**TRƯỜNG ĐẠI HỌC GIAO THÔNG VẬN TẢI**

**PHÂN HIỆU TẠI THÀNH PHỐ HỒ CHÍ MINH**

**BỘ MÔN CÔNG NGHỆ THÔNG TIN**



**BÁO CÁO THỰC TẬP CHUYÊN MÔN**

**ĐỀ TÀI**

**XÂY DỰNG ỨNG DỤNG TRÌNH DUYỆT WEB BẰNG WINFORM C#**

Giảng viên hướng dẫn: PHẠM THỊ MIÊN

Sinh viên thực hiện: LÊ NGUYỄN THÁI KHANG

Lớp: CÔNG NGHỆ THÔNG TIN K59

Khóa: K59

Tp.Hồ Chí Minh, năm 2021

**TRƯỜNG ĐẠI HỌC GIAO THÔNG VẬN TẢI**

**PHÂN HIỆU TẠI TP. HỒ CHÍ MINH**

**BỘ MÔN CÔNG NGHỆ THÔNG TIN**



**BÁO CÁO THỰC TẬP CHUYÊN MÔN**

**ĐỀ TÀI**

**XÂY DỰNG ỨNG DỤNG TRÌNH DUYỆT WEB BẰNG WINFORM C#**

Giảng viên hướng dẫn: PHẠM THỊ MIÊN

Sinh viên thực hiện: LÊ NGUYỄN THÁI KHANG

LỚP CÔNG NGHỆ THÔNG TIN K59

KHÓA: K59

TP. Hồ Chí Minh, năm 2021

|  |  |
| --- | --- |
| **TRƯỜNG ĐẠI HỌC GIAO THÔNG VẬN TẢI**  **PHÂN HIỆU TẠI TP. HỒ CHÍ MINH**  **BỘ MÔN CÔNG NGHỆ THÔNG TIN**  **---------------------------------------** | **CỘNG HOÀ XÃ HỘI CHỦ NGHĨA VIỆT NAM**  **Độc lập – Tự do - Hạnh phúc**  **---------------------------------------------------** |

**THIẾT KẾ TỔNG QUAN ĐỀ TÀI**

-------\*\*\*-------

**Mã sinh viên: 5951071043**

**Khóa: K59**

**Họ Tên SV: Lê Nguyễn Thái Khang**

**Lớp:** **Công Nghệ Thông Tin K59**

1. **Tên đề tài**

Xây dựng ứng dụng trình duyệt web bằng Winform C#

1. **Giáo viên và cán bộ hướng dẫn**

Giảng viên hướng dẫn: ThS. Phạm Thị Miên

1. **Nhiệm vụ thực tập chuyên môn**

* Tìm hiểu về ngôn ngữ lập trình C#
* Tìm hiểu về Http request và thư viện xNet.
* Hoàn thành sản phẩm về trình duyệt web.

**4. Ngày bắt đầu thực tập chuyên môn:** Ngày 10/05/2021

**5. Ngày hoàn thành báo cáo thực tập chuyên môn:** Ngày 12/06/2021

**LỜI CẢM ƠN**

Trong quá trình thực hiện đề tài này, em đã được nhận nhiều sự quan tâm, giúp đỡ từ các thầy cô đang công tác tại trường Đại học Giao Thông Vận Tải Phân hiệu tại thành phố Hồ Chí Minh. Em cám ơn chân thành đến cô Phạm Thị Miên đã tạo điều kiện để em có thể hoàn thành bài thực tập chuyên môn. Nhờ đó mà em có cơ hội học tập, trau dồi những kiến thức, kĩ năng và làm quen với việc tìm hiểu các tài liệu. Em chân thành cảm ơn cô đã tận tâm chỉ bảo, lắng nghe những ý tưởng, tạo điều kiện tốt nhất có thể để giúp em hoàn thành đề tài.

Những thiếu sót trong đề tài nghiên cứu này không thể tránh khỏi do những hạn chế về kiến thức, kinh nghiện thực tế, em mong được nhận những lời chỉ dẫn, góp ý của cô để đề tài thêm phần hoàn thiện hơn.

Lời cuối, em xin chúc cô luôn thành công trong công việc giảng dạy và luôn hạnh phúc trong cuộc sống.

**NHẬN XÉT CỦA GIẢNG VIÊN HƯỚNG DẪN**

|  |
| --- |
| ***Tp. Hồ Chí Minh, ngày …… tháng …… năm 2021***  **Giảng viên hướng dẫn**  **Phạm Thị Miên** |

mục lục

[**DANH MỤC HÌNH VẼ 6**](#_Toc74316308)

[**LỜI MỞ ĐẦU 7**](#_Toc74316309)

[**TỔNG QUAN 1**](#_Toc74316310)

[1.Tổng quan về đề tài 1](#_Toc74316311)

[**1.1 Web browser – trình duyệt là gì ? 1**](#_Toc74316312)

[**1.2 Lịch sử về web browser 1**](#_Toc74316313)

[**1.3 Chức năng 2**](#_Toc74316314)

[**1.4 Tính năng 3**](#_Toc74316315)

[2. Mục tiêu nghiên cứu 4](#_Toc74316316)

[**CHƯƠNG 1: CƠ SỞ LÝ THUYẾT 5**](#_Toc74316317)

[1.1 Tổng quan về ngôn ngữ lập trình C# 5](#_Toc74316318)

[1.2 Thư viện xNet. 6](#_Toc74316319)

[1.3 Optical Character Recognition (OCR) 6](#_Toc74316320)

[**CHƯƠNG 2: CÁC CHỨC NĂNG CHÍNH CỦA HỆ THỐNG 8**](#_Toc74316321)

[2.1 Giao diện trang chính. 8](#_Toc74316322)

[2.2 Giao diện thanh điều hướng và các chức năng. 8](#_Toc74316323)

[2.3 Giao diện hiển thị lịch sử và mục yêu thích. 8](#_Toc74316324)

[2.5 Giao diện chức năng ORC 11](#_Toc74316325)

[2.6 Chức năng thay đổi Proxy để truy cập mạng từ một IP khác . 11](#_Toc74316326)

[2.7 Một số giao diện chức năng khác 13](#_Toc74316327)

[**2.7.1 Giao diện cài đặt . 13**](#_Toc74316328)

[**2.7.2 Xem mã nguồn html. 13**](#_Toc74316329)

[**2.7.3 Chức năng xem trước và in trang. 13**](#_Toc74316330)

[**KẾT LUẬN 15**](#_Toc74316331)

[1. Kết quả 15](#_Toc74316332)

[2. Tồn tại 15](#_Toc74316333)

[3. Hướng phát triển 15](#_Toc74316334)

[**TÀI LIỆU THAM KHẢO 16**](#_Toc74316335)

**DANH MỤC HÌNH VẼ**

[Hình 1 Giao diện chính của trình duyệt. 8](#_Toc74315884)

[Hình 2 Giao diện các phím điều hướng chính của trình duyệt. 8](#_Toc74315885)

[Hình 3 Giao diện hiển thị lịch sử và mục yêu thích. 8](#_Toc74315886)

[Hình 4 Mục yêu thích được lưu trữ bằng file XML 9](#_Toc74315887)

[Hình 5 Mục lịch sử được lưu trữ bằng file XML 9](#_Toc74315888)

[Hình 6 Giao diện xóa lịch sử 10](#_Toc74315889)

[Hình 7 Giao diện chức năng thêm mục yêu thích . 10](#_Toc74315890)

[Hình 8 Giao diện OCR. 11](#_Toc74315891)

[Hình 9 Code Http request. 11](#_Toc74315892)

[Hình 10 Giao diện thay proxy. 12](#_Toc74315893)

[Hình 11 Local Area Network Setting đã được thay proxy. 12](#_Toc74315894)

[Hình 12 Giao diện cài đặt trang 13](#_Toc74315895)

[Hình 13 Chức năng hiển thị mã nguồn HTML của trang. 13](#_Toc74315896)

[Hình 14 Chức năng xem trước trang in. 14](#_Toc74315897)

[Hình 15 Chức năng in. 14](#_Toc74315898)

**LỜI MỞ ĐẦU**

Ngày nay ứng dụng CNTT và việc tin học hoá được xem là 1 trong những yếu tố mang tính quyết định trong hoạt động của các chính phủ, tổ chức cũng như các công ty , nó đóng vai trò hết sức quan trọng , có thể tạo ra những bước đột phá mạnh mẽ.

