

BỘ GIÁO DỤC VÀ ĐÀO TẠO
ĐẠI HỌC ĐÀ NẴNG

VŨ THỊ DIỆU THU

NGHIÊN CỨU GIẢI PHÁP ỨNG DỤNG
KHO DỮ LIỆU KHAI THÁC DỊCH VỤ BUỒI ĐIỆN
1080

Chuyên ngành: Khoa học máy tính
Mã số: 60.48.01

TÓM TẮT LUẬN VĂN THẠC SĨ KỸ THUẬT

Đà Nẵng - Năm 2012

**Công trình được hoàn thành tại
ĐẠI HỌC ĐÀ NẴNG**

Người hướng dẫn khoa học: **PGS.TS. Phan Huy Khánh**

Phản biện 1: **PGS.TS. Tăng Tấn Chiến**

Phản biện 2: **GS.TS. Nguyễn Thanh Thủy**

Luận văn được bảo vệ tại Hội đồng chấm Luận văn tốt nghiệp

thạc sĩ kỹ thuật tại đại học Đà Nẵng vào ngày 03 tháng 03 năm 2012

Có thể tìm hiểu luận văn tại:

- Trung tâm Thông tin - Học liệu, Đại học Đà Nẵng;
- Trung tâm Học liệu, Đại học Đà Nẵng.

MỞ ĐẦU

1. Lý do chọn đề tài

Xuất phát từ nhu cầu tìm kiếm thông tin ngày càng tăng của người dân, Tổng Công ty BCVT Việt Nam thành lập một “Ngân hàng dữ liệu” thông tin để đáp ứng nhu cầu của khách hàng, trong đó được đề cập nhiều nhất là dịch vụ 1080. Với dịch vụ 1080, ngoài các yếu tố về giá cước cuộc gọi phù hợp với người sử dụng và yếu tố về chất lượng phục vụ (thái độ, giọng nói...), thì yếu tố chất lượng dịch vụ cần phải được chú trọng.

Chất lượng dịch vụ 1080 là thông tin, là nguồn cung cấp tri thức cho khách hàng. Thông tin cung cấp này không những chỉ đơn thuần là giải đáp thông tin tìm kiếm một số nhà, một số điện thoại, một chỉ dẫn đơn giản mà còn cả nguồn tri thức kinh tế, xã hội, giáo dục...rất phong phú và đa dạng. Nguồn tri thức đang ngày càng được làm giàu, đa dạng và mở rộng, do đó không phải khai thác viên nào cũng hiểu biết tất cả, nhớ hết mọi điều. Rất may đã có sự trợ giúp của Công nghệ thông tin. Những hệ chương trình trên máy tính đã trợ giúp khai thác, cập nhật, lưu trữ, tra cứu thông tin cho dịch vụ 1080. Vì vậy, hiệu quả của dịch vụ 1080 không những phụ thuộc vào các khai thác viên (giọng nói, tri thức, khả năng tìm kiếm dữ liệu...) mà còn phụ thuộc vào cơ sở dữ liệu, tức là vấn đề tổ chức thu thập dữ liệu, cập nhật dữ liệu, chuẩn hóa dữ liệu sao cho tăng tốc độ tìm kiếm cung cấp thông tin chính xác cho các khai thác viên.

Nhưng hiện tại, các hệ thống cung cấp dịch vụ 1080 hầu hết được tổ chức ở dạng cơ sở dữ liệu quan hệ, nguồn tài nguyên được trình bày ở nhiều định dạng khác nhau.. mà không có một qui tắc đặc tả thống nhất. Đồng thời, chính việc tổ chức kho dữ liệu như trên

cũng gây khó khăn cho các chuyên gia đặc tả kinh nghiệm chia sẻ tri thức cho người dùng, hay tập kết quả tri thức trả về quá lớn có tính sàng lọc không cao dẫn đến rất khó nhận biết chính xác đâu là kết quả mình cần, nhất là các vấn đề cần phải qua vài bước xử lý mới đến được kết quả. Hơn nữa khi kho dữ liệu ngày càng lớn việc phối hợp tất cả tài nguyên đã có trong kho vào việc xử lý để trả lời một yêu cầu nào đó cũng khó khăn. Mặt khác, trong môi trường cạnh tranh, người ta ngày càng cần có nhiều thông tin với tốc độ nhanh, cung cấp tri thức đáng giá để trợ giúp việc ra quyết định, giải quyết công việc kịp thời.

