

ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ



ỨNG DỤNG NHẬN DẠNG VĂN BẢN

Thực hiện:

Nguyễn Hải Thành – 20020717

Giảng viên hướng dẫn: PGS.TS. Hoàng Văn Xiêm

Môn học: Nhập môn tương tác người - máy

HÀ NỘI - 2023

1. Tổng quan

Ứng dụng nhận dạng văn bản dựa trên công nghệ nhận dạng ký tự quang học (OCR – Optical Character Recognition). OCR là một công nghệ cho phép máy tính nhận dạng và chuyển đổi văn bản từ các hình ảnh hoặc tài liệu in thành dữ liệu có thể xử lý được.

Nhận dạng ký tự quang học bắt nguồn từ các công nghệ liên quan đến điện báo và tạo ra các thiết bị đọc cho người mù. Các phiên bản đầu tiên được phát triển với hình ảnh của từng ký tự và làm việc trên một phong chữ tại một thời điểm. Các phiên bản tiên tiến hơn có khả năng tạo ra độ chính xác cao cho hầu hết phong chữ hiện đã phổ biến và có hỗ trợ nhiều định dạng của tệp dữ liệu đầu vào.

Công cụ OCR được phát triển riêng thành nhiều loại ứng dụng dành riêng cho mỗi lĩnh vực như OCR biên nhận, OCR hóa đơn, OCR chứng từ thanh toán hợp pháp, OCR hỗ trợ cho người mù và khiếm thị, ...

Nguyên lý hoạt động của OCR bao gồm các bước sau:

- Tiền xử lý: Ảnh đầu vào được chuẩn bị và xử lý trước khi vào quá trình nhận dạng. Các bước tiền xử lý bao gồm cân bằng độ sáng, làm mịn ảnh, loại bỏ nhiễu và phân vùng văn bản.
- Nhận dạng ký tự: Trong quá trình này, các ký tự trên ảnh được nhận dạng bằng cách so khớp với các mẫu ký tự đã được học trước đó. Các thuật toán nhận dạng ký tự có thể sử dụng các phương pháp như k-means clustering, neural networks hoặc SVM (Support Vector Machine).
- Hậu xử lý: Sau khi nhận dạng ký tự, kết quả được xử lý để cải thiện độ chính xác. Các bước hậu xử lý có thể bao gồm việc sửa các lỗi nhận dạng, gom nhóm các ký tự thành từ và câu, xử lý định dạng văn bản và tách lớp nền.

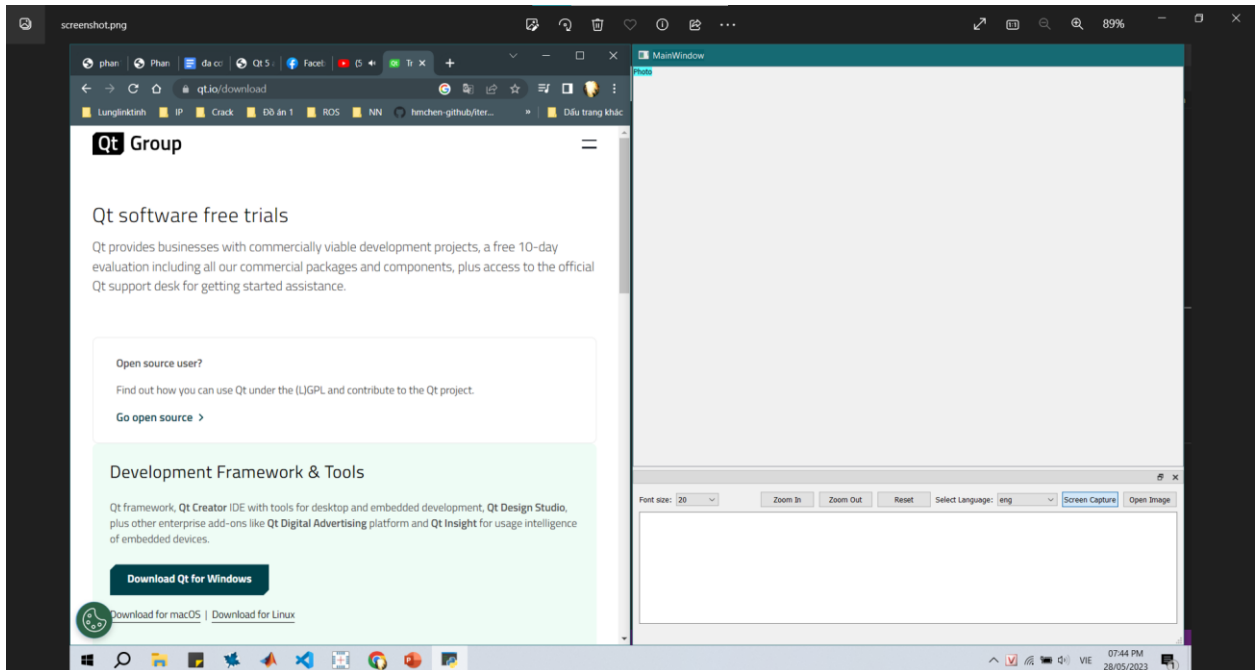
Ứng dụng nhận dạng văn bản của tôi được phát triển với hai tính năng chính:

