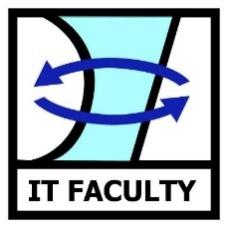


**ĐẠI HỌC ĐÀ NẴNG**

**TRƯỜNG ĐẠI HỌC BÁCH KHOA**

 **KHOA CÔNG NGHỆ THÔNG TIN**



**BÁO CÁO**

**ĐỒ ÁN CÔNG NGHỆ PHẦN MỀM**

**ĐỀ TÀI: Xây dựng Website tổng hợp và quản lý thông tin việc làm**

**GIẢNG VIÊN HƯỚNG DẪN**

**TS. Nguyễn Tấn Khôi**

**SINH VIÊN THỰC HIỆN**

|  |  |  |
| --- | --- | --- |
| **HỌ VÀ TÊN** | **MÃ SINH VIÊN** | **NHÓM** |
| Từ Đức Mạnh | 102220328 | 2 |
| Trần Quang Thắng | 102220338 | 2 |

**Đà Nẵng, 05/2024**

# LỜI MỞ ĐẦU

Trong thời đại công nghệ phát triển mạnh mẽ như hiện nay, các ứng dụng di động ngày càng trở nên phổ biến và được nhiều người sử dụng. Các ứng dụng này mang lại nhiều tiện ích cho người dùng, giúp họ có thể dễ dàng thực hiện các công việc khác nhau mà nhanh gọn chính xác.

Trong hoạt động tìm việc và tuyển việc làm đang được tối ưu hóa công nghệ ngày càng nhiều. Các trang web, ứng dụng tìm việc làm giúp cho người tìm việc và nhà tuyển dụng có thể nhanh chóng tiếp cận nhau một cách hiệu quả nhanh chóng hơn, đồng thời một phần nào đó giúp cản thiện hiện tượng thất nghiệp trong thời gian gần đây.

Chúng em xin chân thành cảm ơn thầy Nguyễn Tấn Khôi đã cung cấp, hỗ trợ cho chúng em những kiến thức cần thiết, cũng như tinh thần học tập và sáng tạo trong học phần này.

Trong quá trình thực hiện đồ án, mặc dù đã cố gắng hết sức song do thời gian và kiến thực có hạn nên chúng em không thể tránh khỏi những thiếu sót. Vì vậy chúng em mong nhận được sự thông cảm và giúp đỡ của thầy cô.

Mục tiêu thực hiện đồ án là sử dụng kiến thức lập trình hướng đối tượng và cấu trúc dữ liệu vào các dự án thực tế nhằm hiểu, tích lũy kinh nghiệm và trau dồi khả năng phân tích thiết kế ứng dụng.

Sơ lược đồ án gồm 5 chương chính:

Chương 1: Tổng quan đề tài.

Chương 2: Cơ sở lý thuyết.

Chương 3: Phân tích và thiết kế hệ thống.

Chương 4: Xây dựng website.

Chương 5: Kết luận và hướng phát triển.

MỤC LỤC

[LỜI MỞ ĐẦU 2](#_Toc166707103)

[CHƯƠNG 1: TỔNG QUAN ĐỀ TÀI 5](#_Toc166707104)

[**1.1. Tên đề tài** 5](#_Toc166707105)

[**1.2. Giới thiệu và lý do chọn đề tài** 5](#_Toc166707106)

[**1.3. Mục tiêu** 5](#_Toc166707107)

[**1.4. Phương pháp nghiên cứu** 5](#_Toc166707108)

[CHƯƠNG 2: CƠ SỞ LÝ THUYẾT 6](#_Toc166707109)

[**2.1. Giới thiệu Python** 6](#_Toc166707110)

[2.1.1. Crawl dữ liệu 6](#_Toc166707111)

[2.1.2. Selenium vs BeautifulSoup 7](#_Toc166707112)

[**2.2 XAMPP/MySQL** 8](#_Toc166707113)

[**2.3 NodeJS** 9](#_Toc166707114)

[**2.4 ReactJS** 10](#_Toc166707115)

[CHƯƠNG 3: PHÂN TÍCH VÀ THIẾT KẾ HỆ THỐNG 12](#_Toc166707116)

[**3.1.** **Phân tích thuật toán** 12](#_Toc166707117)

[3.1.1. Thu tập dữ liệu 12](#_Toc166707118)

[3.1.2. Thuật toán crawler 12](#_Toc166707119)

[3.1.3. Kiểm tra trùng lặp dữ liệu 15](#_Toc166707120)

[**3.2. Chức năng hệ thống** 16](#_Toc166707121)

[3.2.1. Admin 16](#_Toc166707122)

[3.2.2. Nhà tuyển dụng (Công ty - Company) 16](#_Toc166707123)

[3.2.3. Người tìm việc 16](#_Toc166707124)

[**3.3 Cơ sở dữ liệu** 17](#_Toc166707125)

[**3.4. Sơ đồ** 19](#_Toc166707126)

[**3.5. Đặc tả UseCase** 21](#_Toc166707127)

[3.5.1. Đăng nhập 21](#_Toc166707128)

[3.5.2. Đăng ký 21](#_Toc166707129)

[3.5.3. Tìm kiếm việc làm 22](#_Toc166707130)

[3.5.4. Nhận việc làm qua Gmail 22](#_Toc166707131)

[3.5.5. Xoá công việc đăng tuyển 23](#_Toc166707132)

[3.5.6 Thêm công việc đăng tuyển 24](#_Toc166707133)

[3.5.7 Sửa công việc đăng tuyển 24](#_Toc166707134)

[3.5.8. Xoá tài khoản người dùng (Admin) 25](#_Toc166707135)

[3.5.9. Sửa thông tin tài khoản người dùng 25](#_Toc166707136)

[3.5.10. Tìm kiếm thông tin tài khoản người dùng 26](#_Toc166707137)

[3.5.11. Phần quyền từng chức năng cho mỗi tài khoản 27](#_Toc166707138)

[CHƯƠNG 4: XÂY DỰNG WEBSITE 28](#_Toc166707139)

[**4.1. Giao diện đăng nhập, đăng kí** 28](#_Toc166707140)

[**4.2. Giao diện người dùng** 30](#_Toc166707141)

[**4.3. Giao diện tài khoản Admin** 34](#_Toc166707142)

[**4.4. Giao diện tài khoản công ty** 36](#_Toc166707143)

[**4.5. Nhận xét** 39](#_Toc166707144)

[CHƯƠNG 5: KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN 40](#_Toc166707145)

[**5.1. Kết luận** 40](#_Toc166707146)

[**5.2. Hướng phát triển** 40](#_Toc166707147)

[TÀI LIỆU THAM KHẢO 41](#_Toc166707148)

Danh sách hình ảnh

Hình 1: BeautifulSoup Python

Hình 2: Trang chính MySQL

Hình 3: NodeJS

Hình 4: ReactJS

Hình 5: Dữ Liệu khi được lưu ở file CSV

Hình 6: Dữ liệu khi được đẩy lên MySQL

Hình 7: Mô hình cấu trúc hệ thống hướng đối tượng

Hình 8: Usecase người dùng (User)

Hình 9: Usercase Company

Hình 10: Usecase Admin

Hình 11: Giao diện đăng nhập

Hình 12: Giao diện đăng kí Người dùng

Hình 13: Giao diện đăng kí Công ty

Hình 14: Giao diện trang Home

Hình 15: Giao diện trang công việc

Hình 16: Giao diện chi tiết công việc

Hình 17: Giao diện trang Công ty

Hình 18: Giao diện chi tiết Công ty

Hình 19: Giao diện chỉnh sửa thông tin người dùng

Hình 20: Tạo CV xin việc cho người dùng

Hình 21: Giao diện tài khoản Admin

Hình 22: Giao diện quản lí tài khoản

Hình 23: Giao diện thêm quyền

Hình 24: Giao diện quản lí nhóm quyền

Hình 25: Giao diện quản lí công việc của Admin

Hình 26: Giao diện quản lí công ty của Admin

Hình 27: Giao diện của tài khoản công ty

Hình 28: Giao diện quản lí công việc của công ty

Hình 29: Giao diện thêm công việc

Hình 30: Giao diện chỉnh sửa công việc

Hình 31: Giao diện quản lí người ứng tuyển

Hình 32: Giao diện thông tin Công ty

Hình 33: Giao diện gmail công ty thông báo cho người ứng tuyển

# TỔNG QUAN ĐỀ TÀI

## **1.1. Tên đề tài**

“Xây dựng Website tổng hợp và quản lý thông tin việc làm từ các trang Web tuyển dụng “

## **1.2. Giới thiệu và lý do chọn đề tài**

Trong bối cảnh mạng lưới tuyển dụng ngày càng phát triển và đa dạng, việc tìm kiếm thông tin về việc làm có thể trở nên phức tạp và mất nhiều thời gian. Để giải quyết vấn đề này, chúng tôi đã đề xuất và triển khai một dự án xây dựng một trang web tổng hợp và quản lý thông tin việc làm từ các trang web tuyển dụng hàng đầu.

