

使用說明書

CSR REPORT CRAWLER v1.1

公開資訊觀測站 CSR 報告書自動下載工具

更 新 日 期 ： 2 0 2 4 年 2 月 1 6 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.	突發狀況之解決方法.....	9
5.	版本更新.....	11
6.	聯絡資訊.....	12

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。

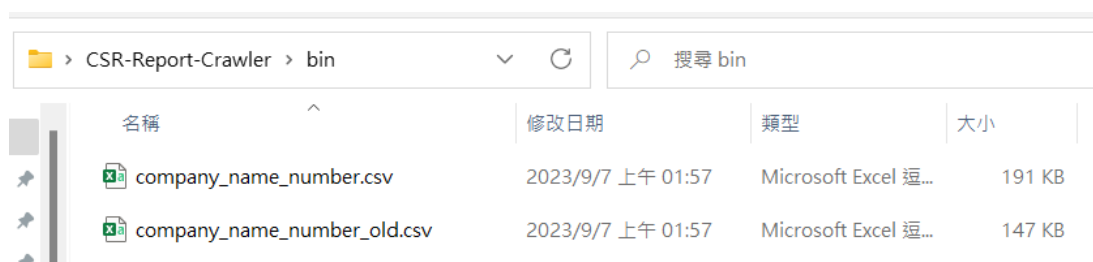


圖 2.3. bin 資料夾須包含 company_name_number.csv

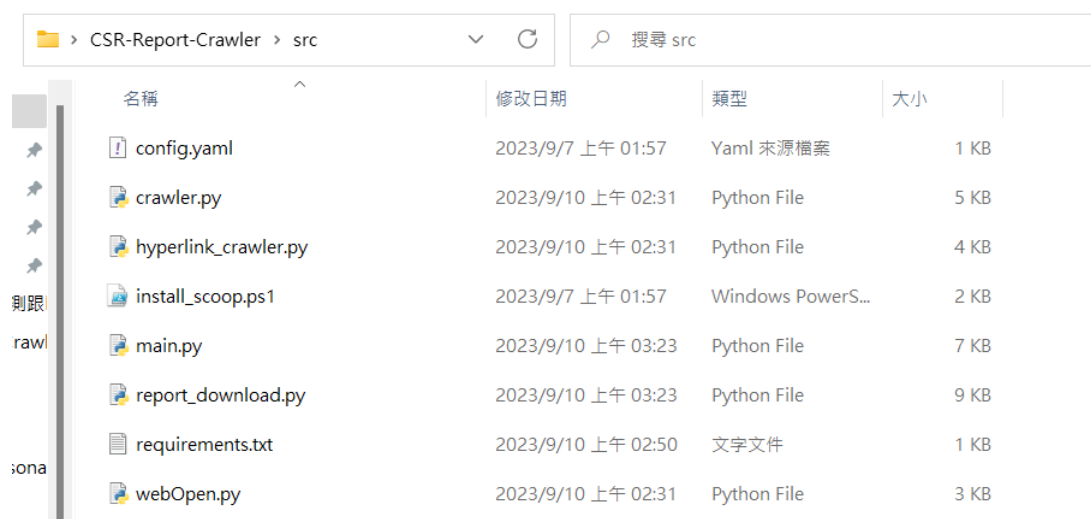


