**TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT THÀNH PHỐ HỒ CHÍ MINH**

**KHOA ĐÀO TẠO CHẤT LƯỢNG CAO**



**ĐỒ ÁN TỐT NGHIỆP**

**XÂY DỰNG TRANG WEB TỔNG HỢP VIỆC LÀM**

**SVTH1: Trần Phương Nam SVTH2: Nguyễn Thị Ngọc Trân**

**MSSV: 16110163 MSSV:**

**Khóa: K16**

**Ngành: Công nghệ thông tin**

**GVHD: TS. Huỳnh Xuân Phụng**

TP. Hồ Chí Minh, tháng 06 năm 2020

**TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT THÀNH PHỐ HỒ CHÍ MINH**

**KHOA ĐÀO TẠO CHẤT LƯỢNG CAO**



**ĐỒ ÁN TỐT NGHIỆP**

**XÂY DỰNG TRANG WEB TỔNG HỢP VIỆC LÀM**

**SVTH1: Trần Phương Nam SVTH2: Nguyễn Thị Ngọc Trân**

**MSSV: 16110163 MSSV:**

**Khóa: K16**

**Ngành: Công nghệ thông tin**

**GVHD: TS. Huỳnh Xuân Phụng**

TP. Hồ Chí Minh, tháng 06 năm 2020

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM

Độc lập – Tự do – Hạnh phúc

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

Tp. Hồ Chí Minh, ngày 28 tháng 06 năm 2020

# NHIỆM VỤ ĐỒ ÁN TỐT NGHIỆP

Họ và tên sinh viên 1: MSSV:

Họ và tên sinh viên 2: MSSV:

Ngành: Lớp:

Giảng viên hướng dẫn: ĐT:

Ngày nhận đề tài: Ngày nộp đề tài:

1. Tên đề tài:

2. Các số liệu, tài liệu ban đầu:

3. Nội dung thực hiện đề tài:

4. Sản phẩm:

TRƯỞNG NGÀNH GIẢNG VIÊN HƯỚNG DẪN

**CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM**

**Độc lập – Tự do – Hạnh phúc**

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

# PHIẾU NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN

Họ và tên SV1: MSSV:

Họ và tên SV2: MSSV:

Ngành:

Tên đề tài:

Họ và tên Giáo viên hướng dẫn:

**NHẬN XÉT**

1. Về nội dung và khối lượng công việc đã thực hiện

2. Ưu điểm

3. Khuyết điêm

4. Đề nghị cho bảo vệ hay không?

5. Đánh giá loại

6. Điểm: (Bằng chữ: )

Tp*. Hồ Chí Minh, ngày tháng năm 20…*

Giáo viên hướng dẫn

*(Ký & ghi rõ họ tên)*

**CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM**

**Độc lập – Tự do – Hạnh phúc**

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

# PHIẾU NHẬN XÉT CỦA GIÁO VIÊN PHẢN BIỆN

Họ và tên SV1: MSSV:

Họ và tên SV2: MSSV:

Ngành:

Tên đề tài:

Họ và tên Giáo viên phản biện:

**NHẬN XÉT**

1. Về nội dung và khối lượng công việc đã thực hiện

2. Ưu điểm

3. Khuyết điêm

4. Đề nghị cho bảo vệ hay không?

5. Đánh giá loại

6. Điểm: (Bằng chữ: )

Tp*. Hồ Chí Minh, ngày tháng năm 20…*

Giáo viên phản biện

*(Ký & ghi rõ họ tên)*

# LỜI CẢM ƠN

Nhóm xin chân thành cảm ơn khoa Đào tạo Chất Lượng Cao của trường đại học Sư Phạm Kỹ Thuật TP. Hồ Chí Minh đã tạo điều kiện tốt nhất cho nhóm hoàn thành báo cáo này.

Nhóm cũng xin gửi lời tri ân sâu sắc đến toàn thể giảng viên, những người thầy cô đã giảng dạy, trang bị cho nhóm những kiến thức, nền tảng vững chắc để có thể thực hiện đồ án này. Ngoài ra, nhóm xin gửi lời cảm ơn và biết ơn sâu sắc đến thầy Huỳnh Xuân Phụng, người đã hướng dẫn tận tình, theo dõi nhóm và đưa ra những lời nhận xét, đánh giá bổ ích để nhóm đạt được kết quả tốt nhất.

Tuy nhóm đã cố gắng hoàn thiện đồ án nhiều nhất có thể nhưng do thời gian có hạn nên không thể tránh khỏi những sai sót. Nhóm hy vọng sẽ nhận được những góp ý, nhận xét và sự cảm thông từ quý Thầy Cô và các bạn.

Một lần nữa, nhóm xin gửi lời chân thành cảm ơn.