Cùng với sự phát triển không ngừng về kỹ thuật máy tính và mạng điện tử, CNTT cũng được những công nghệ cấp cao chinh phục hết đỉnh cao này đến đỉnh cao khác. Mạng Internet là một trong những sản phẩm có giá trị hết sức lớn lao và ngày càng trở thành nền tảng chính cho sự truyền tải, trao đổi thông tin toàn cầu.

Là một sinh viên IT – được đào tạo những cơ sở lý luận, được cung cấp những kiến thức từ cơ bản đến phức tạp về về khoa học tự nhiên, kiến thức cơ bản như mạng máy tính, hệ thống thông tin, lập trình, phần mềm… đã giúp em nâng cao trình độ hiểu biết của mình.

Xuất phát từ nhận thức trên và dựa trên kiến thức học được, em đã lựa chọn đi sâu nghiên cứu rõ hơn về đề tài : ***“Xây dựng ứng dụng trình duyệt web bằng Winform C#”***

Mặc dù đã cố gắng hết khả năng của mình nhưng do trình độ kiến thức và kinh nghiệm còn hạn chế, nên không tránh khỏi có những sơ sót. Em rất mong được sự nhận xét, đánh giá, đóng góp ý kiến của thầy để đề tài này được hoàn thiện hơn.

**TỔNG QUAN**

1. **Tổng quan về đề tài**
   1. **Web browser – trình duyệt là gì ?**

**Một** **web browser** là một phần mềm ứng dụng để truy xuất, trình diễn và chuyển các nguồn thông tin (information resource) trên mạng hệ thống mạng toàn cầu (World Wide Web). Một nguồn thông tin được nhận dạng bởi một Uniform Resource Identifier (URI) và có thể là một trang web, phim - video, hình ảnh (images) hoặc các mẫu thông tin khác.

Mặc dù các trình duyệt với mục đích là để truy cập vào hệ thống mạng toàn cầu, các trình duyệt còn được sử dụng để truy cập các thông tin được cung cấp bởi các web servers (máy chủ web) trong hệ thống mạng riêng hoặc các tài liệu (files) đến các hệ thống file (file system). Hoặc cũng được dùng để tiết kiệm tài nguyên thông tin cho các hệ thống lưu trữ file.

* 1. **Lịch sử về web browser**

**Lịch sử của các trình duyệt Web** bắt đầu vào cuối thập niên 1980, khi một loạt các công nghệ đặt nền móng cho các trình duyệt Web đầu tiên, WorldWideWeb, do Tim Berners-Lee vào năm 1991. Sự phát triển này dựa trên các công nghệ phần mềm và phần cứng hiện có, ngoài ra cũng mở ra nhiều công nghệ mới bắt đầu tư đây.

Ted Nelson và Douglas Engelbart phát triển khái niệm siêu văn bản trước khi Berners-Lee và CERN khá lâu. Web browser đã trở thành cốt lõi của World Wide Web. Berners-Lee thừa nhận sự đóng góp của Engelbart.

Việc giới thiệu trình duyệt Mosaic NCSA Web vào năm 1993 - một trong những trình duyệt web với giao diện đồ họa đầu tiên - đã dẫn tới sự bùng nổ trong việc dùng web. Marc Andreessen, người lãnh đạo của đội Mosaic tại NCSA, sớm bắt đầu công ty riêng của mình, đặt tên là Netscape, và phát hành các phiên bản Mosaic Netscape Navigator vào năm 1994, và sau đó nhanh chóng trở thành trình duyệt phổ biến nhất thế giới, chiếm 90% của tất cả các sử dụng Web đỉnh cao của nó.

Microsoft đã đáp trả bằng trình duyệt Internet Explorer của mình trong năm 1995 (cũng bị ảnh hưởng nhiều từ Mosaic), và đầu cuộc chiến trình duyệt đầu tiên. Bằng việc kết hợp Internet Explorer với Windows, Microsoft đã có thể tận dụng ưu thế của nó trong thị trường hệ điều hành để quản lý thị trường trình duyệt Web; Internet Explorer sử dụng đạt vị trí trên 95% người dùng vào năm 2002. Internet Explorer chiếm 60% người sử dụng trình duyệt vào Tháng Tư năm 2010 theo thống kê của Net Applications.

Opera đầu tiên xuất hiện vào năm 1996; mặc dù nó đã không bao giờ đạt được sử dụng rộng rãi, chiếm 2% thị phần trình duyệt vào tháng tư năm 2010, tuy nhiên Opera chiếm một phần đáng kể của thị trường đang phát triển nhanh là trình duyệt Web điện thoại di động, Oepra được cài đặt sẵn trên trên 40.000.000 điện thoại. Opera cũng có sẵn trên một số hệ thống nhúng khác, bao gồm video game console Wii của Nintendo.

Năm 1998, Netscape triển khai Mozilla Foundation trong một nỗ lực để sản xuất một trình duyệt cạnh tranh bằng cách sử dụng mô hình phần mềm mã nguồn mở. Đó là trình duyệt sau nay phát triển thành Firefox, sau đó thì trình duyệt này được phát triển khá tốt trong khi vẫn còn trong giai đoạn beta, ngay sau khi phát hành Firefox 1.0 vào cuối năm 2004, Firefox (mọi phiên bản) đã chiếm 7,4% thị phần sử dụng trình duyệt. Và đến tháng tư năm 2010, Firefox đã có một thị phần 25%.

Safari của Apple đã phát hành phiên bản beta đầu tiên vào tháng một năm 2003; vào tháng 10 năm 2009, đã chi phối thị phần của trình duyệt web dựa trên việc phát triển từ các ứng dụng của Apple, chiếm dưới 5% thị trường trình duyệt toàn bộ vào tháng 4 năm 2010. Nó được xem như là công cụ cho web (còn được gọi là WebKit) để trình diễn các tính năng và là nền tảng được ứng dụng nhiều trên thiết bị di động, bao gồm cả hệ điều hành iPhone, Google Android, Nokia S60 và Palm WebOS.

Vào tháng 9 năm 2008, Google bắt đầu nhảy vào cuộc chiến các trình duyệt bằng trình duyệt web Google Chrome. Đến tháng 4 năm 2010, Google Chrome đã chiếm 7% thị phần.

* 1. **Chức năng**

**Mục đích chính của một trình duyệt web** là để mang lại nguồn thông tin cho người dùng. Quá trình này bắt đầu khi một người sử dụng nhập vào (URI) hay tạm gọi là một đường dẫn, Ví dụ như <http://www.khoahocviet.org/>. Các tiền tố của URI sẽ xác định cho phương thức truy cập và dữ liệu nhận được sẽ được biên dịch như thế nào. Các loại URI thông dụng nhất bắt đầu với http: định một nguồn tài nguyên để được lấy dựa trên Hypertext Transfer Protocol (HTTP) (tôi tạm gọi là giao thức truyền dữ liệu cho các siêu văn bản). Nhiều trình duyệt cũng hỗ trợ một loạt các tiền tố khác, chẳng hạn như https: (cũng tương tự như HTTP nhưng dữ liệu được mã hóa và bảo mật hơn) hay ftp: cho File Transfer Protocol (giao thức truyền tải file), và file: cho các tập tin lưu trữ nội bộ. Các trình duyệt web có thể không trực tiếp xử lý mà thường chuyển qua các úng dụng khác xử lý. Ví dụ, mailto: URI thường sẽ được chuyển qua chương trình ứng dụng mail mặc định mà người dùng đang sử dụng, hoặc news: sẽ gọi đến các chương trình đọc tin tức.