圖 2.4. src 資料夾須包含以上八個檔案

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。

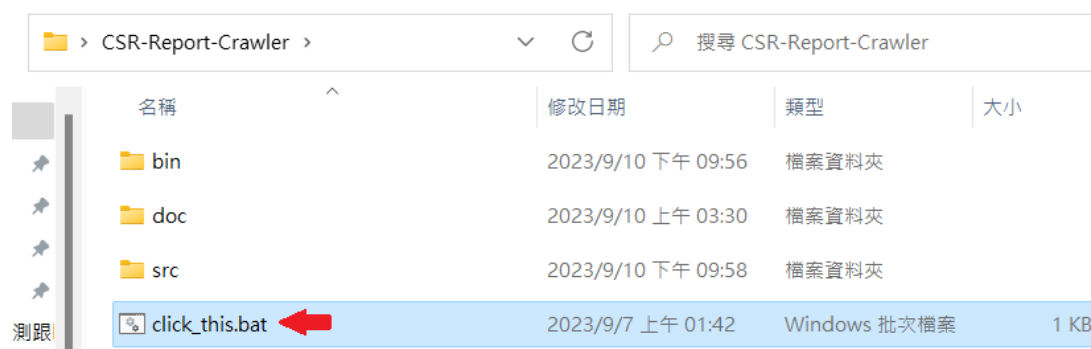


圖 2.5. 雙擊 click_this.bat



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。

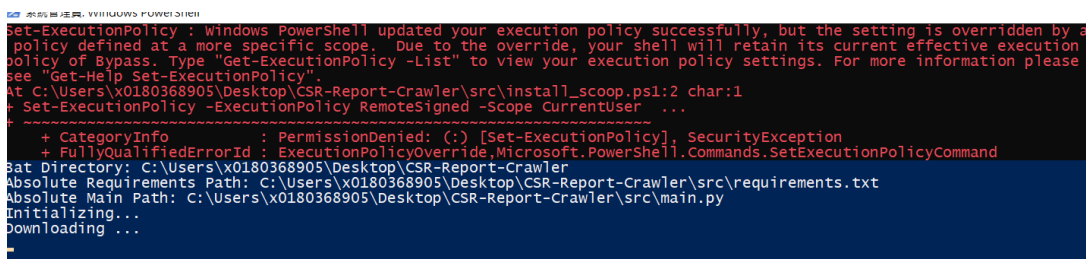


圖 2.7. 載入需要的套件

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

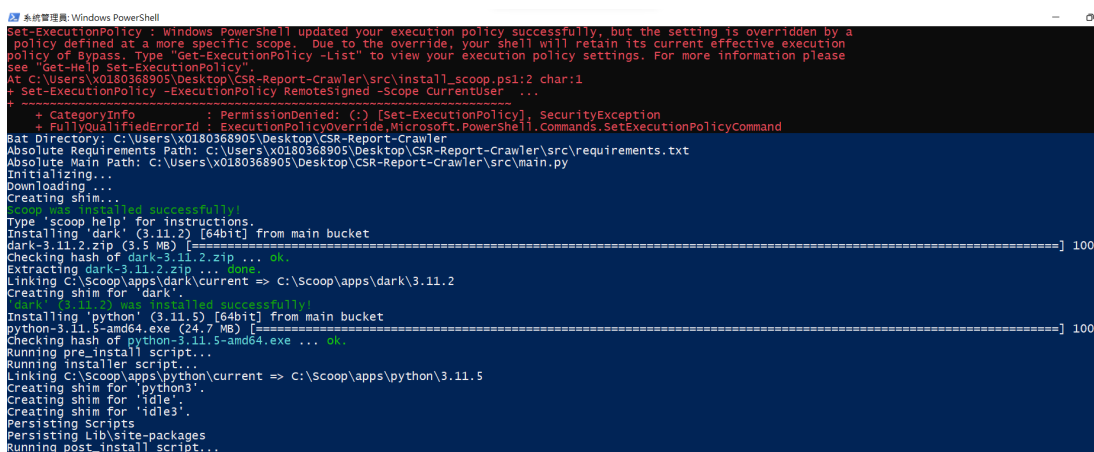


圖 2.8. 等待環境安裝

