|  | BỘ GIÁO DỤC VÀ ĐÀO TẠO  **TRƯỜNG ĐẠI HỌC CÔNG NGHỆ TP. HCM** |
| --- | --- |

**ĐỒ ÁN MÔN HỌC**

**<TÌM HIỂU VỀ CÔNG CỤ SELENIUM VÀ MONGODMONGODB>**

Ngành: **<KHOA HỌC DỮ LIỆU >**

Chuyên ngành: **<KHOA HỌC DỮ LIỆU >**

Giảng viên hướng dẫn :ThS. LÊ NHẬT TÙNG

Sinh viên thực hiện :

2286400003-NGUYỄN HOÀNG ANH

2286400041-BẠCH QUANG TÙNG

2286400004-NGUYỄN TUẤN ANH

Lớp: 22DKHA1

TP. Hồ Chí Minh, <2024>

NHẬN XÉT CỦA GIẢNG VIÊN HƯỚNG DẪN . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .. . . . . . . . . . . . . . . TPHCM, Ngày. . . . . . tháng. . . . . . năm 2024 Giáo viên hướng dẫn

(Ký tên, đóng dấu)

**LỜI CAM ĐOAN**

Chúng tôi, Nguyễn Tuấn Anh, Bạch Quang Tùng và Nguyễn Hoàng Anh, xin cam đoan rằng:

Tất cả thông tin và nghiên cứu được trình bày trong bài báo cáo này là chính xác, trung thực và khách quan. Dữ liệu đã được thu thập và phân tích một cách cẩn thận từ các nguồn tin cậy và chính thống.

Mọi thông tin hoặc quan điểm được trích dẫn từ nguồn khác đều được ghi rõ nguồn gốc và tuân thủ quy định về trích dẫn. Chúng tôi cam kết rằng không có bất kỳ hành vi sao chép hoặc sử dụng thông tin không hợp lệ nào từ các nguồn bên ngoài.

Bài báo cáo này là kết quả nghiên cứu độc lập của chúng tôi và chưa từng được công bố ở bất kỳ đâu trước đây. Chúng tôi đã thực hiện đầy đủ các quy tắc và quy định của môn học, bao gồm cả việc tham khảo và sử dụng công cụ nghiên cứu.

Chúng tôi mong rằng bài báo cáo này sẽ mang lại cái nhìn tổng quan rõ ràng và toàn diện về chủ đề “Dạy học lập trình trên trang web titv”, góp phần vào sự phát triển của lĩnh vực nghiên cứu này.

TPHCM, ngày 09 tháng 06 năm 2024  
Sinh viên  
Nguyễn Tuấn Anh  
Bạch Quang Tùng  
Nguyễn Hoàng Anh

**Mục Lục**

1. **Tổng Quan**  
   Giới thiệu  
   1.1. Mục tiêu của báo cáo  
   1.2. Phương pháp nghiên cứu
2. **Giới thiệu khái quát về công cụ Selenium và MongoDB**  
   2.1. Định nghĩa và lịch sử phát triển  
   2.2. Các thành phần của Selenium  
   • 2.2.1. Selenium WebDriver  
   • 2.2.2. Selenium IDE  
   • 2.2.3. Selenium Grid  
   2.3. Tính năng nổi bật  
   2.4. Ứng dụng của Selenium trong thu thập dữ liệu  
   2.5. Giới thiệu về MongoDB và ứng dụng trong lưu trữ dữ liệu
3. **Tổng quan về website TITV**  
   3.1. Lịch sử hình thành và phát triển  
   3.2. Mô hình kinh doanh và dịch vụ  
   3.3. Thị trường và đối thủ cạnh tranh
4. **Phương pháp thu thập dữ liệu**  
   4.1. Xác định mục tiêu thu thập dữ liệu  
   4.2. Thiết kế quy trình thu thập dữ liệu  
   4.3. Sử dụng Selenium trong thu thập dữ liệu từ TITV  
   • 4.3.1. Cài đặt và cấu hình Selenium  
   • 4.3.2. Viết mã tự động hóa  
   • 4.3.3. Xử lý và lưu trữ dữ liệu bằng MongoDB
5. **Kết quả thu thập dữ liệu**  
   5.1. Phân tích dữ liệu thu thập được  
   5.2. Đánh giá chất lượng dữ liệu  
   5.3. So sánh với dữ liệu nguồn khác  
   5.4. Phân tích dữ liệu bằng MongoDB
6. **Thảo luận**  
   6.1. Những thách thức trong quá trình thu thập dữ liệu  
   6.2. Đề xuất cải tiến quy trình thu thập dữ liệu
7. **Kết luận**  
   7.1. Tóm tắt kết quả nghiên cứu (tích hợp video cào dữ liệu)  
   7.2. Định hướng nghiên cứu tiếp theo

