BÁO CÁO ĐỀ TÀI:

**Xây dựng ứng dụng phân tích và thu thập dữ liệu**

**trên Web việc làm**

MỤC LỤC

[CHƯƠNG 1: TỔNG QUAN 6](#_Toc27651735)

[1.1. Định nghĩa – mô tả bài toán 6](#_Toc27651736)

[1.1.1. Tên đề tài 6](#_Toc27651737)

[1.1.2. Giới thiệu đề tài 6](#_Toc27651738)

[1.1.3. Lý do chọn đề tài 6](#_Toc27651739)

[1.1.4. Mục tiêu 7](#_Toc27651744)

[1.1.5. Phương pháp nghiên cứu 7](#_Toc27651745)

[CHƯƠNG 2: CƠ SỞ LÝ THUYẾT 7](#_Toc27651752)

[2.1. Giới thiệu Python 7](#_Toc27651753)

[2.1.1. BeautifulSoup 7](#_Toc27651749)

[2.1.2. Selenium 8](#_Toc27651751)

[2.2. MYSQL/ MYSQL Workbench 9](#_Toc27651754)

[2.2.1. MySQL 10](#_Toc27651751)

[2.2.2. MySQL Workbench 10](#_Toc27651751)

[2.3. NodeJS 11](#_Toc27651754)

[2.4. VueJS 11](#_Toc27651754)

[2.5. Docker 12](#_Toc27651754)

[CHƯƠNG 3: PHÂN TÍCH VÀ THIẾT KẾ HỆ THỐNG](#_Toc27651755) 12

[3.1. Cào dữ liệu 13](#_Toc27651756)

[3.1.1. Thu tập data từ Vieclam24h.vn 13](#_Toc27651757)

[3.1.2. Thuật toán crawl dữ liệu 13](#_Toc27651758)

[3.1.3. Kiểm tra trùng lặp dữ liệu 16](#_Toc27651758)

[3.1.4. Kiểm tra API 16](#_Toc27651758)

[3.2. Chức năng hệ thống 19](#_Toc27651759)

[3.2.1. Admin 16](#_Toc27651758)

[3.2.2. Client 16](#_Toc27651758)

[3.3. Cơ sở dữ liệu (Database) 17](#_Toc27651759)

[3.4. Sơ đồ Usecase 19](#_Toc27651759)

[3.5. Đặc tả usecase 24](#_Toc27651759)

[3.5.1. Đăng nhập 24](#_Toc27651758)

[3.5.2. Đăng ký 24](#_Toc27651758)

[3.5.3. Tìm kiếm việc làm 26](#_Toc27651758)

[3.5.4. Nhận việc qua gmail 26](#_Toc27651758)

[3.5.5. Xoá tài khoản người dùng 27](#_Toc27651758)

[3.5.6. Sửa thông tin tài khoản người dùng 27](#_Toc27651758)

[3.5.7. Tìm kiếm tài khoản người dùng 27](#_Toc27651758)

[3.6. Sequence Diagram 27](#_Toc27651759)

[3.6.1. Đăng nhập 27](#_Toc27651758)

[3.6.2. Đăng kí 27](#_Toc27651758)

[3.6.3. Xoá người dùng 28](#_Toc27651758)

[3.6.4. Cập nhật thông tin 28](#_Toc27651758)

[3.6.5. Tìm kiếm người dùng 29](#_Toc27651758)

[3.6.6. Tìm kiếm việc làm 30](#_Toc27651758)

[3.7. Activity Diagram](#_Toc27651759) 31

[3.7.1. Đăng nhập 33](#_Toc27651758)

[3.7.2. Đăng kí 33](#_Toc27651758)

[CHƯƠNG 4: XÂY DỰNG WEBSITE 33](#_Toc27651760)

[4.1. Thiết kế giao diện](#_Toc27651761) 33

[4.1.1. Giao diện cho khách hàng 34](#_Toc27651762)

[4.1.2. Giao diện quản lý Admin 36](#_Toc27651763)

[4.2. Giao diện Admin](#_Toc27651777) 37

[4.3. Kết luận 40](#_Toc27651780)

[4.3. Hướng phát triển 40](#_Toc27651780)

**DANH SÁCH HÌNH ẢNH**

[Hình 1. Beautiful Soup 10](#_Toc27651471)

[Hình 2. NodeJS 12](#_Toc27651472)

[Hình 3. Vue 13](#_Toc27651473)

[Hình 4. Cơ sở dữ liệu 17](#_Toc27651474)

[Hình 5. Table CSDL 19](#_Toc27651475)

[Hình 6. Usecase người dùng 20](#_Toc27651476)

[Hình 7. Usecase nhà tuyển dụng 21](#_Toc27651477)

[Hình 8. Usecase quản lí người dùng 21](#_Toc27651478)

[Hình 9. Sơ đồ tuàn tự 22](#_Toc27651479)

[Hình 10. Usecase tìm kiếm việc làm 23](#_Toc27651480)

[Hình 11. Sơ đồ tuần tự 23](#_Toc27651481)

[Hình 12. Đăng kí 27](#_Toc27651482)

[Hình 13. Đăng nhập 27](#_Toc27651483)

[Hình 14. Trang Index 28](#_Toc27651484)

[Hình 15. Thông tin chi tiết 28](#_Toc27651485)

[Hình 16. Gửi gmail 29](#_Toc27651486)

[Hình 17. Giao diện list user 30](#_Toc27651487)

# TỔNG QUAN

## 1.Định nghĩa và mô tả bài toán

### Tên đề tài

### “Xây dựng ứng dụng phân tích và thu thập dữ liệu trên Web việc làm “

### 1.1.2 Giới thiệu đề tài

Trong thời đại công nghệ 4.0 đang diễn ra song hành với nền kinh tế không ngừng vươn mình mạnh mẽ, đòi hỏi mỗi doanh nghiệp phải có những giải pháp chiến lược phù hợp đúng đắn để có thể đứng vững và phát triển. Một giải pháp được đặt ra là ứng dụng tiến bộ của công nghệ thông tin vào công việc tìm kiếm việc làm. Những công việc hằng ngày không còn là thao tác thủ công với một đống hỗn độn tài liệu nữa mà được thay bằng những chương trình hỗ trợ cho mọi công việc. Nó sẽ giảm thiểu được những khó khăn trong công việc tìm kiếm việc làm như: nhân lực, thời gian, độ chính xác của thông tin…Giúp cho các doanh nghiệp có thể tuyển dụng nhânsự cũng như các người dùng tìm kiếm việc là một cách dễ dàng hơn. Do đó dự án xây dựng và phát triển một chương trình hoàn chỉnh ứng dụng những kỹ thuật và công nghệ mới nhất để giúp cho việc tìm kiếm nhân sự cũng như việc làm dễ dàng trên toàn quốc.

Website cung cấp thông tin người dùng, doanh nghiệp, giúp cho công ty có thể tìm kiếm nhân sự và ngược lại các người dùng có thể lựa chọn công ty phù hợp với mình.

### 1.1.3 Lý do chọn đề tài

Để tìm kiếm việc làm phù hợp không phải là một việc dễ dàng, nhu cầu của nhà tuyển dụng, nhu cầu của người tìm việc. Bên cạnh đó thì việc quản lý tìm kiếm và đăng việc như thế nào cũng là một câu hỏi lớn cần giải quyết. Vậy việc quản lý việc làm như thế nào và bằng cách gì? Quản lý việc tìm kiếm như thế nào? Hồ sơ và việc làm lưu trữ như thế nào? Quản lý như thế nào được gọi là tiện lợi, ít tốn công sức nhưng lại cho ra kết quả một cách hiệu quả. Trang web hỗ trợ việc tìm kiếm việc làm cho sinh viên ngàng công nghệ thông tin, đăng thông tin tuyển dụng. Trang Web đáp ứng nhu cầu tiềm việc làm hiện nay. Tạo cơ hội cho sinh viên tìm được việc làm thích hợp cũng như nhà tuyển dụng tìm được nhân tài phù hợp.

### 1.1.4 Mục tiêu

Xây dựng các chức năng cần thiết của một website tìm kiếm việc làm. Website có thể cung cấp các chức năng cần thiết cho người dùng để tìm kiếm việc làm hay tuyển dụng cũng như các chức năng cần thiết dành cho quản trị viên để quản lí website. Ngoài ra giao diện website cũng phải được bố trí hợp lí dễ nhìn, thân thiện với người dùng, phù hợp với nhiều loại thiết bị.

### 1.1.5 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.



BeautifulSoup Python

## 2.2 MySQL / MySQL Workbench

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 ungó 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 ungó trở thành một lựa chọn tuyệt vời cho các ứng dụng web và doanh nghiệp.

MySQL Workbench là một công cụ phát triển cơ sở dữ liệu đồ họa (GUI) dành cho MySQL. Nó cung cấp một loạt các tính năng giúp cho việc thiết kế, phát triển và quản lý cơ sở dữ liệu MySQL trở nên dễ dàng hơn.

Tính năng chính của MySQL Workbench :

* Thiết kế cơ sở dữ liệu: MySQL Workbench cung cấp một giao diện đồ họa trực quan để giúp bạn tạo và chỉnh sửa lược đồ cơ sở dữ liệu.
* Quản lý cơ sở dữ liệu: MySQL Workbench cung cấp các công cụ để giúp bạn quản lý cơ sở dữ liệu của mình, chẳng hạn như tạo và quản lý người ung, tạo và quản lý các nhiệm vụ bảo trì và khôi phục cơ sở dữ liệu.
* Truy vấn cơ sở dữ liệu: MySQL Workbench cung cấp một trình soạn thảo truy vấn tích hợp cho phép bạn tạo và chạy các truy vấn SQL.
* Mô phỏng cơ sở dữ liệu: MySQL Workbench cung cấp một công cụ mô phỏng cơ sở dữ liệu cho phép bạn kiểm tra các truy vấn SQL của mình trong một môi trường an toàn.

Chính vì những lợi ích trên mà nhóm chúng em quyết định chọn hệ quản trị CSDL và công cụ quản lí cơ sở dữ liệu là MySQL và MySQL Workbench.

## 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. Điều này có thể giúp ứng dụng xử lý được nhiều người dùng hơn mà không bị quá tải.



NodeJS

## 2.4 VueJS

Hiện nay có nhiều ngôn ngữ được ứng dụng để code Front-End website điển hình như là: TypeScript, CSS, VueJS, React… Trong đó nhóm chúng em chọn VueJS làm ngôn ngữ để code Fronend cho website lần này vì những ưu điểm sau đây:

* Cộng đồng người dùng lớn: Vue.js có một cộng đồng phát triển lớn và năng động. Điều này làm cho nó trở nên dễ dàng tìm kiếm trợ giúp và tham gia các dự án cộng đồng.
* Dễ học: Vue.js được thiết kế để dễ học. Nó có cú pháp đơn giản và dễ hiểu.
* Hiệu suất: Vue.js có thể được mở rộng bằng các thành phần tùy chỉnh. Điều này làm cho nó trở nên linh hoạt và có thể tùy chỉnh.