## **1.3. Mục tiêu**

Mục tiêu chính của dự án là tạo ra một nền tảng trực tuyến giúp người tìm việc dễ dàng truy cập và tìm kiếm thông tin về các cơ hội việc làm từ nhiều nguồn khác nhau. Bằng cách tổng hợp và cập nhật thông tin việc làm một cách tự động, chúng tôi mong muốn giúp cải thiện quá trình tìm kiếm việc làm cho người dùng và cung cấp giải pháp hiệu quả cho các doanh nghiệp tuyển dụng.

## **1.4. Phương pháp nghiên cứu**

Trên cơ sở những phương pháp và kỹ thuật liên quan đến đề tài như các phương pháp nghiên cứu tài liệu, phương pháp kham thảo các website tìm kiếm việc làm thực tế. Để tìm ra cách giải quyết các yêu cầu của đề tài, cũng như phát triển thêm các tính năng phù hợp với đề tài.

CHƯƠNG 2: CƠ SỞ LÝ THUYẾT

## **2.1. Giới thiệu Python**

### **2.1.1. Crawl dữ liệu**

Cào dữ liệu (data scraping) là một quá trình thu thập dữ liệu từ các trang web hoặc các nguồn khác bằng cách sử dụng các thuật toán tự động. Cào dữ liệu thường được sử dụng để thu thập dữ liệu từ các trang web có cấu trúc hoặc có thể truy cập được bằng API.

Cào dữ liệu có thể được sử dụng cho nhiều mục đích khác nhau, bao gồm:

* Thu thập dữ liệu cho phân tích: Cào dữ liệu có thể được sử dụng để thu thập dữ liệu từ các nguồn khác nhau, chẳng hạn như các trang web tin tức, mạng xã hội, hoặc các cửa hàng trực tuyến. Dữ liệu này sau đó có thể được sử dụng để phân tích xu hướng thị trường, dự đoán hành vi người dùng, hoặc đánh giá hiệu suất của các chiến dịch marketing.
* Tự động hóa các tác vụ: Cào dữ liệu có thể được sử dụng để tự động hóa các tác vụ thường được thực hiện thủ công, chẳng hạn như thu thập thông tin sản phẩm hoặc dịch vụ từ các trang web thương mại điện tử. Điều này có thể giúp tiết kiệm thời gian và công sức cho người dùng.
* Tạo các ứng dụng mới: Cào dữ liệu có thể được sử dụng để tạo các ứng dụng mới, chẳng hạn như các ứng dụng thông tin, các ứng dụng phân tích, hoặc các ứng dụng trò chơi.

Cào dữ liệu có thể được thực hiện bằng nhiều cách khác nhau, bao gồm:

* Sử dụng các công cụ cào dữ liệu: Có nhiều công cụ cào dữ liệu có sẵn miễn phí hoặc trả phí. Các công cụ này cung cấp các tính năng và chức năng khác nhau để giúp người dùng cào dữ liệu dễ dàng hơn.
* Sử dụng các ngôn ngữ lập trình: cào dữ liệu cũng có thể được thực hiện bằng cách sử dụng các ngôn ngữ lập trình, chẳng hạn như Python. Ngôn ngữ lập trình này cung cấp các thư viện và API giúp người dùng cào dữ liệu dễ dàng hơn.

Trong đồ án lần này chúng em dùng ngôn ngữ Python để tương tác với dữ liệu

### **2.1.2. Selenium vs BeautifulSoup**

Selenium và BeautifulSoup là hai thư viện Python phổ biến được sử dụng để thao tác với dữ liệu web. Selenium cho phép bạn tương tác với các trang web như một người dùng thực, trong khi BeautifulSoup cho phép bạn phân tích mã HTML của một trang web.

Selenium là một bộ công cụ tự động hóa trình duyệt mã nguồn mở. Nó cho phép bạn điều khiển các trình duyệt web phổ biến như Chrome, Firefox và Edge. Selenium có thể được sử dụng để thực hiện các tác vụ như đăng nhập vào trang web, điền vào biểu mẫu và nhấp vào các liên kết.

BeautifulSoup là một thư viện Python để phân tích mã HTML và XML. Nó cho phép bạn truy cập và thao tác với dữ liệu được lưu trữ trong mã HTML dưới dạng text. Từ đó xác định được các đối tượng cần thao tác để làm việc.

Mysql.connector là một trình điều khiển Python cho phép bạn kết nối và tương tác với cơ sở dữ liệu MySQL. Nói đơn giản, nó giống như một chiếc cầu nối giữa ngôn ngữ Python và hệ quản trị cơ sở dữ liệu MySQL.

Tính năng chính:

* Quản lý kết nối: Thiết lập và đóng kết nối tới máy chủ MySQL.
* Thực thi truy vấn: Thực hiện các câu lệnh SQL như SELECT, INSERT, UPDATE và DELETE.
* Lấy kết quả: Trả về kết quả truy vấn theo nhiều định dạng khác nhau, như bộ tup, từ điển và các đối tượng tùy chỉnh.

Trong đồ án lần này chúng em sử dụng BeautifulSoup version 4.12.2 và Selenium version 4.15.2 để hỗ trợ cho việc cào dữ liệu về. Và mysql.connector version 8.2.0 để tương tác với cơ sở dữ liệu.

Để cài đặt các thư viện này chúng ta dử dụng lệnh pip install trong command line.



Hình 1: BeautifulSoup Python

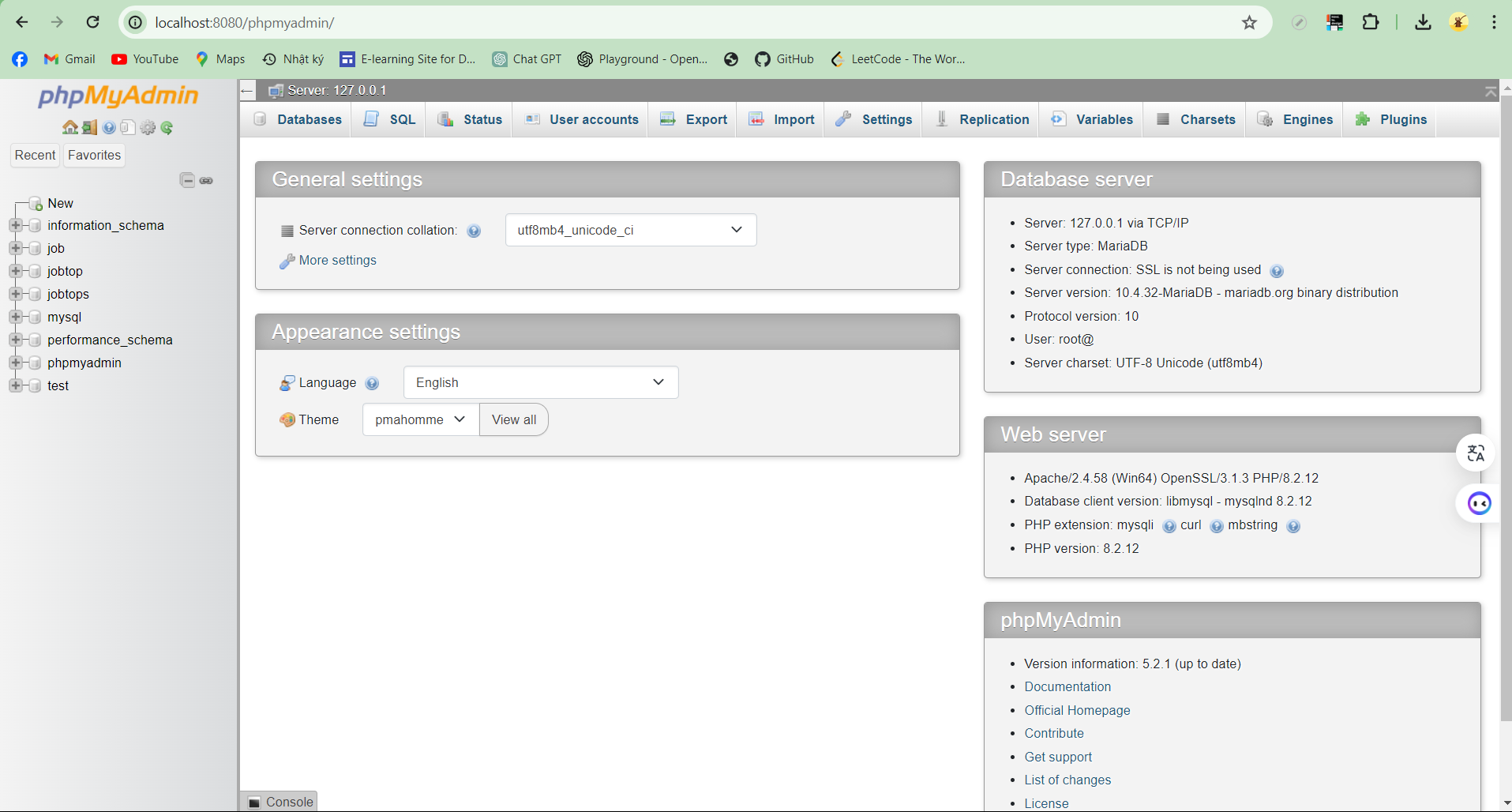
## **2.2 XAMPP/MySQL**

XAMPP là một ứng dụng mã nguồn mở được phát triển bởi Apache Friends, cho phép bạn tạo một môi trường phát triển web đầy đủ trên máy tính cá nhân của mình. Tên "XAMPP" được viết tắt từ "X" (dành cho các hệ điều hành khác nhau như Windows, Linux, Mac), "Apache" (một trong những máy chủ web phổ biến nhất), "MySQL" (hệ quản trị cơ sở dữ liệu), "PHP" (ngôn ngữ lập trình phổ biến cho việc phát triển web), và "Perl" (ngôn ngữ lập trình thông dịch đa năng). XAMPP cung cấp một cách nhanh chóng và dễ dàng để cài đặt và cấu hình môi trường phát triển web trên máy tính cá nhân, giúp cho việc phát triển và thử nghiệm ứng dụng web trở nên thuận tiện hơn.