Để có những đánh giá cần thiết về sử dụng dịch vụ 1080, chúng ta cần khảo sát sơ bộ CSDL hiện đang sử dụng. CSDL dịch vụ 1080 bao gồm các dữ liệu chính : Dữ liệu Danh bạ, dữ liệu Thông tin KTXH ...

Dữ liệu Danh bạ :

Dữ liệu Danh bạ được thu thập từ hai nguồn dữ liệu chính : Dữ liệu danh bạ điện thoại của BĐ TPĐN và dữ liệu danh bạ điện thoại từ các Bru điện tỉnh (thành), các Công ty trong và ngoài ngành gửi đến. Dịch vụ 1080 không quan tâm đến việc cập nhật dữ liệu này, vì việc cập nhật đã được các bộ phận phát triển dịch vụ ở các đơn vị thực hiện. Tại dịch vụ 1080, yêu cầu cần thiết là phải biên tập lại các nguồn dữ liệu để lưu trữ, phục vụ công tác tra cứu.

Chẳng hạn, dữ liệu Danh bạ BĐ TPĐN hiện có hơn 150.000 mẫu tin. Sau khi bộ phận cập nhật danh bạ chính, sửa chữa, dữ liệu được chuyển giao cho bộ phận biên tập 1080 để biên tập dữ liệu. Nội dung Table có cấu trúc chi tiết hình 1.

Field Name	Data Type	
TEN_CQ	Text	
DIA_CHI	Text	
GHICHU	Text	
TBT	Text	
DCT	Text	
MA_SO	Number	
NOIDUNG	Memo	
NGUOINHAP	Text	
DINHHDANG	Number	

Field Properties

Hình 1. Cấu trúc bảng dữ liệu Danh bạ

Dữ liệu thông tin kinh tế xã hội :

Dữ liệu này được sưu tầm, cập nhật qua sách, báo, TV, internet ... hoặc từ các nguồn tin thông tin liên kết với các bưu điện tỉnh thành khác hoặc các đơn vị cung cấp thông tin chuyên ngành. Mỗi một lĩnh vực dữ liệu được quản lý trong một Table. Chẳng hạn, lĩnh vực du lịch có cấu trúc :

Số mẫu tin trong bảng tùy thuộc vào dữ liệu thu thập được.

Field Name	Data Type	
MA_SO	Number	
TIEUDE	Text	
NOIDUNG	Text	
NGUOINHAP	Text	
DINHHDANG	Number	

Field Properties

Hình 2. Cấu trúc bảng dữ liệu KTXH

Đồng thời xét về yếu tố cạnh tranh trong xã hội công nghệ hiện nay dịch vụ 1080 muốn tồn tại bắt buộc triển khai tốt trên tất cả mọi lĩnh vực, đặt mục tiêu trở thành từ điển sống cho tất cả mọi người dân, mọi đối tượng từ nhiều lứa tuổi, nhiều trình độ khác nhau,

cung cấp những kỹ năng thiết yếu cho mọi người, truyền kinh nghiệm để giải quyết vấn đề, đặc biệt trên ngành giáo dục và đối tượng là học sinh, sinh viên

Do vậy, kho tri thức này phải được tổ chức sao cho có thể dễ dàng cập nhật, đồng thời cung cấp những cơ chế truy xuất thông minh để đáp ứng được nhu cầu đa dạng của người dùng. Việc cải tiến, xây dựng kho tri thức đủ mạnh và những hệ thống trợ giúp dịch vụ 1080 thông minh hơn luôn là yêu cầu quan trọng trong việc nâng cao chất lượng dịch vụ 1080.

Đó cũng là những lý do để tôi chọn đề tài luận văn theo hướng từng bước nghiên cứu, tìm hiểu, nhu cầu tra cứu, giải đáp thông tin trong thực tiễn. Đề tài luận văn mang tên: ***“Nghiên cứu giải pháp ứng dụng kho dữ liệu khai thác dịch vụ Bưu điện 1080”***

2. Mục đích và nhiệm vụ của đề tài

Đề tài sẽ nghiên cứu các cơ sở lý thuyết về tổ chức dữ liệu, cách quản lý và khai thác tri thức của mô hình hoạt động cung cấp dịch vụ 1080 đang được triển khai. Từ đó, tạo ra mô hình kho dữ liệu vật lý có kiến trúc mở cho phép quản lý, cập nhật, khai thác dễ dàng, nhằm phục vụ công tác giải đáp thông tin, cung cấp nguồn tri thức cho khách hàng.