VueJS

## 2.5 Postman

Postman là một công cụ phổ biến được sử dụng để phát triển, kiểm thử, và tương tác với các API (Application Programming Interface). Postman hỗ trợ nhiều phương thức HTTP như GET, POST, PUT, DELETE và nhiều phương thức khác. Có khả năng kiểm thử API với khả năng chạy các bộ kiểm thử tự động và kiểm thử đơn với giao diện dễ sử dụng. Postman là một công cụ mạnh mẽ và linh hoạt giúp tăng cường quá trình phát triển và kiểm thử API, đồng thời giúp cộng đồng lập trình viên chia sẻ và tương tác với nhau qua các bộ sưu tập.



## 2.6 Docker

Docker là một nền tảng mã nguồn mở cho phép bạn tạo, chạy và quản lý các ứng dụng trong các môi trường ảo hóa được gọi là container. Container là các gói phần mềm nhẹ, độc lập, chứa tất cả các thành phần cần thiết để chạy một ứng dụng, bao gồm mã, thư viện, cấu hình và các file hệ thống.

Docker được sử dụng rộng rãi trong các môi trường phát triển, sản xuất và thử nghiệm. Nó được sử dụng bởi các doanh nghiệp lớn và nhỏ để chạy các ứng dụng web, ứng dụng di động, ứng dụng máy tính để bàn và ứng dụng thời gian thực.

Docker có thể được sử dụng để tạo container cho các ứng dụng được viết bằng bất kỳ ngôn ngữ lập trình nào.



Docker

**CHƯƠNG 3: PHÂN TÍCH VÀ THIẾT KẾ HỆ THỐNG**

## Phân tích thuật toán

### 3.1.1. Thu tập data từ Vieclam24h.vn

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 khả thi.

* 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.

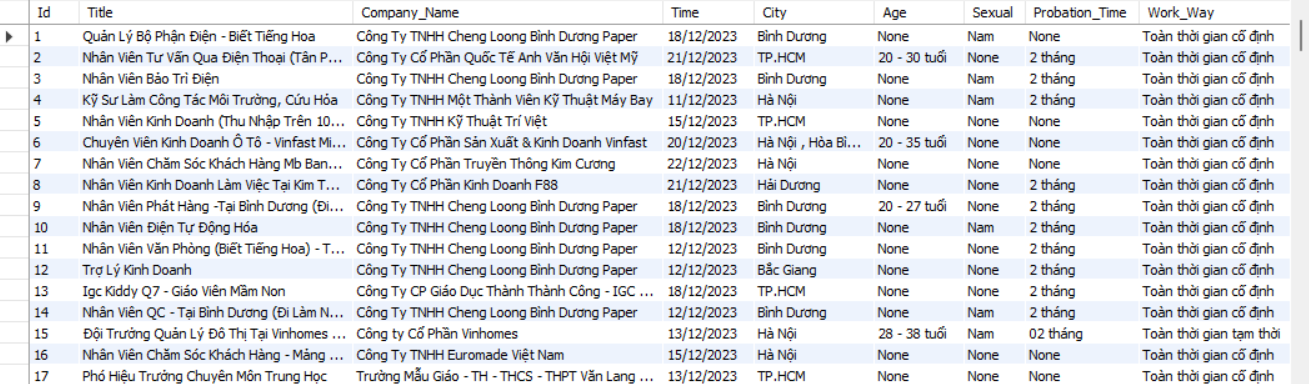
+) Các tham số được sử dụng như sau:' ': Đây là ký tự hoặc chuỗi được sử dụng để nối các đoạn văn bản khi chúng được trích xuất. Trong trường hợp này, đoạn văn bản sẽ được nối bằng khoảng trắng. Strip=True: Thiết lập strip thành True để loại bỏ các khoảng trắng và ký tự xuống dòng ở đầu và cuối đoạn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

### 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.

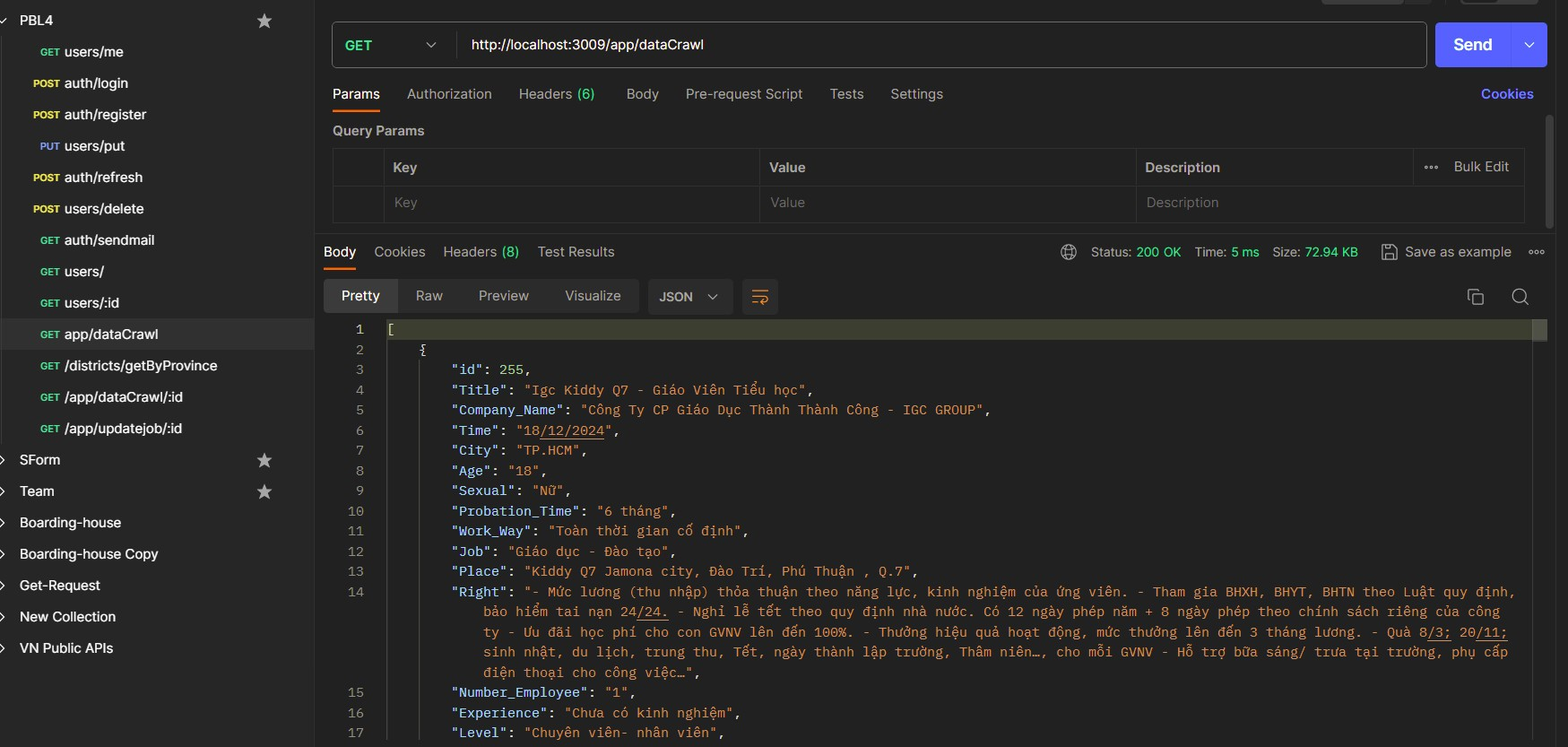
* Đầu tiên những dữ liệu mới cào về mình cho vào 1 mảng.
* Tiếp theo viết hàm để lấy tất cả các dữ liệu từ DB.
* Mình sẽ cho dữ liệu lần lượt kiểm tra qua 7 cột đầu tiên. Nếu dữ liệu giống nhau hết cả 7 cột thì sẽ kết luận đây là dữ liệu đã có từ trước và bỏ qua nó.
* Những dữ liệu còn lại sẽ tiếp tục được thêm vào Database mà không sợ bị rơi vào tình trạng trùng lặp dữ liệu.
* Tại sao lại là 7 cột. Vì trong quá trình cào dữ liệu em nhận thấy có những link việc làm trùng cả tên title bài viết, trùng tên công ty, trùng vị trí cần tuyển dụng. Nhưng nó vẫn được tính là 1 data hoàn toàn mới vì được truyển ở những thành phố khác nhau chẳng hạn …



Cơ sở dữ liệu

### 3.1.4 Kiểm thử API/ HTTP bằng Postman

Dưới đây là hình ảnh kiểm thử thành công 1 bản ghi trong dữ liệu (tương ứng với 1 công việc được up lên website việc làm sau khi được cào về từ Vieclam24h.vn)



## 3.2. Chức năng hệ thống

### 3.2.1. Admin

Admin

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

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

- Chức năng xem thông tin tài khoản của admin.