MySQL là một hệ quản trị cơ sở dữ liệu quan hệ mã nguồn mở và miễn phí. Nó là một trong những hệ quản trị cơ sở dữ liệu phổ biến nhất trên thế giới, được sử dụng bởi các doanh nghiệp, tổ chức và cá nhân. MySQL được sử dụng để lưu trữ và quản lý dữ liệu, chẳng hạn như thông tin khách ung, sản phẩm, đơn ung, …

Tính năng chính của MySQL:

* Cơ sở dữ liệu quan hệ: MySQL là một hệ quản trị cơ sở dữ liệu quan hệ, có nghĩa là dữ liệu được lưu trữ trong các bảng. Các bảng được liên kết với nhau bằng các khóa ngoại, cho phép bạn truy vấn dữ liệu từ nhiều bảng một cách dễ dàng.
* Mã nguồn mở: MySQL là mã nguồn mở, có nghĩa là bạn có thể tải xuống và sử dụng nó miễn phí. Điều này làm cho nó trở thành một lựa chọn tuyệt vời cho các doanh nghiệp và tổ chức có ngân sách hạn chế.
* Mức độ bảo mật cao: MySQL cung cấp một số tính năng bảo mật để giúp bảo vệ dữ liệu của bạn, chẳng hạn như xác thực người ung, mã hóa và kiểm soát truy cập.
* Hiệu suất cao: MySQL là một hệ quản trị cơ sở dữ liệu hiệu quả, có thể xử lý lượng lớn dữ liệu. Điều này làm cho nó trở thành một lựa chọn tuyệt vời cho các ứng dụng web và doanh nghiệp.



Hình 2: Trang chính MySQL

## **2.3 NodeJS**

Hiện nay có rất nhiều ngôn ngữ được dung để code BackEnd cho 1 website như là: Java, Python, PHP, Ruby, …

Trong đó nhóm chúng em quyết định chọn NodeJS làm ngôn ngữ cho đồ án lần này vì những lợi ích sau:

* Cộng đồng phát triển lớn: Node.js có một cộng đồng phát triển lớn và năng động. Có rất nhiều thư viện và công cụ có sẵn cho Node.js, giúp việc phát triển ứng dụng trở nên dễ dàng hơn.
* Hiệu suất cao: Node.js được xây dựng dựa trên V8, trình thông dịch JavaScript mã nguồn mở của Google. V8 được thiết kế để chạy JavaScript một cách nhanh chóng và hiệu quả, điều này làm cho Node.js trở nên lý tưởng cho các ứng dụng thời gian thực.
* Trong các ứng dụng web có lưu lượng truy cập cao, Node.js có thể được sử dụng để xử lý nhiều yêu cầu cùng lúc.



Hình 3: NodeJS

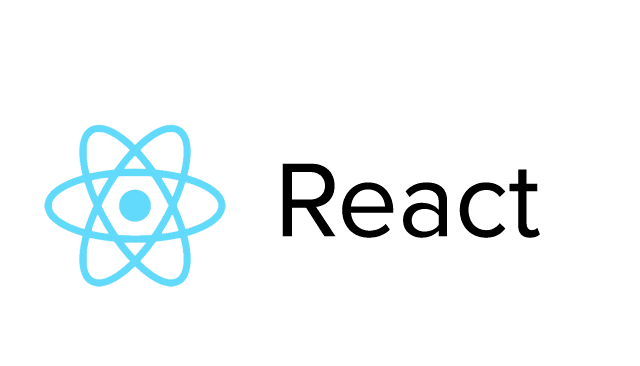
## **2.4 ReactJS**

ReactJS là một thư viện JavaScript mã nguồn mở được phát triển bởi Facebook, chủ yếu được sử dụng để xây dựng giao diện người dùng cho các ứng dụng web hiệu quả và linh hoạt. ReactJS tập trung vào việc xây dựng các thành phần UI có thể tái sử dụng và quản lý trạng thái của các thành phần này một cách dễ dàng, giúp cho việc phát triển ứng dụng web trở nên nhanh chóng và dễ dàng bảo trì.

Tính năng của ReactJS:

* Thành phần UI tái sử dụng.
* Virtual DOM.
* Quản lí trạng thái.
* Hỗ trợ JSX.

ReactJS là một lựa chọn lý tưởng cho việc phát triển ứng dụng web đơn trang (Single Page Applications - SPAs) và các ứng dụng web có yêu cầu tương tác cao. Với cộng đồng lớn và một hệ sinh thái phong phú, ReactJS cho phép người phát triển xây dựng các ứng dụng web phức tạp và mở rộng dễ dàng.



Hình 4: ReactJS

# CHƯƠNG 3: PHÂN TÍCH VÀ THIẾT KẾ HỆ THỐNG

## **Phân tích thuật toán**

### **3.1.1. Thu tập dữ liệu**

Hiện nay có rất nhiều trang web hỗ trợ tìm kiếm việc làm khi mà nhu cầu việc làm ngày một lớn dần. Ta có thể kể đến một số trang web như sau: indeed, VietnamWork, TopCV, … Nhưng phụ thuộc vào nhu cầu và ý tưởng cào dữ liệu của nhóm chúng em. Thì chúng em quyết định chọn trang web Vieclam24h.vn là website để thực hiện đồ án lần này vì nó có đầy đủ thông tin để chúng em có thể hoàn thành ý tưởng của nhóm.

Sau đây là một số thông tin về trang web mà chúng em muốn giới thiệu: Vieclam24h.vn là một trang web tìm kiếm việc làm trực tuyến hàng đầu tại Việt Nam. Trang web được thành lập vào năm 2007 và hiện có hơn 10 triệu lượt truy cập mỗi tháng.

Vieclam24h.vn cung cấp thông tin tuyển dụng từ các nhà tuyển dụng lớn và nhỏ trên khắp Việt Nam. Trang web có nhiều tính năng hữu ích giúp người dùng tìm kiếm việc làm phù hợp, bao gồm:

* Tìm kiếm theo ngành nghề, vị trí, địa điểm, mức lương, v.v.
* Lọc kết quả tìm kiếm theo các tiêu chí khác nhau.
* Tạo hồ sơ xin việc trực tuyến.
* Gửi hồ sơ xin việc trực tiếp đến nhà tuyển dụng.
* Nhận thông báo về các tin tuyển dụng mới.

Vieclam24h.vn là một lựa chọn tuyệt vời cho cả người dùng và nhà tuyển dụng. Trang web cung cấp một nền tảng thuận tiện và hiệu quả để tìm kiếm và tìm được việc làm phù hợp.

### **3.1.2. Thuật toán crawler**

Việc cào dữ liệu từ 1 trang web sẽ có rất nhiều cách để áp dụng:

* Cách thủ công là cách đơn giản nhất, nhưng cũng là cách tốn thời gian và công sức nhất. Cách này bao gồm việc truy cập vào trang web bằng trình duyệt web, sau đó sử dụng các công cụ như phím tắt, chuột hoặc các công cụ khác để trích xuất dữ liệu từ trang web.
* Cách thủ công có thể được sử dụng để cào dữ liệu từ các trang web có cấu trúc đơn giản, chẳng hạn như các trang web chỉ có một số lượng nhỏ dữ liệu tĩnh. Tuy nhiên, cách này sẽ không thể áp dụng cho các trang web có cấu trúc phức tạp hơn, chẳng hạn như các trang web có nhiều dữ liệu động hoặc các trang web sử dụng JavaScript.
* Cách tự động là cách hiệu quả hơn so với cách thủ công, nhưng cũng đòi hỏi nhiều kỹ năng lập trình hơn. Cách này sử dụng các chương trình máy tính để truy cập vào trang web và trích xuất dữ liệu.
* Có nhiều cách khác nhau để tự động cào dữ liệu từ trang web. Một cách phổ biến là sử dụng các thư viện lập trình có sẵn, chẳng hạn như thư viện requests trong Python hoặc thư viện Selenium trong Java. Các thư viện này cung cấp các hàm và phương thức giúp lập trình viên dễ dàng truy cập vào trang web và trích xuất dữ liệu.
* Một cách khác để tự động cào dữ liệu từ trang web là sử dụng các công cụ cào dữ liệu chuyên dụng, chẳng hạn như Web Scraper hoặc Octoparse. Các công cụ này cung cấp một giao diện trực quan giúp lập trình viên dễ dàng tạo các quy tắc để trích xuất dữ liệu từ trang web.