- Trích xuất văn bản từ hình ảnh
- Truy cập nội dung từ màn hình máy tính (chụp màn hình)

2.1. Trích xuất văn bản từ hình ảnh

2.2. Truy cập nội dung màn hình (chụp màn hình)

Tính năng này cho phép người dùng có thể chụp lại ảnh hiện tại của màn hình máy tính sau khi nhấn nút Screen Capture. Hình ảnh sẽ được lưu với tên “Screenshot.png” để có thể sử dụng làm ảnh đầu vào.



Hình 2.2. Ảnh chụp màn hình được thực hiện bởi ứng dụng

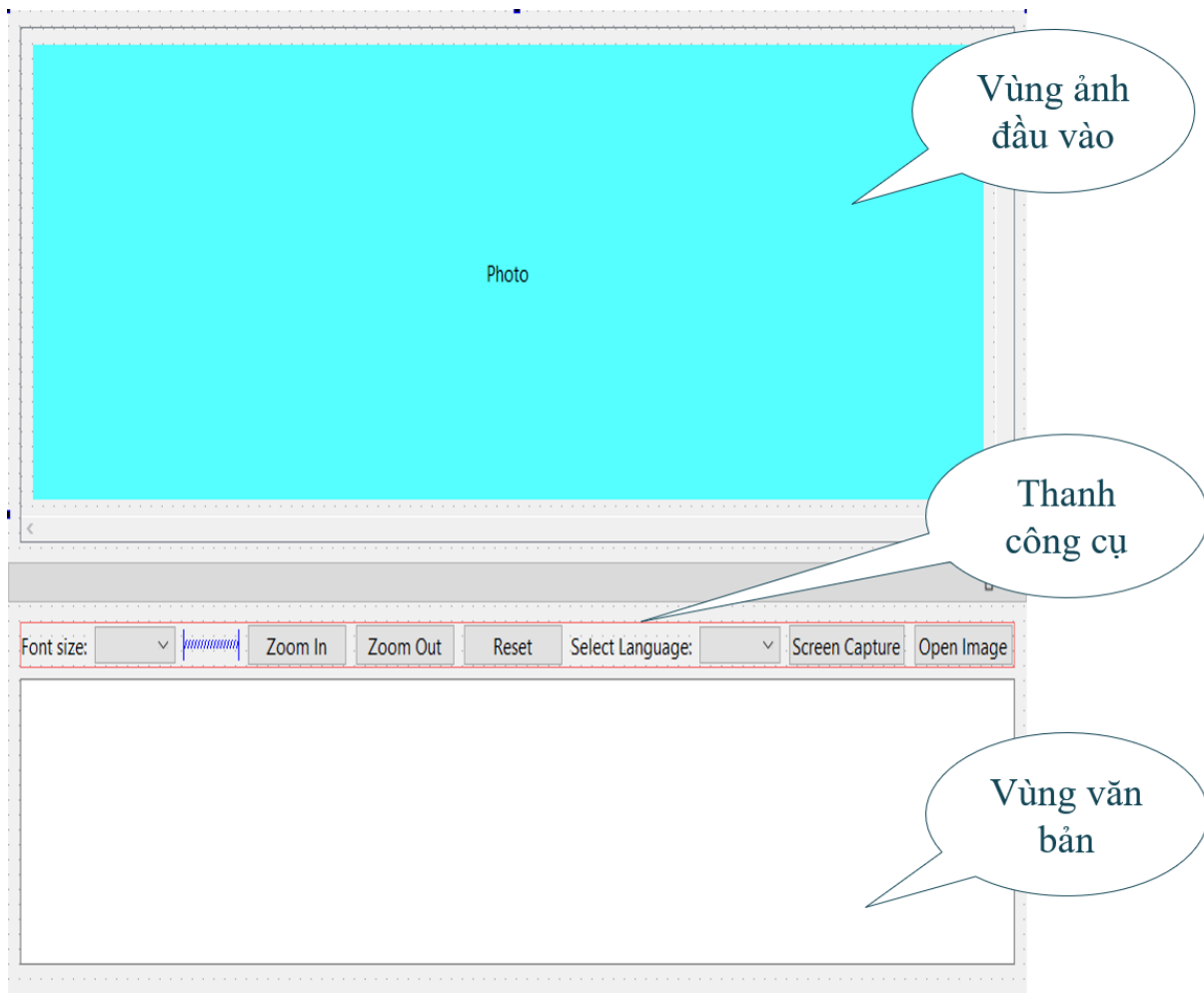