Nhóm sinh viên

# TÓM TẮT

# MỤC LỤC

[NHIỆM VỤ ĐỒ ÁN TỐT NGHIỆP i](#_Toc43809815)

[PHIẾU NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN ii](#_Toc43809816)

[PHIẾU NHẬN XÉT CỦA GIÁO VIÊN PHẢN BIỆN iii](#_Toc43809817)

[LỜI CẢM ƠN 1](#_Toc43809818)

[TÓM TẮT 2](#_Toc43809819)

[MỤC LỤC 3](#_Toc43809820)

[DANH MỤC CÁC CHỮ VIẾT TẮT 6](#_Toc43809821)

[DANH MỤC CÁC BẢNG BIỂU 7](#_Toc43809822)

[DANH MỤC CÁC HÌNH ẢNH, BIỂU ĐỒ 8](#_Toc43809823)

[Chương 1. TỔNG QUAN 9](#_Toc43809824)

[1.1. Khảo sát hiện trạng 9](#_Toc43809825)

[1.2. Tính cấp thiết của đề tài 9](#_Toc43809826)

[1.3. Mục tiêu và nhiệm vụ của đề tài 9](#_Toc43809827)

[1.3.1. Mục tiêu 9](#_Toc43809828)

[1.3.2. Nhiệm vụ 9](#_Toc43809829)

[1.4. Công nghệ lập trình 9](#_Toc43809830)

[1.5. Phạm vi nghiên cứu 9](#_Toc43809831)

[1.6. Phương pháp nghiên cứu 9](#_Toc43809832)

[Chương 2. CƠ SỞ LÝ THUYẾT 9](#_Toc43809833)

[2.1. Spring Framework 9](#_Toc43809834)

[2.1.1. Giới thiệu về Spring Framework 9](#_Toc43809835)

[2.1.2. Kiến trúc của Spring Framework 10](#_Toc43809836)

[2.1.3. Ưu điểm của Spring Framework 10](#_Toc43809837)

[2.2. MySQL 11](#_Toc43809838)

[2.2.1. Giới thiệu về MySQL 11](#_Toc43809839)

[2.2.2. Ưu điểm của MySQL 11](#_Toc43809840)

[2.2.3. Nhược điểm của MySQL 11](#_Toc43809841)

[2.2.4. Tại sao chọn MySQL? 11](#_Toc43809842)

[2.3. AngularJs 11](#_Toc43809843)

[2.3.1. Giới thiệu về AngularJs 11](#_Toc43809844)

[2.3.2. Ưu điểm 11](#_Toc43809845)

[2.3.3. Nhược điểm 11](#_Toc43809846)

[2.3.4. Tại sao chọn AngularJs? 11](#_Toc43809847)

[2.4. Web crawling 11](#_Toc43809848)

[2.4.1. Giới thiệu về web crawling 11](#_Toc43809849)

[2.4.2. Ưu điểm của web crawling 11](#_Toc43809850)

[2.4.3. Nhược điểm của web crawling 11](#_Toc43809851)

[2.4.4. Crawler4j 11](#_Toc43809852)

[Chương 3. XÁC ĐỊNH YÊU CẦU BÀI TOÁN 11](#_Toc43809853)

[3.1. Phân tích bài toán 11](#_Toc43809854)

[3.2. Quy trình nghiệp vụ 11](#_Toc43809855)

[3.3. Mô hình hóa chức năng 11](#_Toc43809856)

[Chương 4. THIẾT KẾ PHẦN MỀM 11](#_Toc43809857)

[4.1. Thiết kế kiến trúc 11](#_Toc43809858)

[4.1.1. Front-end 11](#_Toc43809859)

[4.1.2. Server Rest 11](#_Toc43809860)

[4.2. Thiết kế dữ liệu 11](#_Toc43809861)

[4.2.1. Database schema 11](#_Toc43809862)

[4.2.2. Mô tả sơ đồ quan hệ 11](#_Toc43809863)

[4.3. Thiết kế lược đồ tuần tự 11](#_Toc43809864)

[Chương 5. THIẾT KẾ GIAO DIỆN 12](#_Toc43809865)

[5.1. Screen flow cho ứng dụng bên doanh nghiệp 12](#_Toc43809866)

[5.2. Screen flow cho ứng dụng bên ứng viên 12](#_Toc43809867)

[5.3. Screen flow cho ứng dụng bên admin 12](#_Toc43809868)

[5.4. Đặc tả chi tiết màn hình 12](#_Toc43809869)

[Chương 6. KẾT QUẢ SO SÁNH, THỰC NGHIỆM, PHÂN TÍCH 12](#_Toc43809870)

[6.1. Hệ thống back-end 12](#_Toc43809871)

[6.2. Hệ thống front-end 12](#_Toc43809872)

[Chương 7. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN 12](#_Toc43809873)

[7.1. Kết quả đạt được 12](#_Toc43809874)

[7.2. Hạn chế 12](#_Toc43809875)

[7.3. Hướng phát triển 12](#_Toc43809876)

[TÀI LIỆU THAM KHẢO 12](#_Toc43809877)

[Tiếng Việt: 12](#_Toc43809878)

[Tiếng Anh: 12](#_Toc43809879)

[PHỤ LỤC 12](#_Toc43809880)