Trong đồ án lần này chúng em sử dụng thư viện Selenium và BeautifulSoup để thực hiện việc cào dữ liệu từ website. Selenium là một công cụ tự động hóa trình duyệt web. Nó có thể mô phỏng hành vi người dùng trên trình duyệt (như Chrome, Firefox, etc.) và thậm chí thao tác với các trang web có sử dụng JavaScript động. BeautifulSoup là một thư viện Python để phân tích HTML và XML. Nó giúp bạn trích xuất thông tin từ các trang web đã được tải về.

* Đầu tiên Selenium sẽ mở 1 trình duyệt mô phỏng Chrome lên.

Vd: Selenium.webdriver.Chrome(options=chrome\_options)

* Tiếp theo là truy cập vào đường link trang web ‘https://vieclam24h.vn/tim-kiem-viec-lam-nhanh’.   
  Vd: Driver.get(URL).
* BeautifulSoup sẽ được dùng để truy cập vào trang HTML dưới dạng văn bản của website để tìm và trích xuất thông tin từ các thẻ bằng phương thức find(), find\_all(), select(), …

Vd: link = page\_source.find\_all('a', class\_=class\_name)

Trong ví dụ này đường link của từng trang công việc sẽ được lấy từ trang chủ vieclam24h.vn thông qua thẻ ‘a’ có tên lớp là class\_name.

* Như vậy thì code sẽ trực quan hơn và dễ dàng cho người lập trình trong việc xác định vị trí các đối tượng cần lấy thông tin
* Sau khi đã xác định được vị trí của đối tượng muốn lấy dữ liệu. Phương thức get\_text() sẽ được dùng để lấy đoạn văn bản về DataBase.

Vd: div[2].get\_text(' ', strip=True)

+) get\_text(' ', strip=True): Gọi phương thức get\_text trên phần tử đã chọn để trích xuất văn bản từ bên trong phần tử.

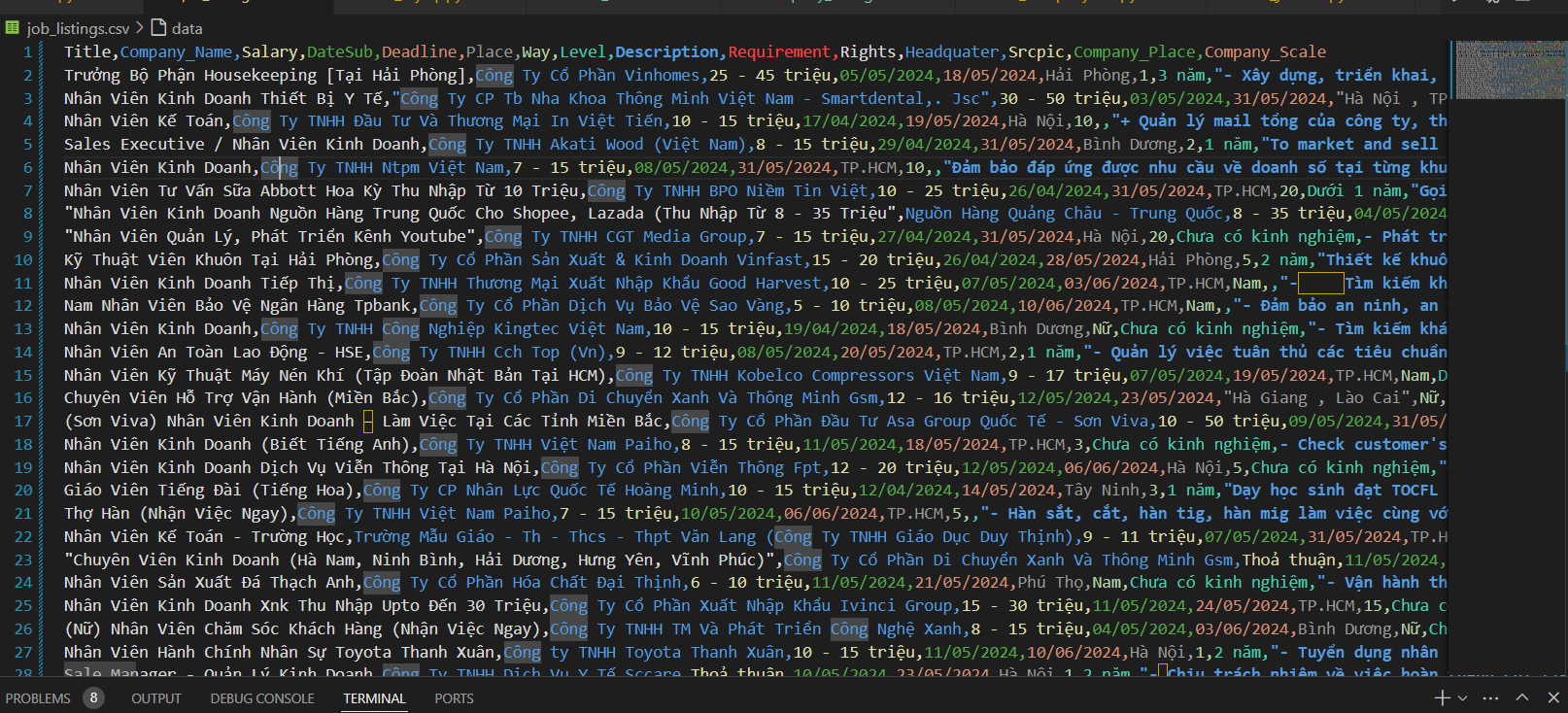
+) Các tham số được sử dụng như sau:' ': Đây là ký tự hoặc chuỗi được sử dụng để nối các đoạn văn bản khi chúng được trích xuất. Trong trường hợp này, đoạn văn bản sẽ được nối bằng khoảng trắng. Strip=True: Thiết lập strip thành True để loại bỏ các khoảng trắng và ký tự xuống dòng ở đầu và cuối đoạn văn bản.

* Sau khi truy cập vào trang web chính, em sẽ lần lượt lấy hết tất cả các link thành phần ở trang 1 (trang mặc định) bằng BS4. Sau đó truy cập vào từng link thành phần thông qua Selenium để lấy tất cả các thông tin cần thiết ở từng link như là: Tên công ty, địa chỉ, số người, yêu cầu bằng cấp, giới tính, …Tiếp theo trang 1 em sẽ lần lượt cho tăng dần số trang lên tuỳ vào mong muốn của admin sẽ cào bao nhiêu trang. Công việc cào sẽ lặp đi lặp lại và dữ liệu lấy được sẽ đưa xuống DataBase.

### **3.1.3. Kiểm tra trùng lặp dữ liệu**

Sau khi dữ liệu đã được lưu trữ về database. Mình sẽ thực hiện các thao tác để cho việc dữ liệu cào về không bị trùng lặp với dữ liệu đã tồn tại trước đó.

* Đầu tiên trước khi crawl dữ liệu thì chúng ta sẽ đọc dữ liệu trong file csv trước (Đây là file chứa các dữ liệu từ những lần crawl trước).
* Tiếp theo chúng ta sẽ bắt đầu với việc crawl lần mới.
* Mình sẽ kiểm tra lần crawl này có trùm với dữ liệu đã được đọc từ đầu không. Những dữ liệu trùng lặp là những dữ liệu mà dữ liệu của tất cả các cột trong bảng là giống nhau.
* Những dữ liệu mà không trùng lặp sẽ tiếp tục được thêm vào Database.



Hình 5: Dữ Liệu khi được lưu ở file CSV

A screenshot of a computer

Description automatically generated

Hình 6: Dữ liệu khi được đẩy lên MySQL

## **3.2. Chức năng hệ thống**

### **3.2.1. Admin**

- Chức năng quản lý, thêm, sửa, xoá các tài khoản người dùng.

- Chức năng quản lý / chỉnh sửa việc làm đăng trên web.

- Chức năng quản lý / chỉnh sửa công ty.

- Chỉnh sửa quyền cho tài khoản người dùng.