### 3.2.2. Client

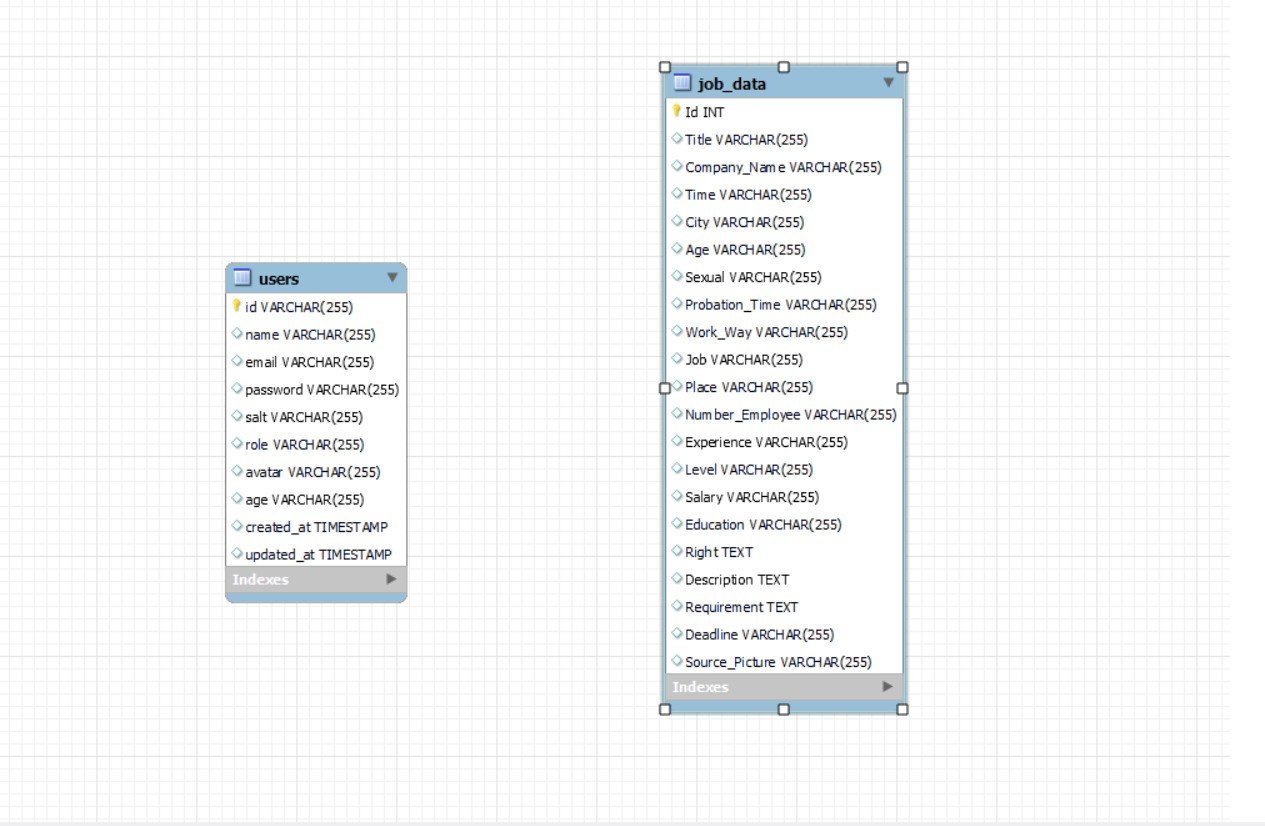
Client

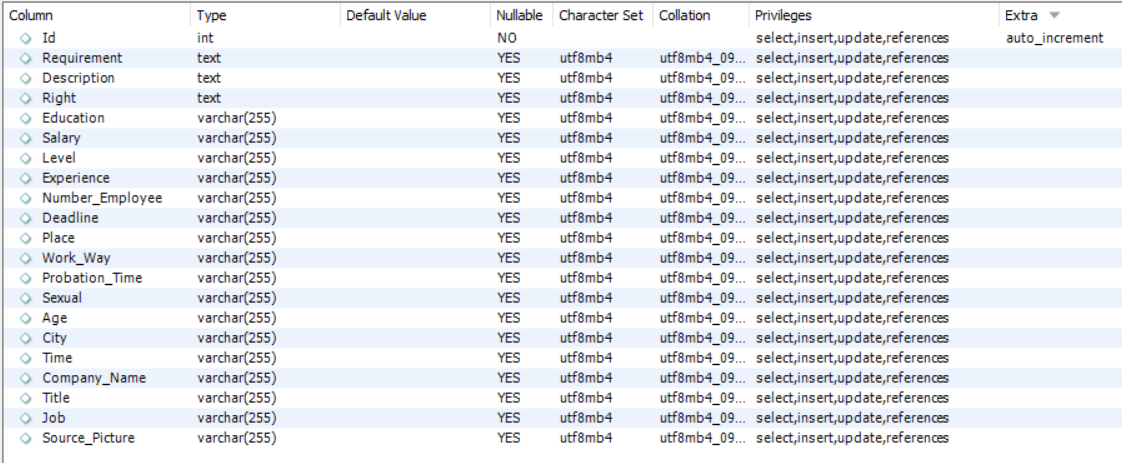
* Chức năng đăng nhập.
* Chức năng đăng ký.
* Chức năng xem thông tin chi tiết công việc.
* Chức năng quản lý tài khoản của mình.
* Chức năng tìm kiếm công việc theo từ khoá, địa điểm.
* Được gửi thông tin công việc về gmail cá nhân.

## 3.3 Cơ sở dữ liệu

- Job\_data: Table này gồm 21 column bao gồm `Title`, `Company\_Name`, `Time`, `City`,`Age`,`Sexual`,`Probation\_Time`,`WorkWay`,`Job`,`Place`,`Number\_Employee`,`Experience`,`Level`,`Salary`,`Education`,`Right`,`Description`,`Requirement`, `Deadline`, `Source\_Picture`. 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\_Name: Tên công ty đăng thông tin tuyển dụng hoặc cung cấp việc làm.
3. Time: Thời điểm đăng thông tin tuyển dụng hoặc thời gian công việc được đăng.
4. City: Địa điểm hoặc thành phố nơi công việc đặt tại.
5. Age: Độ tuổi yêu cầu hoặc độ tuổi ưu tiên cho người dùng.
6. Sexual: Yêu cầu giới tính của người dùng (Nam, Nữ, hoặc không yêu cầu).
7. Probation Time: Thời gian thử việc (nếu có).
8. Work Way: Hình thức làm việc (toàn thời gian, bán thời gian, làm việc từ xa, etc.).
9. Job: Chức vụ hoặc công việc cụ thể.
10. Place: Địa điểm làm việc cụ thể.
11. Number\_Employee: Số lượng nhân viên cần tuyển dụng.
12. Experience: Yêu cầu kinh nghiệm làm việc.
13. Level: Cấp bậc công việc (ví dụ: Junior, Intermediate, Senior).
14. Salary: Mức lương được cung cấp cho công việc.
15. Education: Yêu cầu trình độ học vấn của người dùng.
16. Right: Quyền lợi và phúc lợi cho nhân viên.
17. Description: Mô tả công việc hoặc thông tin chi tiết về công ty và vị trí tuyển dụng.
18. Requirement: Các yêu cầu khác đối với người dùng.
19. Deadline: Hạn chót nộp đơn ứng tuyển.
20. Source\_Picture: Liên kết hoặc đường dẫn đến hình ảnh liên quan đến công việc hoặc công ty.

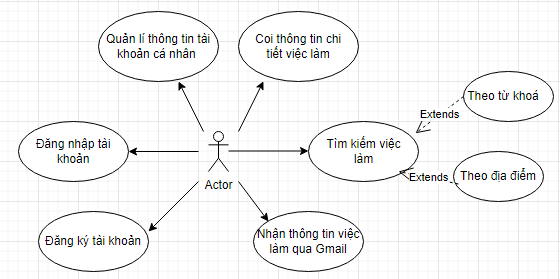




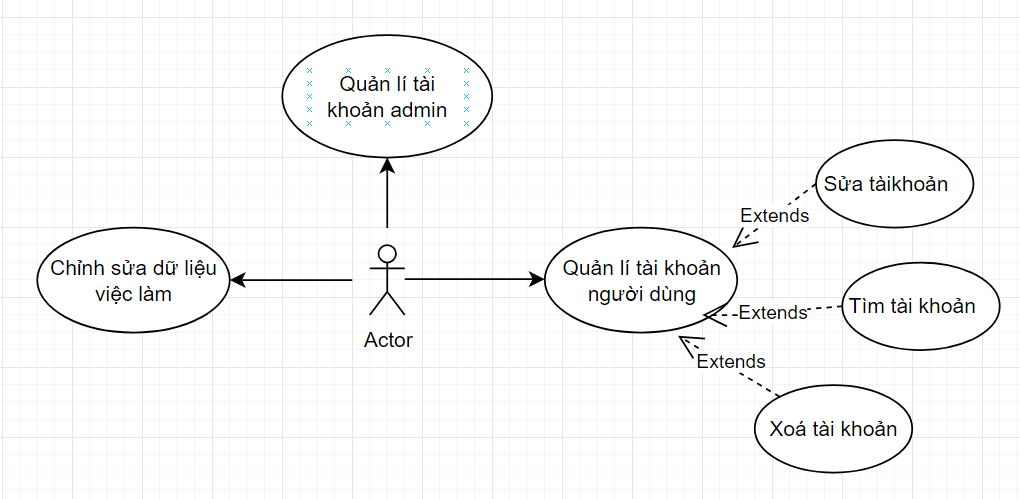
## Trong đó có 3 trường là ‘Requirement, Description, Right’ có kiểu dữ liệu là Text. Bởi vì TEXT là kiểu dữ liệu có kích thước lớn hơn so với CHAR hay là VARCHAR. Vậy nên trường hợp này chúng em dùng TEXT để biểu diễn cho 3 trường trên. Bởi vì đặc thù của 1 trang web việc làm 24h.vn thì phần mô tả,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

## 3.4. Sơ đồ

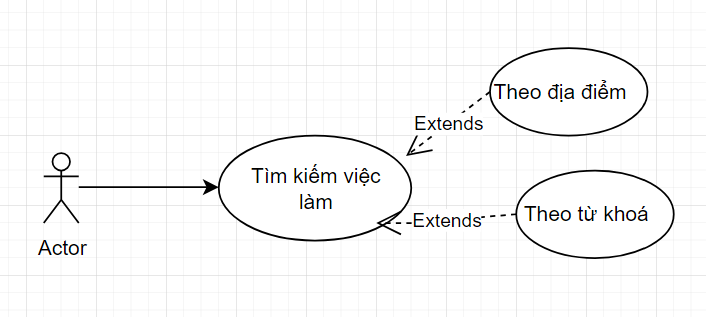
- Sơ đồ usecase



Usecase người dùng (User)



Usecase Admin



Usecase tìm kiếm việc làm

## 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 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 |

### 3.5.6 Sửa thông tin tài khoản người dùng

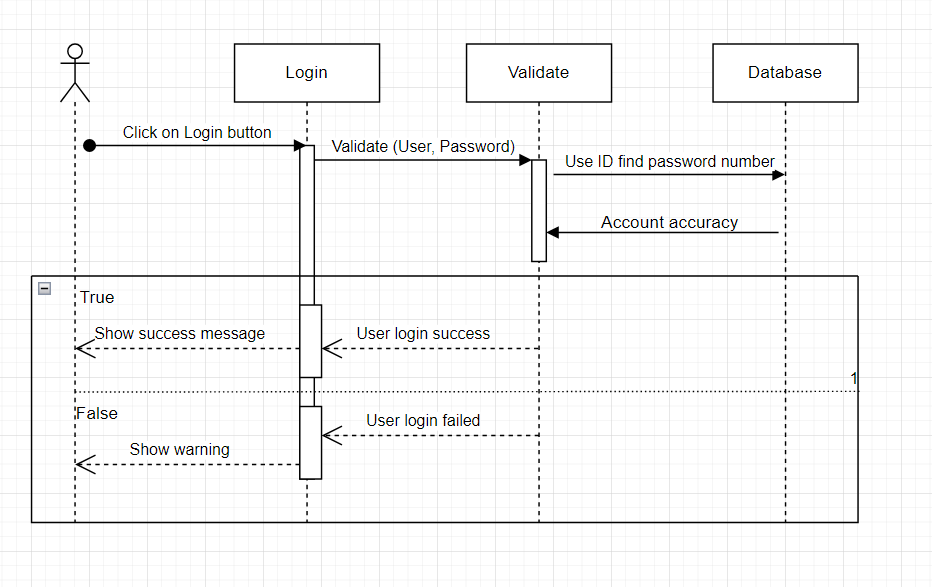
|  |  |
| --- | --- |
| Tên usecase | Usecase sửa tài khoản người dùng |
| Tác nhân | Admin |
| Mô tả | Admin muốn sửa 1 tài khoản của user trong hệ thống data quản lý tài khoản |
| Điều kiện tiên quyết | Hệ thống được đăng nhập vào tải khoản của admin |
| Dòng dự kiện chính | 1.Admin đăng nhập vào tài 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 |