# DANH MỤC CÁC CHỮ VIẾT TẮT

# DANH MỤC CÁC BẢNG BIỂU

# DANH MỤC CÁC HÌNH ẢNH, BIỂU ĐỒ

# TỔNG QUAN

## Khảo sát hiện trạng

Với số lượng người tìm việc và nhu cầu tìm các ứng viên tiềm năng, đạt yêu cầu của các nhà tuyển dụng ngày càng tăng lên, việc quảng bá các bài đăng tuyển dụng của doanh nghiệp đến những ứng viên tiềm năng không phải là một bài toán khó nhờ những ứng dụng của công nghệ thông tin. Nhờ những trang web việc làm, các nhà tuyển dụng có thể dễ dàng tìm kiếm những ứng viên đạt tiêu chuẩn cho các vị trí còn thiếu trong công ty họ; đối với những người tìm việc, đây là một nền tảng vô cùng hữu ích để họ có thể tìm kiếm những công việc họ muốn làm dựa trên sở thích và đam mê, kinh nghiệm sẵn có.

Thế nhưng, với sự tồn tại của vô số trang web việc làm, từ những trang web lớn như Vietnamworks, timviecnhanh, … cho đến các trang web chỉ tập trung vào một mảng phụ như itviec (công nghệ thông tin – IT), … , việc tìm kiếm một việc làm phù hợp với nhu cầu của người tìm việc trở nên khó khăn hơn. Không những thế, đối với các doanh nghiệp nhỏ, việc tiếp cận những ứng viên tiềm năng cũng trở nên khắc nghiệt hơn do phải cạnh tranh với các doanh nghiệp, tập đoàn lớn.

**Vấn đề cần phải giải quyết:**

* Đối với doanh nghiệp, làm sao các nhà tuyển dụng có thể tìm những ứng viên phù hợp, tiềm năng cho những vị trí còn trống trong công ty họ một cách công bằng nhất có thể?
* Đối với những người tìm việc, làm sao những người tìm việc có thể tìm được những công việc, vị trí phù hợp với năng lực, sở thích của họ một cách dễ dàng nhất?

## Mục tiêu và nhiệm vụ của đề tài

### Mục tiêu

Tìm hiểu về Spring Framework, MySQL, Angular 9 và cách xây dựng crawler.

Xây dựng ứng dụng tổng hợp việc làm từ các trang web lớn hiện nay:

* Cho phép các doanh nghiệp đăng ký và tạo bài đăng tuyển dụng.
* Cho phép các người tìm việc đăng ký và tìm kiếm công việc mong muốn.
* Giới thiệu (recommend) những ứng viên tiềm năng dành cho doanh nghiệp và những công việc, vị trí phù hợp với năng lực dành cho ứng viên.
* Tổng hợp tất cả những bài đăng việc làm trên các trang web việc làm lớn như TimViecNhanh, CareerBuilder, Jobsgo, …

### Nhiệm vụ

Một số nhiệm vụ mà nhóm cần thực hiện trong quá trình thực hiện đồ án:

* Kháo sát thực trạng của các trang web việc làm hiện nay.
* Xây dựng và thiết kế hệ thống.
* Phân tích và thiết kế hệ quản trị cơ sở dữ liệu.
* Thiết kế giao diện người dùng cho các nhóm người dùng khác nhau.
* Thiết kế hệ thống crawler.
* Kéo các bài đăng tuyển dụng từ các trang web khác một cách hợp pháp.
* Triển khai ứng dụng trên nền tảng Google Cloud.

## Công nghệ lập trình

Sử dụng Spring Framework và hệ quản trị cơ sở dữ liệu MySQL để thiết kế back-end server và Angular 9 để xây dựng front-end.

Sử dụng các thư viện Jsoup và Crawler4j để thiết kế hệ thống crawler.

## Phạm vi nghiên cứu

Tìm hiểu và ứng dụng các công nghệ đã tìm hiểu vào đồ án.

* Spring Framework: Tìm hiểu và xây dựng trang web theo module Spring Boot của Spring Framework.
* Angular 9: Tìm hiểu và xây dựng giao diện người dùng.
* Crawler4j và Jsoup: Tìm hiểu và trích xuất các dữ liệu việc làm từ các trang web việc làm phổ biến.
* MySQL: Phân tích và thiết kế hệ cơ sở dữ liệu.
* Các trang web tuyển dụng: TimViecNhanh, CareerBuilder, Jobsgo, Timviec24h.

## Phương pháp nghiên cứu

Nhóm đã thực hiện đồ án thông qua các phương pháp nghiên cứu sau:

* Tìm hiểu các nghiệp vụ cơ bản của đồ án thông qua các trang web việc làm lớn (Vietnamworks, TimViecNhanh, Jobsgo, CareerBuilder).
* Tìm kiếm các kiến thức về những thư viện, framework đã sử dụng trong đồ án qua các trang web lập trình và các project, thư viện mã nguồn mở trên GitHub.

