USER MANUAL

1. Cấu trúc folder crawl\_taxCode\_company gồm có 5 folders, 1 file .gitignore và guide.docx. User chỉ cần quan tâm đến 3 folders:
   1. Folder \_script

* Folder này chứa các file script (code), gồm có:
* func.py ( lưu các function cần dùng)
* main.ipynb : user có thể sử dụng file này để chạy trên nền tảng jupyter, hoặc sử dụng để phát triển thêm cho app
* crawl\_enterprise\_tax\_code.py: chứa script để chạy toàn bộ chương trình.
  1. Folder \_raw\_data
* Chứa 1 sub-folder \_old và một file Excel store data cần check.
  + Folder \_old lưu những file cũ đã chạy trước đây. Sau mỗi lần sử dụng, copy file data\_excel cũ vào folder \_old. Nếu không cut vào folder \_old, script sẽ không phân biệt đc file excel nào cần chạy.
  + File Excel chứa dữ liệu cần kiểm tra. Lưu ý file cần có format như sau.

A black and white line

Description automatically generated with medium confidence

* 1. Folder\_crawled
* Sau khi app chạy thành công. Kết quả dữ liệu trả về sẽ được lưu theo các folder, folder được đặt tên theo ngày chạy. User chỉ cần vào đúng folder để lấy file Excel và sử dụng.

Lưu ý: Nếu cần file .csv sẽ không thể sử dụng Excel để xem file, mà cần sử dụng Notepad. Click phải vào file, chọn Open with và chọn Notepad.

A screenshot of a computer

Description automatically generated

1. Cách sử dụng.
   1. Kiểm tra folder \_raw\_data

* Đảm bảo file Excel cũ đã được xóa hoặc đưa vào thư mục \_old. Thư mục \_raw\_data chỉ CHỨA DUY NHẤT 1 FILE EXCEL cần kiểm tra
  1. Chạy script ở thư mục \_script
* User có 2 cách để chạy script.
  + - Sử dụng file **main.ipynb** để chạy, tất cả đường dẫn đến file raw\_data, data\_crawled đã được thiết lập nên chỉ cần chạy run file mà kh chỉnh sửa gì thêm.

( Yêu cầu máy đã được cài đặt python và library liên quan)

* + - Sử dụng file crawl\_enterprise\_tax\_code.py. Tất cả việc bạn cần làm là 1 chiếc double click :D Double click để chạy file và đợi kết quả.

Trong lúc chạy file sẽ có 1 cửa sổ hiển thị, KHÔNG ĐƯỢC TẮT cho đến khi c.trình chạy xong, cửa sổ này sẽ tự tắt.

A screenshot of a computer

Description automatically generated

* 1. Vào folder \_crawled để lấy kết quả, file sẽ được export ở folder theo ngày chạy app.