### 3.5.7 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 |

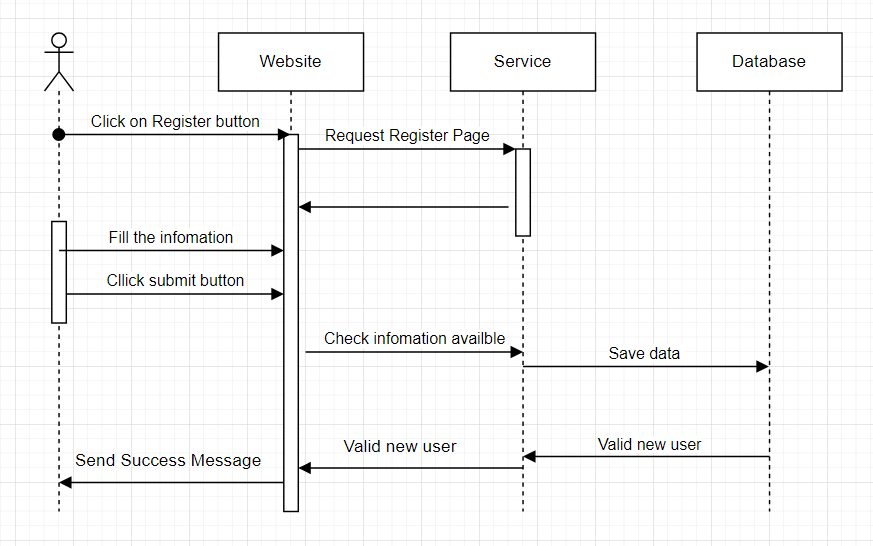
## 3.6 Sequence Diagram (Sơ đồ tuần tự)