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。

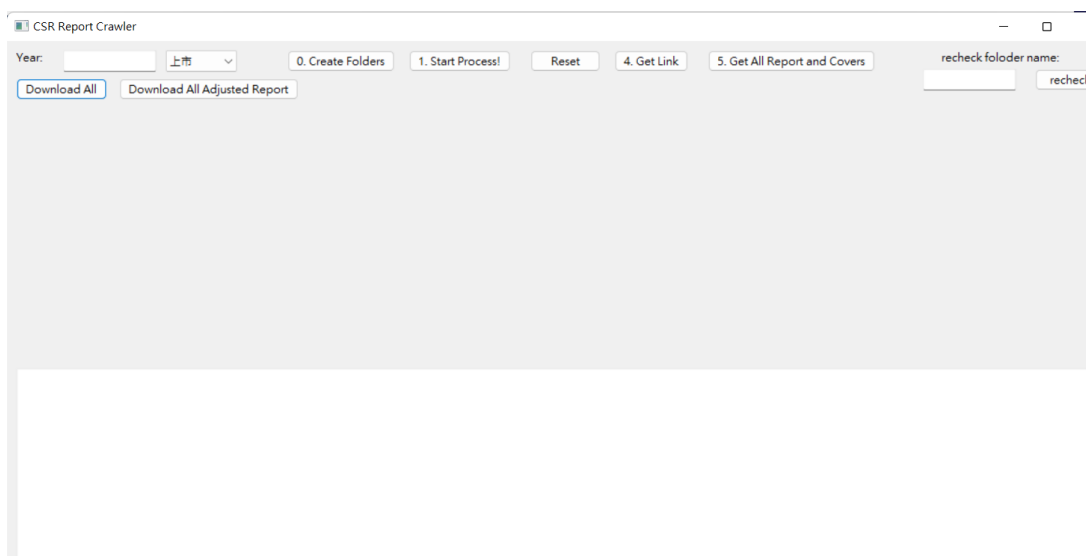


圖 2.9. 安裝完畢後，CSR Report Crawler 程式視窗會自動跳出

3. 主要功能介紹與操作指南

如圖 3.1，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。

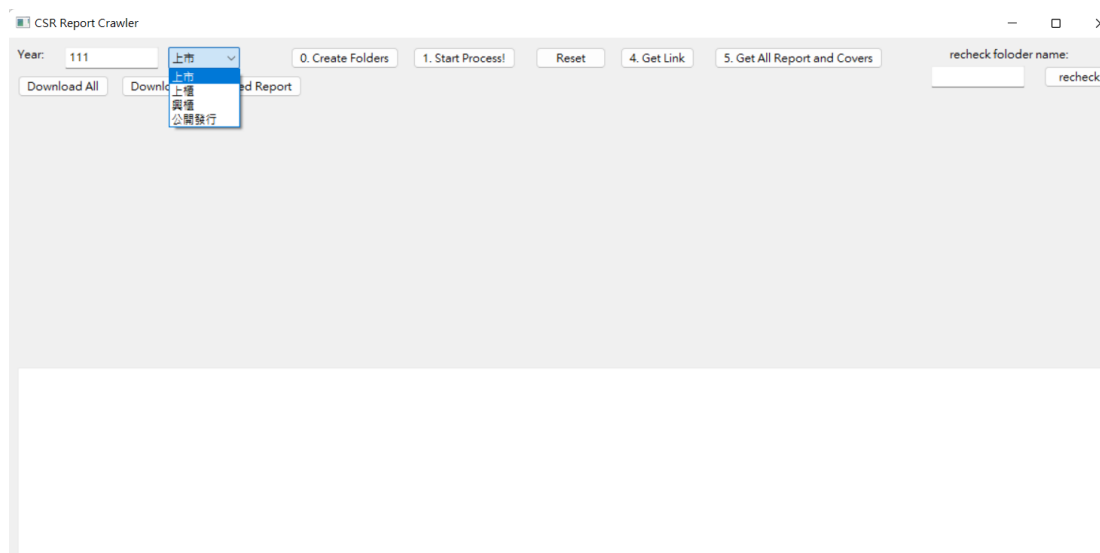


圖 3.1. 輸入條件，點擊“Download All”即可下載當年度的所有報告書

在當年度尾，若需要檢查所有「中文版永續報告書(修正後版本)」欄位，是否有包含更新的中文版報告書，則點擊“Download All Adjusted Report”。