### **3.2.2. Nhà tuyển dụng (Công ty - Company)**

* Chức năng đăng nhập.
* Chức năng đăng ký.
* Chức năng xem thông tin chi tiết công việc.
* Chức năng tìm kiếm công việc theo từ khoá, địa điểm.
* Chức năng quản lý người ứng tuyển vào từng công việc.
* Chức năng quản lý tài khoản của mình.
* Được gửi thông tin công việc về gmail cá nhân.

### **3.2.3. Người tìm việc**

* Chức năng đăng nhập.
* Chức năng đăng ký.
* Chức năng xem thông tin chi tiết công việc.
* Chức năng xem thông tin chi tiết công ty.
* Chức năng quản lý tài khoản của mình.
* Chức năng tìm kiếm công việc theo từ khoá, địa điểm.
* Tạo CV xin việc, nộp CV

## **3.3 Cơ sở dữ liệu**

- Jobinfo: Table này gồm 13 column bao gồm ‘Title’, ‘Company\_Id’, ‘Career\_Id’, ‘Address’ , ‘numberEmployee, ‘Level’, ‘Salary’, ‘Description’, ‘Requirement’, ‘Rights’, ‘DateSub’, ‘Deadline’, ‘sourcePicture. Để biểu diễn được hết trọn vẹn dữ liệu được cào về từ website việc làm.

1. Title: Tiêu đề công việc hoặc thông báo tuyển dụng.
2. Company\_Id: Id công ty đăng thông tin tuyển dụng hoặc cung cấp việc làm.
3. Career\_Id: Id ngành nghề của công việc.
4. Address: Địa điểm làm công việc.
5. numberEmployee: Số lượng nhân viên cần tuyển.
6. Level: Năng lực yêu cầu.
7. Salary: Mức lương của công việc đăng tuyển.
8. Description: Mô tả công việc hoặc thông tin chi tiết về công ty và vị trí tuyển dụng.
9. Requirement: Các yêu cầu khác đối với người dùng.
10. Rights: Phúc lợi khi ứng tuyển công việc.
11. DateSub: Thời điểm đăng thông tin tuyển dụng hoặc thời gian công việc được đăng.
12. Deadline: Hạn nộp hồ sơ của công việc.
13. sourcePicture: Hình ảnh của công việc.

=> Trong đó có 3 trường là ‘Requirement, Description, Rights’ có kiểu dữ liệu là Text. Bởi vì TEXT là kiểu dữ liệu có kích thước lớn hơn so với CHAR hay là VARCHAR. Vậy nên trường hợp này chúng em dùng TEXT để biểu diễn cho 3 trường trên. Bởi vì đặc thù của 1 trang web việc làm 24h.vn thì phần mô tả, lợi ích và yêu cầu luôn chứa lượng thông tin lớn hơn nhiều so với các thông tin còn lại trong trang.

Company\_listings: ‘Name’, ‘GT’, ‘DC’, ‘QM’

Name: Tên công ty.

GT: Giới thiệu công ty.

DC: Địa chỉ công ty.

QM: Quy mô của công ty.

User: ‘Username’, ‘Password’, ‘Quyen’

Username

Password

Quyen

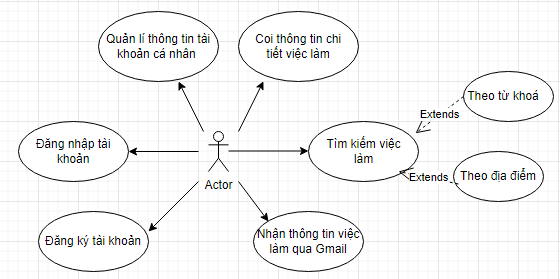
## **3.4. Sơ đồ**

- Mô hình cấu trúc hệ thống hướng đối tượng A screenshot of a computer

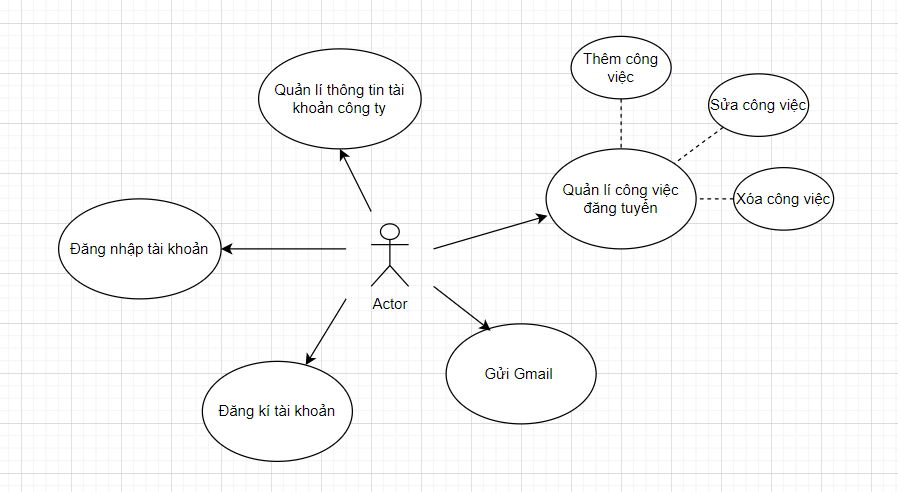
Description automatically generated

Hình 7: Mô hình cấu trúc hệ thống hướng đối tượng

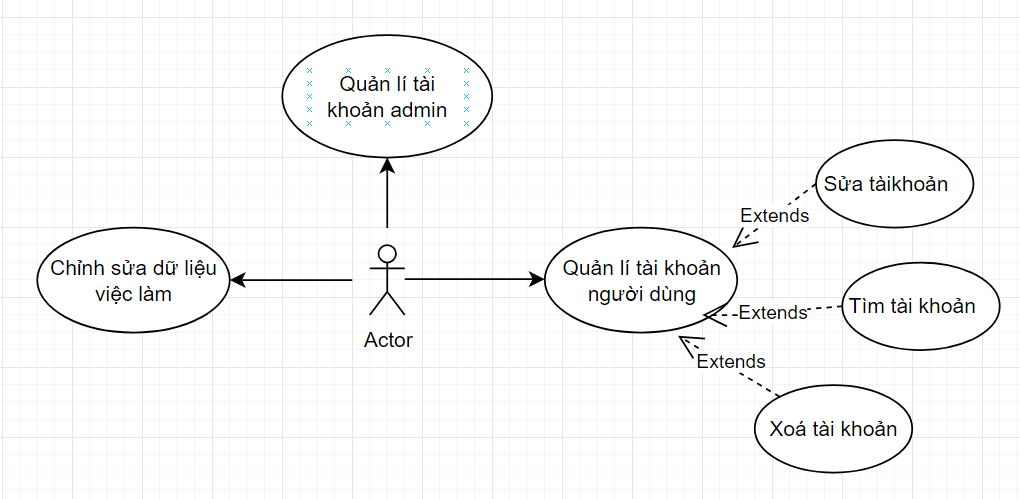
- Sơ đồ usecase



Hình 8: Usecase người dùng (User)



Hình 9: Usercase Company



Hình 10: Usecase Admin

## **3.5. Đặc tả UseCase**

### **3.5.1. Đăng nhập**

|  |  |
| --- | --- |
| Tên usecase | Usecase đăng nhập |
| Tác nhân | User |
| Mô tả | User thực hiện việc đăng nhập tài khoản để sử dụng hệ thống |
| Điều kiện tiên quyết | Hệ thống đã có dữ liệu tài khoản |
| Dòng dự kiện chính | 1. Người dùng chạy chương trình  2. Người dùng nhập tài khoản và mật khẩu  3. Bấm vào nút đăng nhập |
| Kết quả | Nếu thành công sẽ chuyển tới trang chủ |

### **3.5.2. Đăng ký**

|  |  |
| --- | --- |
| Tên usecase | Usecase đăng ký |
| Tác nhân | User |
| Mô tả | User thực hiện việc đăng ký tài khoản mới để sử dụng hệ thống |
| Điều kiện tiên quyết | Hệ thống được chạy |
| Dòng dự kiện chính | 1. Người dùng chạy chương trình  2. Người dùng nhấn vào nút đăng ký  3. Nhập gmail, tài khoản, mật khẩu và nhắc lại mật khẩu  4. Nhấn vào nút đăng ký |
| Kết quả | Nếu thành công tài khoản sẽ được tạo, dữ liệu được thêm vào hệ thống |

### **3.5.3. Tìm kiếm việc làm**

|  |  |
| --- | --- |
| Tên usecase | Usecase tìm kiếm việc làm |
| Tác nhân | User |
| Mô tả | User thực hiện việc tìm kiếm thông tin việc làm mà mình muốn theo địa điểm và ngành nghề |
| Điều kiện tiên quyết | Hệ thống được đăng nhập vào tải khoản của user |
| Dòng dự kiện chính | 1. Người dùng đăng nhập vào tài khoản của mình  2. Người dùng chọn thành phố hoặc ngành nghề mình muốn trong list có sẵn  3. Nhấn vào nút tìm kiếm |
| Kết quả | Hệ thống sẽ hiển thị những việc làm với thông tin như người dùng muốn tìm kiếm |