Mục tiêu cụ thể của đề tài bao gồm:

- Xây dựng được mô hình kho dữ liệu có khả năng cập nhật và truy xuất linh hoạt, có chiều sâu về tri thức.
- Xây dựng cơ chế và hệ thống cập nhật dữ liệu thường xuyên cho kho dữ liệu.

3. Đối tượng và phạm vi nghiên cứu

Đối tượng nghiên cứu của luận văn: Nghiên cứu các mô hình tổ chức và khai thác dữ liệu, đặc biệt là mô hình kho dữ liệu

(Datawarehousing); nghiên cứu các mô hình quản lý tri thức (Knowledge Management) và về mô hình hoạt động cung cấp dịch vụ 1080 ngành Bưu điện.

Phạm vi nghiên cứu của luận văn là các công cụ để xây dựng hệ thống khai thác dữ liệu cho dịch vụ 1080. Phạm vi ứng dụng chú trọng cho chuyên ngành Tin học để trợ giúp triển khai mô hình hoạt động chuyên sâu tri thức của dịch vụ 1080.

4. Ý nghĩa khoa học và thực tiễn của đề tài

Đề tài ***“Nghiên cứu giải pháp ứng dụng kho dữ liệu khai thác dịch vụ bưu điện 1080”*** có ý nghĩa khoa học và thực tiễn:

- Xây dựng được mô hình kho dữ liệu trợ giúp khai thác dữ liệu thông minh hơn
- Có thể là mô hình tham khảo cho các đơn vị mong muốn xây dựng Trung tâm khai thác – giải đáp thông tin

5. Bố cục của luận văn

Ngoài phần mở đầu, kết luận, tài liệu tham khảo và phụ lục trong luận văn được trình bày gồm 3 chương như sau :

Chương 1 : Cơ sở lý thuyết

Trình bày về cơ sở lý thuyết với một số lý luận cần thiết về công nghệ tri thức và các kỹ thuật biểu diễn tri thức, xử lý thuyết tri thức.

Chương 2 : Phân tích thiết kế hệ thống

Tiếp cận sâu hơn về các vấn đề của đề tài cần nghiên cứu như là phân tích hiện trạng và đề xuất giải pháp cho hệ thống, trình bày cách thức xây dựng một ứng dụng Ontology, tiến hành xây dựng Ontology cho ứng dụng và đề xuất được mô hình về cấu trúc dữ liệu của hệ thống cần xây dựng.

Chương 3 : Cài đặt hệ thống thử nghiệm.

Đưa ra ứng dụng demo để cài đặt hệ thống thử nghiệm, đánh giá kết quả đạt được

CHƯƠNG 1 - CƠ SỞ LÝ THUYẾT

1.1. Tìm hiểu về tri thức

1.1.1. *Khái niệm về tri thức và quản lý tri thức*

Tri thức là nhận thức và hiểu biết về một sự việc, sự thật hay thông tin được thu thập ở dạng kinh nghiệm, học tập hay thông qua những suy luận, suy ngẫm. Tri thức là sự đánh giá về việc sở hữu, xâu chuỗi những chi tiết liên quan đến một vấn đề mà nếu để riêng lẻ chúng sẽ ít có giá trị hơn.

Tri thức là cái mà con người ta cần để xử lý tình huống, áp dụng vào công việc hàng ngày. Tất cả mọi người đều tìm kiếm tri thức, chỉ có điều mỗi người quan tâm đến một lĩnh vực, một khía cạnh khác nhau. Trình độ mỗi người khác nhau cũng làm cho việc tiếp nhận tri thức khác nhau.

Quản lý tri thức là tập hợp các công việc bao gồm tạo ra, giữ lại và sử dụng những tri thức nhằm nâng cao hiệu quả công việc của một tổ chức, một cá nhân. Quản lý tri thức hầu hết được thực hiện thông qua hai loại công việc chính. Thứ nhất là tập hợp những tài liệu, kinh nghiệm cá nhân của mỗi người một cách thích hợp sau đó phổ biến đến mọi người. Thứ hai là những công việc như chuyển giao việc sử dụng các công cụ như phần mềm, hệ thống e-mail .v.v...