Tuy nhiên trình duyệt cũng có cơ chế mở và được hỗ trợ bởi các plugins của nó để có thể hiển thị được các loại tài liệu trực tiếp trên web browser. Hầu hết các trình duyệt có thể hiển thị hình ảnh, âm thanh, video, và các tập tin XML, và thường có một số plug-in mặc định để hỗ trợ các ứng dụng Flash và Java applet. Khi gặp phải một tập tin không được hỗ trợ thì tập tin này sẽ được tải về thay vì hiển thị trực tiếp trên web, lúc này trình duyệt sẽ nhắc người dùng để lưu tập tin vào đĩa.

Tính tương tác trong một trang web có thể được ứng dụng bởi javascript mà thường không đòi hỏi sự hỗ trợ thêm của plugin. javascript có thể được sử dụng cùng với các công nghệ khác để cho phép tương tác trực tiếp với máy chủ trang web, ví dụ: tương tác thông qua AJAX.

Các thông tin trên trang web có thể chứa siêu liên kết (hyperlinks) tới các nguồn thông tin khác. Mỗi liên kết có chứa các URI để đi đến các nguồn thông tin khác. Khi một liên kết được nhấp, trình duyệt điều hướng người dùng tới trang thông tin khác.

* 1. **Tính năng**

Các trình duyệt web có sẵn bao gồm các tính năng từ giao diện người dùng tối thiểu, dựa trên văn bản với hỗ trợ xương sống cho các giao diện người dùng HTML phong phú hỗ trợ nhiều định dạng và giao thức.

Tất cả các trình duyệt web chính cho phép người dùng mở nhiều tài nguyên thông tin cùng lúc, hoặc trong các cửa sổ trình duyệt khác nhau hoặc trong các [tab](https://en.wikipedia.org/wiki/Tabbed_document_interface) khác nhau của cùng một cửa sổ. Các trình duyệt chính cũng bao gồm các trình chặn cửa sổ [bật lên](https://en.wikipedia.org/wiki/Pop-up_blockers) để ngăn các cửa sổ không mong muốn xuất hiện khi không có sự đồng ý của người dùng.

Hầu hết các trình duyệt web đều có thể hiển thị danh sách các trang web mà người dùng đã [đánh dấu](https://en.wikipedia.org/wiki/Bookmark_(World_Wide_Web)) để có thể nhanh chóng quay lại với họ. Dấu trang cũng được gọi là "Mục ưa thích" trong Internet Explorer . Ngoài ra, tất cả các trình duyệt web lớn đều có một số [trình](https://en.wikipedia.org/wiki/Web_feed)[kết hợp](https://en.wikipedia.org/wiki/News_aggregator)[nguồn cấp dữ liệu web](https://en.wikipedia.org/wiki/Web_feed) được xây dựng sẵn . Trong [Firefox](https://en.wikipedia.org/wiki/Firefox) , nguồn cấp dữ liệu web được định dạng là "dấu trang trực tiếp" và hoạt động giống như một thư mục các dấu trang tương ứng với các mục gần đây trong nguồn cấp dữ liệu.

Hơn nữa, hầu hết các trình duyệt có thể được mở rộng qua các [trình cắm thêm](https://en.wikipedia.org/wiki/Browser_extension) , các thành phần có thể tải xuống cung cấp các tính năng bổ sung.

## 2. Mục tiêu nghiên cứu

Hầu hết các trình duyết web lớn đều có những yếu tố giao diện người dùng chung:

* Nút quay lại và chuyển tiếp để quay lại tài nguyên trước và tương ứng.
* Một nút Refresh hoặc Reload để tài lại tài nguyên hiện hành.
* Một nút Stop để hủy bỏ tài các tài nguyên. Trong một số trình duyệt nút dừng thường được tích hợp với nút tải lại.
* Một nút Home để để trở về trang mà người dùng đặt làm trang mặt định khi mở trình duyệt.
* Một thanh địa chỉ để nhập vào Unifom Resource Identifier (URI) của tài nguyên mà mình mong muốn hiển thị.
* Khả năng xem mã nguồn HTML của một trang.
* Thêm xóa sửa lịch sử , mục ưa thích . Lịch sử và mục ưa thích được lưu trữ bằng file XML.
* ….

Ngoài những chức năng cơ bản trên, dựa vào những kiến thức đã học trên lớp cùng với những kiến thức tiền hiểu trên mạng và qua sách hướng dẫn về lập trình c# em đã nghiên cứu và tích hợp vào trình duyệt những tính năng cũng như ứng dụng sau:

* Thay đổi proxy của máy trực tiếp bằng trình duyệt cụ thể là địa chỉ IP và Port bằng địa chỉ IP và Port do mình lựa chọn.
* Sử dụng thư viện Xnet để lấy dữ liệu về IP và Port trực tiếp từ trang web: <https://spys.one/free-proxy-list/VN/> . Việc sử dụng thư viện Xnet nên việc cập nhật dữ liệu IP và Port được diễn ra thường xuyên dựa theo sự cập nhật của trang web.
* Ứng dụng nhận dạng ký tự quan học Optical Character Recognitioncơ bản để dễ dàng chuyển văn bảng dưới dạng hình ảnh ra văn bản dưới dạng file có thể chỉnh sửa được.

# CƠ SỞ LÝ THUYẾT

**1.1 Tổng quan về ngôn ngữ lập trình C#**

C# là ngôn ngữ lập trình hiện đại được pháp triển bởi Microsoft và được phê duyệt bởi European Computer Organization (ISO). C# được thiết kế cho các ngôn ngữ chung cơ sở hả tầng, trong đó bao gồm các mã (Excutable Code) và môi trường thực thi cho phép sử dụng các ngôn ngữ cấp cao khác nhau trên nền tảng máy tính và kiến trúc khác nhau.

* **Đặc trưng của C#.**

Các đặc điểm để làm cho C# là ngôn ngữ lập trình được sử dụng rộng rãi:

* + C# là ngôn ngữ đơn giản

Như ta đã biết thì ngôn ngữ C# dựng trên nền tảng C++ và Java nên ngôn ngữ C# khá đơn giản. Nếu chúng ta thân thiện với C và C++ hoậc thậm chí là Java, chúng ta sẽ thấy C# khá giống về diện mạo, cú pháp, biểu thức, toán tử và những chức năng khác được lấy trực tiếp từ ngôn ngữ C và C++, nhưng nó đã được cải tiến để làm cho ngôn ngữ đơn giản hơn. Một vài trong các sự cải tiến là loại bỏ các dư thừa, hay là thêm vào những cú pháp thay đổi.

* + C# là ngôn ngữ hiện đại

Một vài khái niệm khá mới mẻ khá mơ hồ với các bạn vừa mới học lập trình, như xử lý ngoại lệ, những kiểu dữ liệu mở rộng, bảo mật mã nguồn..v..v… Đây là những đặc tính được cho là của một ngôn ngữ hiện đại cần có. Và C# chứa tất cả các đặt tính ta vừa nêu trên. Các bạn sẽ dần tìm hiểu được các đặt tính trên qua các bài học trong series này.

* + C# là một ngôn ngữ lập trình thuần hướng đối tượng

Lập trình hướng đối tượng(tiếng Anh: Object-oriented programming, viết tắt: OOP) là một phương pháp lập trình có 4 tính chất. Đó là tính trừu tượng (abstraction), tính đóng gói (encapsulation), tính đa hình (polymorphism) và tính kế thừa (inheritance). C# hỗ trợ cho chúng ta tất cả những đặc tính trên. Và để hiểu rõ hơn thì chúng ta sẽ có một chương trình bày về phần này.

* + C# là một ngôn ngữ ít từ khóa

C được sử dụng để mô tả thôn# là ngôn ngữ sử dụng giới hạn những từ khóa (gồm khoảng 80 từ khóa và mười mấy kiểu dữ liệu xây dựng sẵn). Nếu bạn nghĩ rằng ngôn ngữ có càng nhiều từ khóa thì sẽ càng mạnh mẽ hơn. Điều này không phải sự thật, lấy ví dụ ngôn ngữ C# làm điển hình nhé. Nếu bạn học sâu về C# bạn sẽ thấy rằng ngôn ngữ này có thể được sử dụng để làm bất cứ nhiệm vụ nào.