如圖 3.2，下載完畢後，跟目錄會出現一個 output 資料夾。進入 output 資料夾，裡面會包含系統桶根據創建時間和欲下載公司類型命名的資料夾。

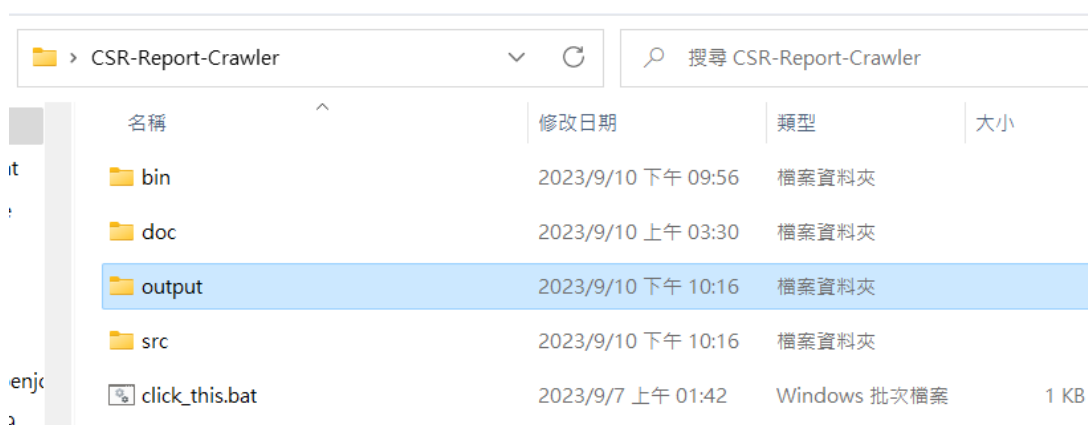


圖 3.2. 根目錄會出現一個 output 資料夾

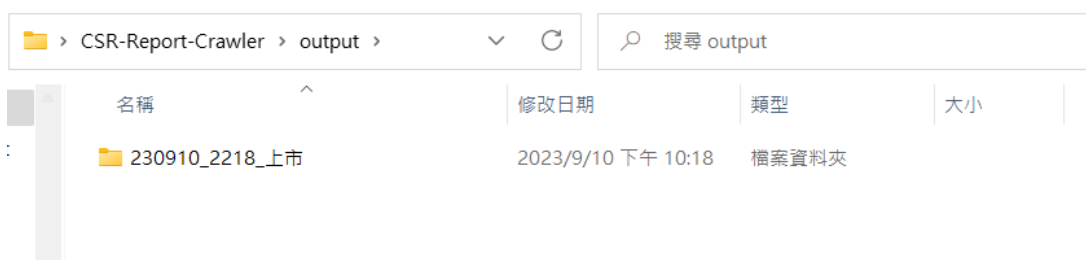


圖 3.3. 系統會根據創建時間和欲下載的公司類型來命名資料夾

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。

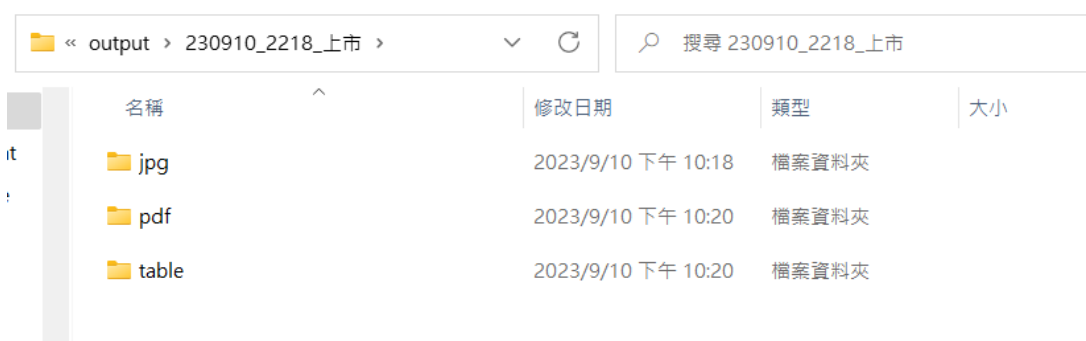


圖 3.4. 當中會包含三個子資料夾

如圖 3.4.，「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。

接下來，系統會根據條件啟動程式。如圖 3.5.，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。

同時，如圖 3.6.，藍色的終端機會顯示進度條，可查看當前的報告書與封面頁下載進度。



圖 3.5. Chrome 受到自動測試軟體控制

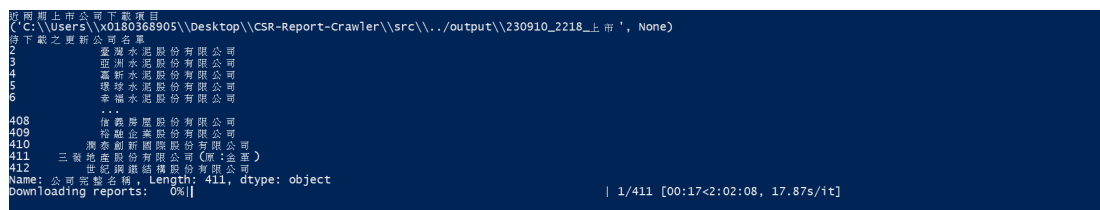


圖 3.6. 終端機會顯示進度條，可查看當前的進度

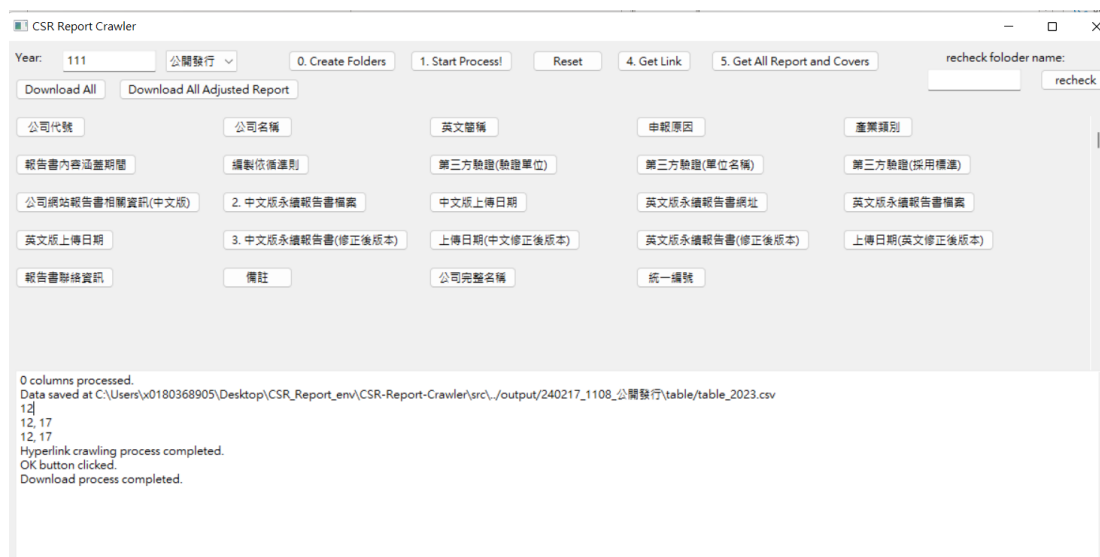


圖 3.7. 下載完報告書與封面後，文字框會顯示如上

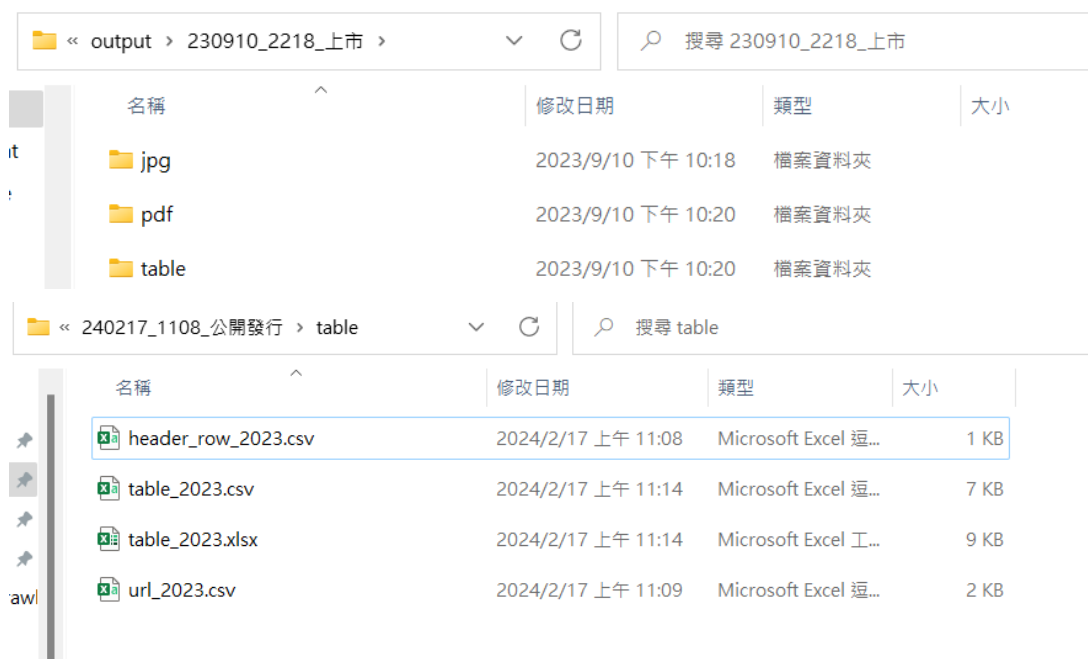


圖 3.8. 完整的公司資料、統編、報告書連結會被存在 table 資料夾

- “header_row_2023.csv”：儲存了 2023 年公開資訊觀測站中永續報告書的欄位資訊。