### **3.5.4. Nhận việc làm qua Gmail**

|  |  |
| --- | --- |
| Tên usecase | Usecase nhận việc làm thông qua gmail |
| Tác nhân | User |
| Mô tả | User thực muốn biết thêm về thông tin việc làm thông qua chi tiết được gửi về gmail |
| Điều kiện tiên quyết | Hệ thống được đăng nhập vào tải khoản của user, user vào trang chi tiết của việc làm |
| Dòng dự kiện chính | 1. Người dùng đăng nhập vào tài khoản của mình  2. Người dùng nhấn vào một công việc mà mình quan tâm ở trang chủ để vào trang chi tiết  3. Nhấn vào nút ‘Gửi Gmail’  4. Nhập email và sô điện thoại theo yêu cầu  5. Nhấn vào nút ‘Send Mail’ |
| Kết quả | Hệ thống sẽ gửi đường link chi tiết về gmail cho người dùng |

### **3.5.5. Xoá công việc đăng tuyển**

|  |  |
| --- | --- |
| Tên usecase | Usecase xoá công việc đăng tuyển |
| Tác nhân | Company, Admin |
| Mô tả | Company, Admin muốn xoá 1 công việc được đăng tuyển khỏi hệ thống data quản lý công việc |
| Điều kiện tiên quyết | Hệ thống được đăng nhập vào tải khoản của company, admin |
| Dòng dự kiện chính | 1. Company, Admin đăng nhập vào tài khoản của mình  2. Trong menu chọn trang Job  3. Chọn 1 công việc bất kì mà muốn xoá  4. Nhấn vào nút Delete |
| Kết quả | Công việc đó đó sẽ bị xoá khỏi dữ liệu của hệ thống |

### **3.5.6 Thêm công việc đăng tuyển**

|  |  |
| --- | --- |
| Tên usecase | Usecase thêm công việc đăng tuyển |
| Tác nhân | Company, Admin |
| Mô tả | Company, Admin muốn thêm 1 công việc được đăng tuyển khỏi hệ thống data quản lý công việc |
| Điều kiện tiên quyết | Hệ thống được đăng nhập vào tải khoản của company, admin |
| Dòng dự kiện chính | 1. Company, Admin đăng nhập vào tài khoản của mình  2. Trong menu chọn trang Job  3. Nhấn vào nút thêm công việc  4. Điền thông tin công việc và đăng lên |
| Kết quả | Công việc sẽ được thêm lên dữ liệu của hệ thống |

### **3.5.7 Sửa công việc đăng tuyển**

|  |  |
| --- | --- |
| Tên usecase | Usecase sửa công việc đăng tuyển |
| Tác nhân | Company |
| Mô tả | Company muốn sửa 1 công việc được đăng tuyển khỏi hệ thống data quản lý công việc |
| Điều kiện tiên quyết | Hệ thống được đăng nhập vào tải khoản của company |
| Dòng dự kiện chính | 1. Company đăng nhập vào tài khoản của mình  2. Trong menu chọn trang Job  3. Chọn 1 công việc bất kì mà muốn sửa  4. Nhấn nút sửa và điền thông tin rồi thêm |
| Kết quả | Công việc đó sẽ được sửa và đẩy lên dữ liệu của hệ thống |

### **3.5.8. Xoá tài khoản người dùng (Admin)**

|  |  |
| --- | --- |
| Tên usecase | Usecase xoá tài khoản người dùng |
| Tác nhân | Admin |
| Mô tả | Admin muốn xoá 1 tài khoản của user khỏi hệ thống data quản lý tài khoản |
| Điều kiện tiên quyết | Hệ thống được đăng nhập vào tải khoản của admin |
| Dòng dự kiện chính | 1. Admin đăng nhập vào tài khoản của mình  2. Trong menu chọn quản lí người dùng  3. Chọn 1 tài khoản bất kì mà admin muốn xoá  4. Nhấn vào nút Delete |
| Kết quả | Tài khoản đó sẽ bị xoá khỏi dữ liệu của hệ thống |

### **3.5.9. Sửa thông tin tài khoản người dùng**

|  |  |
| --- | --- |
| Tên usecase | Usecase sửa tài khoản người dùng |
| Tác nhân | Admin |
| Mô tả | Admin muốn sửa 1 tài khoản của user trong hệ thống data quản lý tài khoản |
| Điều kiện tiên quyết | Hệ thống được đăng nhập vào tải khoản của admin |
| Dòng dự kiện chính | 1. Admin đăng nhập vào tài khoản của mình  2. Trong menu chọn quản lí người dùng  3. Chọn 1 tài khoản bất kì mà admin muốn sửa  4. Nhấn vào nút Update |
| Kết quả | Tài khoản đó sẽ được sửa lại dữ liệu của hệ thống |

### **3.5.10. Tìm kiếm thông tin tài khoản người dùng**

|  |  |
| --- | --- |
| Tên usecase | Usecase tìm kiếm tài khoản người dùng |
| Tác nhân | Admin |
| Mô tả | Admin muốn tìm kiếm 1 tài khoản của user trong hệ thống data quản lý tài khoản |
| Điều kiện tiên quyết | Hệ thống được đăng nhập vào tải khoản của admin |
| Dòng dự kiện chính | 1. Admin đăng nhập vào tài khoản của mình  2. Trong menu chọn quản lí người dùng  3. Nhập thông tin cần tìm vào ô tìm kiếm  4. Nhấn vào nút Enter |
| Kết quả | Các tài khoản có từ khoá trùng sẽ hiện ra |

### **3.5.11. Phần quyền từng chức năng cho mỗi tài khoản**

|  |  |
| --- | --- |
| Tên usecase | Usecase tìm kiếm tài khoản người dùng |
| Tác nhân | Admin |
| Mô tả | Admin muốn phân quyền 1 tài khoản của user trong hệ thống data quản lý tài khoản |
| Điều kiện tiên quyết | Hệ thống được đăng nhập vào tải khoản của admin |
| Dòng dự kiện chính | 1. Admin đăng nhập vào tài khoản của mình  2. Trong menu chọn roles  3. Chọn quyền muốn thay đổi  4. Nhấn vào nút Enter |
| Kết quả | Các tài khoản có từ khoá trùng sẽ hiện ra |

