程式架構說明

本組應用程式開發邏輯以全面自動化之想法為主，適時加入可以人機互動的元素，以增加程式執行之彈性而幫助ESG報告書檢索作業可以更加順利，以下會分為主要三部分程式段。

1. 清單攫取:   
   每年五月底開始，各公司會陸續上傳各家ESG報告書於公開資訊觀測站平台以利大眾利害關係人閱取，故可從該網站讀取公司清單，以避免遇到公司沒有上傳ESG報告書的情況。
2. 主程式段

|  |  |  |
| --- | --- | --- |
| 模組 | 相關工具 | 簡述 |
| 下載  報表 | Python、Uipath | 透過http request直接下載公司清單中之企業ESG報告書 |
| 表格  擷取 | Python、Uipath (pdfplimber) | 根據報告書中的所有表格查詢GRI準則相關關鍵字，並回傳該準則對應之頁碼欄位。若無法完整擷取(此初以表格列數小於5定義為非完整之表格)，會納入無法擷取清單 |
| 轉換  頁碼 | VBA、Uipath | 因PDF頁碼與報告書本身頁碼有相異之可能性，故以特定位置(如頁眉最右端或是頁尾)擷取頁碼相減算出頁碼差異，以利準確切割PDF頁面。若無法判別之報告書會跳出警示視窗，可人力輔以輸入。 |
| 表格  清理 | VBA、Uipath | 清除非必要之符號與資訊，並透過頁碼差異轉換表格中之索引頁碼。 |
| 報告  切割 | Uipath | 依據擷取完善之檢索表格，以準則對應到之頁碼形成該準則之一份PDF。 |
| Google Drive 上傳 | Uipath | 上傳各準則PDF，並加上準則超連結至雲端硬碟，串接Google workspace使用者介面。 |

1. 無法擷取清單:

在表格擷取模組段，將無法完整擷取之企業回傳，可以透過人機協作的方式處理程式無法產出的部分，此部分亦為日後可以繼續發展改進之地方，降低無法擷取的比例。