Ngoài những đặc điểm trên thì còn một số ưu điểm nổi bật của C#

* + C# có cấu trúc khá gần gũi với các ngôn ngữ lập trình truyền thống, nên cũng khá dể dàng tiếp cận và học nhanh với C#.
  + C# có thể biên dịch trên nhiều nền tảng máy tính khác nhau.
  + C# được xây dựng trên nền tảng của C++ và Java nên nó được thừa hưởng những ưu điểm của ngôn ngữ đó.
  + C# là một phần của .NET Framework nên được sự chống lưng khá lớn đến từ bộ phận này.
  + C# có IDE Visual Studio cùng nhiều plug-in vô cùng mạnh mẽ.

## 1.2 Thư viện xNet.

xNet là một thư viện của .NET Framework, bao gồm:

* + Lớp proxy: HTTP, Socks4 (a), Socks5, Chain.
  + Các lớp học được sử dụng để xử lý giao thức HTTP 1.0/1.1: duy trì hoạt động, gzip, lật, khối hóa, SSL, proxy, v.v.

Link tải thư viện chính chủ: [X-rus/xNet: xNet - class library for .NET Framework (github.com)](https://github.com/X-rus/xNet)

HttpRequest được sử dụng để gửi yêu cầu. Trong đó, bạn có thể đặt các cài đặt khác nhau: tiêu đề, thời gian chờ, liệu bạn có cần thực hiện theo chuyển hướng hay không, có nên giữ kết nối liên tục hay không và các cài đặt khác. Khi bạn gửi một yêu cầu, trước tiên nó chấp nhận các tiêu đề phản hồi bằng cách sử dụng HttpResponse, sau đó trả về một liên kết đến đối tượng này để bạn có thể tải xuống nội dung của thư. Bạn có thể tải nội dung thư bằng một trong các phương pháp đặc biệt.

* 1. **Optical Character Recognition (OCR)**

ORC là loại phần mềm máy tính được tạo ra để chuyển các hình ảnh của chữ viết tay hoặc chữ đánh máy (thường được quét bằng máy scanner) thành các văn bản tài liệu.

OCR được hình thành từ một lĩnh vực nghiên cứu về nhận dạng mẫu, trí tuệ nhận tạo và machine vision. Mặc dù công việc nghiên cứu học thuật vẫn tiếp tục, một phần công việc của OCR đã chuyển sang ứng dụng trong thực tế với các kỹ thuật đã được chứng minh.

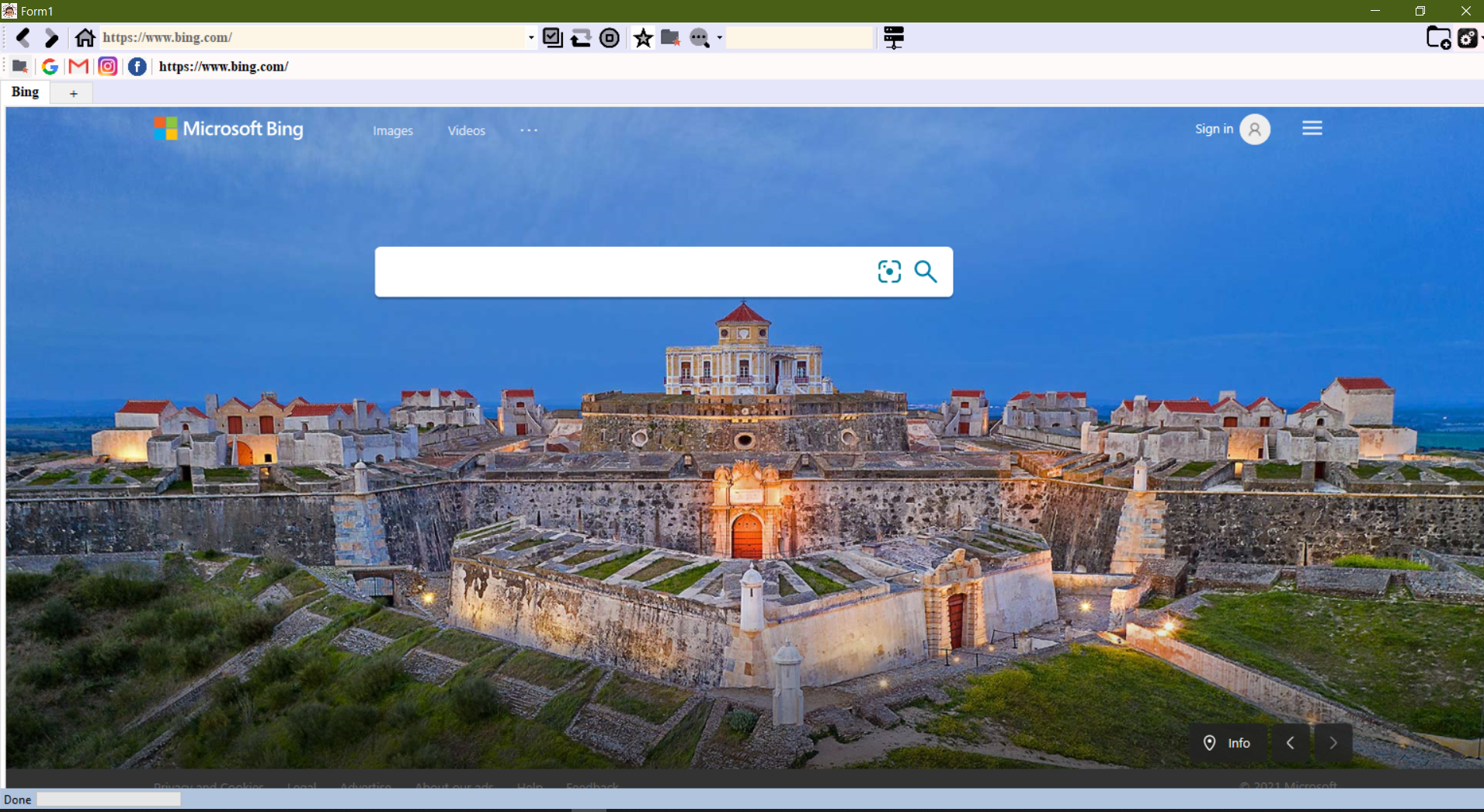
Nhận dạng ký tự quang học (dùng các kỹ thuật quang học chẳng hạn như gương và ống kính) và nhận dạng ký tự số (sử dụng máy quét và các thuật toán máy tính) lúc đầu được xem xét như hai lĩnh vực khác nhau. Bởi vì chỉ có rất ít các ứng dụng tồn tại với các kỹ thuật quang học thực sự, bởi vậy thuật ngữ Nhận dạng ký tự quang học được mở rộng và bao gồm luôn ý nghĩa nhận dạng ký tự số.

Đầu tiên hệ thống nhận dạng yêu cầu phải được huấn luyện với các mẫu của các ký tự cụ thể. Các hệ thống "thông minh" với độ chính xác nhận dạng cao đối với hầu hết các phông chữ hiện nay đã trở nên phổ biến. Một số hệ thống còn có khả năng tái tạo lại các định dạng của tài liệu gần giống với bản gốc bao gồm: hình ảnh, các cột, bảng biểu, các thành phần không phải là văn bản.

ORC được sử dụng và phát triển dựa trên nền tảng mã nguồn mở tesseract-ocr do Google tài trợ. Có khả năng nhận dạng chữ Việt rất tốt. Đây là một chương trình nguồn mở Java/.NET, hỗ trợ nhận dạng cho các dạng ảnh PDF, TIFF, JPEG, GIF, PNG, và BMP.