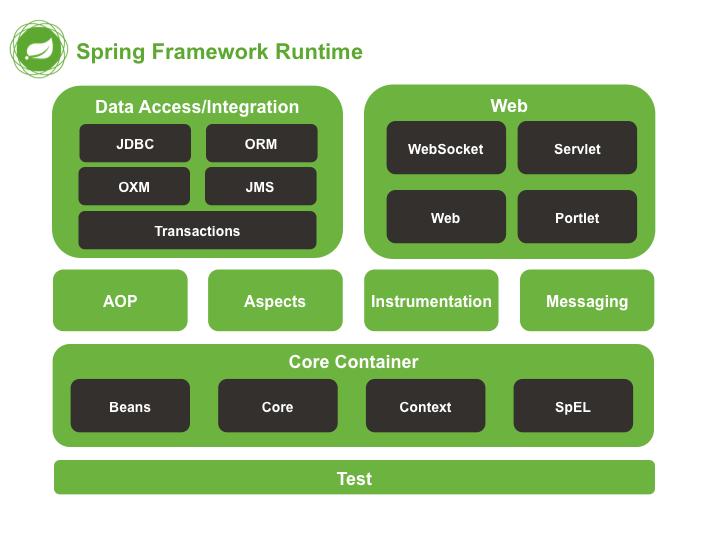
# CƠ SỞ LÝ THUYẾT

## Spring Framework

### Giới thiệu về Spring Framework

Spring Framework là một framework ứng dụng và bộ chứa đảo ngược điều khiển cho nền tảng Java. Chức năng tính của framework này có thể áp dụng cho bất kỳ ứng dụng Java nào, tuy vậy, nhiều bản mở rộng dành cho việc xây dựng ứng dụng nền web cũng được phát triển trên nền tảng Java EE. Tuy framework không ấn định lên một mô hình lập trình cụ thể nào, Spring Framework đã trở nên phổ biến trong cộng đồng Java như một phương án song hành, thay thế, hoặc thậm chí bổ sung cho mô hình Enterprise JavaBeans (EJB). Đây là framework mã nguồn mở.

Spring Framework cung cấp các giải pháp cho nhiều thách thức kỹ thuật mà các nhà phát triển phần mềm Java và các tổ chức muốn tạo ứng dụng dựa trên nền tảng Java phải đối mặt với Java platform. Bởi vì số lượng lớn chức năng của Spring Framework, việc phân loại những bộ phận cơ bản kiến tạo nên framework là một việc rất khó. Spring Framework không chỉ được liên kết riêng biệt với Java Enterprise platform mặc dù sự tích hợp sâu của Spring trong Java EE là một yếu tố quan trọng cho sự phổ dụng, nổi tiếng của nó.



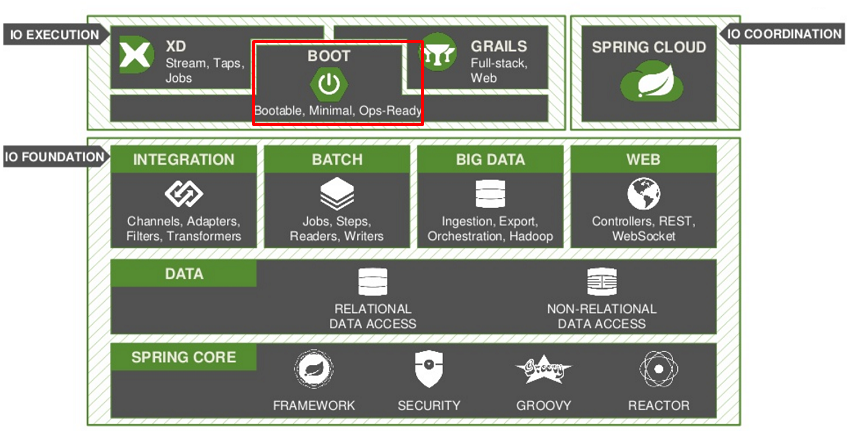
### Ưu điểm của Spring Framework

Spring cho phép lập trình viên sử dụng POJOs. Việc sử dụng POJOs giúp người dùng không phải làm việc với EJB, ứng dụng, các luồng chạy, cấu hình… đơn giản hơn rất nhiều. Ngoài ra, Spring được tổ chức theo kiểu mô đun. Số lượng các gói và các lớp khá nhiều, nhưng người dùng chỉ cần quan tâm đến những gì họ cần và không cần quan tâm đến phần còn lại.

Ngoài những tính chất trên, Spring hỗ trợ sử dụng khá nhiều công nghệ như ORM Framework, các logging framework, JEE, các thư viện tạo lịch trình (Quartz và JDK timer), … Module Web của Spring được thiết kế theo mô hình MVC nên nó cung cấp đầy đủ các tính năng giúp thay thế các web framework khác như Struts.

### Spring Boot

Spring Boot là một Project nằm trên tầng IO Execution (tầng thực thi) của Spring IO Framework. Spring Boot là một bước tiếp theo của Spring, để làm cho Spring dễ dàng hơn trong việc thiết lập và phát triển ứng dụng. Với Spring Boot các cấu hình Spring được giảm thiểu tối đa. Spring Boot hỗ trợ các bộ chứa nhúng (embedded containers) điều này cho phép các ứng dụng web có thể chạy độc lập mà không cần phải trên khai lên các web server.