- ”table_2023.csv”：記錄了 2023 年以 utf-8 編碼的上市公司完整表格。
- ”table_2023.xlsx”：包含了 2023 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_2023.csv”：存放了 2023 年上市公司的報告書連結。

圖 3.9. table_2023.xlsx

名稱	修改日期	類型	大小
2023_03557311.pdf	2024/2/17 上午 11:13	Microsoft Edge P...	13,951 KB
2023_03700301.pdf	2024/2/17 上午 11:12	Microsoft Edge P...	31,649 KB
2023_03707901.pdf	2024/2/17 上午 11:11	Microsoft Edge P...	35,999 KB
2023_03750168.pdf	2024/2/17 上午 11:12	Microsoft Edge P...	3,842 KB
2023_03787101.pdf	2024/2/17 上午 11:14	Microsoft Edge P...	15,502 KB
2023_03795904.pdf	2024/2/17 上午 11:14	Microsoft Edge P...	22,701 KB
2023_27236155.pdf	2024/2/17 上午 11:10	Microsoft Edge P...	19,225 KB

圖 3.10. 本次下載的報告書

如圖 3.10.所示，若選擇的公司種類為「公開發行」，所下載的報告書會被儲存於「公開發行」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」。

名稱	修改日期	類型	大小
230910_2218_上市	2023/9/10 下午 10:18	檔案資料夾	