### 3.6.1 Đăng nhập



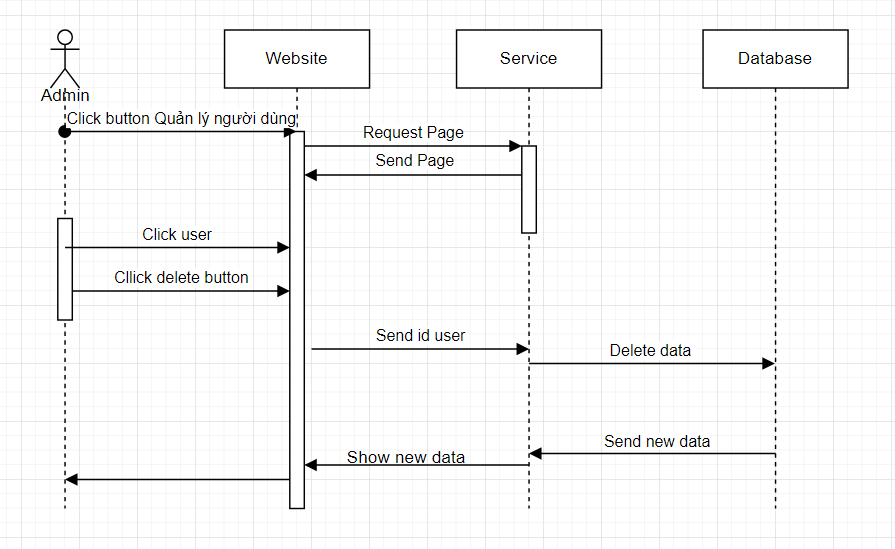
Sequence Đăng nhập

### 3.6.2 Đăng kí

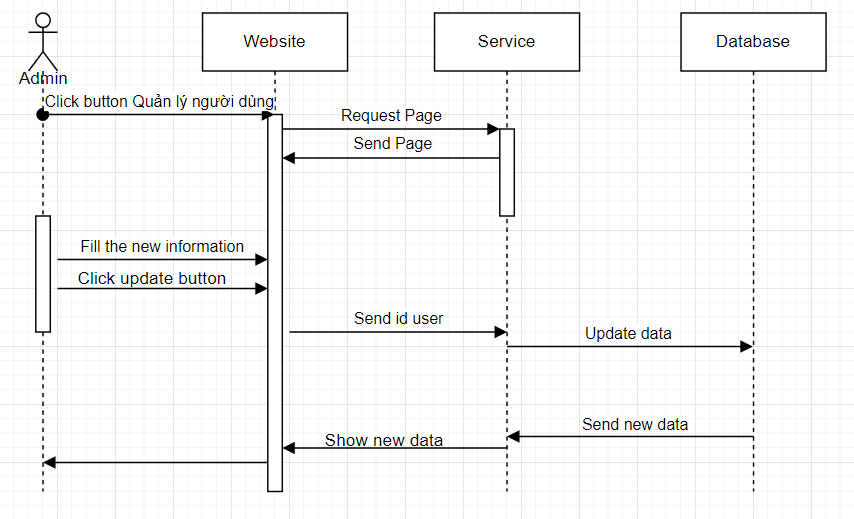


­Sequence đăng kí

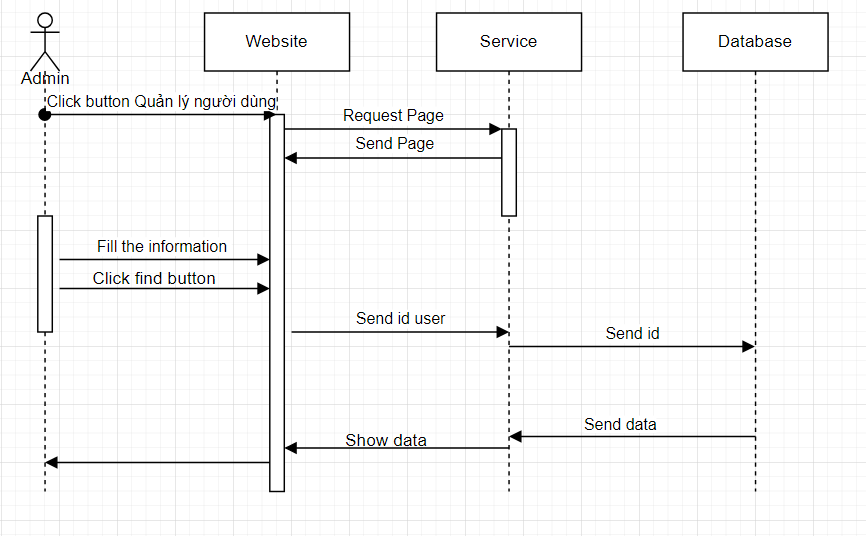
### 3.6.3 Xoá người dùng



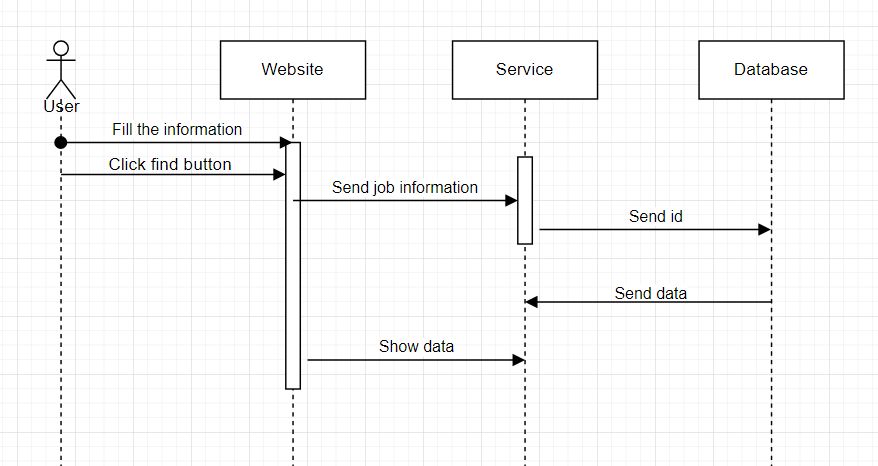
### 3.6.4 Cập nhật thông tin người dùng



### 3.6.5 Tìm kiếm người dùng

****

### 3.6.6 Tìm kiếm công việc

****

## 3.7 Activity Diagram

### 3.7.1 Đăng ký

A screenshot of a diagram

Description automatically generated

### 3.7.2 Đăng nhập

A screenshot of a diagram

Description automatically generated

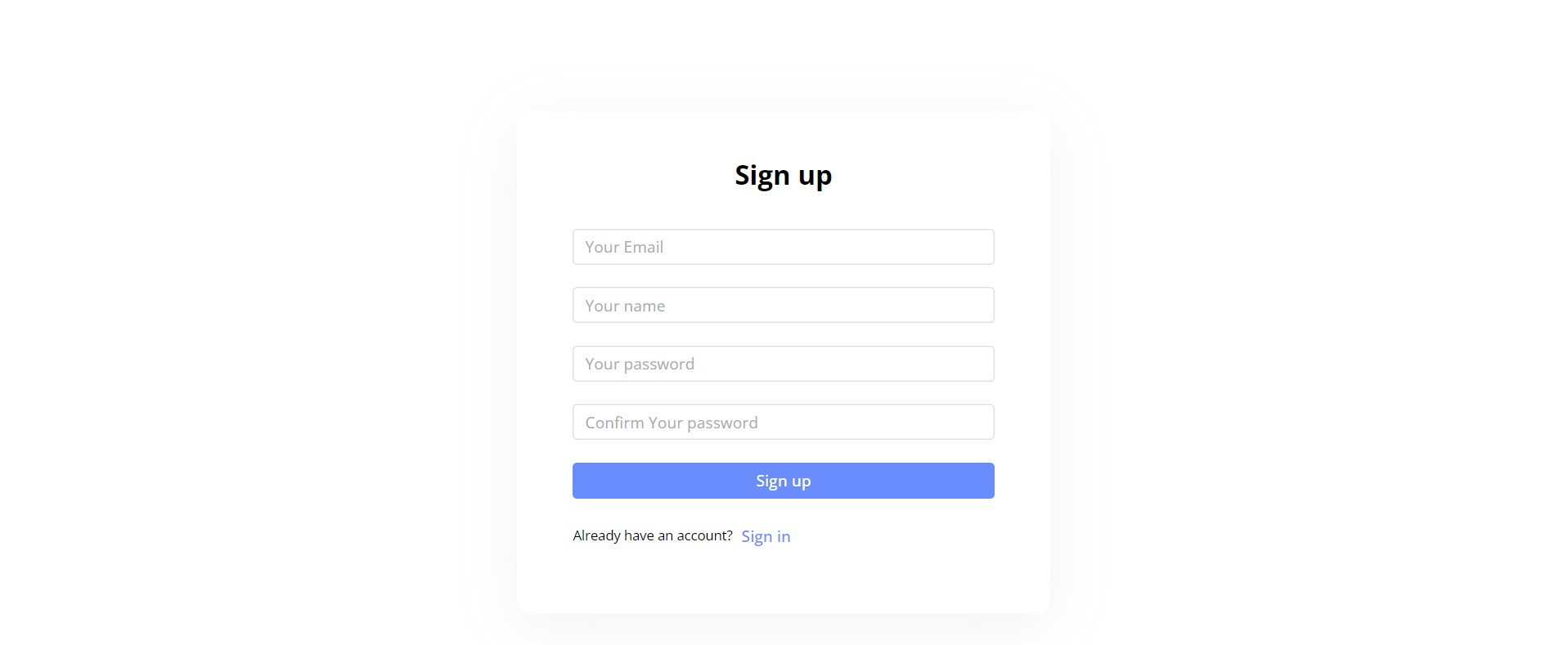
**CHƯƠNG 4: XÂY DỰNG WEBSITE**

## 4.1 Thiết kế giao diện người dùng (User)

### 4.1.1 Giao diện đăng kí

Đây là trang mà người có thể tự đăng ký tài khoản cho mình. Người dùng chỉ cần đăng ký một tài khoản hoàn toàn miễn phí tại đây là đã có toàn quyền sử dụng những tính năng giúp cho việc tìm kiếm người dùng và tìm kiếm việc làm trở nên dễ dàng hơn.

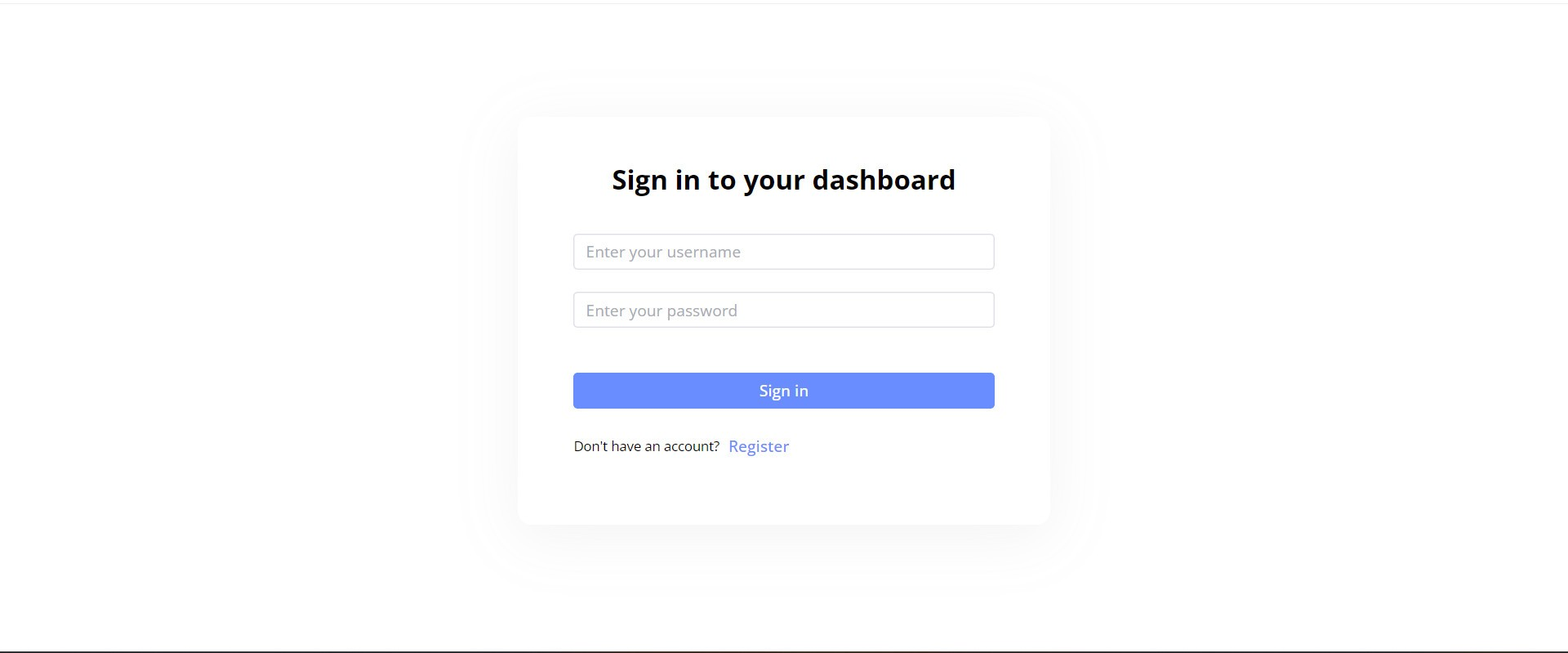
Để tạo được tài khoản, người dùng cần phải nhập đầy đủ thông tin mà hệ thống yêu cầu bao gồm: Tên đăng nhập, mật khẩu, địa chỉ email. Sau khi hoàn tất việc nhập thông tin trên, người dùng bấm chọn nút “đăng ký” nằm ngay phía dưới để hoàn tất việc đăng ký. Trong quá trình lưu dữ liệu nếu có lỗi sẽ có thông báo gửi đến người dùng, ngược lại người dùng sẽ được đưa đến trang đăng nhập kèm thông báo đăng ký thành công.



Đăng ký

### 4.1.2 Giao diện đăng nhập

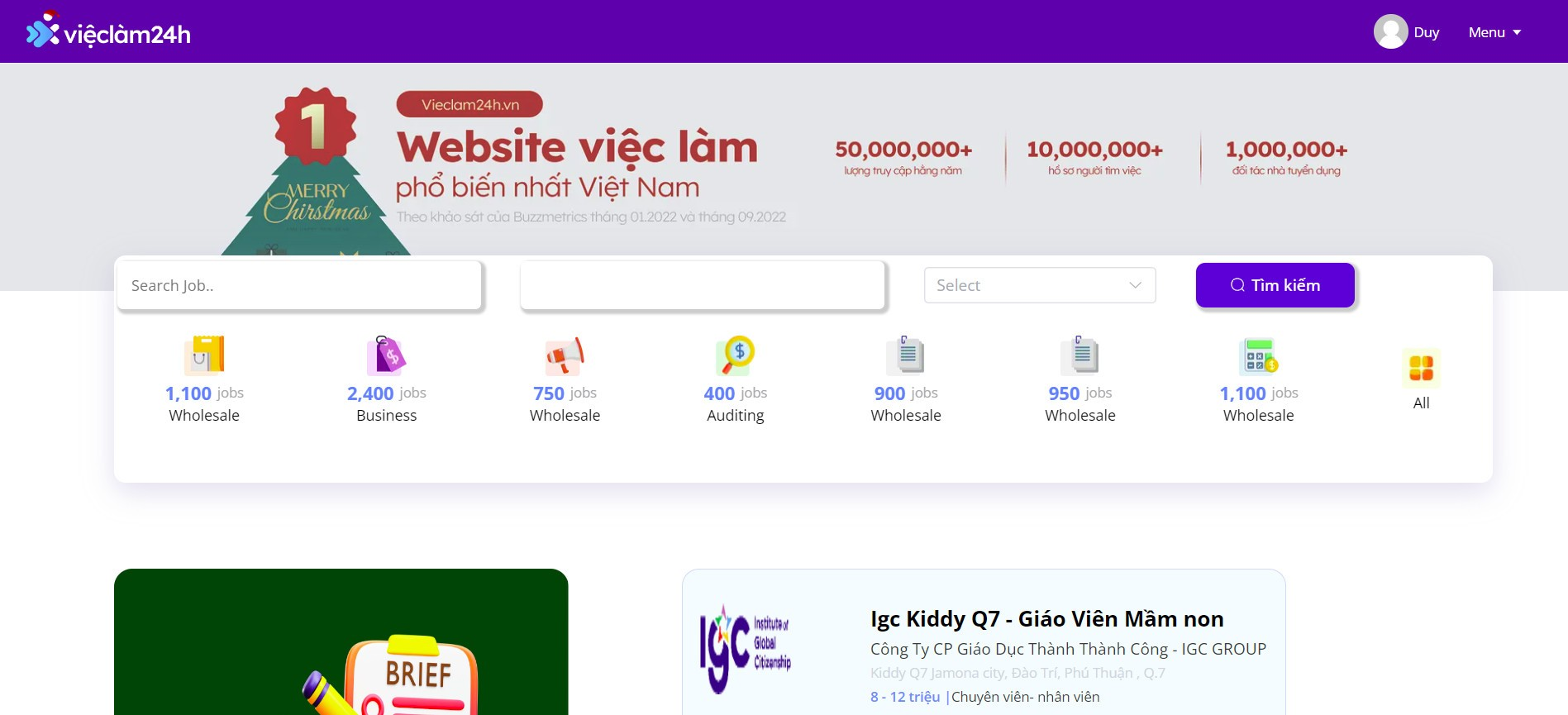
Người dùng sau khi đã có tài khoản trong hệ thống sẽ đăng nhập tại trang này, các thông tin cần nhập bao gồm tên đăng nhập và mật khẩu đã được đăng ký trước đó. Tại đây còn có tùy chọn. Lưu mật khẩu giúp cho người dùng giảm bớt thao tác khi vào hệ thống, nếu quan trọng về vấn đề bảo mật thì không nên kích hoạt tùy chọn này.