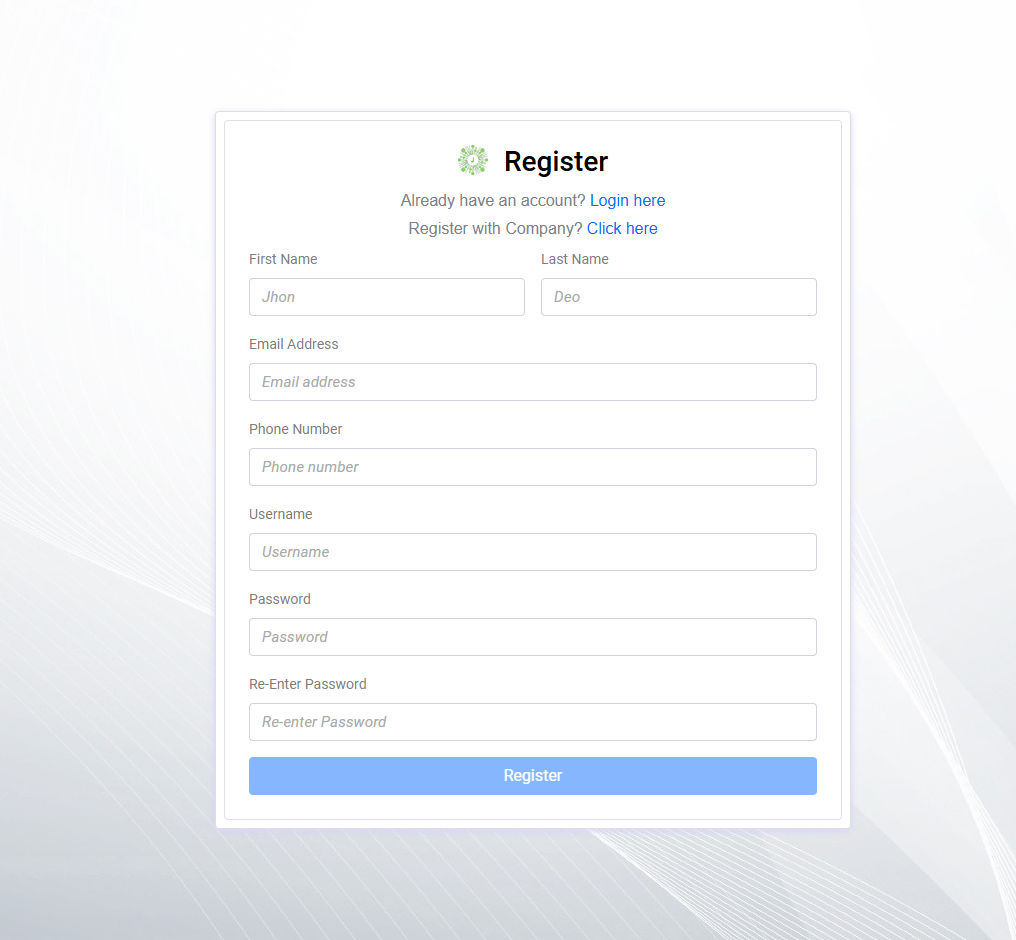
# CHƯƠNG 4: XÂY DỰNG WEBSITE

## **4.1. Giao diện đăng nhập, đăng kí**

### 



Hình 11: Giao diện đăng nhập



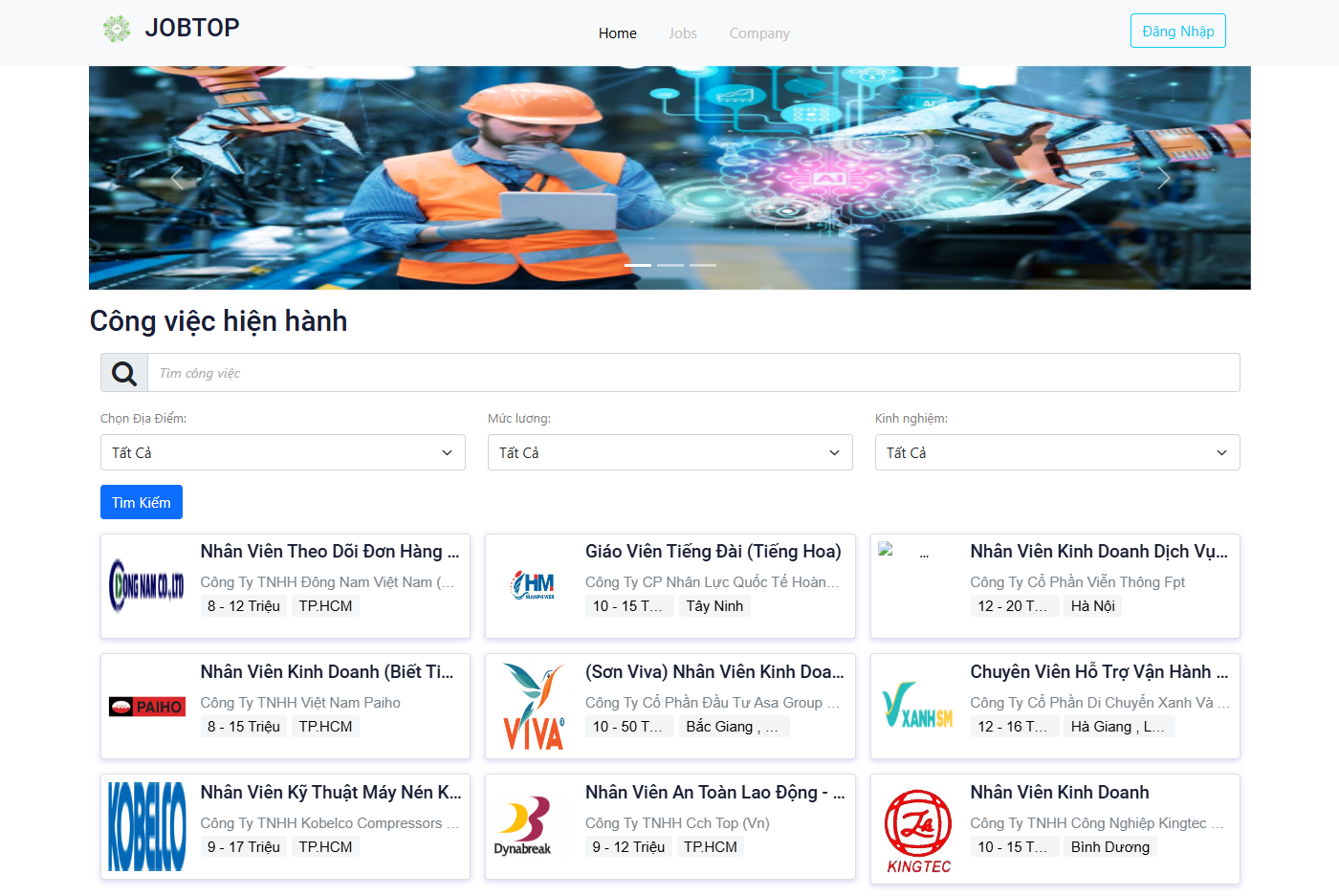
Hình 12: Giao diện đăng kí Người dùng

A screenshot of a computer

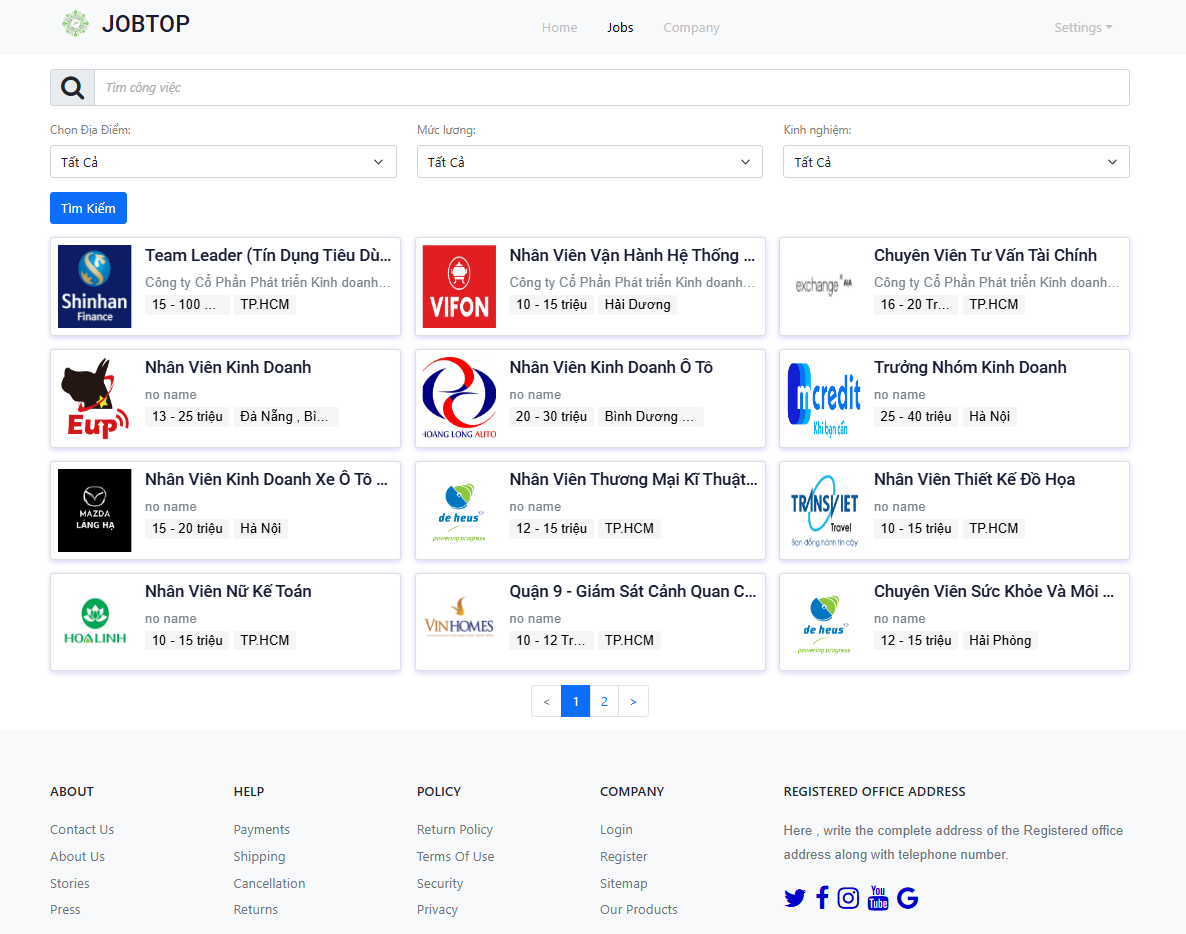
Description automatically generated

Hình 13: Giao diện đăng kí Công ty

## **4.2. Giao diện người dùng**



Hình 14: Giao diện trang Home



Hình 15: Giao diện trang công việc

A screenshot of a computer

Description automatically generated

Hình 16: Giao diện chi tiết công việc

A screenshot of a computer

Description automatically generated

Hình 17: Giao diện trang Công ty

A screenshot of a computer

Description automatically generated

Hình 18: Giao diện chi tiết Công ty

A screenshot of a computer

Description automatically generated

Hình 19: Giao diện chỉnh sửa thông tin người dùng

A screenshot of a computer

Description automatically generated

Hình 20: Tạo CV xin việc cho người dùng

## **4.3. Giao diện tài khoản Admin**

A colorful background with white text

Description automatically generated

Hình 21: Giao diện tài khoản Admin

A screenshot of a computer

Description automatically generated

Hình 22: Giao diện quản lí tài khoản

A screenshot of a computer

Description automatically generated

Hình 23: Giao diện thêm quyền

A screenshot of a computer

Description automatically generated

Hình 24: Giao diện quản lí nhóm quyền

A screenshot of a computer

Description automatically generated

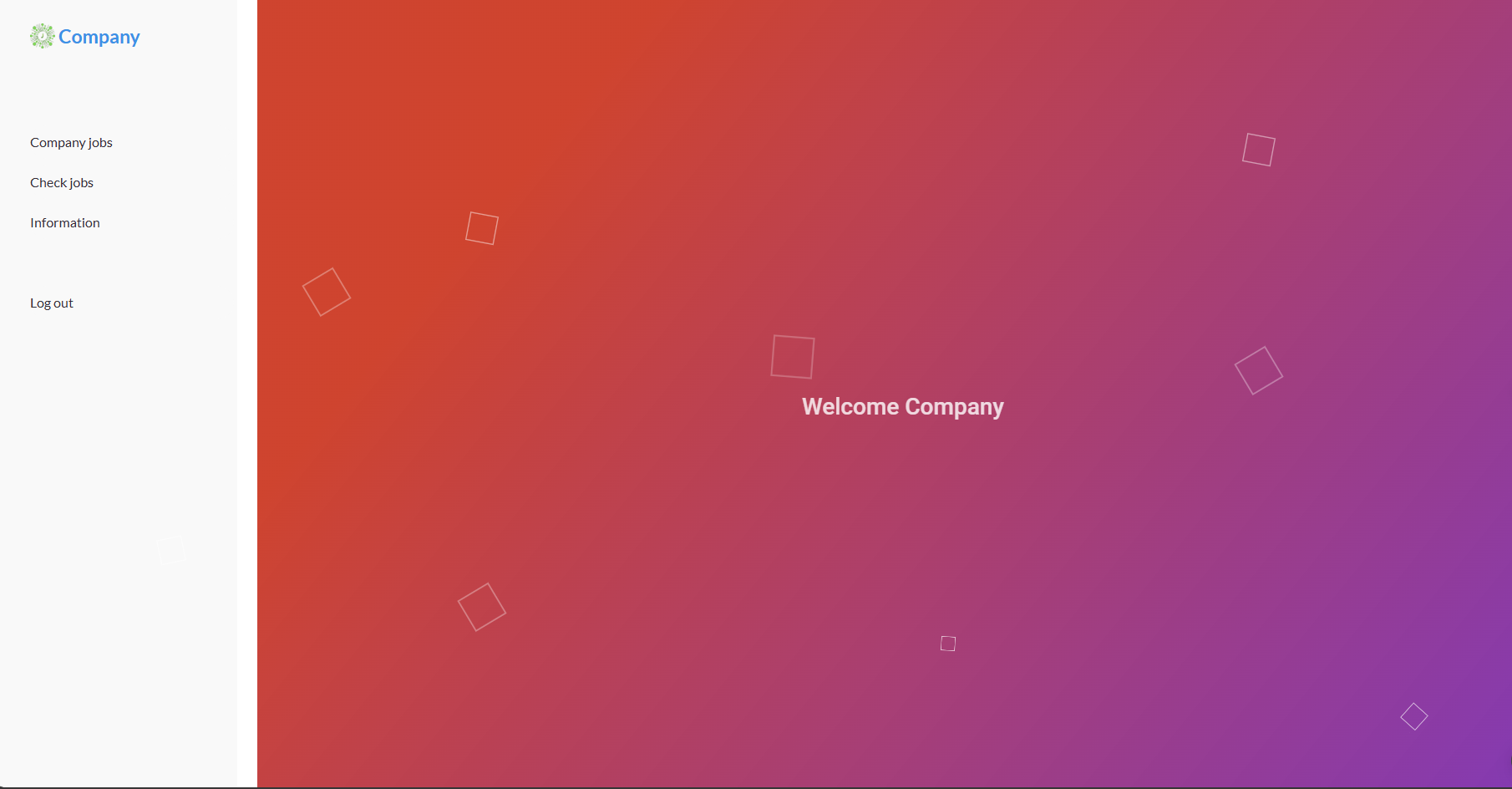
Hình 25: Giao diện quản lí công việc của Admin

A screenshot of a computer

Description automatically generated

Hình 26: Giao diện quản lí công ty của Admin

## **4.4. Giao diện tài khoản công ty**



Hình 27: Giao diện của tài khoản công ty

A screenshot of a computer

Description automatically generated

Hình 28: Giao diện quản lí công việc của công ty

A screenshot of a computer

Description automatically generated

Hình 29: Giao diện thêm công việc

A screenshot of a computer

Description automatically generated

Hình 30: Giao diện chỉnh sửa công việc

A screenshot of a computer

Description automatically generated

Hình 31: Giao diện quản lí người ứng tuyển

A screenshot of a computer

Description automatically generated

Hình 32: Giao diện thông tin Công ty

A screenshot of a computer

Description automatically generated

Hình 33: Giao diện gmail công ty thông báo cho người ứng tuyển

## **4.5. Nhận xét**

Triển khai ứng dụng theo cấu trúc hệ thống hướng đối tượng trên có những lợi ích như là:

* Tính mô-đun: Cấu trúc hệ thống hướng đối tượng giúp tách biệt các chức năng của ứng dụng thành các thành phần nhỏ, độc lập với nhau. Điều này giúp cho việc phát triển và bảo trì ứng dụng dễ dàng hơn.