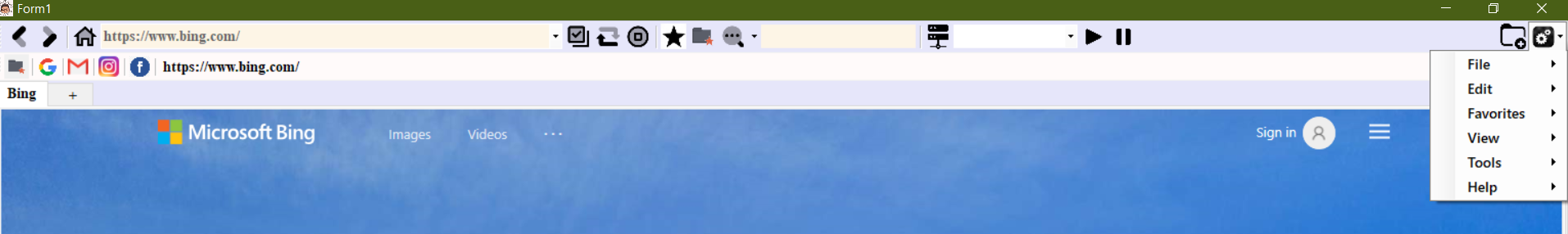
# CHƯƠNG 2: CÁC CHỨC NĂNG CHÍNH CỦA HỆ THỐNG

**2.1 Giao diện trang chính.**



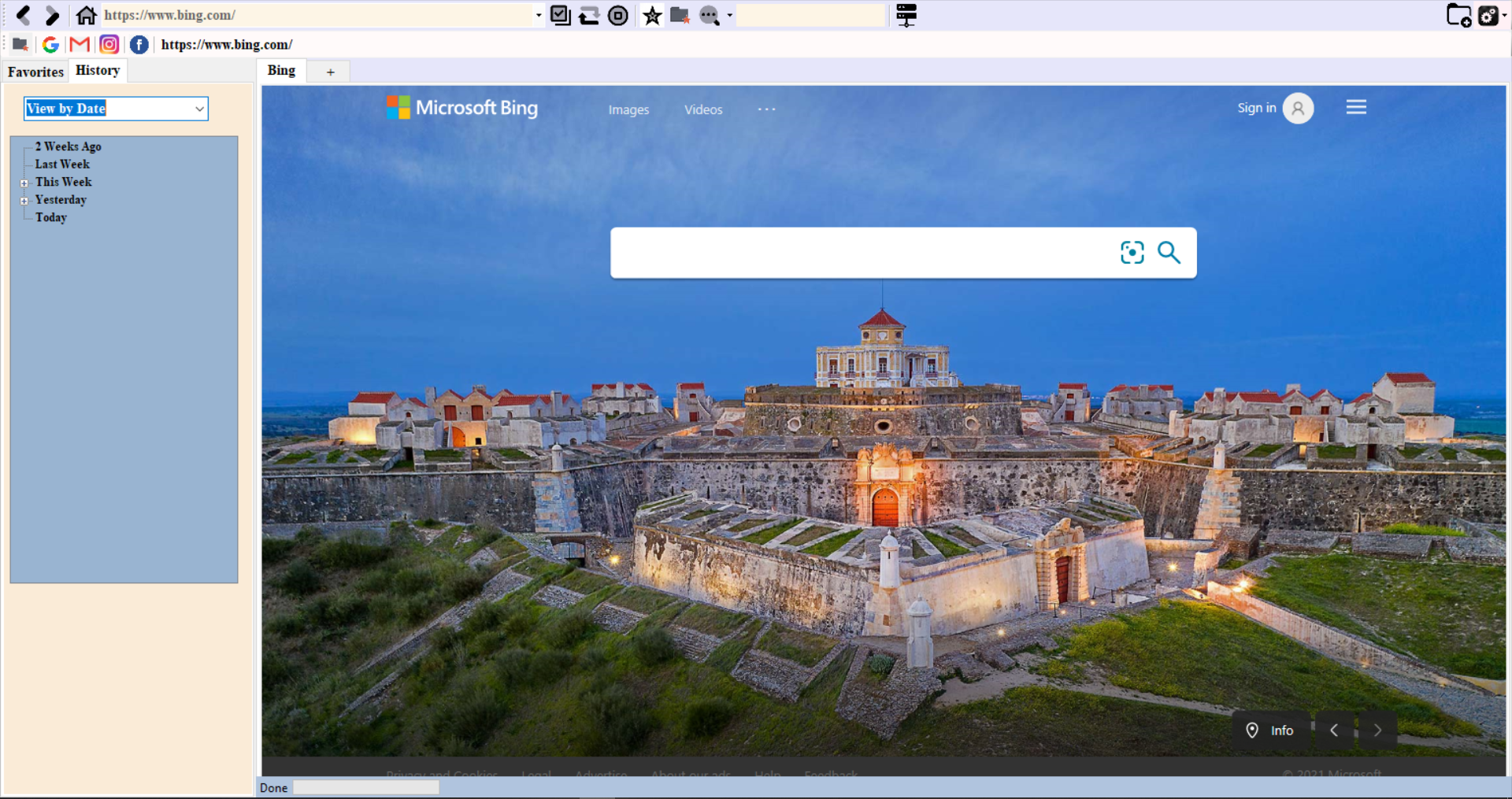
Hình 1 Giao diện chính của trình duyệt.

## 2.2 Giao diện thanh điều hướng và các chức năng.

****

Hình 2 Giao diện các phím điều hướng chính của trình duyệt.

## 2.3 Giao diện hiển thị lịch sử và mục yêu thích.



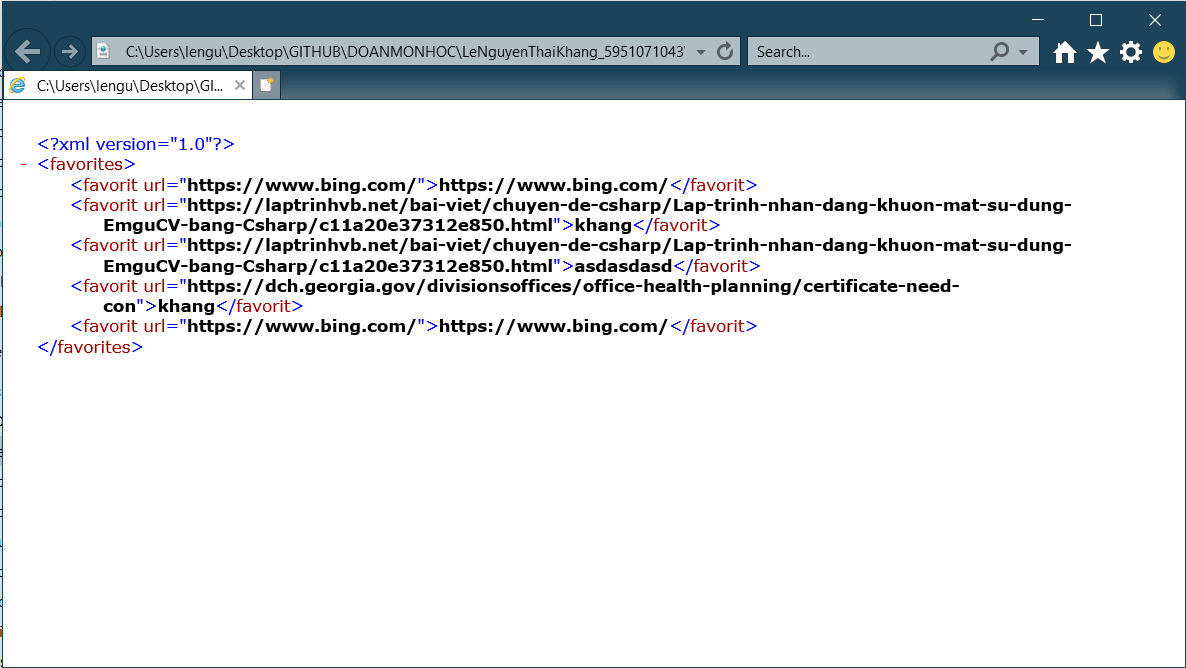
Hình 3 Giao diện hiển thị lịch sử và mục yêu thích.

Sau khi người dùng chọn vào nút theo hướng mũi tên chỉ ở hình trên thì giao diện lịch sử và mục ưa thích sẽ hiện ra với lần lượt hai phần là mục yêu thích và lịch sử .