2.3. Các tính năng khác

Ngoài ra, ứng dụng còn cho phép người dùng có thể phóng to, thu nhỏ ảnh đầu vào hay đặt nó về kích thước ban đầu. Đối với vùng văn bản đầu ra, người dùng có thể tự do chỉnh sửa kích cỡ font chữ, thay đổi nội dung hoặc thực hiện các thao tác đơn giản khác như Copy, Cut, Paste, Undo.

3. Giao diện ứng dụng

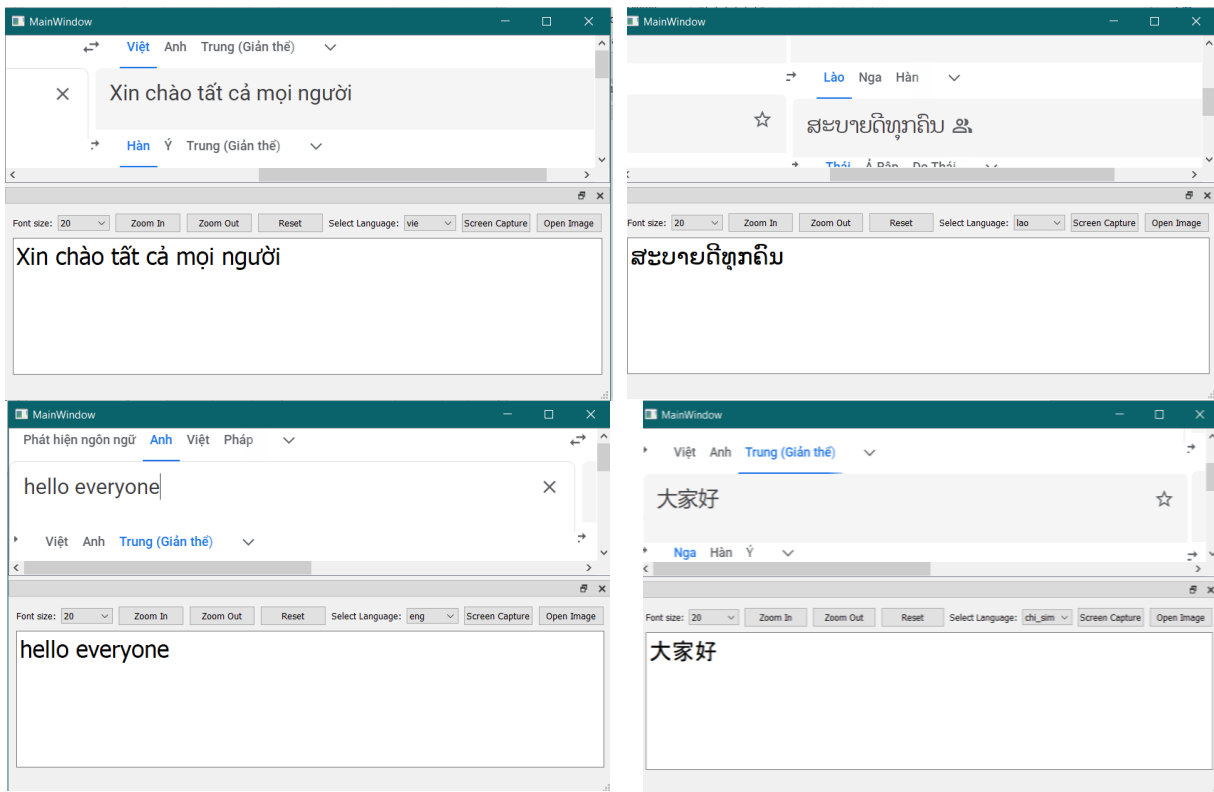
Với các tính năng trên của ứng dụng, giao diện của ứng dụng được thiết kế với phần mềm Qt Creator gồm ba vùng:

- Vùng ảnh đầu vào: Vùng này hiển thị ảnh đầu vào được sử dụng trong ứng dụng, người dùng có thể thao tác bằng chuột để chọn vùng hình ảnh cần xuất sang văn bản tại đây
- Vùng văn bản đầu ra: Đây là vùng hiển thị đoạn văn bản sau khi đã được trích xuất từ vùng ảnh đầu vào, mọi chỉnh sửa đoạn văn bản cũng được hiển thị ở đây.
- Thanh công cụ: Chứa các nút thay đổi kích cỡ chữ, phóng to, thu nhỏ, đặt lại kích thước của ảnh đầu vào, chọn ngôn ngữ cần nhận dạng, nút chụp màn hình, mở ảnh.

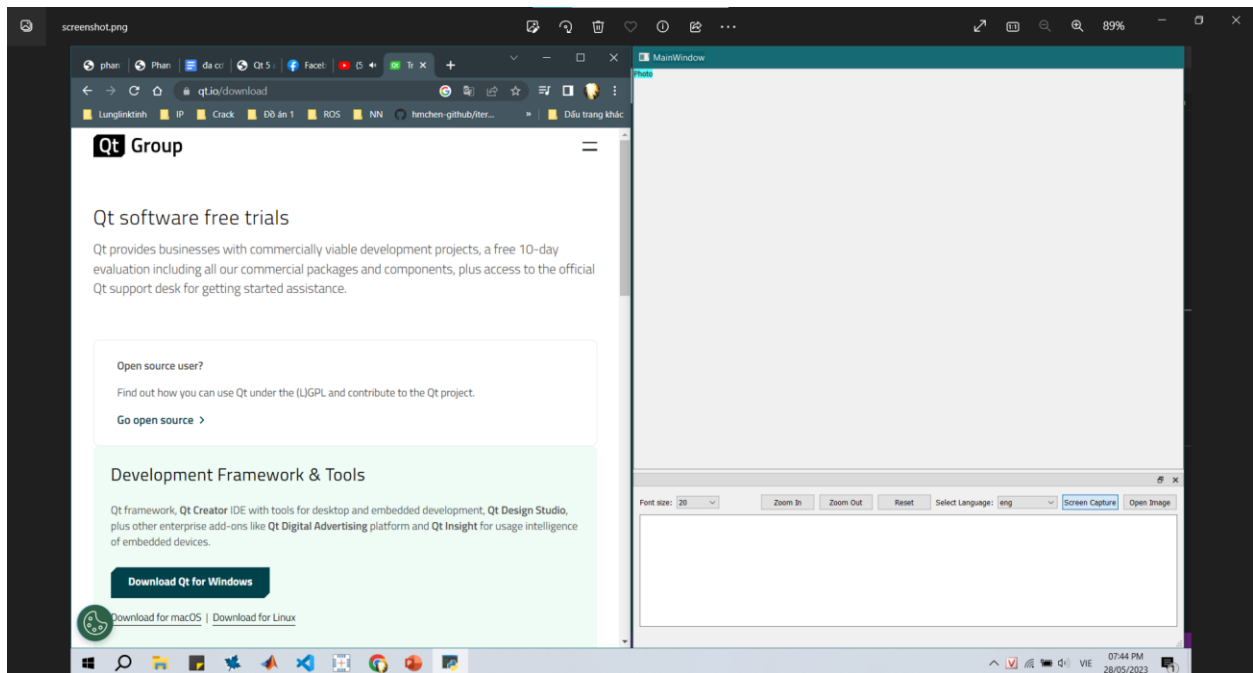


Hình 3.1. Bố cục giao diện của ứng dụng

4. Kết quả



Hình 4.1. Ảnh kết quả trích xuất từ một số ngôn ngữ khác nhau



Hình 4.2. Ảnh chụp màn hình bởi ứng dụng

Video demo:

<https://drive.google.com/file/d/16PxKYioDxBzcc5HoHcejTZMb296ffNq/view?usp=sharing>

5. Kết luận

Hiện tại, ứng dụng có thể nhận diện hầu hết các dạng ngôn ngữ, tuy nhiên mức độ chính xác còn chưa cao. Giao diện có thể đáp ứng được các tính năng được đề ra nhưng vẫn còn đơn điệu.

Trong tương lai, ứng dụng sẽ được phát triển thêm các bộ dịch thuật, tối ưu độ chính xác của việc nhận dạng và tối ưu được giao diện.