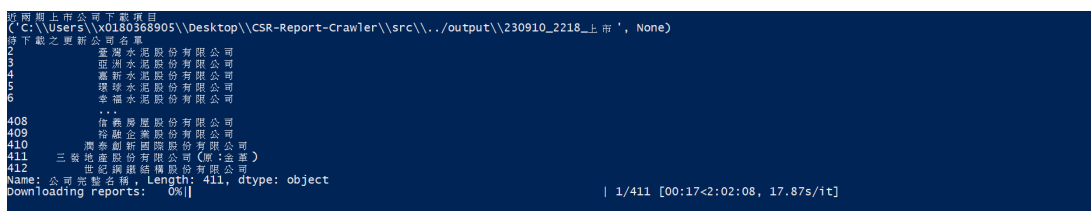


圖 3.11. 終端機會顯示進度條，可查看當前的進度

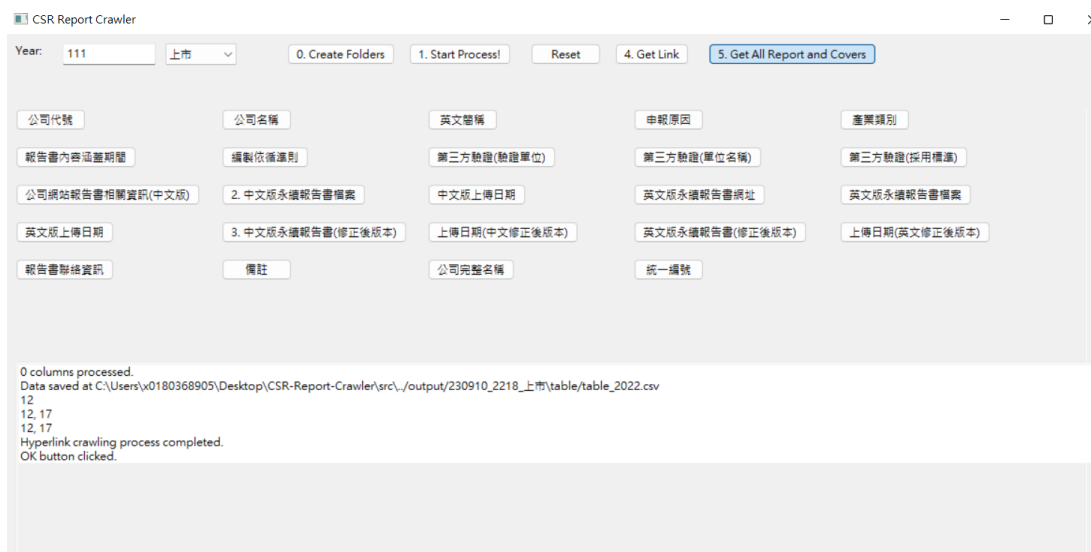


圖 3.12. 下載完報告書與封面後，文字框會顯示如上

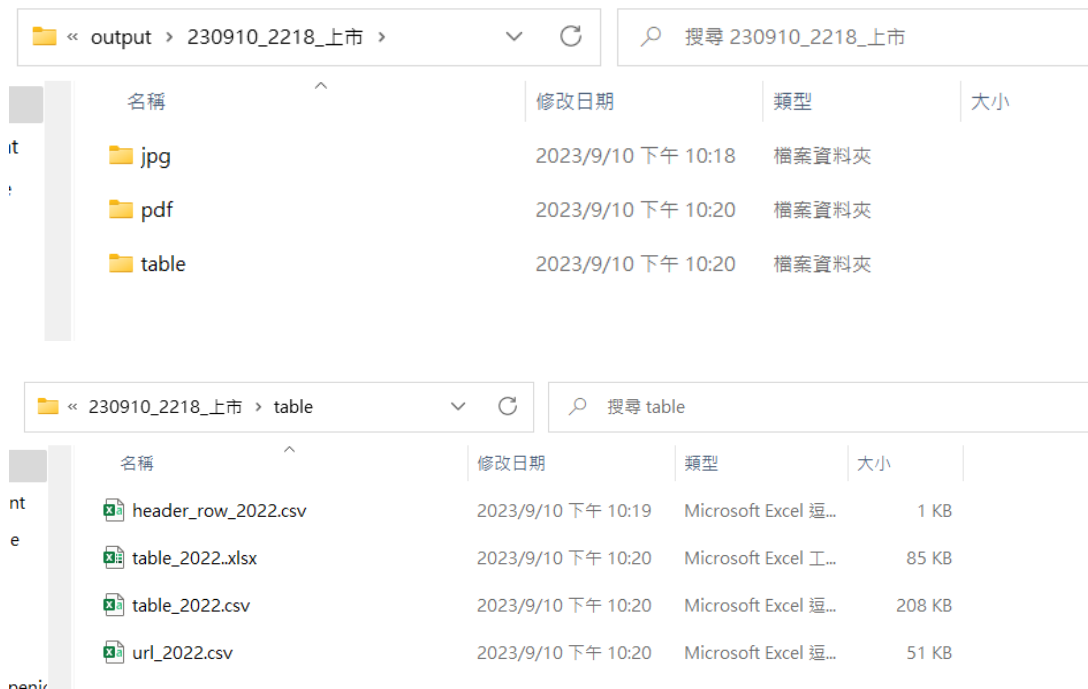


圖 3.13. 完整的公司資料、統編、報告書連結會被存在 table 資料夾

- “header\_row\_2022.csv”：儲存了 2022 年公開資訊觀測站中永續報告書的欄位資訊。
- ”table\_2022.csv”：記錄了 2022 年以 utf-8 編碼的上市公司完整表格。
- ”table\_2022.xlsx”：包含了 2022 年的上市公司表格，格式設定為可以直接打開而不會顯示亂碼。
- ”url\_2022.csv”：存放了 2022 年上市公司的報告書連結。

圖 3.14. table\_2022.xlsx

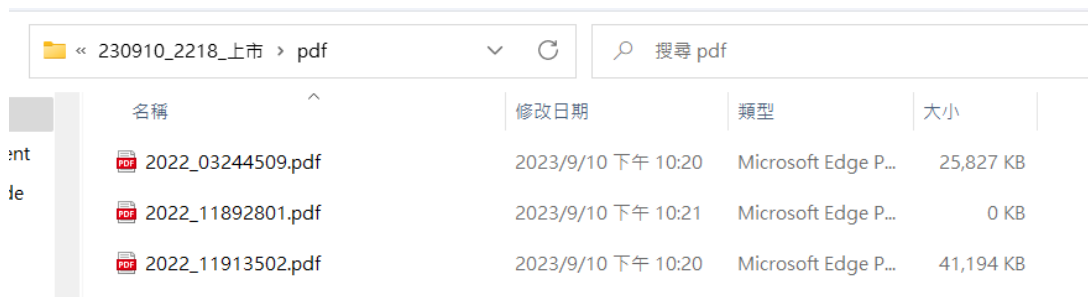


圖 3.15. 下載下來的報告書

所下載的報告書會被儲存於「上市」資料夾，下載下來的報告書檔名格式設定為「西元年\_公司統編」。