Ưu điểm của Spring Boot:

* Có các tính năng của Spring Framework.
* Tạo ứng dụng độc lập, có thể chạy bằng java -jar (cho cả java web).
* Nhúng trực tiếp các ứng dụng server (Tomcat, Jetty…) do đó không cần phải triển khai file WAR.
* Cấu hình ít, tự động cậu hình bất kì khi nào có thể (Giảm thời gian viết code, tăng năng suất).
* Không yêu cầu XML config.
* Cung cấp nhiều plugin.
* Chuẩn cho Microservices (cloud support; giảm việc setup, config; các thư viện hỗ trợ…)

## MySQL

### Giới thiệu về MySQL

MySQL là hệ quản trị cơ sở dữ liệu tự do nguồn mở phổ biến nhất thế giới và được các nhà phát triển rất ưa chuộng trong quá trình phát triển ứng dụng. Vì MySQL là hệ quản trị cơ sở dữ liệu tốc độ cao, ổn định và dễ sử dụng, có tính khả chuyển, hoạt động trên nhiều hệ điều hành cung cấp một hệ thống lớn các hàm tiện ích rất mạnh. Với tốc độ và tính bảo mật cao, MySQL rất thích hợp cho các ứng dụng có truy cập CSDL trên internet. Người dùng có thể tải về MySQL miễn phí từ trang chủ. MySQL có nhiều phiên bản cho các hệ điều hành khác nhau: phiên bản Win32 cho các hệ điều hành dòng Windows, Linux, Mac OS X, Unix, FreeBSD, NetBSD, Novell NetWare, SGI Irix, Solaris, SunOS.

MySQL là một trong những ví dụ rất cơ bản về Hệ Quản trị Cơ sở dữ liệu quan hệ sử dụng Ngôn ngữ truy vấn có cấu trúc (SQL).

MySQL được sử dụng cho việc bổ trợ NodeJs, PHP, Perl, và nhiều ngôn ngữ khác, làm nơi lưu trữ những thông tin trên các trang web viết bằng NodeJs, PHP hay Perl, ...

### Ưu điểm của MySQL

Dễ sử dụng: MySQL là cơ sở dữ liệu tốc độ cao, ổn định, dễ sử dụng và hoạt động trên nhiều hệ điều hành cung cấp một hệ thống lớn các hàm tiện ích rất mạnh.

Độ bảo mật cao: MySQL rất thích hợp cho các ứng dụng có truy cập CSDL trên Internet khi sở hữu nhiều nhiều tính năng bảo mật thậm chí là ở cấp cao.

Đa tính năng: MySQL hỗ trợ rất nhiều chức năng SQL được mong chờ từ một hệ quản trị cơ sở dữ liệu quan hệ cả trực tiếp lẫn gián tiếp.

Khả năng mở rộng và mạnh mẽ: MySQL có thể xử lý rất nhiều dữ liệu và hơn thế nữa nó có thể được mở rộng nếu cần thiết.

Nhanh chóng: Việc đưa ra một số tiêu chuẩn cho phép MySQL để làm việc rất hiệu quả và tiết kiệm chi phí, do đó nó làm tăng tốc độ thực thi.

### Nhược điểm của MySQL

Giới hạn: Theo thiết kế, MySQL không có ý định làm tất cả và nó đi kèm với các hạn chế về chức năng mà một vào ứng dụng có thể cần.

Độ tin cậy: Cách các chức năng cụ thể được xử lý với MySQL (ví dụ tài liệu tham khảo, các giao dịch, kiểm toán, …) làm cho nó kém tin cậy hơn so với một số hệ quản trị cơ sở dữ liệu quan hệ khác.

Dung lượng hạn chế: Nếu số bản ghi của người dùng lớn dần lên thì việc truy xuất dữ liệu của bạn là khá khó khăn, khi đó người dùng sẽ phải áp dụng nhiều biện pháp để tăng tốc độ truy xuất dữ liệu như là chia tải database này ra nhiều server, hoặc tạo cache MySQL.

### Tại sao chọn MySQL?

Nhóm quyết định chọn MySQL làm hệ quản trị cơ sở dũ liệu với những lý do sau đây:

* MySQL là hệ quản trị cơ sở dữ liệu mà nhóm có nhiều kinh nghiệm làm việc nhất trong tất cả các hệ quản trị CSDL mà nhóm đã có cơ hội tiếp xúc và làm việc với.
* Nhóm đã thiết kế CSDL dựa trên những lợi ích, ưu điểm mà hệ quản trị cơ sở dữ liệu SQL cung cấp (sự thống nhất dữ liệu, tính toàn vẹn của dữ liệu, …).

## Angular

### Giới thiệu về Angular

Angular là một web framework mã nguồn mở dựa trên TypeScript do nhóm Angular tại Google và bởi một cộng đồng của các cá nhân và tập đoàn. Angular là một bản viết lại hoàn chỉnh từ cùng một nhóm đã xây dựng AngularJs.

### Angular và AngularJs

Angular được thiết kế như là một bản viết lại của Angular Js.

