|  |  |
| --- | --- |
| logo (CMYK)-01 | BỘ GIÁO DỤC VÀ ĐÀO TẠO  **TRƯỜNG ĐẠI HỌC CÔNG NGHỆ TP. HCM** |

**ĐỒ ÁN MÔN HỌC**

**TÌM HIỂU SELENIUM VÀ MONGODB, THU THẬP DỮ LIỆU SẢN PHẨM NHÀ THUỐC PHARMACITY**

Ngành: **KHOA HỌC DỮ LIỆU**

Môn học: **MÃ NGUỒN MỞ TRONG KHOA HỌC DỮ LIỆU**

Giảng viên hướng dẫn : THS. LÊ NHẬT TÙNG

Sinh viên thực hiện :

Hà Thế Anh MSSV: 2286400002

Nguyễn Nhật Nam MSSV: 2286400019

Hoàng Quang Minh MSSV: 2286400017

Lớp: 22DKHA1

TP. Hồ Chí Minh, 2024

|  |  |
| --- | --- |
| logo (CMYK)-01 | BỘ GIÁO DỤC VÀ ĐÀO TẠO  **TRƯỜNG ĐẠI HỌC CÔNG NGHỆ TP. HCM** |

**ĐỒ ÁN MÔN HỌC**

**TÌM HIỂU SELENIUM VÀ MONGODB, THU THẬP DỮ LIỆU SẢN PHẨM NHÀ THUỐC PHARMACITY**

Ngành: **KHOA HỌC DỮ LIỆU**

Môn học: **MÃ NGUỒN MỞ TRONG KHOA HỌC DỮ LIỆU**

Giảng viên hướng dẫn : THS. LÊ NHẬT TÙNG

Sinh viên thực hiện :

Hà Thế Anh MSSV: 2286400002

Nguyễn Nhật Nam MSSV: 2286400019

Hoàng Quang Minh MSSV: 2286400017

Lớp: 22DKHA1

TP. Hồ Chí Minh, 2024

# LỜI CAM ĐOAN

Chúng tôi, Hà Thế Anh, Hoàng Quang Minh và Nguyễn Nhật Nam xin cam đoan rằng:

Mọi thông tin và nghiên cứu được trình bày trong bài báo cáo này là trung thực và khách quan được thu thập và phân tích một cách cẩn thận dựa trên các nguồn chính thống và đáng tin cậy.

Bất kỳ thông tin hoặc ý kiến nào được trích dẫn từ các nguồn khác đều được nêu rõ nguồn gốc và được trích dẫn theo đúng quy định. Chúng tôi xin cam đoan rằng không có bất kỳ sự sao chép hoặc sử dụng thông tin không đúng đắn nào từ các nguồn khác.

Bài báo cáo này là công trình nghiên cứu độc lập của chúng tôi chưa từng được công bố ở bất kỳ nơi nào khác. Tôi cam đoan rằng đã tuân thủ đầy đủ các quy tắc và quy định của môn học bao gồm cả việc tham khảo và sử dụng công cụ nghiên cứu.

Tôi hy vọng rằng bài báo cáo này sẽ cung cấp một cái nhìn tổng quan rõ ràng và toàn diện về chủ đề “Tìm hiểu Selenium và MongoDB, thu thập dữ liệu sản phẩm nhà thuốc Pharmacity" và sẽ đóng góp một phần nhỏ vào lĩnh vực nghiên cứu này.

TP.HCM,Ngày.....tháng.....năm 2024

**Sinh viên**

Hà Thế Anh

Hoàng Quang Minh

Nguyễn Nhật Nam

# NHẬN XÉT CỦA GIẢNG VIÊN HƯỚNG DẪN

*TP. Hồ Chí Minh, ngày tháng năm 2023*

**Giáo viên hướng dẫn**

**(**Ký tên,đóng dấu**)**

# MỤC LỤC

[LỜI CAM ĐOAN i](#_Toc180357459)

[NHẬN XÉT CỦA GIẢNG VIÊN HƯỚNG DẪN ii](#_Toc180357460)

[MỤC LỤC iii](#_Toc180357461)

[DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT VÀ TỪ KHÓA v](#_Toc180357462)

[CHƯƠNG 1. TỔNG QUAN 1](#_Toc180357463)

[1.1. Giới thiệu đề tài 1](#_Toc180357464)

[1.2. Nhiệm vụ của đồ án 1](#_Toc180357465)

[*1.2.1.* *Tính cấp thiết của đề tài* 1](#_Toc180357466)

[*1.2.1.* *Ý nghĩa khoa học và thực tiễn của đề tài* 2](#_Toc180357467)

[1.3. Mục tiêu 3](#_Toc180357468)

[*1.3.1.* *Mục tiêu tổng quan* 3](#_Toc180357469)

[*1.3.2.* *Mục tiêu cụ thể* 3](#_Toc180357470)