**Tài liệu tham khảo**

**Phụ lục**  
9.1. Mã nguồn mẫu sử dụng Selenium  
9.2. Biểu đồ và bảng số liệu liên quan

**Nội dung :**

**1. Tổng Quan**

**Giới thiệu**

Trong thời đại công nghệ số hiện nay, việc thu thập và phân tích dữ liệu từ các nguồn học tập trực tuyến đang trở nên vô cùng quan trọng. Website TITV (Trang dạy lập trình và công nghệ thông tin) đã trở thành một nền tảng phổ biến cung cấp các khóa học và tài nguyên học tập về lập trình. Nghiên cứu và thu thập dữ liệu từ TITV không chỉ giúp nâng cao chất lượng giảng dạy mà còn tạo ra những hiểu biết sâu sắc về nhu cầu và hành vi của người học.

Báo cáo này sẽ tập trung vào việc sử dụng các công cụ như Selenium và MongoDB để thu thập và quản lý dữ liệu từ TITV. Selenium, với khả năng tự động hóa các thao tác trên trình duyệt, cho phép chúng ta thu thập dữ liệu một cách hiệu quả từ các trang web động. Trong khi đó, MongoDB sẽ hỗ trợ trong việc lưu trữ và phân tích dữ liệu thu thập được.

**1.1. Mục tiêu của báo cáo**

Mục tiêu của báo cáo này là:

* Phân tích và áp dụng Selenium trong việc thu thập dữ liệu từ website TITV.
* Nghiên cứu ứng dụng của MongoDB như một giải pháp lưu trữ và quản lý dữ liệu hiệu quả.
* Đánh giá chất lượng dữ liệu thu thập từ TITV và đề xuất các phương pháp cải tiến trong quy trình thu thập.

**1.2. Phương pháp nghiên cứu**

Để đạt được các mục tiêu nêu trên, báo cáo sẽ sử dụng phương pháp nghiên cứu kết hợp giữa định tính và định lượng. Phương pháp định tính sẽ bao gồm phân tích tài liệu và nghiên cứu trường hợp từ các báo cáo trước đây. Phương pháp định lượng sẽ tập trung vào việc thu thập dữ liệu thực tế từ TITV bằng công cụ Selenium, sau đó lưu trữ và phân tích dữ liệu bằng MongoDB.

**2. Giới thiệu khái quát về công cụ Selenium và MongoDB**

**2.1. Định nghĩa và lịch sử phát triển**

Selenium là một bộ công cụ mã nguồn mở được thiết kế để tự động hóa các trình duyệt web. Ra mắt lần đầu vào năm 2004, Selenium đã nhanh chóng trở thành một trong những công cụ phổ biến nhất trong lĩnh vực kiểm thử phần mềm và tự động hóa web. Qua thời gian, Selenium đã phát triển với nhiều thành phần khác nhau như Selenium WebDriver, Selenium IDE và Selenium Grid, cho phép người dùng linh hoạt trong việc tự động hóa quy trình kiểm thử và thu thập dữ liệu.

MongoDB, một hệ quản trị cơ sở dữ liệu NoSQL, được phát triển vào năm 2007. MongoDB cho phép lưu trữ dữ liệu theo dạng tài liệu, giúp việc xử lý và quản lý dữ liệu trở nên linh hoạt hơn so với các hệ quản trị cơ sở dữ liệu quan hệ truyền thống. Sự phát triển của MongoDB đã phản ánh xu hướng ngày càng tăng của việc sử dụng dữ liệu phi cấu trúc và bán cấu trúc trong các ứng dụng hiện đại.

**2.2. Các thành phần của Selenium**

* **2.2.1. Selenium WebDriver**: Đây là thành phần chính của Selenium, cho phép lập trình viên tương tác với các trình duyệt web một cách trực tiếp. WebDriver hỗ trợ nhiều ngôn ngữ lập trình như Java, Python, C#, và Ruby, giúp linh hoạt hơn trong việc phát triển các bài kiểm thử.
* **2.2.2. Selenium IDE**: Là một công cụ đơn giản, cho phép người dùng ghi lại các tương tác trên trình duyệt và phát lại chúng. Selenium IDE là lựa chọn tốt cho những người mới bắt đầu làm quen với tự động hóa kiểm thử.
* **2.2.3. Selenium Grid**: Selenium Grid cho phép chạy các bài kiểm thử trên nhiều máy và trình duyệt khác nhau cùng một lúc, giúp tiết kiệm thời gian kiểm thử và tăng cường hiệu suất.