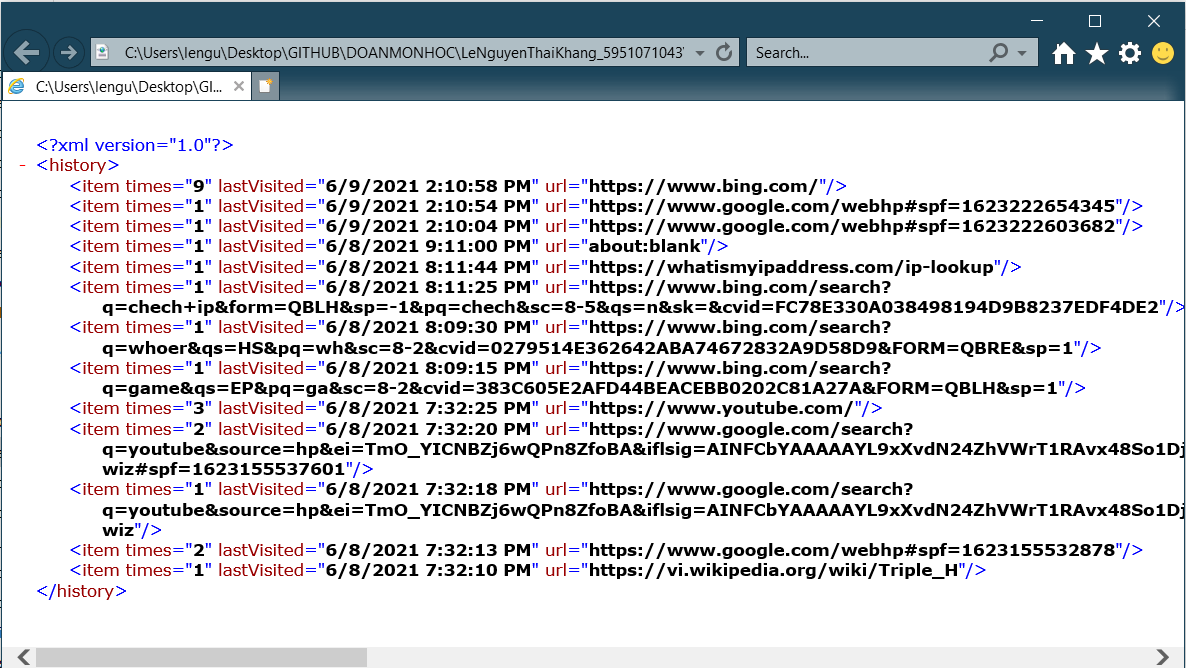
Mục yêu thích được dùng để lưu trữ những trang mình thích dưới dạng dười link mà khi ta nhấn vào dường linh đó ta có thể trực tiếp đến được trang web đó .

Mục lịch sử là nơi để lưu trữ lại những đưởng link mà chúng ta đã truy cập qua , chug1 được sắp xếp theo ngày .

Những dữ liệu về lịch sử và mục yêu thích sẽ được lưu trữ bằng file XML .



Hình 4 Mục yêu thích được lưu trữ bằng file XML

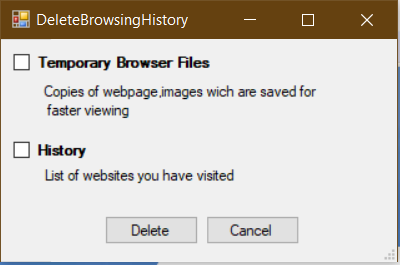


Hình 5 Mục lịch sử được lưu trữ bằng file XML

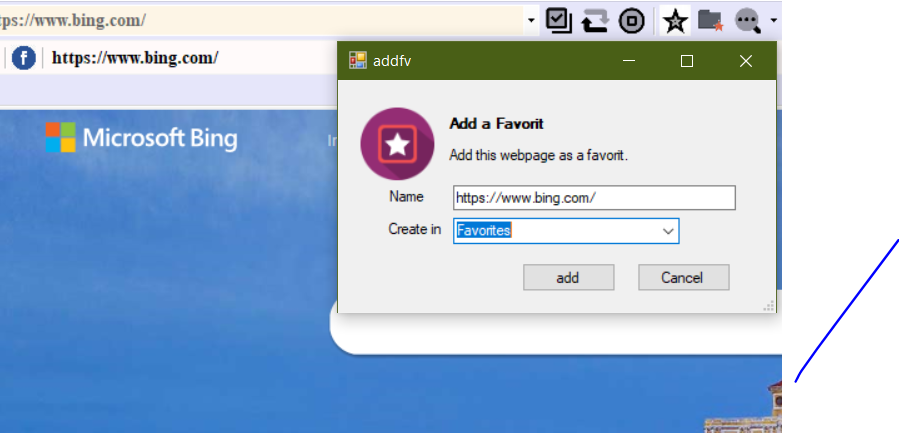
Ngoài ra khi chúng ta muốn xóa đi một đường link nào đó trong mục yêu thích thì chúng ta có thể nhấn chuột phải trực tiếp vào dường link cần xóa và nhấn delete.

Để xóa lịch sử truy cập thì chúng ta có thể truy cập vào Setting => Tools=>

Delete Brower History.



Hình 6 Giao diện xóa lịch sử

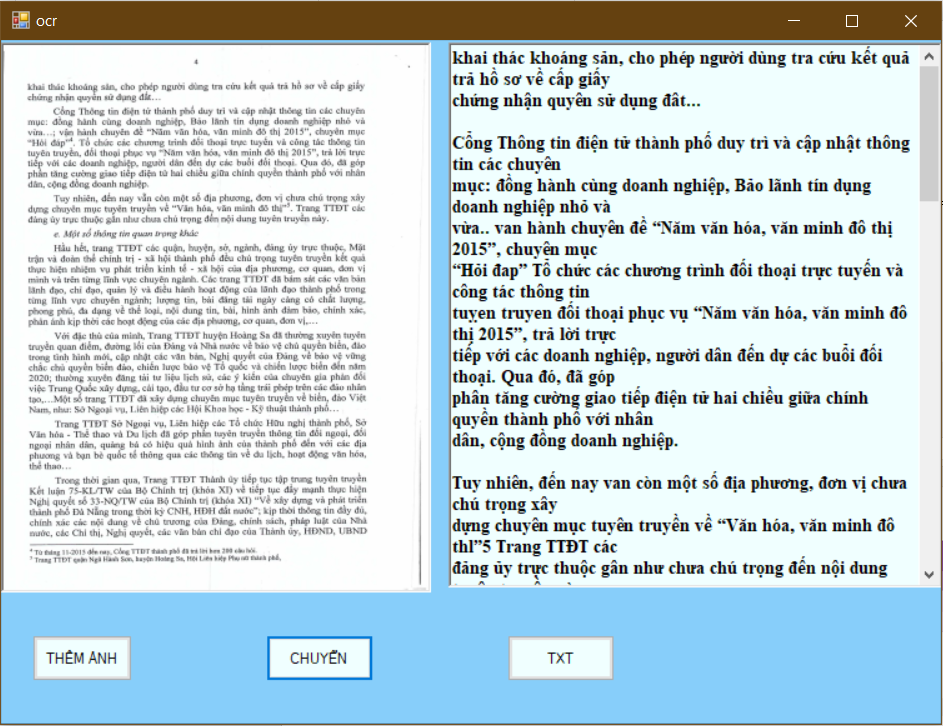
**2.4 Giao diện thêm dường dẫn vào mục ưa thích.** 

Hình 7 Giao diện chức năng thêm mục yêu thích .

Khi truy cập vào một trang nào đó mà chúng ta cảm thấy ưa thích hoặc chúng ta cảm thấy nó quan trọng và cần phải truy cập lại trang đó nhiều lần thì chúng ta có thể nhấn vào nút theo hướng mũi tên chỉ ở hình trên. Khi đó trình duyệt sẽ hiện ra một form để xác nhận yêu cầu. Khi nhấn “OK” thì đường link trên sẽ trực tiếp được lưu vào data và hiển thị trong mục ưa thích .

## 2.5 Giao diện chức năng ORC

Khi người dùng truy cập theo dường dẫn Setting => Tools => OCR , một giao diện sẽ được hiện ra như hình bên dưới .



Hình 8 Giao diện OCR.