230910_2224_上市	2023/9/10 下午 10:24	檔案資料夾	

圖 3.11. 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_2023.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。

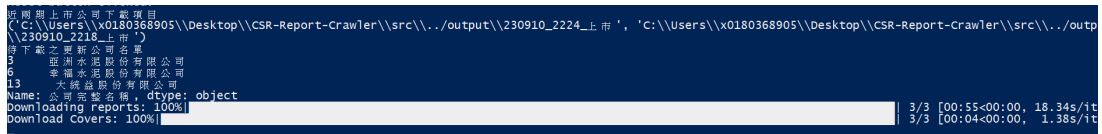


圖 3.12. 系統只會下載更新的報告書

終端機會展示即將下載的公司名單。如圖 3.6 所示，為首次下載時的進度條顯示；而圖 3.12 則顯示當進行更新時，終端機上的進度條狀態。

4. 突發狀況之解決方法

在下载的过程中，可能會發生一些不可抗力之問題。可以嘗試下列方式，進行初步障礙排解。

4.1. 檢查「本次開啟系統所建立的資料夾（其檔名命名格式為 YYMMDD_hhmm_公司類型）。」，若是 table/table_{年份}.xlsx 為完全正常的：

將「本次開啟系統所建立的資料夾」檔名複製到“recheck folder name:”下方的輸入匡，並點擊 recheck 按鈕，進行重新下載。



圖 4.1. 系統會根據 table 資料夾中的資料，繼續下載檔案

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

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

4.3. 常見之終端機報錯

- UnicodeDecodeError: 'utf-8' codec can't decode byte 0xa4 in position 0: invalid start byte

若有更新公司名單“company_name_number.csv”，請用記事本打開，檢查底下編碼是否為“UTF-8”。可嘗試另存新檔（control+shift+s），