* Tính tái sử dụng: Các lớp của hệ thống có thể được tái sử dụng trong các ứng dụng khác. Điều này giúp tiết kiệm thời gian và chi phí phát triển.
* Tính dễ hiểu: Cấu trúc hệ thống hướng đối tượng giúp cho ứng dụng dễ hiểu và dễ sử dụng hơn.
* Chương trình đã thực hiện chính xác mục đích yêu cầu ban đầu.

# CHƯƠNG 5: KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

## **5.1. Kết luận**

Trong quá trình thực hiện dự án xây dựng Website tổng hợp và quản lý thông tin việc làm từ các trang Web tuyển dụng, chúng tôi đã đạt được nhiều thành tựu đáng kể. Website của chúng tôi đã trở thành một nguồn thông tin đáng tin cậy và hữu ích cho cả người tìm việc và doanh nghiệp tuyển dụng.

Chúng tôi đã xây dựng một giao diện người dùng thân thiện và dễ sử dụng, cung cấp các tính năng tìm kiếm và lọc thông tin việc làm một cách linh hoạt và tiện lợi. Bằng cách tự động thu thập và cập nhật dữ liệu từ các trang Web tuyển dụng hàng đầu, chúng tôi đã giúp người tìm việc tiết kiệm thời gian và công sức trong quá trình tìm kiếm việc làm.

Ngoài ra, việc tích hợp hệ thống đăng ký và nộp hồ sơ trực tuyến cũng đã tạo ra một cơ hội thuận lợi cho các ứng viên nộp hồ sơ và cho các doanh nghiệp tuyển dụng tiếp cận ứng viên phù hợp.

## **5.2. Hướng phát triển**

Dự án của chúng tôi vẫn còn nhiều tiềm năng phát triển. Dưới đây là một số hướng phát triển tiềm năng:

* Mở rộng phạm vi: Khả năng mở rộng phạm vi của dự án để bao gồm nhiều nguồn thông tin việc làm hơn, từ các trang Web tuyển dụng đa quốc gia đến các ngành nghề và lĩnh vực khác nhau.
* Tích hợp tính năng xác thực và bảo mật cao: Đảm bảo rằng dữ liệu người dùng được bảo vệ an toàn và cung cấp tính năng xác thực đáng tin cậy cho cả người tìm việc và doanh nghiệp tuyển dụng.
* Tối ưu hóa trải nghiệm người dùng di động: Phát triển ứng dụng di động hoặc tối ưu hóa giao diện người dùng cho các thiết bị di động để cung cấp trải nghiệm tốt nhất cho người dùng trên mọi nền tảng.
* Phát triển cộng đồng người dùng: Tạo ra một cộng đồng hoạt động tích cực của người tìm việc và doanh nghiệp tuyển dụng để chia sẻ kinh nghiệm và tạo ra một môi trường hỗ trợ và tương tác.

Với các hướng phát triển này, chúng tôi hy vọng dự án sẽ tiếp tục phát triển và mang lại giá trị đối với cả cộng đồng người tìm việc và doanh nghiệp tuyển dụng.

# TÀI LIỆU THAM KHẢO

1. "Fluent Python" của Luciano Ramalho, 2015
2. "Node.js Design Patterns" của Mario Casciaro, 2014.
3. <https://www.w3schools.com/nodejs/>
4. [https://www.w3schools.com/reactjs/](https://www.w3schools.com/vue/)
5. “Python Testing with Selenium: Learn to Implement Different Testing Techniques Using” của Sujay Raghavendra,