Khi chúng ta thêm một văn bản dạng ảnh vào thì chưng trình sẽ sử dụng các thuật toán thuộc thư viện Tesseract để so sánh với những gì được lưu trữ trong file Vietnamese.trainddata và xuất ra được đoạn văn bản chúng ta cần chuyển đổi.

Hiện chức năng này chỉ mới có thể quét được văn bản dạng chữ đánh máy. Không thể quét được chữ viết tay và những ký tự phức tạp cho nên ứng dụng này chủ yếu dùng để chuyển đổi những ảnh có văn bản đơn giản sang dạng văn bản có thể chình sửa và xuất ra file txt.

## 2.6 Chức năng thay đổi Proxy để truy cập mạng từ một IP khác .

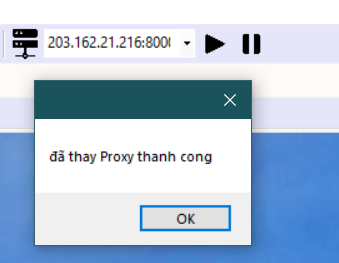
Khi chúng ta nhấn vào biểu tượng Proxy có trên thanh menu thì thông qua Http request của thư viện xNet sẽ lấy được dữ liệu là địa chỉ ip, port .IP và port sẽ được hiển thị trong combobox để cho chúng ta chọn lựa .



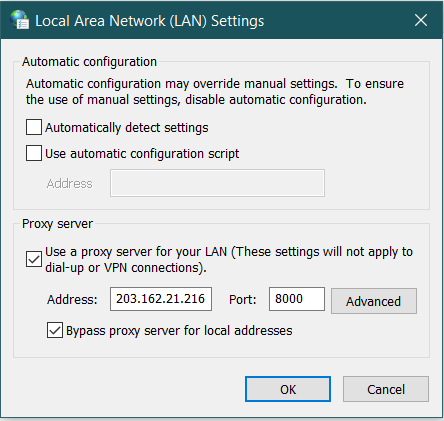
Hình 9 Code Http request.

Hình trên là đoạn code được viết để lấy dữ liệu dựa trên Http request từ trang: https://www.proxynova.com/proxy-server-list/country-vn

Khi đã chọn xong chỉ cần nhấn nút play thì ngay lập tức ip vs port sẽ được thay thế trực tiếp vào máy tính. Khi đó máy tính sẽ truy cập theo ip và port đó .



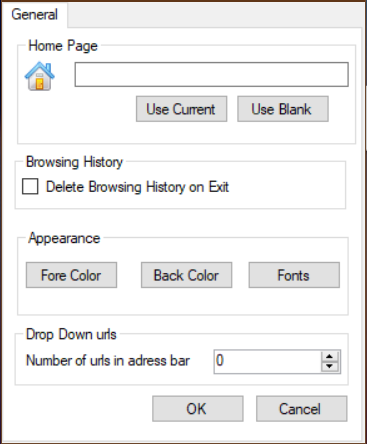
Hình 10 Giao diện thay proxy.



Hình 11 Local Area Network Setting đã được thay proxy.

## 2.7 Một số giao diện chức năng khác

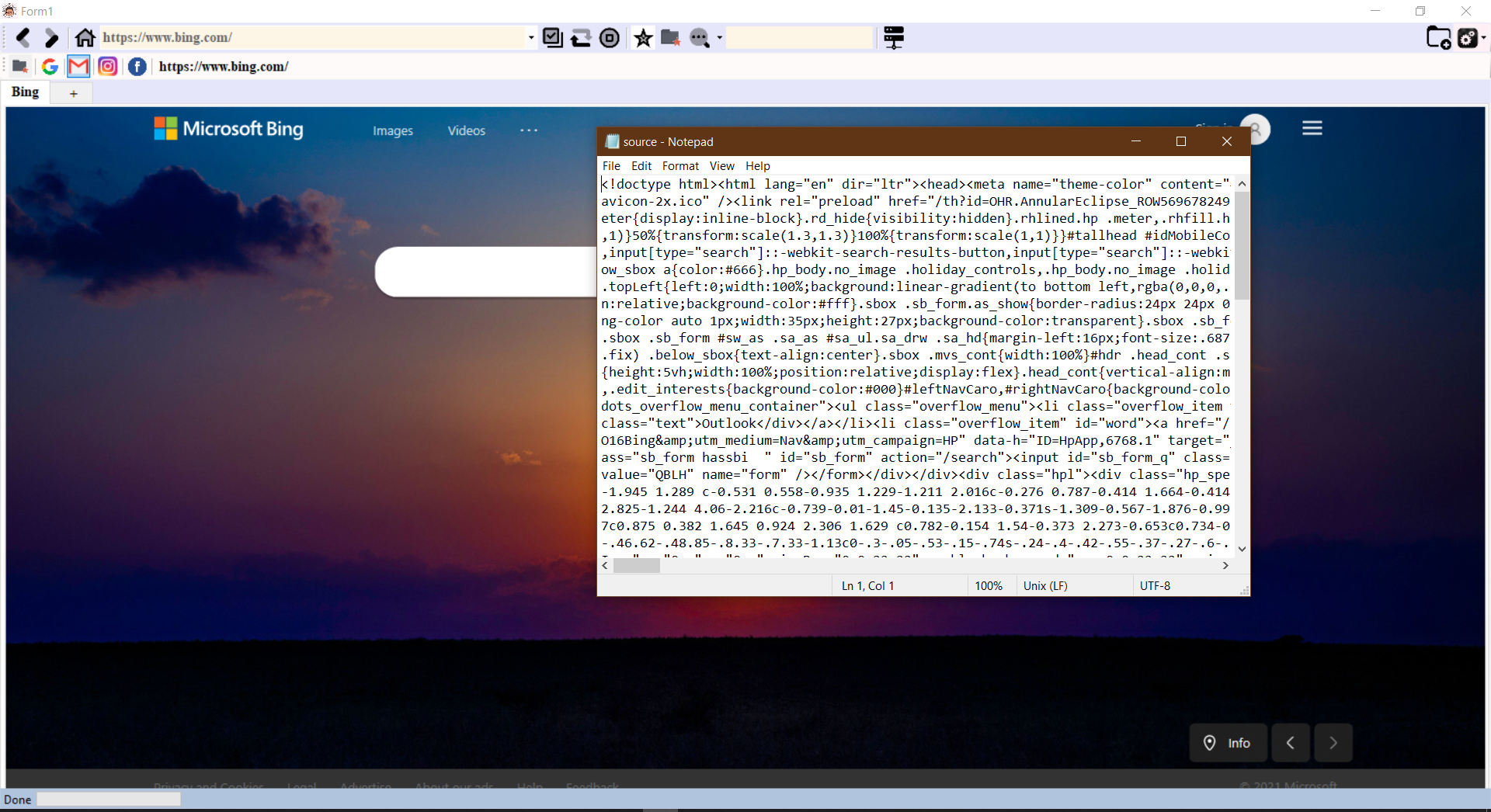
### 2.7.1 Giao diện cài đặt .



