



**ĐẠI HỌC
HOA SEN**

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC HOA SEN
KHOA KHOA HỌC VÀ CÔNG NGHỆ**

KHOÁ LUẬN TỐT NGHIỆP

BÁO CÁO CÁ NHÂN

Tên đề tài:

XÂY DỰNG KIẾN TRÚC CÔNG THÔNG TIN TÌM VIỆC

Giảng viên hướng dẫn	:	TS. Trần Vũ Bình	
Sinh viên thực hiện	:	Phùng Chí Nguyên	(09023L)
Lớp	:	QL092L	

Tháng 12 năm 2011

LỜI CẢM ƠN

Trước tiên, tôi xin chân thành cảm ơn TS. Trần Vũ Bình đã tận tình hướng dẫn, góp ý và định hướng nhóm chúng tôi trong quá trình thực hiện khoá luận tốt nghiệp này. Khoá luận đã được hoàn thành cũng chính là nhờ sự nhắc nhở, đôn đốc và sự giúp đỡ nhiệt tình của thầy.

Xin chân thành cảm ơn các thầy cô Khoa Khoa Học & Công Nghệ đã nhiệt tình giúp đỡ tôi trong suốt thời gian học tập tại trường. Chính thầy cô đã xây dựng cho tôi kiến thức nền tảng và những kiến thức chuyên môn để nhóm hoàn thành khoá luận này.

Tôi cũng xin cảm ơn gia đình và bạn bè đã luôn bên cạnh, cổ vũ và động viên tôi những lúc khó khăn để có thể vượt qua và hoàn thành tốt khoá luận này.

Mong rằng các kiến thức cũng như các kinh nghiệm tôi nhận được trong thời gian làm khoá luận vừa qua sẽ giúp ích cho tôi nhiều hơn trong công việc sau này.

Trân trọng.

MỤC LỤC

LỜI CẢM ƠN.....	i
MỤC LỤC	ii
1. Những đóng góp của cá nhân trong đề án	1
1.1. Tìm hiểu và phân tích nghiệp vụ tuyển dụng trực tuyến	1
1.2. Lý thuyết về Tag và Taxonomy	1
1.3. Tìm hiểu cây quyết định	2
1.4. Xử lý dữ liệu bằng Pivot Transformation	3
1.5. Công việc khác.....	3
2. Đánh giá kết quả thực hiện và hướng phát triển.....	4
2.1. Đánh giá kết quả đạt được	4
2.2. Những hạn chế	4
2.3. Hướng phát triển	4
3. Những kinh nghiệm có được sau khi thực hiện khoá luận	5
3.1. Về chuyên môn	5
3.2. Về phong cách làm việc	6

1. Những đóng góp của cá nhân trong đề án

1.1. Tìm hiểu và phân tích nghiệp vụ tuyển dụng trực tuyến

Tìm hiểu các trang tìm việc trực tuyến như LinkedIn, SimplyHired, Monster, Vietnamworks... để phân tích chức năng và cấu trúc thông tin, từ đó so sánh ưu khuyết điểm giữa chúng và đưa ra nhu cầu và chức năng cần xây dựng trong đề tài. Đồng thời phân tích đặc điểm chung của các hệ thống đã có là thông tin tuyển dụng được đăng tải dưới dạng văn bản thô sơ, hồ sơ ứng viên có khuôn khổ nhất định, thiếu sự linh động, việc tìm kiếm công việc theo từ khóa (full-text search), không hỗ trợ so khớp giữa hồ sơ xin việc và yêu cầu tuyển dụng để hỗ trợ nhà tuyển dụng hay ứng viên đưa ra quyết định phù hợp.

1.2. Lý thuyết về Tag và Taxonomy

Do mục tiêu của đề tài là xây dựng kiến trúc có tổ chức thông tin với cấu trúc linh hoạt, khả năng so khớp giữa các đối tượng thông tin. Tôi được phân công tìm hiểu lý thuyết về Tag và Taxonomy nhằm mục đích rút ra cơ sở và phương pháp tổ chức thông tin có cấu trúc linh hoạt, đáp ứng được khả năng phân loại và tìm kiếm thông tin nhanh chóng, hiệu quả, dễ dàng so khớp.