Đăng nhập

### 4.1.3 Giao diện trang chủ

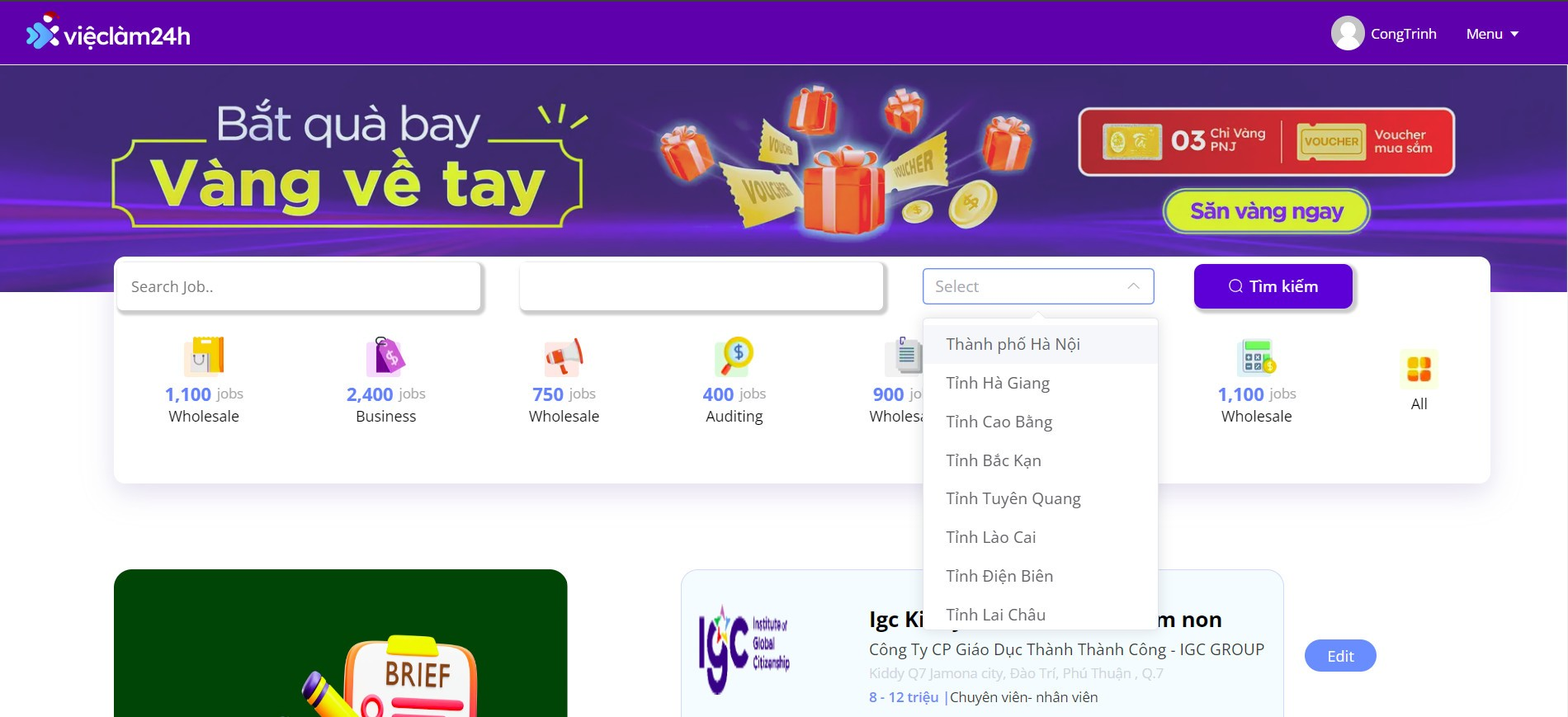
Sau khi đăng nhập thành công, người dùng sẽ được đưa đến trang chủ của ứng dụng. Ở đây sẽ có mục tìm kiếm theo địa điểm cũng như theo ngành nghề để các người dùng có thể thoải mái lựa chọn.



Trang index

### 4.1.4 Tìm kiếm việc làm theo thành phố / loại công việc

Sau khi đăng nhập thành công. Tại trang chủ của website user có thể tìm kiếm việc làm theo mục tiêu của bản thân dựa trên loại việc làm hoặc địa điểm mà user đó muốn tìm việc.



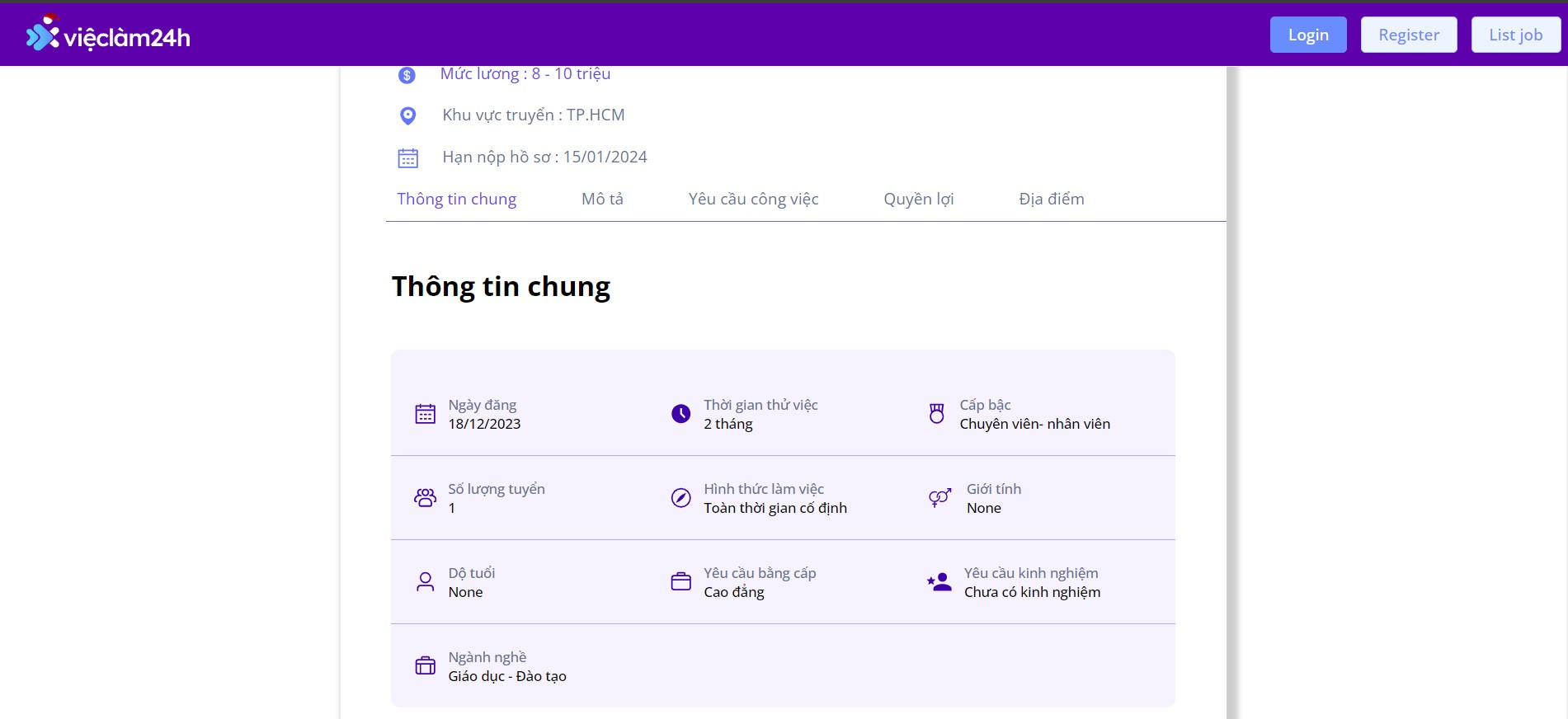
Tìm kiếm việc theo thành phố cụ thể

### 4.1.5 Thông tin chi tiết của 1 việc làm

Dưới đây là thông tin chi tiết của một công việc sau khi mà người dùng nhấn vào để xem các thông tin chi tiết hơn về việc làm mà mình quan tâm.

Ở trang chi tiết này sẽ gồm có:

1. Ngày đăng bài viết (18/12/2023)
2. Thời gian thử việc (2 tháng)
3. Cấp bậc (Chuyên viên - Nhân viên)
4. Số lượng nhân viên được tuyển dụng (1)
5. Hình thức làm việc (Toàn thời gian cố định)
6. Giới tính (None – Không yêu cầu)
7. Độ tuổi (None – Không yêu cầu)
8. Bằng cấp yêu cầu (Cao đẳng)
9. Kinh nghiệm yêu cầu (Chưa có kinh nghiệm)
10. Ngành nghề của việc làm (Giáo dục – Đào tạo)



Trang thông tin chi tiết

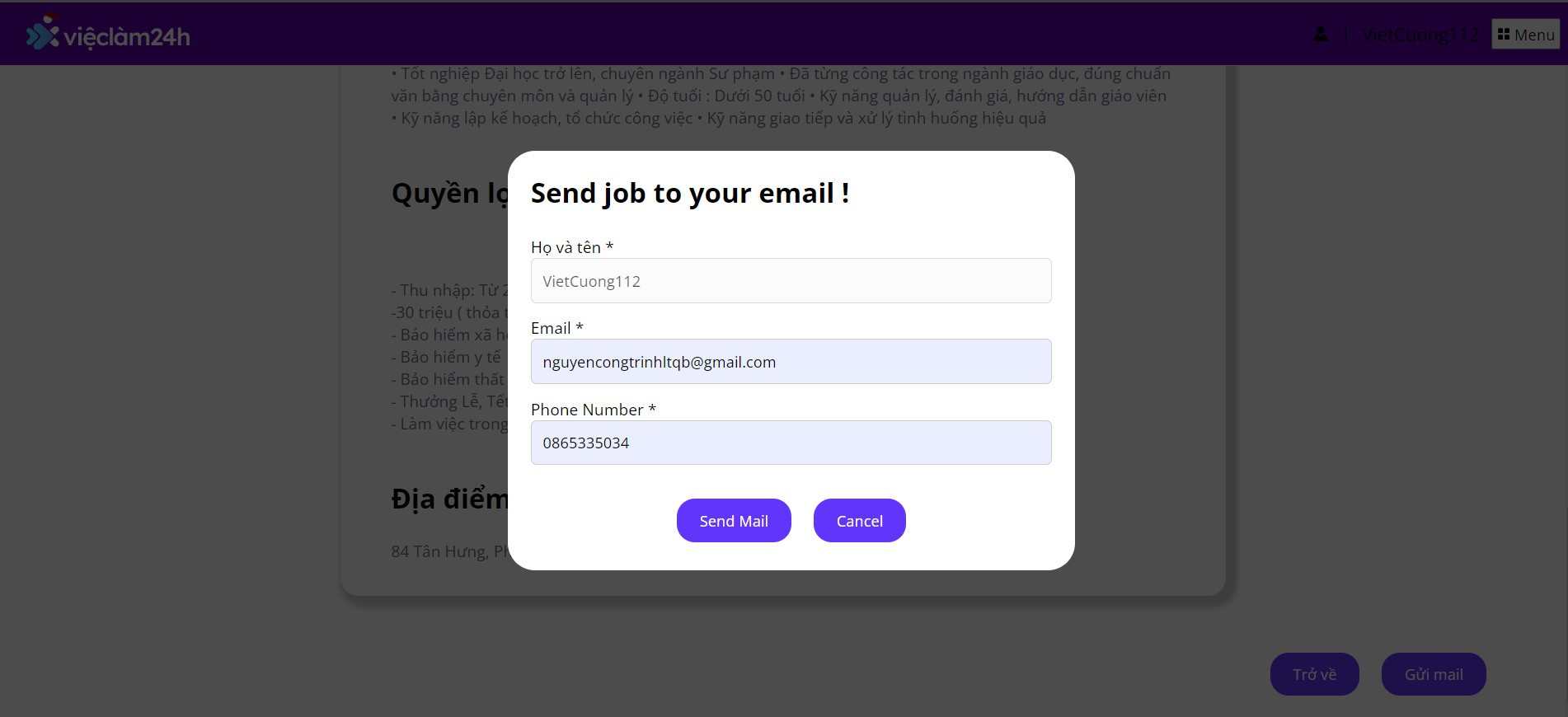
### 4.1.6 Gmail được gửi về

Sau khi user quyết định tìm hiểu thêm về công việc nào đó thì chỉ cần nhấn vào nút ‘Gửi gmail’ để thông tin chi tiết được gửi về gmail của user đó.