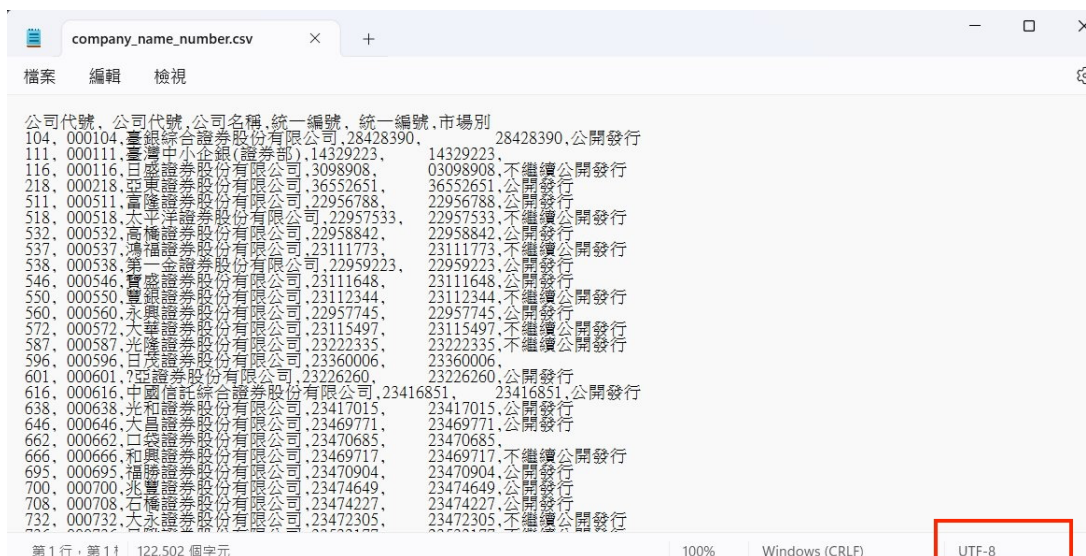


圖 4.2. 檢查右下角之檔案編碼

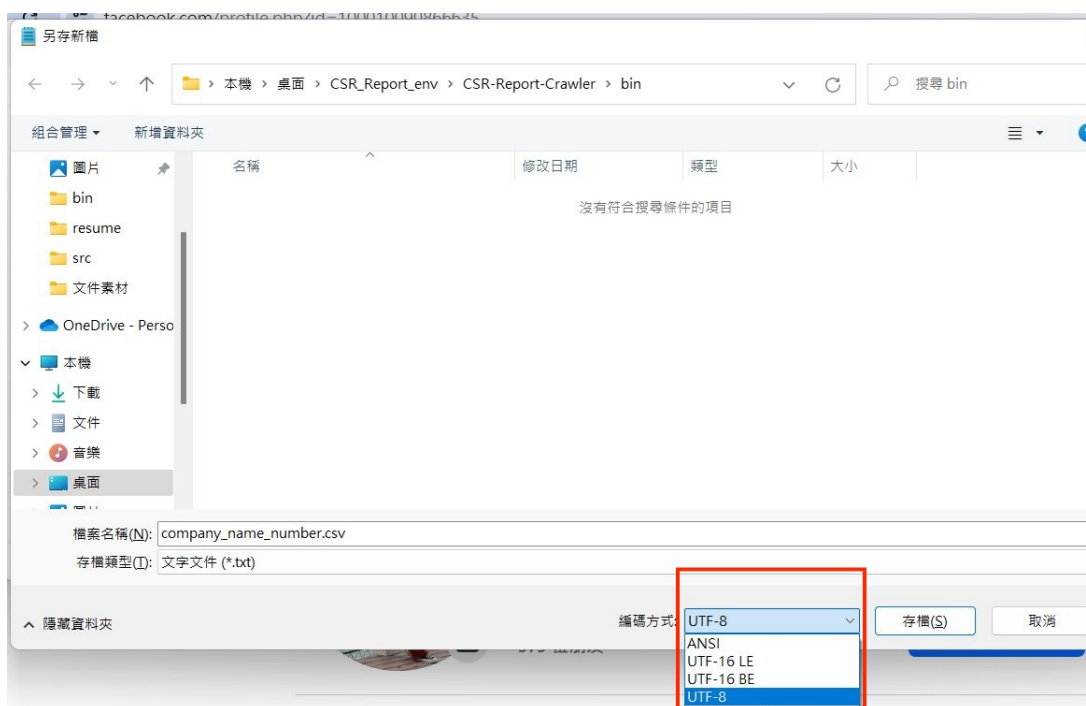
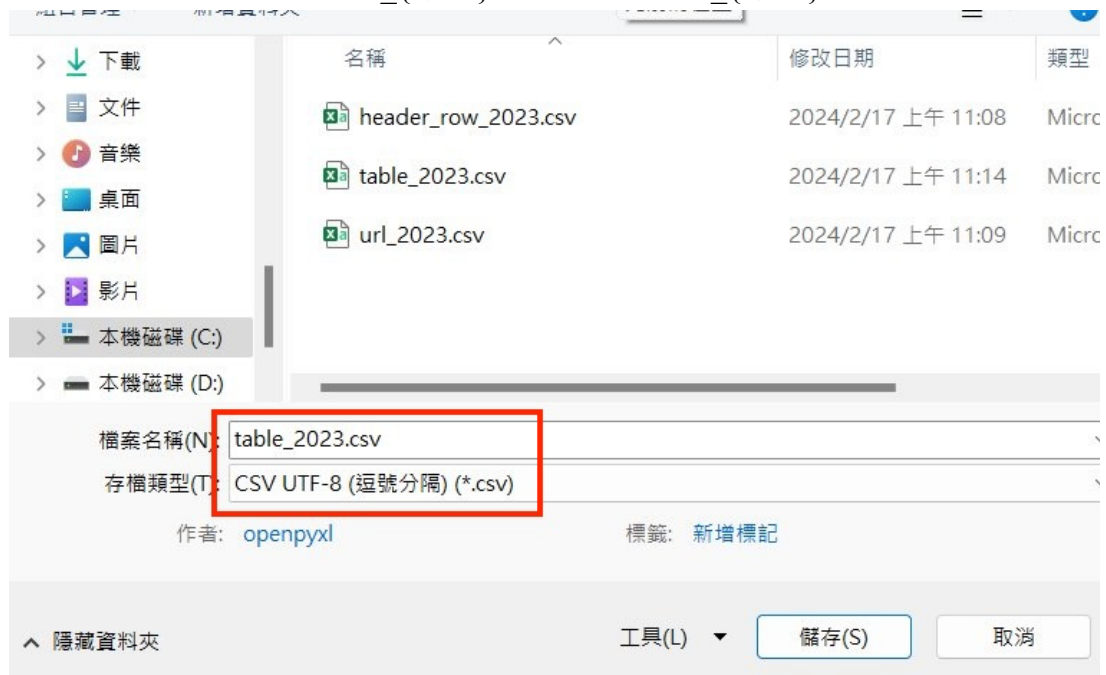


圖 4.3. 以 UTF-8 編碼另存新檔

若是在更新檔案時，有操作到“table_{年份}.xlsx”或是“table_{年份}.csv”，可以打開“table_{年份}.xlsx”另存新檔編碼正確的“table_{年份}.csv”。

名稱	修改日期	類型	大小
header_row_2023.csv	2024/2/17 上午 11:08	Microsoft Excel 逗...	1 KB
table_2023.csv	2024/2/17 上午 11:14	Microsoft Excel 逗...	7 KB
table_2023.xlsx	2024/2/17 上午 11:14	Microsoft Excel 工...	9 KB
url_2023.csv	2024/2/17 上午 11:09	Microsoft Excel 逗...	2 KB

圖 4.4. “table_{年份}.xlsx” 與 “table_{年份}.csv”



- FileNotFoundError: [Errno 2] No such file or directory: 'D:\\CSR-Report-Crawler\\src\\../output\\231101_1351_公開發行\\table\\table_2022.csv'

請確認先前下載的檔案位置是否跟本次下載的位置是相同的。

- KeyError: '公司完整名稱'

請以正確的編碼另存新檔 “company_name_number.csv”。

5. 版本更新

本系統支援更新前的功能，有數字編號的按鈕為舊版系統之功能，可參見「使用說明書 v1.0」做使用。

6. 聯絡資訊

開發人員：沈健好

信箱：110306010@nccu.edu.tw