[1.4. Đối tượng và phạm vi 3](#_Toc180357471)

[*1.4.1.* *Đối tượng* 4](#_Toc180357472)

[*1.4.2.* *Phạm vi* 4](#_Toc180357473)

[1.5. Phương pháp nghiên cứu 4](#_Toc180357474)

[*1.5.1.* *Phương pháp nghiên cứu sơ bộ* 4](#_Toc180357475)

[*1.5.2.* *Phương pháp nghiên cứu tài liệu* 4](#_Toc180357476)

[*1.5.3.* *Phương pháp nghiên cứu thống kê* 4](#_Toc180357477)

[*1.5.4.* *Phương pháp thực nghiệm* 4](#_Toc180357478)

[*1.5.5.* *Phương pháp đánh giá* 4](#_Toc180357479)

[CHƯƠNG 2. CƠ SỞ LÝ THUYẾT 5](#_Toc180357480)

[CHƯƠNG 3. KẾT QUẢ THỰC NGHIỆM 6](#_Toc180357481)

[TÀI LIỆU THAM KHẢO 7](#_Toc180357482)

# DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT VÀ TỪ KHÓA

# TỔNG QUAN

## Giới thiệu đề tài

Trong bối cảnh cộng nghệ thông tin ngày càng phát triển, việc thu thập và quản lý dữ liệu sản phẩm một cách hiệu quả là yếu tố then chốt đối với các doanh nghiệp, đặc biệt là trong lĩnh vực dược phẩm. Để tối ưu hóa được quy trình này, công cụ tự động như Selenium được sử dụng với công việc thu thập dữ liệu từ các website một cách tự động, trong khi MongoDB đóng vai trò là một kho lưu trữ và quản lý được khối lượng thông tin sản phẩm một cách linh hoạt. Áp dụng hai công cụ này lại với nhau giúp các doanh nghiệp và trong đó có nhà thuốc Pharmacity có thể dễ dàng theo dõi và quản lý sản phẩm, nhanh chóng cập nhật được thông tin, từ đó nâng cao hiệu quả kinh doanh và cải thiện khả năng cạnh tranh.

## Nhiệm vụ của đồ án

Nhiệm vụ của đề tài “Tìm hiểu Selenium và MongoDB, thu thập dữ liệu sản phẩm nhà thuốc Pharmacity” là áp dụng được Selenium để tự động quá trình thu thập dữ liệu sản phẩm từ website của nhà thuốc Pharmacity và lưu trữ, quản lý bằng MongoDB. Thông qua quá trình thu thập và quản lý dữ liệu này, doanh nghiệp sẽ có cái nhìn sâu sắc hơn về cái sản phẩm từ đó tối ưu hóa chiến lược bán hàng, nâng cao chất lượng dịch vụ và tăng khả năng cạnh tranh trên thị trường.

### *Tính cấp thiết của đề tài*

Trong thời đại số hóa, việc thu thập và quản lý dữ liệu hiểu quả là yếu tố then chốt để doanh nghiệp tồn tại và phát triển. Nhà thuốc Pharmacity là một chuỗi của hàng bán thuốc lẻ lớn đang phải đối mặt với sự cạnh tranh khốc liệt từ nhiều đối thủ. Để đáp ứng được nhu cầu của thị trường và duy trì cạnh tranh thì doanh nghiệp cần phải ứng dụng được công nghệ tự động hóa quy thu thập và quản lý dữ liệu sản phẩm.

Dự án “Tìm hiểu Selenium và MongoDB, thu thập dữ liệu sản phẩm nhà thuốc Pharmacity” được triển khai nhằm giải quyết được bài toán thu thập và quản lý dữ liệu sản phẩm. Cụ thể việc thu thập và quản lý dữ liệu mang lại cho nhà thuốc Pharmacitiy những lợi ích sau:

+ Tối ưu hóa được chiến lược kinh doanh: Việc thu thập dữ liệu sẽ giúp doanh nghiệp hiểu rõ hơn về thị trường, từ đó đưa ra được chiến lược marketing và bán hàng chính xác tối ưu hóa hiệu quả kinh doanh.

+ Cải thiện dịch vụ và trải nghiệm của người dùng: Với dữ liệu được cập nhật tự động hóa liên tục, doanh nghiệp có thể nhanh chóng bắt kịp xu hướng và đáp ứng nhu cầu của khách hàng, nâng cao sự trải nghiệm dịch vụ của người dùng.

+ Tối ưu hóa vận hành: Dữ liệu chính xác về sản phẩm giúp doanh nghiệp quản lý kho hàng hiệu quả hơn, giảm lãng phí thời gian và tối ưu hóa quá trình phân phối.

+ Đưa ra quyết định dựa trên lữ liệu: Doanh nghiệp có thể dựa trên dữ liệu chính xác để đưa ra quyết định thay đổi chiến lược một cách đúng đắn.