### 

Đăng nhập để được gửi mail

* Người dùng nhập vào email và số điện thoại của mình



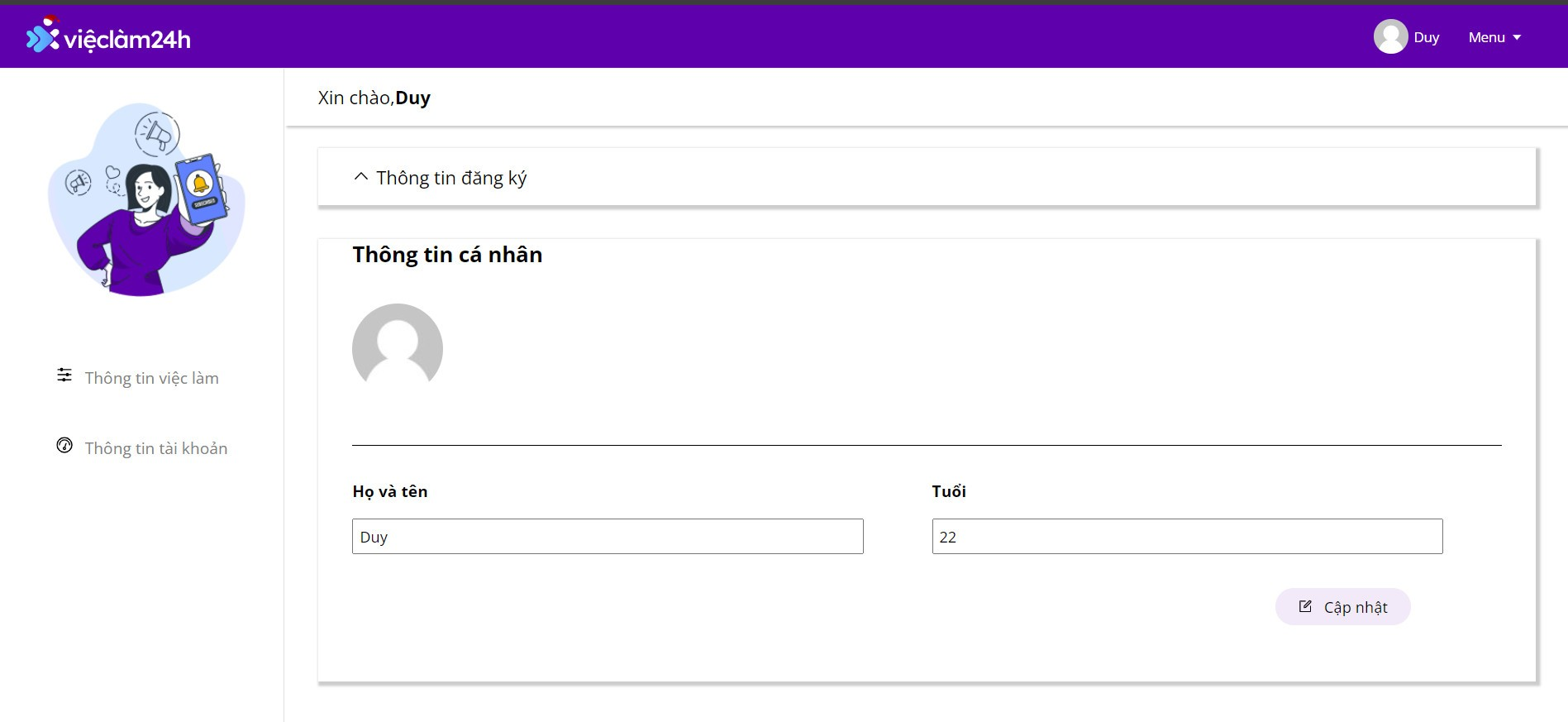
Thông tin việc làm sẽ được gửi về gmail của user

### 4.1.7 Xem thông tin chi tiết của tài khoản user

Mỗi user sẽ xem được thông tin tài khoản cá nhân của bản thân bao gồm:

* Họ tên
* Tuổi

Ngoài ra họ còn có thể cập nhật các thông tin trên nếu muốn.