**2.3. Tính năng nổi bật**

Selenium cung cấp nhiều tính năng nổi bật như khả năng tương thích với nhiều trình duyệt khác nhau, hỗ trợ nhiều ngôn ngữ lập trình, và khả năng mở rộng linh hoạt. Việc có thể chạy kiểm thử song song thông qua Selenium Grid cũng là một điểm mạnh lớn, giúp giảm thời gian kiểm thử tổng thể.

**2.4. Ứng dụng của Selenium trong thu thập dữ liệu**

Selenium không chỉ được sử dụng trong kiểm thử phần mềm mà còn được ứng dụng rộng rãi trong việc thu thập dữ liệu từ các trang web động. Bằng cách tự động hóa các thao tác trên trình duyệt, Selenium cho phép thu thập thông tin từ các trang web mà không cần phải can thiệp thủ công, tiết kiệm thời gian và công sức.

**2.5. Giới thiệu về MongoDB và ứng dụng trong lưu trữ dữ liệu**

MongoDB là một giải pháp lưu trữ dữ liệu mạnh mẽ, cho phép lưu trữ dữ liệu theo dạng tài liệu JSON. Sự linh hoạt trong cấu trúc dữ liệu của MongoDB giúp cho việc mở rộng và quản lý dữ liệu trở nên dễ dàng hơn. Trong báo cáo này, MongoDB sẽ được sử dụng để lưu trữ và quản lý dữ liệu thu thập từ TITV, giúp dễ dàng phân tích và truy xuất thông tin khi cần thiết.

**3. Tổng Quan về website TITV**

**3.1 Lịch sử hình thành và phát triển**

Website TITV (<https://titv.vn/>) ra đời với mục tiêu trở thành nền tảng học tập và đào tạo công nghệ thông tin, đặc biệt là trong lĩnh vực lập trình. Ban đầu, TITV chỉ tập trung vào các khoá học lập trình cơ bản, nhưng qua thời gian, nó đã mở rộng thành một hệ sinh thái học tập toàn diện với các khoá học phong phú về công nghệ và lập trình, từ cơ bản đến nâng cao. Các khóa học được thiết kế để phục vụ từ người mới bắt đầu đến những lập trình viên chuyên nghiệp, hỗ trợ người học cập nhật kiến thức và kỹ năng mới nhất trong ngành. Sự phát triển không ngừng của TITV đã đưa nó trở thành một trong những trang học tập trực tuyến hàng đầu tại Việt Nam trong lĩnh vực này.

**3.2 Mô hình kinh doanh và dịch vụ**

TITV hoạt động theo mô hình freemium, kết hợp giữa các tài nguyên miễn phí và các khoá học trả phí. Người dùng có thể truy cập vào một số tài liệu học tập miễn phí, nhưng với các khóa học nâng cao hoặc chuyên sâu, họ cần trả phí để học. Điều này mang đến sự linh hoạt cho người học, giúp họ tiếp cận kiến thức cơ bản mà không cần tốn kém, và lựa chọn trả phí khi muốn đào sâu vào các chủ đề đặc biệt. TITV cũng cung cấp các dịch vụ tư vấn nghề nghiệp và hỗ trợ học tập thông qua các dự án thực tế, giúp người học ứng dụng kiến thức vào thực tiễn. Thông qua cách tiếp cận này, TITV không chỉ tạo thu nhập bền vững mà còn xây dựng cộng đồng học tập năng động và kết nối người học với các nhà tuyển dụng tiềm năng.

**3.3 Thị trường và đối thủ cạnh tranh**

TITV hiện đang cạnh tranh trong thị trường giáo dục trực tuyến ngày càng phát triển tại Việt Nam, với các đối thủ nổi bật như Udemy, Coursera, Kyna và Edumall. Điểm mạnh của TITV nằm ở khả năng cung cấp nội dung học tập bằng tiếng Việt với giá cả phải chăng và phù hợp với người học trong nước. Ngoài ra, TITV đã điều chỉnh nội dung để đáp ứng nhu cầu cụ thể của người dùng Việt Nam, như cung cấp các khóa học về lập trình căn bản cho người mới hoặc các khoá đào tạo nâng cao cho lập trình viên chuyên nghiệp. TITV cũng tích hợp các dự án thực tế vào khóa học của mình, giúp người học có cái nhìn sâu sắc về ứng dụng thực tế của kiến thức lập trình.