+ Nâng cao khả năng cạnh tranh: Ứng dụng công nghệ vào thu thập dữ liệu, doanh nghiệp nhanh chóng thích ứng được xự thay đổi của thị trường và thay đổi chiến lược cho phù hợp, tăng cường vị thế cạnh tranh.

Với những lợi ích trên dự án không chỉ giúp doanh nghiệp nâng cao hiểu quả mà còn tăng cường khả năng và phát triển bền vững trong lĩnh vực dược phẩm.

### *Ý nghĩa khoa học và thực tiễn của đề tài*

**Ý nghĩa khoa học:** Dự án của nhóm chúng tôi sẽ đóng góp vào lĩnh vực tự động hóa thu thập dữ liệu bằng cách sử dụng công cụ Selenium và MongoDB. Việc áp dụng hai công cụ này giúp đơn giản hóa quá trình thu thập dữ liệu từ website, đặc biệt là dữ liệu về sản phẩm từ hệ thống nhà thuốc. Dự án không chỉ mở rộng kiến thức về cách tự động hóa trong thu thập dữ liệu mà còn giúp tích hợp những phương pháp xử lý dữ liệu cơ bản để khai thác thông tin hữu ích từ các nguồn dữ liệu trực tuyến. Điều này giúp các nhà phân tích và chuyên gia những công cụ hiện đại hơn trong công việc xử lý dữ liệu, tiềm năng ứng dụng được trong các dự án lớn.

**Ý nghĩa thực tiễn:** Dự án này mang lại nhiều lợi ích cho nhà thuốc Pharmacity trong việc tự động hóa thư thập và quản lý dữ liệu sản phẩm. Nhờ đó, dữ liệu sản phẩm được cập nhật liên tục giúp doanh nghiệp hiểu rõ hơn về xu hướng tiêu thụ trên thị trường, đẩy nhanh được chiến lược marketing và cải thiện dịch vụ đối với khách hàng. Việc áp dụng công cụ này giúp doanh nghiệp giảm bớt được thời gian và nguồn nhân lực, nắm bắt nhanh chóng trước những thay đổi của thị trường. Điều này đặc biệt quan trọng khi doanh nghiệp cần phải duy trì được lợi thế cạnh tranh trong môi trường phát triển hiện nay. Phương pháp thu thập và quản lý dữ liệu từ dự án có thể được mở rộng ra nhiều ngành hàng khác . Giúp nâng cao hiệu quả quản lý sản phẩm trong kinh doanh.

Ngoài ra, kết quả của dự án mang lại giúp doanh nghiệp có thể đưa ra những chiến lược dựa trên dữ liệu một cách chính xác thay vì dựa trên cảm tính. Điều này góp phần hỗ trở các nhà quản lý trong việc phát triển bền vững trong hoạt động kinh doanh. Nhờ sự kết hợp của lý thuyết và tự động hóa vào trong ứng dụng thực tiễn, dự án không chỉ mang lại giá trị khoa học mà còn đóng góp quan trọng vào việc nâng cao hiệu và phát triển bền vững của doanh nghiệp.

## Mục tiêu

### *Mục tiêu tổng quan*

Dự án này nhằm phát triển phương pháp tự động thu thập dữ liệu và quản lý dữ liệu sản phẩm nhà thuốc Pharmacity bằng công cụ Selenium để thu thập tự động dữ liệu, lưu trữ và quản lý dữ bằng MongoDB. Mục tiêu giúp doanh nghiệp nắm bắt nhanh chóng và chính xác thông tin sản phẩm, từ đó đưa ra được chiến lược kinh doanh hiệu quả, cải thiện hoạt động và tăng cường khả năng cạnh tranh trên thị trường

### *Mục tiêu cụ thể*

(chưa viết)

## Đối tượng và phạm vi

### *Đối tượng*

Đối tượng nghiên cứu của dự án là các sản phẩm đang được bán trên hệ thống trực tuyến của nhà thuốc Pharmacity. Dữ liệu sẽ bao gồm thông tin về sản phẩm, giá cả đánh giá cảu khách hàng. Mục tiêu của dự án là thu thập và phân tích dữ liệu để hiểu rõ hơn về những yếu tố ảnh hưởng đến hoạt động kinh doanh của Pharmacity, từ đó để xuất các giả pháp giúp cái thiện hiệu suất và tối ưu hóa quá trình vẫn hành.

### *Phạm vi*

Phạm vi phân tích tập trung vào

## Phương pháp nghiên cứu

### *Phương pháp nghiên cứu sơ bộ*

### *Phương pháp nghiên cứu tài liệu*

### *Phương pháp nghiên cứu thống kê*

### *Phương pháp thực nghiệm*

### *Phương pháp đánh giá*

# CƠ SỞ LÝ THUYẾT

# KẾT QUẢ THỰC NGHIỆM

# TÀI LIỆU THAM KHẢO