Tagging là một phương pháp tổ chức, phân loại và chia sẻ một lượng dữ liệu lớn, nó là *một cách đơn giản và trực quan để tổ chức nguồn thông tin, cho phép người dùng gắn các từ khóa hoặc tag (thẻ) vào thông tin hoặc đối tượng dữ liệu*. Trong đó, “**Tag** là một từ khóa không có thứ tự hay một thuật ngữ nhằm chỉ một mẫu thông tin (như một bookmark trên internet, hình ảnh kỹ thuật số, hoặc một tập tin máy tính). Loại dữ liệu biến đổi này giúp miêu tả một mục tin và cho phép người sử dụng tìm lại mục đó bằng cách duyệt hay dò tìm.”¹

Việc lưu trữ thông tin dưới dạng tag giúp dễ dàng trong việc phân loại dữ liệu, tuy nhiên để cải thiện được tốc độ của việc tìm kiếm và so khớp, thì các tag thông tin này phải được tổ chức theo dạng cây, lưu trữ và phân các tag, hiệu quả hơn trong việc

¹ Nguồn: <http://vi.wikipedia.org/wiki/Tag>

tìm kiếm và đặc biệt hơn là việc so khớp. Dựa theo tài liệu của Hypertext² chúng tôi định nghĩa taxonomy như sau: *Taxonomy là sự phân loại toàn bộ thông tin trong một hệ thống có phân cấp, sự phân loại này theo một mối quan hệ có trước của các thực thể trong thế giới thực mà nó biểu diễn.*

Taxonomy thể hiện ưu điểm trong việc phân lớp thực thể thông tin theo ngữ nghĩa, chúng thiết lập một quan hệ ngữ nghĩa đơn giản để phân biệt giữa các đối tượng trong một miền thông tin. Nó được sử dụng chủ yếu để giúp cho việc tìm kiếm và duyệt thông tin thuận lợi và nhanh chóng hơn.

1.3. Tìm hiểu cây quyết định

Từ kiến trúc tổ chức thông tin có cấu trúc linh hoạt, khả năng so khớp thông tin đa dạng của kiến trúc, để đưa ra những thông tin có tính quyết định nhằm hỗ trợ người dùng. Nên từ sự linh hoạt này cần được sử dụng để tạo ra các tri thức mới, các tri thức tiềm năng trong nguồn dữ liệu đã có. Vì vậy nhóm áp dụng cây quyết định vào vấn đề này, đòi hỏi phải hiểu rõ giải thuật và cách ứng dụng cây quyết định.

Cây quyết định là một trong những thuật toán điển hình hỗ trợ phân loại và hồi quy, sử dụng rất tốt các mô hình dự đoán. Thuật toán này sẽ khảo sát sự ảnh hưởng của mỗi thuộc tính trong tập dữ liệu và kết quả của thuộc tính dự đoán, sau đó sử dụng các thuộc tính đầu vào với các quan hệ rõ ràng để tạo thành một nhóm phân hoá (các node). Sự liên kết với nhau theo mức độ phụ thuộc giữa các node với nhau thiết lập nên cấu trúc dạng cây

❖ Đánh giá về cây quyết định

• Điểm mạnh

- Cây quyết định sinh ra các quy tắc hiệu được
- Cây quyết định có thể thực thi trong những lĩnh vực hướng quy tắc
- Dễ dàng tính toán trong khi phân lớp
- Cây quyết định xử lý với cả thuộc tính liên tục và thuộc tính rời rạc
- Thể hiện rõ ràng những thuộc tính quan trọng nhất cho việc dự đoán phân lớp

² Tham khảo: <http://www.hipertext.net/english/pag1011.htm#origenNota2>

- **Điểm yếu**

Mặc dù có những điểm mạnh nổi bật trên, cây quyết định vẫn không tránh khỏi những điểm yếu. Cây quyết định không thích hợp lắm với những bài toán mục tiêu, như là dự đoán giá trị của thuộc tính liên tục: thu nhập, huyết áp, lãi suất ngân hàng... Bên cạnh đó, cây quyết định cũng khó giải quyết với những dữ liệu thời gian liên tục.

- Cây quyết định dễ xảy ra lỗi khi có nhiều phân lớp
- Chi phí tính toán để đào tạo cao

1.4. Xử lý dữ liệu bằng Pivot Transformation

Để giải quyết bài toán sử dụng kiến trúc tổ chức thông tin linh hoạt hỗ trợ ra quyết định, chúng tôi quyết định áp dụng kỹ thuật khai thác dữ liệu với cây quyết định. Trong đó, tôi đảm nhận việc xử lý dữ liệu để đưa vào quá trình khai thác, cụ thể là chuyển đổi các thuộc tính có cấu trúc linh hoạt theo dạng tag thành kiểu Pivot table. Vì vậy tôi đã đưa ra giải pháp sử dụng Pivot Transformation³ trong Microsoft SQL Server, viết các Store procedure để thực thi việc xử lý dữ liệu, chi tiết phương pháp thực hiện đã được nêu rõ trong báo cáo chính của nhóm.