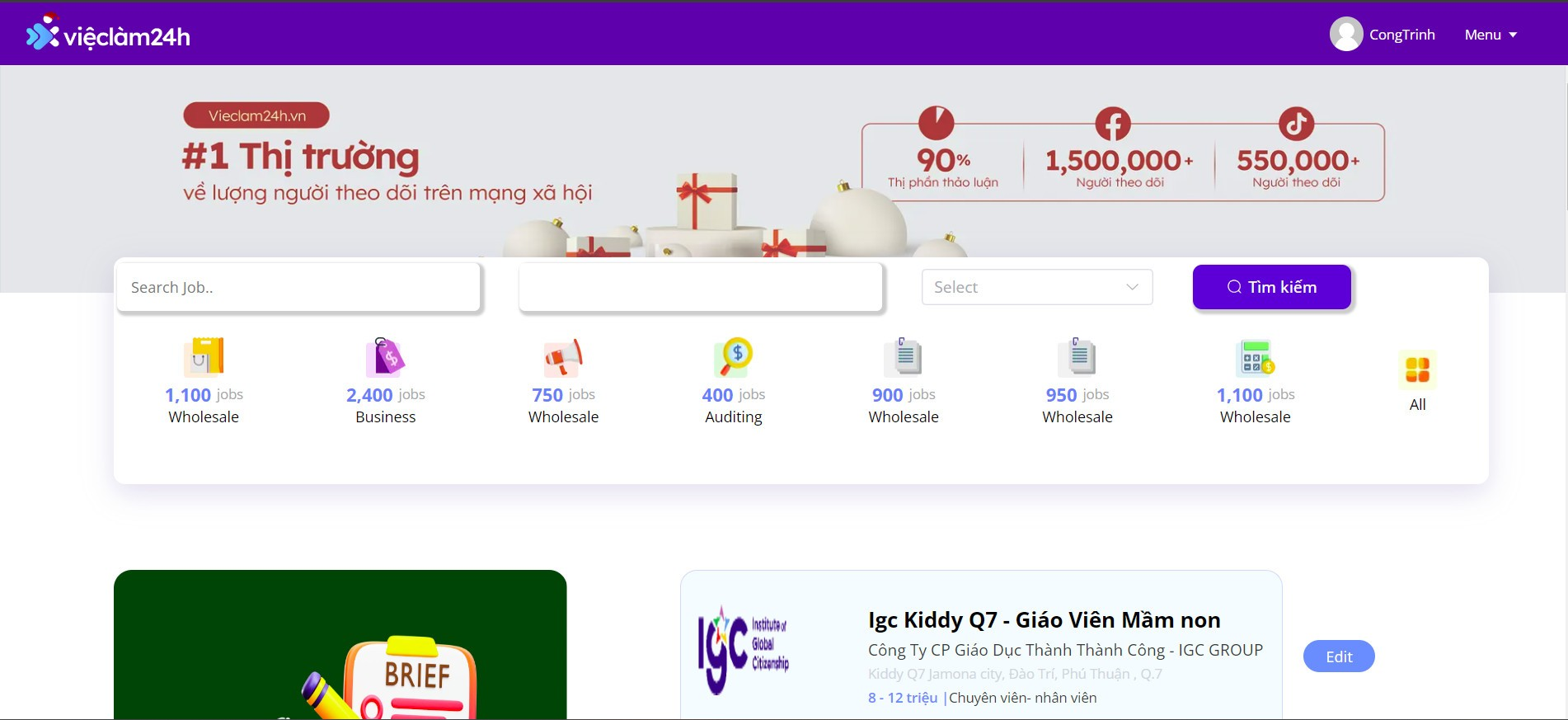


User xem thông tin cá nhân của mình

## 4.2 Giao diện Admin

### 4.2.1 Giao diện trang chủ admin

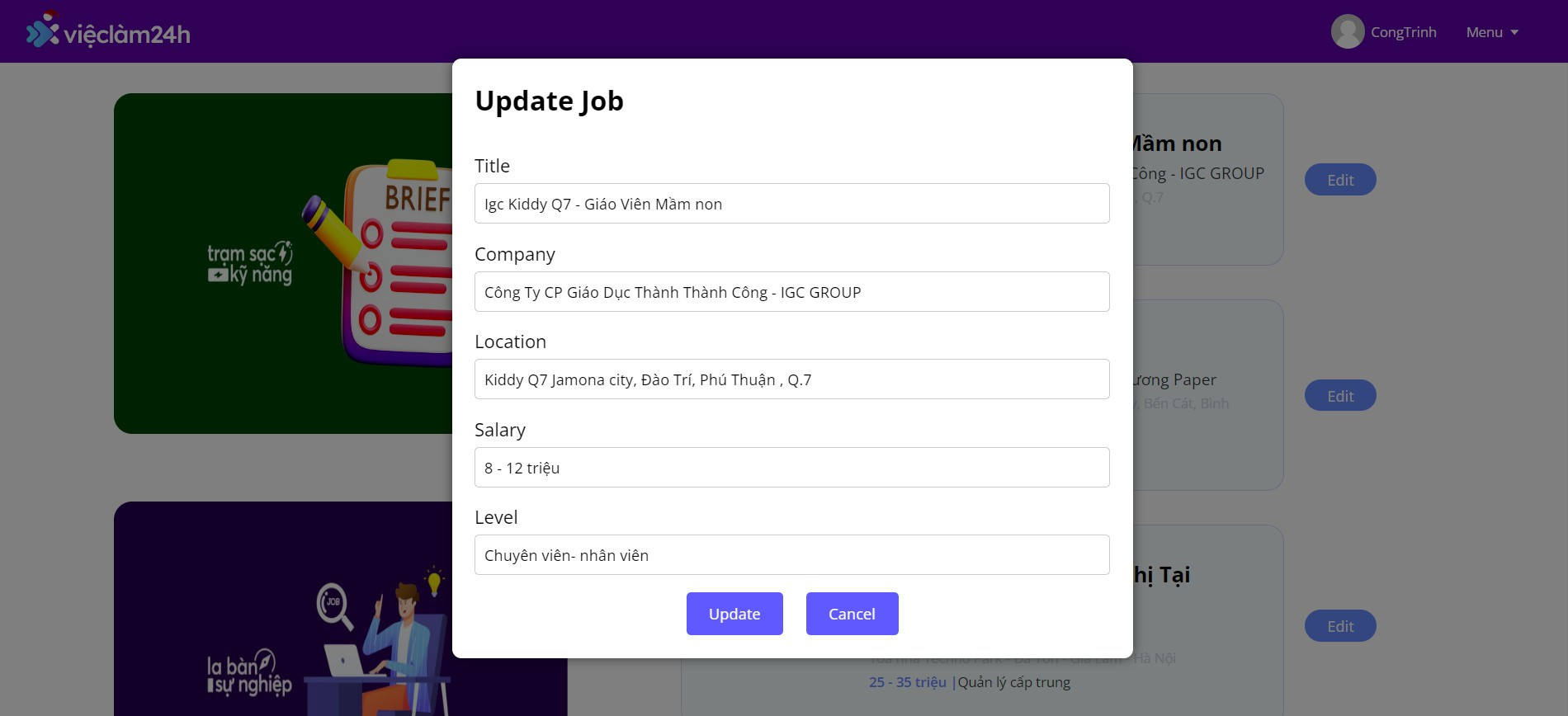
Sau khi đăng nhập thành công vào tài khoản Admin, thì sẽ được đưa đến trang chủ của website như hình sau



Trang index Admin

### 4.2.2 Update lại việc làm

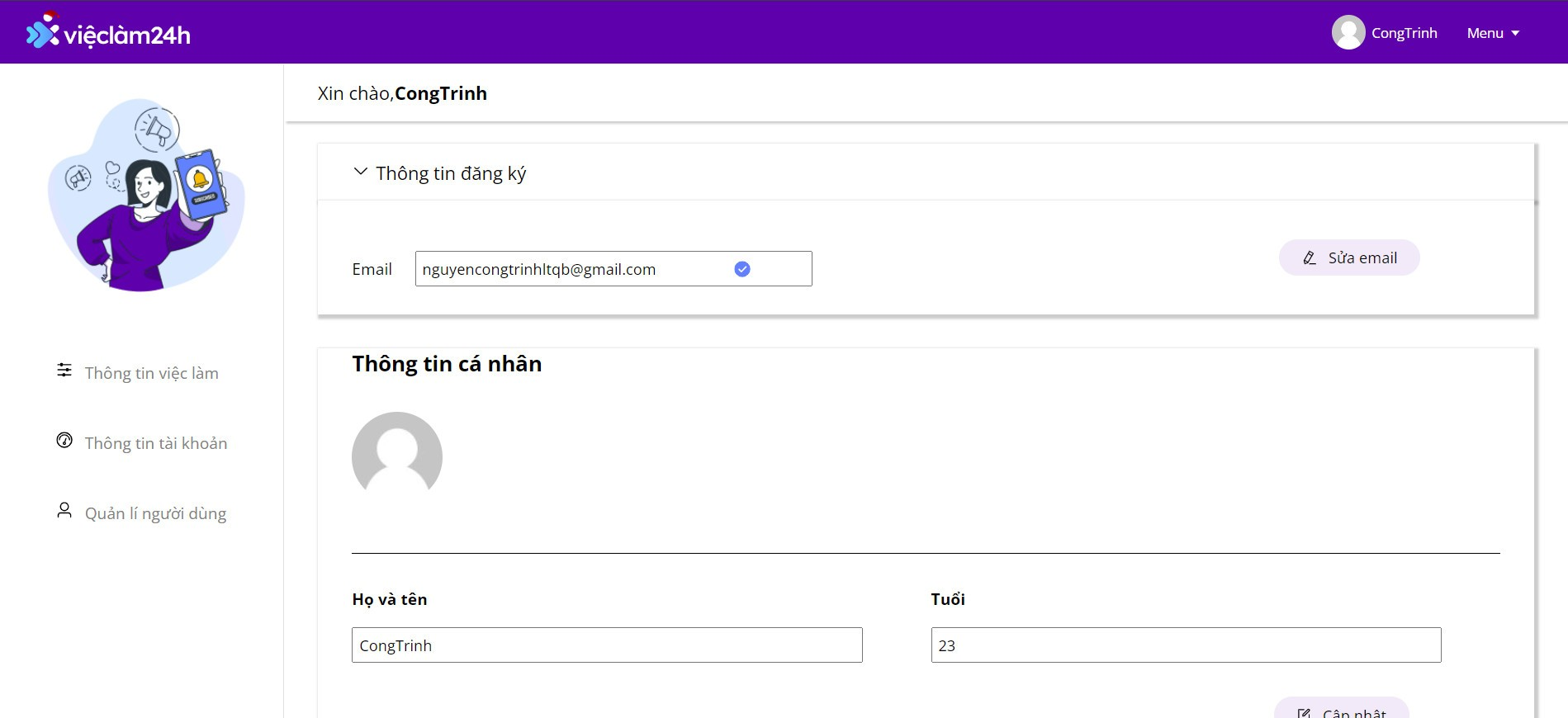
Admin sẽ có chức năng chỉnh sửa lại các thông tin việc làm trong trường hợp thông tin chưa chính xác hoặc không phù hợp với yêu cầu.



Giao diện chỉnh sửa lại thông tin việc làm

### 4.2.3 Quản lý thông tin tài khoản cá nhân

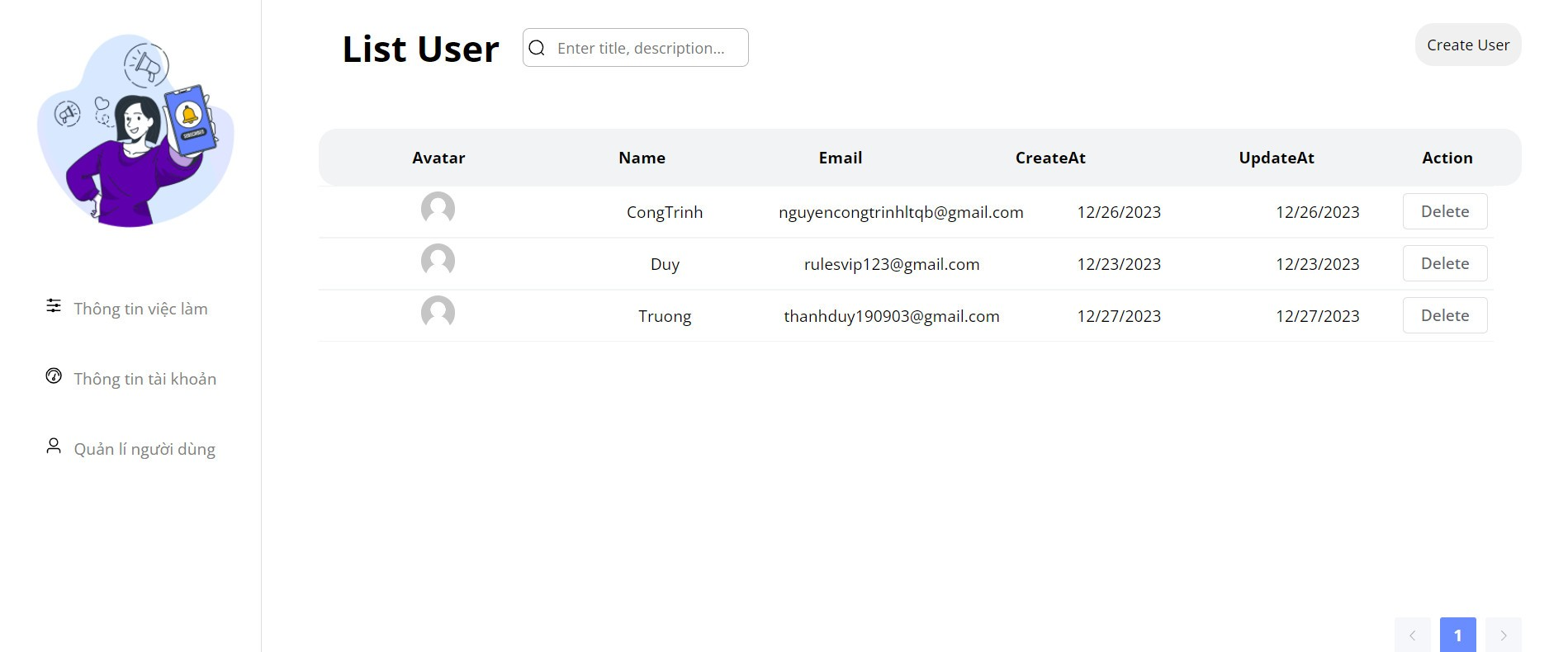
Admin có chức năng cập nhật lại gmail cũng như các thông tin của mình như họ tên, tuổi tại trang chỉnh sửa thông tin cá nhân của admin.



Trang quản lí tài khoản admin

### 4.2.4 Quản lý thông tin người dùng website

Ngoài việc quản lí tài khoản của mình, admin có thể quản lí và xem được tài khoản của các user. Chức năng tìm kiếm theo từ khoá, xoá tài khoản được đưa vào.



Trang quản lí tài khoản người dùng

## 4.3 Kết luận

Kết quả đạt được: Website đã hoàn thành được các chức năng như:

\* Admin:

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

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

- Chức năng xem thông tin tài khoản của admin.

\* Khách hàng:

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

## 4.4 Hướng phát triển

- Thêm thêm bình luận và đánh giá của user.

**TÀI LIỆU THAM KHẢO**

1. "Fluent Python" của Luciano Ramalho, 2015
2. "Node.js Design Patterns" của Mario Casciaro, 2014.
3. "Vue.js Up and Running" của Callum Macrae, 2021.
4. <https://www.w3schools.com/nodejs/>
5. <https://www.w3schools.com/vue/>
6. “Python Testing with Selenium: Learn to Implement Different Testing Techniques Using” của Sujay Raghavendra, 2020.

**SOURCE CODE**

1. <https://github.com/MyCrabs/TopDev_Data>
2. <https://github.com/trinhnguyenx/web_aws_terraform.git>
3. <https://github.com/trinhnguyenx/FE-Crawl-Data.git>