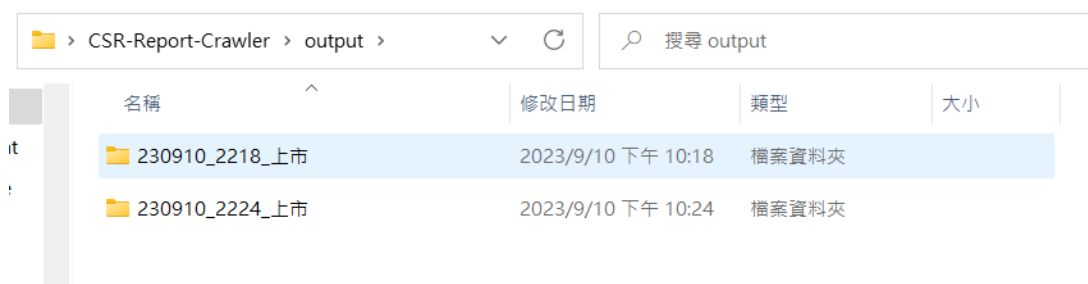


圖 3.16. 下載新增的報告書

若需要確認報告書是否有更新，請重新開啟 CSR Report Crawler 程式。每次操作，資料夾會按照當前時間自動創建一個新的。若選擇的公司類別已存在於 output 資料夾中，系統會自動與最近創建的同類型資料夾中的”table\_2022.csv”進行比對。系統只會針對新版本的表格中有，但先前版本中缺失的公司，進行下載。

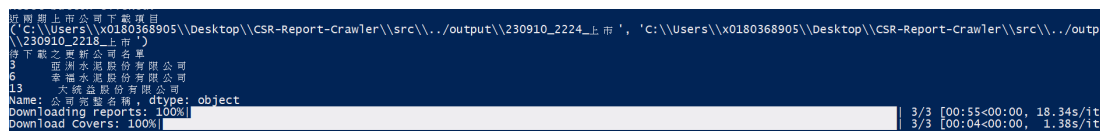


圖 3.17. 系統只會下載更新的報告書

終端機會展示即將下載的公司名單。如圖 3.11 所示，為首次下載時的進度條顯示；而圖 3.17 則顯示當進行更新時，終端機上的進度條狀態。

#### 4. 常見問題

若載入表格時發生錯誤，可以嘗試重新開啟程式。中斷後重新開啟程式需要完成下列步驟：

- 4.1. 關閉 CSR Report Crawler 系統視窗。
- 4.2. 關閉藍色終端機視窗。
- 4.3. 若 Chrome 的自動測試控制視窗仍在運行，請將其關閉。

4.4. 刪除此次使用系統時所建立的資料夾（其檔名命名格式為 YYMMDD\_hhmm\_公司類型）。

## 5. 聯絡資訊

開發人員：沈健好

信箱：110306010@nccu.edu.tw