* Angular không có một khái niệm của “scope” hoặc controller, thay vào đó nó sử dụng hệ gia phả các thành phần (component) như một tính chất cấu trúc chính.
* Angular có một expression syntax khác so với AngularJs, sử dung “[ ]” cho property binding và “( )” cho event binding.
* Tính mô đun – đa số các tính năng chính đã chuyển về các mô đun.
* Anuglar chuyển sang sử dụng ngôn ngữ Typescrpit của Microsoft, trong đó có các tính chất như là static typing và annotation.

### Tại sao chọn Angular 9?

## Web crawling

### Giới thiệu về web crawling

Web crawling là quá trình tự động trích xuất các thông tin từ các trang web và lưu trữ nó dưới một định dạng phù hợp. Chương trình mà thực hiện công việc này gọi là web crawler.

Thông thường, khi muốn lấy một số thông tin từ các trang web, chúng ta sẽ dùng các API mà các trang đó cung cấp. Đây là cách đơn giản, tuy nhiên không phải trang web nào cũng cung cấp sẵn API cho chúng ta sử dụng. Do đó chúng ta cần một kĩ thuật để lấy các thông tin từ các trang web đó mà không thông qua API.

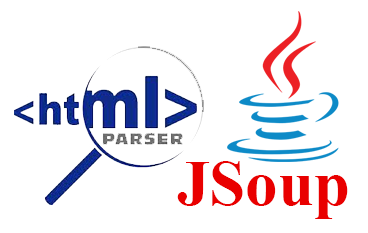
### Web crawling và web scraping

Web crawling và web scraping có những điểm khác nhau sau đây:

|  |  |
| --- | --- |
| Web scraping | Web crawling |
| Trích xuất dữ liệu từ nhiều nguồn khác nhau trên web. | Tải các page từ trên web xuống. |
| Có thể thực hiện trên bất cứ quy mô nào. | Đa phần thực hiện trên quy mô lớn. |
| Chống trùng lặp có thể không có. | Chống trùng lặp là một phần quan trọng. |
| Cần crawl agent và parser | Chỉ cần crawl agent. |

### Jsoup

Jsoup là Java HTML Parser. Nó là một thư viện mã nguồn mở, được sử dụng để phân tích nội dung HTML. Jsoup cung cấp các API dùng để lấy và thao tác dữ liệu từ URL hoặc từ tập tin HTML. Jsoup sử dụng các phương thức giống với DOM, CSS, JQuery để lấy dữ liệu và thao tác với dữ liệu.



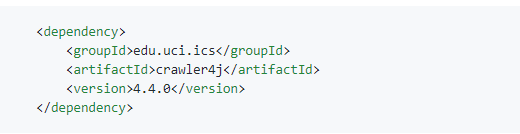
Jsoup bao gồm nhiều class, tuy nhiên có 3 class quan trọng nhất, bao gồm:

* org.jsoup.Jsoup
* org.jsoup.nodes.Document
* org.jsoup.nodes.Element

### Crawler4j

Crawler4j là một web crawler mã nguồn mở dành cho Java, cung cấp một giao diện đơn giản cho việc crawl các trang web. Khi sử dụng, người dùng có thể setup một web crawler đa luồng trong vài phút.

Để sử dụng Crawler4j, người dùng cần thêm dependency dưới đây vào file pom.xml:



Một web crawler sẽ cần có hai class chính:

* Crawler class: class này sẽ extend interface WebCrawler của Crawler4j và có chức năng quyết định các URL nào sẽ được crawl về và xử lý các page đã được tải xuống. Trong đó có hai hàm chính sẽ được override: shouldVisit (quyết định URL sẽ được crawl hay không) và visit (được gọi ngay sau khi nội dung của URL đã được crawl được tải xuống. Người dùng có thể dễ dàng lấy các thông tin từ trang đã được tải xuống như url, văn bản, đường link, html, …).
* Controller class: định nghĩa các seed của crawler, thư mục mà dữ liệu crawl được sẽ được lưu trữ và số lượng các luồng song song.

# XÁC ĐỊNH YÊU CẦU BÀI TOÁN

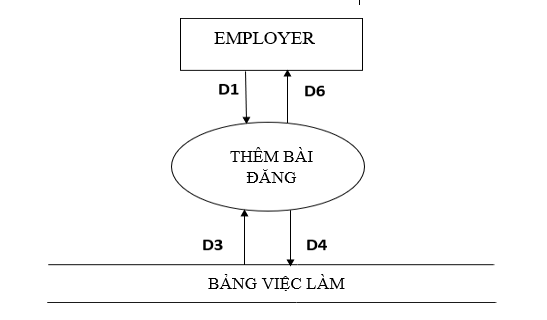
## Phân tích bài toán

Với sự xuất hiện của vô số trang web tuyển dụng việc làm hiện nay, đã nảy sinh ra rất nhiều vấn đề. Đối với những người tìm việc, số lượng các trang web tuyển dụng lớn khiến cho họ gặp nhiều khó khăn trong việc chọn lựa công việc mà họ muốn làm. Khi họ không tìm thấy công việc mong muốn trên trang web này, họ sẽ phải chuyển qua trang web khác và họ phải lại thực hiện việc tìm kiếm việc làm từ đầu. Ngoài ra, mỗi trang web việc làm đều có những công cụ tìm kiếm khác nhau, có các doanh nghiệp và bài đăng công việc khác nhau; hơn thế nữa, có thể có trường hợp một bài đăng gần như giống nhau của một công ty nhưng lại có những điều khoản khác nhau, khiến cho người tìm việc bối rối. Vì thế, nhóm đã đưa ra một giải pháp cho vấn đề này.