Ngoài ra, tôi còn nhận thấy phương pháp Pivot này giúp cho việc thống kê dữ liệu, thể hiện dữ liệu dưới một góc nhìn trực quan hơn. Ví dụ như được sử dụng để thống doanh số bán hàng của từng sản phẩm theo tháng.

1.5. Công việc khác

- Cùng với các thành viên khác nghiên cứu giải pháp so khớp thông tin trong kiến trúc lưu trữ và so khớp thông tin linh hoạt.
- Nhập dữ liệu tranining set về hồ sơ ứng viên và yêu cầu tuyển dụng để thực hiện khai thác dữ liệu bằng cây quyết định.
- Hỗ trợ các thành viên khác xây dựng Data mining engine trong Job Zoom framework.

³ Tham khảo <http://msdn.microsoft.com/en-us/library/ms140308.aspx>

2. Đánh giá kết quả thực hiện và hướng phát triển

2.1. Đánh giá kết quả đạt được

- Kiến trúc giải quyết được nhu cầu lưu trữ thông tin nhiều, đa dạng, dễ dàng trong việc tìm kiếm và so khớp.
- Tổ chức thông tin linh động dưới dạng tag kết hợp taxonomy. Thông tin được lưu trữ dưới dạng cây đa cấp và có mức độ tương quan giữa các tag.
- Hệ thống so khớp linh hoạt, kết hợp mức độ tương quan giữa các tag và cây quyết định
- Kiến trúc dễ dàng triển khai, mở rộng, đồng thời hiệu năng của hệ thống vẫn được đảm bảo khi áp dụng cây quyết định và hệ thống so khớp để gợi ý cho actor xây dựng cây thông tin nhu cầu và “sản phẩm đáp ứng nhu cầu” một cách nhanh chóng và hiệu quả.
- Ứng dụng kiến trúc vào bài toán JobZoom nhằm đánh giá khả năng tổ chức thông tin linh hoạt, hiệu quả của hệ thống so khớp, đồng thời đánh giá tính mở rộng, dễ dàng triển khai và tái sử dụng của kiến trúc khi áp dụng trong các lĩnh vực cần lưu trữ và so khớp thông tin.

2.2. Những hạn chế

- Chưa được áp dụng vào thực tế nên bài toán kiến trúc vẫn còn ở mức trừu tượng, cần điều chỉnh để đảm bảo tính khách quan và linh hoạt của hệ thống so khớp.
- Chưa triển khai được kiến trúc JobZoom framework trong thực tế nên khó kiểm tra tính khả thi và tốc độ tính toán, so khớp và xây dựng cây quyết định do nhóm phát triển khi triển khai với số lượng người dùng lớn.

2.3. Hướng phát triển

- Xác thực độ tin cậy của người dùng, cụ thể đối với công thông tin tìm việc là các hồ sơ ứng viên và yêu cầu tuyển dụng.
- Xây dựng công cụ hỗ trợ lựa chọn ứng viên như phỏng vấn trực tuyến, thuyết trình trực tuyến... để hỗ trợ nhà tuyển dụng và ứng viên tiết kiệm chi phí và thời gian.

- Mở rộng và phát triển xử lý độ tương quan giữa các tag theo các góc nhìn thông tin khác nhau nhằm hỗ trợ người dùng tìm kiếm thông tin nhanh chóng, chuẩn xác và khả năng so khớp chính xác hơn.
- Để đảm bảo hiệu năng hệ thống khi dữ liệu người dùng tăng lên thì phương pháp phân tán dữ liệu sẽ là giải pháp cho vấn đề hiệu năng hệ thống.
- Áp dụng quy trình tuyển dụng vào hệ thống, phát triển hệ thống có quy trình quản lý từ lúc ứng viên nộp hồ sơ đến khi ứng viên trở thành nhân viên của công ty, giúp nhà tuyển dụng dễ dàng trong việc quản lý quy trình tuyển dụng của mình.

3. Những kinh nghiệm có được sau khi thực hiện khoá luận

3.1. Về chuyên môn

3.1.1. Xác định và phân tích vấn đề

Với dữ liệu ban đầu và mục tiêu đề tài, tôi và các thành viên trong nhóm đã xác định bài toán cần giải quyết, sau đó phân tích và chia nhỏ bài toán thành các vấn đề nhỏ. Từ đó tiến hành phân tích rõ và chi tiết từng vấn đề. Dựa vào kiến thức đã học và tìm hiểu các phương pháp để đưa ra giải pháp cho các vấn đề. Thông qua đề tài này giúp tôi nâng cao khả năng xác định và phân tích vấn đề, giải quyết vấn đề tốt hơn.