Hình 12 Giao diện cài đặt trang

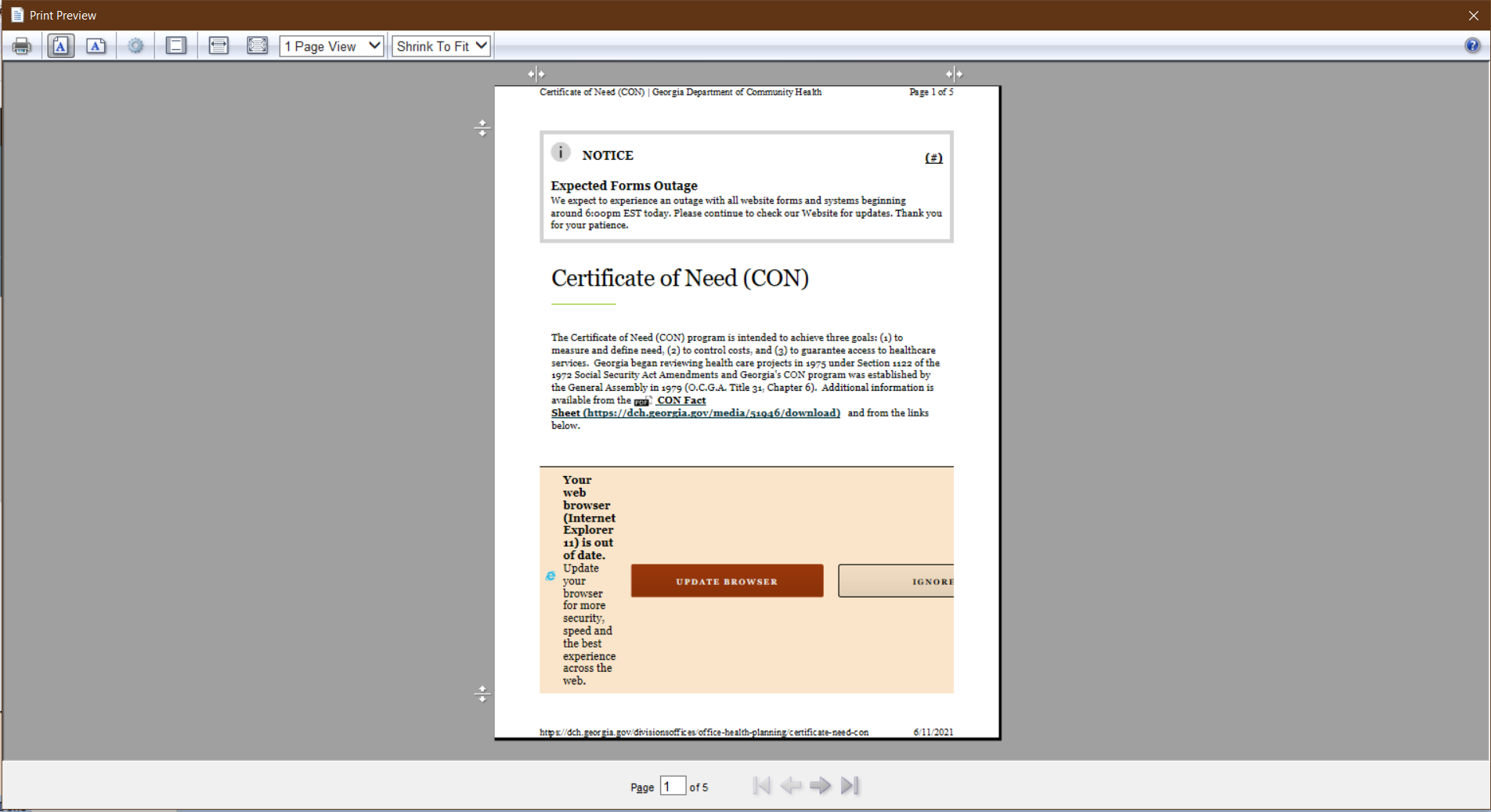
### 2.7.2 Xem mã nguồn html.

Khi nhấn phím F12 thì 1 file TXT sẽ hiện ra và trên đó hiển thị toàn bộ mã nguồn html của trang mà mình vừa truy cập.

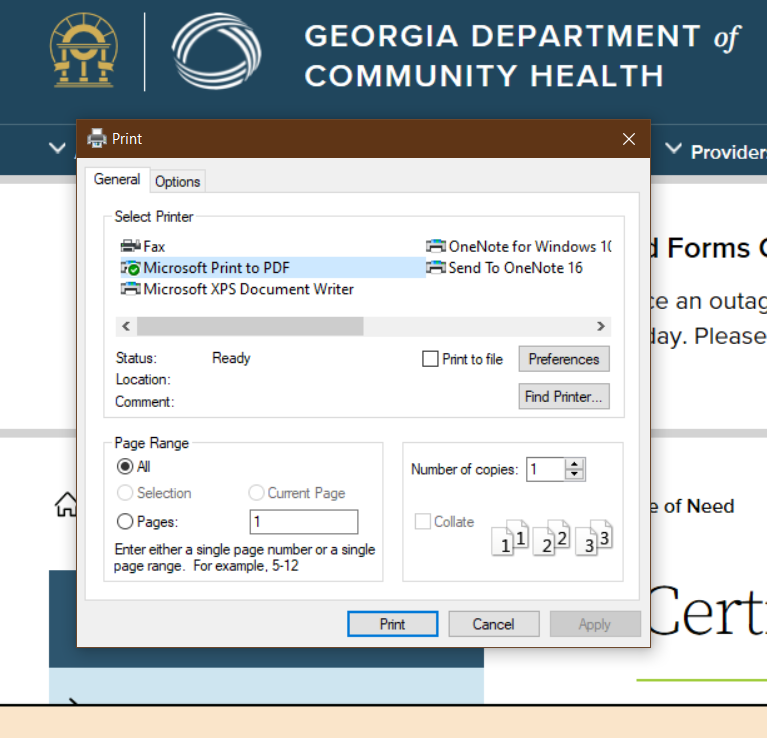


Hình 13 Chức năng hiển thị mã nguồn HTML của trang.

***2.7.3 Chức năng xem trước và in trang.***



Hình 14 Chức năng xem trước trang in.



Hình 15 Chức năng in.

#### Tìm Phương trình đường thẳng hồi quy tuyến tính

#### Đầu tiên chúng ta cần thay đổi tên các cột trong dataframe,tên không có kí tự đặc biệt.

# KẾT LUẬN

## 1. Kết quả

- Nguyên cứu và thực hiện thành công các ứng dụng của winform C# để xây dựng một trình duyệt web với những chức năng cơ bản cần có.

- Hoàngthành nghiên cứu http request để lấy địa chỉ IP và Port từ một trang web cung cấp proxy về để thực hiện chức năng thay proxy của máy khi truy cập web làm cho máy chủ không xác định được địa chỉ IP.

- Hoàn thành các chức năng ban đầu đề ra kèm theo đó là chức năng nhận dạng ký tự quan học OCR dựa trên thư viện Tesseract.

## 2. Tồn tại

- Tuy nhiên hệ thống quản lý còn gặp nhiều thiếu sót trong việc sử lý các sự kiện cần thiết cũng như gặp khó khăn khi truy cập các trang web động.

- Chưa thực sự dễ sử dụng cho người dùng, nhiều trường hợp thực tế áp dụng vào khi sử dụng vẫn chưa thực sự hiệu quả .

- Thiếu sót một số chức năng tiện lợi so với những trình duyệt thông dụng hiện nay.

- Khi thay đổi Proxy thì chỉ truy cập được mạng bằng những địa chỉ IP trong nước. đối với IP nước ngoài thì vẫn chưa sử dụng được.

## 3. Hướng phát triển

- Trong tương lai khi có đầy đủ thời gian và kiến thức em sẽ xây dựng hệ thống trình duyệt với đầy đủ chức năng.

- Có thể thêm những chức năng tiện ích mà những trình duyệt hiện nay đang có .

- Linh hoạt, thân thiện hơn cho người sử dụng.

# TÀI LIỆU THAM KHẢO

[1] <https://www.howkteam.vn/course/lap-trinh-winform-co-ban-27>, “Lập trình C# winform cơ bản” .

[2] <http://thuvienso.vlute.edu.vn/doc/bai-giang-ve-ngon-ngu-c-269423.html> “ Bài giảng ngôn ngữ lập trình C# trường đại học Sư Phạm Kỹ Thuật Vĩnh Long

[3] [tesseract-ocr/tesseract: Tesseract Open Source OCR Engine (main repository) (github.com)](https://github.com/tesseract-ocr/tesseract)

[4] [Tesseract OCR with C# .NET | Iron OCR (ironsoftware.com)](https://ironsoftware.com/csharp/ocr/tutorials/c-sharp-tesseract-ocr/)

[5] [[C#] 4 cách sử dụng PostRequest bằng thư viện xNet - C# Căn Bản (csharpcanban.com)](https://csharpcanban.com/4-cach-su-dung-postrequest-bang-thu-vien-xnet.html)

[6] [X-rus/xNet: xNet - class library for .NET Framework (github.com)](https://github.com/X-rus/xNet).