

Hướng dẫn sử dụng chương trình nhận diện văn bản từ hình ảnh

1. Ngôn ngữ sử dụng: Python từ 3.6.0 trở lên
2. Thư viện hỗ trợ:

* PyQt5
* tkinter
* pytesseract (cần PIL.Image để hỗ trợ)
* pyperclip (để copy nội dung vào clipboard)

Để cài đặt thư viện : vào Terminal (Windows + R -> cmd) -> pip install « Tên thư viện »

* Ví dụ : pip instal pyqt5

1. Cách dùng

* Tải file lên bằng cách nhấn vào nút “Tải dữ liệu”, có thể lựa chọn các định dạng file .pdf, .png, .jpg,…
* Để đọc văn bản từ file đã chọn, nhấn “Chuyển đổi”, do thuật toán hơi củ chuối nên nhấn phải đợi một tí, hầu như xuất hiện văn bản ngay lập tức.
* Sao chép nội dung văn bản vừa tìm được bằng cách nhấn “Copy to clipboard”
* Lưu văn bản bằng cách nhấn vào “Lưu”, file lưu sẽ được lưu dưới định đạng “detect\_” + fileName + “.txt”
* Muốn thoát chương trình nhấn “Thoát” hoặc “X” trên title bar.

1. Chương trình chưa đọc được file utf-8 (sẽ phát triển trong ít ngày tới).
2. Thay đổi một số đường dẫn mặc định ở đầu chương trình.
3. Chương trình được phát triển trong 3 ngày (2,5 ngày sửa lỗi) nên chắc chắn có nhiều thiếu sót. Mong mọi người đóng góp.
4. File test mẫu được đặt trong thư mục File, mọi người có thể test thử chương trình.
5. Cảm ơn mọi người đã đọc.