Ứng dụng của nhóm là một trang web việc làm, nơi mà các nhà tuyển dụng có thể đăng các bài đăng tuyển dụng và những người kiếm việc có thể kiếm được công việc mà họ mong muốn. Ngoài ra, ứng dụng này sẽ tổng hợp những bài đăng tuyển dụng từ những trang web khác nếu như người tìm việc không tìm được những việc làm mong muốn trên trang này. Hơn nữa, ứng dụng sẽ giới thiệu những bài đăng tuyển dụng mà phù hợp với năng lực và yêu cầu đặc biệt của các ứng viên.

## Quy trình nghiệp vụ

### Quy trình đăng bài tuyển dụng

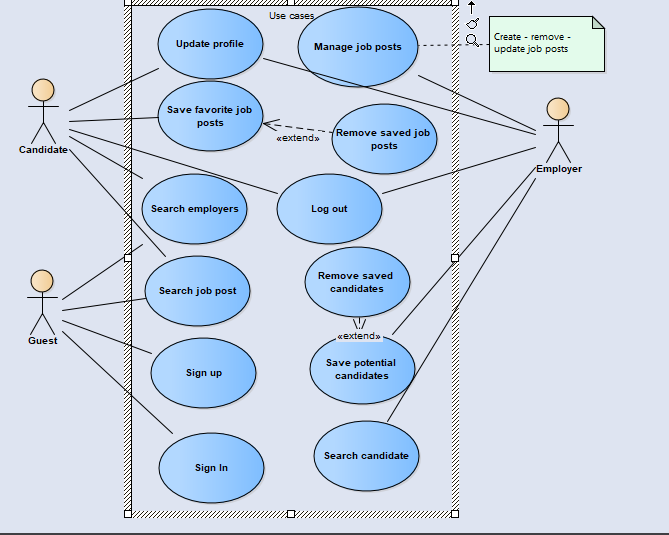


## Mô hình hóa chức năng

### Danh sách các actor

* Guest: người dùng cơ bản.
* Employer: doanh nghiệp.
* Candidate: người tìm việc, ứng viên.

### Sơ đồ use case



### Mô tả sơ đồ usecase

|  |  |  |
| --- | --- | --- |
| **ID** | **Tên** | **Mô tả** |
| UC01 | Sign Up | Đăng ký tài khoản người dùng. |
| UC02 | Sign In | Đăng nhập vào tài khoản người dùng. |
| UC03 | Update Profile | Cập nhật profile. |
| UC04 | Search Job Post | Tìm kiếm bài đăng tuyển dụng. |
| UC05 | Search Employer | Tìm kiếm doanh nghiệp. |
| UC06 | Search Candidate | Tìm kiếm ứng viên. |
| UC07 | Manage Job Posts | Quản lý các bài đăng tuyển dụng. |
| UC08 | Save Potential Candidate | Lưu những hồ sơ ứng viên tiềm năng. |
| UC09 | Save Favorite Job Post | Lưu những bài đăng tuyển dụng quan tâm. |
| UC10 | Remove Saved Job Post | Xóa bài đăng đã lưu. |
| UC11 | Remove Saved Candidate | Xóa hồ sơ ứng viên. |

### Đặc tả một số usecase chính

Usecase “Sign Up” (UC01)

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Sign Up | Code | UC01 |
| Description | Đăng ký vào hệ thống. | | |
| Actor | Guest | | |
| Trigger | Người dùng muốn đăng ký một tài khoản mới. | | |
| Pre-condition | Email đăng ký không trùng lặp với bất cứ email đã đăng ký. | | |
| Post-condition | Success: Sẽ được điều hướng đến homepage.  Fail: Yêu cầu người dùng nhập lại thông tin đăng ký hợp lệ. | | |
| Basic flow | 1. Người dùng nhập email và thông tin cá nhân (nếu là ứng viên) hoặc thông tin công ty (nếu là doanh nghiệp).  2. Người dùng nhấn nút đăng ký  3. Hệ thống kiểm tra email đăng ký đã tồn tại trong cơ sở dữ liệu hay không.  4. Hệ thống sẽ thông báo tạo tài khoảng thành công, điều hướng người dùng đến trang chủ. | | |
| Alternative flow | Không có | | |
| Exception flow | 4a. Hệ thống sẽ thông báo email của người dùng đã tồn tại hoặc email không hợp lệ hoặc người dùng đã điền thiếu các field trong form đăng ký. | | |
| Relationship | Không có. | | |