3.1.2. Cách phân tích những nghiên cứu đã có

Trong quá trình thực hiện đề tài đòi hỏi đọc hiểu nhiều tài liệu chuyên ngành và các tài liệu nghiên cứu đã có, cho nên để tiết kiệm thời gian cần phải đọc và hiểu kỹ phần tóm lược của tài liệu để nắm vững nội dung, xác định tài liệu có liên quan đến vấn đề mình cần nghiên cứu hay không. Đọc mục lục của tài liệu để nắm bắt các điểm trọng tâm, ghi chú lại nội dung chính yếu của các phần theo cách hiểu của mình cùng với số trang tham khảo để sử dụng về sau.

3.1.3. Nghiên cứu và tổng hợp để đưa ra giải pháp

Sau khi phân tích, tìm hiểu nhiều tài liệu khác nhau cần lập bảng so sánh các tài liệu, kiến thức thu nhận được đó theo một bộ tiêu chí cụ thể, từ đó xác định các tiêu

chí mà những nghiên cứu trước đây còn chưa giải quyết triệt để, từ đó xác định tầm vực của nghiên cứu mới cho mình.

Nghiên cứu mới phải căn cứ vào các tài liệu, chứng cứ khoa học cùng những dẫn chứng cụ thể để có thể thuyết phục người đọc, đồng thời phải biết xác định một số hạn chế quan trọng trong các nghiên cứu trước để khắc phục, tránh cách làm dàn trải mà không đi sâu vào một vấn đề nào.

3.1.4. Quản trị cơ sở dữ liệu

Do đề tài đòi hỏi giải quyết bài toán tổ chức thông tin có cấu trúc linh hoạt và áp dụng khai thác dữ liệu vào kiến trúc, qua đó nâng cao kỹ năng xây dựng và quản trị cơ sở dữ liệu, viết các store procedure để xử lý dữ liệu đưa vào khai thác dữ liệu bằng cây quyết định.

3.1.5. Kiến trúc phần mềm

Nghiên cứu các kiến trúc phần mềm, mô hình xây dựng kiến trúc và các design pattern nhằm thiết kế kiến trúc framework đảm bảo tính linh động, dễ mở rộng, hiệu năng hệ thống và ứng dụng vào các lĩnh vực khác nhau. Đạt được kinh nghiệm thiết kế kiến trúc phần mềm và cụ thể xây dựng hệ thống công thông tin tìm việc JobZoom framework.

3.1.6. Mở rộng kiến thức

- Áp dụng kiến thức đã học về các lý thuyết cấu trúc dữ liệu và giải thuật vào thiết kế cấu trúc lưu trữ thông tin linh hoạt.
- Phát triển ứng dụng web trên nền tảng ASP.NET MVC.
- Hiểu thêm lý thuyết xác suất thống kê từ các giải thuật khai thác dữ liệu.

3.2. Về phong cách làm việc

3.2.1. Làm việc nhóm

- Cần thống nhất các nguyên tắc làm việc để đạt hiệu quả cao khi làm việc nhóm.
- Các thành viên trong nhóm tôn trọng ý kiến của nhau, đặt lợi ích tập thể lên hàng đầu, cùng đóng góp để hoàn thành mục tiêu chung của khóa luận, luôn tuân thủ các thời hạn đã đặt ra từ buổi họp trước.

- Cần bình tĩnh trong khi thảo luận nhóm, tránh gây hiểu lầm dẫn đến giảm hiệu quả làm việc.
- Phân chia công việc đúng khả năng, sở trường, sở thích của mỗi người để đạt hiệu quả công việc tốt nhất.
- Phối hợp các thế mạnh của thành viên trong nhóm để bổ sung qua lại cho nhau.
- Học hỏi thêm nhiều kiến thức chưa biết từ các bạn làm chung khóa luận.
- Hỗ trợ các thành viên nhằm hoàn thành mục tiêu chung đề ra.
- Nâng cao tính hợp tác trong làm việc nhóm.
- Nâng cao khả năng làm việc độc lập khi giải quyết công việc được giao.
- Tạo khả năng làm việc nhóm, cùng phân tích, suy nghĩ đưa ra ý kiến về một vấn đề.

3.2.2. Phong cách làm việc

- Học hỏi cách làm việc khoa học, đúng giờ cùng với phương pháp nghiên cứu, tìm hiểu tài liệu và giải quyết vấn đề một cách khoa học của giảng viên hướng dẫn.
- Có những tiến bộ về tiến độ và chất lượng công việc được giao.
- Làm việc nhiệt tình đóng góp ý kiến cho nhóm.
- Chủ động, nghiên cứu tìm hiểu thêm các kiến thức, công nghệ mới.
- Hoà đồng, nhiệt tình, tôn trọng các thành viên trong nhóm.