Usecase “Save Potential Candidate” – UC08

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Sign Up | Code | UC01 |
| Description | Đăng ký vào hệ thống. | | |
| Actor | Guest | | |
| Trigger | Người dùng muốn đăng ký một tài khoản mới. | | |
| Pre-condition | Email đăng ký không trùng lặp với bất cứ email đã đăng ký. | | |
| Post-condition | Success: Sẽ được điều hướng đến homepage.  Fail: Yêu cầu người dùng nhập lại thông tin đăng ký hợp lệ. | | |
| Basic flow | 1. Người dùng nhập email và thông tin cá nhân (nếu là ứng viên) hoặc thông tin công ty (nếu là doanh nghiệp).  2. Người dùng nhấn nút đăng ký  3. Hệ thống kiểm tra email đăng ký đã tồn tại trong cơ sở dữ liệu hay không.  4. Hệ thống sẽ thông báo tạo tài khoảng thành công, điều hướng người dùng đến trang chủ. | | |
| Alternative flow | Không có | | |
| Exception flow | 4a. Hệ thống sẽ thông báo email của người dùng đã tồn tại hoặc email không hợp lệ hoặc người dùng đã điền thiếu các field trong form đăng ký. | | |
| Relationship | Không có. | | |

Usecase “Search Job Post” – UC04

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Sign Up | Code | UC01 |
| Description | Đăng ký vào hệ thống. | | |
| Actor | Guest | | |
| Trigger | Người dùng muốn đăng ký một tài khoản mới. | | |
| Pre-condition | Email đăng ký không trùng lặp với bất cứ email đã đăng ký. | | |
| Post-condition | Success: Sẽ được điều hướng đến homepage.  Fail: Yêu cầu người dùng nhập lại thông tin đăng ký hợp lệ. | | |
| Basic flow | 1. Người dùng nhập email và thông tin cá nhân (nếu là ứng viên) hoặc thông tin công ty (nếu là doanh nghiệp).  2. Người dùng nhấn nút đăng ký  3. Hệ thống kiểm tra email đăng ký đã tồn tại trong cơ sở dữ liệu hay không.  4. Hệ thống sẽ thông báo tạo tài khoảng thành công, điều hướng người dùng đến trang chủ. | | |
| Alternative flow | Không có | | |
| Exception flow | 4a. Hệ thống sẽ thông báo email của người dùng đã tồn tại hoặc email không hợp lệ hoặc người dùng đã điền thiếu các field trong form đăng ký. | | |
| Relationship | Không có. | | |

# THIẾT KẾ PHẦN MỀM

## Thiết kế kiến trúc

Về mặt kiến trúc, hệ thống sẽ gồm hai phần chính:

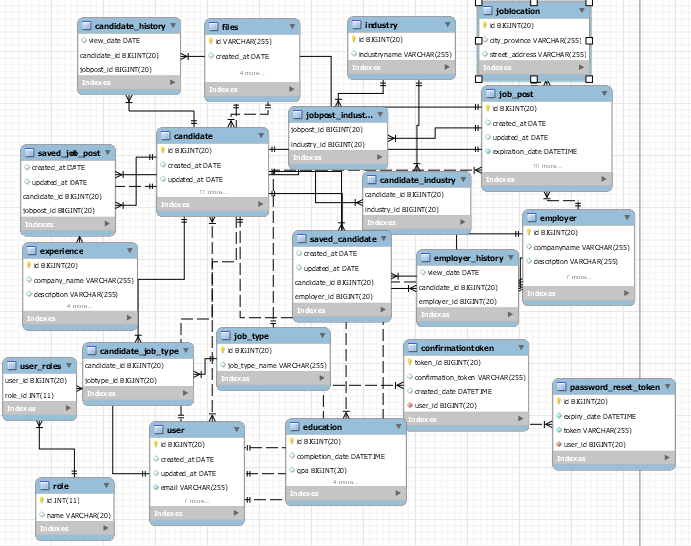
* Front-end: gồm ba phần bên phía client
* Candidate Web: phục vụ các chức năng của actor Candidate (ứng viên).
* Employer Web: phục vụ các chức năng của actor Employer (doanh nghiệp).
* Guest Web: phục vụ các chức năng của actor Guest (khách).
* Back-end: phản hồi các RESTful API được gọi từ Candidate Web, Employer Web và Guest Web.

### Front-end

### Back-end

## Thiết kế dữ liệu

### Database schema



### Mô tả sơ đồ quan hệ

## Thiết kế lược đồ tuần tự

# THIẾT KẾ GIAO DIỆN

## Screen flow cho ứng dụng bên doanh nghiệp

## Screen flow cho ứng dụng bên ứng viên

## Screen flow cho ứng dụng bên admin

## Đặc tả chi tiết màn hình

# KẾT QUẢ SO SÁNH, THỰC NGHIỆM, PHÂN TÍCH

## Hệ thống back-end

## Hệ thống front-end

# KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

## Kết quả đạt được

## Hạn chế

## Hướng phát triển

# TÀI LIỆU THAM KHẢO

## Tiếng Việt:

## Tiếng Anh:

# PHỤ LỤC