1.1.2. *Phương pháp biểu diễn và xử lý tri thức*

Trong thực tế không thể có phương pháp tổng quát giải quyết vấn đề cho mọi bài toán. Có thể phương pháp này phù hợp cho bài toán này nhưng lại không phù hợp cho bài toán khác. Điều này có nghĩa khi nói tới một bài toán, ta phải chú ý đến phương pháp biểu diễn nó cùng với các phương pháp tìm kiếm trong không gian bài toán nhận được. Trong phần này, giới thiệu một số phương pháp

biểu diễn tri thức cơ bản như: *biểu diễn tri thức nhờ logic hình thức, biểu diễn tri thức nhờ luật sinh, biểu diễn tri thức nhờ mạng ngữ nghĩa, biểu diễn tri thức nhờ Frame, biểu diễn tri thức nhờ bộ ba liên hợp OAV, tổng hợp tri thức và chuyển đổi tri thức.*

1.2. Các công cụ biểu diễn tri thức

Biểu diễn tri thức là vấn đề được đưa ra liên quan đến cả khoa học nhận thức và trí tuệ nhân tạo. Trong khoa học nhận thức thì liên quan đến làm thế nào để con người lưu trữ và xử lý thông tin. Còn trong trí tuệ nhân tạo thì vấn đề chính nằm ở chỗ lưu trữ tri thức sao cho chương trình có thể xử lý được và đạt được những trí thông minh tương tự như con người.

Có một số phương pháp biểu diễn tri thức như biểu diễn bằng logic, bằng frames, bằng các luật (rules), mạng ngữ nghĩa, cây tri thức (ontology) .v.v.. những kỹ thuật này đều bắt nguồn từ lý thuyết xử lý thông tin của con người.

Tùy theo nhu cầu cũng như cách thức xử lý tri thức mà ta nên chọn phương thức biểu diễn phù hợp nhất để biểu diễn tri thức.

1.2.1. Cây biểu diễn tri thức (Ontology)

Trong lĩnh vực công nghệ thông tin, Ontology được dùng để thu nhận tri thức về một lĩnh vực theo cách chung nhất, đưa ra cách hiểu chung về lĩnh vực đó. Ontology đưa ra một bộ từ vựng chung về một lĩnh vực và định nghĩa các thuật ngữ cùng với những quan hệ giữa chúng với nhau. Một Ontology là một tập các mô tả hình thức và cụ thể về các khái niệm trong lĩnh vực đang quan tâm.

1.2.2. Ngôn ngữ Ontology

Ngôn ngữ Ontology là dạng ngôn ngữ hình thức được sử dụng để mã hóa ontology. Có rất nhiều ngôn ngữ Ontology cả ở dạng

dành riêng cho một lĩnh vực nào đó và dạng chuẩn cho tất cả các ứng dụng.

1.2.3. Ngôn ngữ SWRL

SWRL (Semantic Web Rule Language) là ngôn ngữ qui tắc, được sử dụng để xây dựng các luật và truy vấn nhằm cập nhật và truy xuất tri thức từ Ontology. SWRL là tập con của ngôn ngữ OWL nên nó thường được biểu diễn cùng các Ontology trong một file OWL.

SWRL sẽ giúp chúng ta thiết lập các luật trên Ontology hoặc thực hiện các truy vấn trên các dữ kiện trên Ontology. Trong phần này, chúng tôi đã tìm hiểu một số luật cơ bản về SWRL.

1.2.4. Các bước xây dựng ứng dụng Ontology

Ontology được gọi là thành công khi và chỉ khi nó được gắn liền với thành công của ứng dụng. Do vậy trước khi đi vào xây dựng Ontology chúng ta xem xét các bước xây dựng một ứng dụng Ontology để từ đó có định hướng tốt hơn trong việc xây dựng Ontology.

CHƯƠNG 2 - PHÂN TÍCH THIẾT KẾ HỆ THỐNG

2.1. Tìm hiểu về dịch vụ 1080

2.1.1. Phân tích hiện trạng

Như đã nêu ở chương mở đầu, nhu cầu tìm kiếm và cập nhật tri thức của xã hội ngày càng cao dẫn đến sự đòi hỏi của mọi người đối với các hệ thống cung cấp tri thức phải đảm bảo các yếu tố nhanh, cung cấp tri thức đáng giá và dễ nhận diện. Xét về phương diện này chúng ta nhận thấy các hệ thống tìm kiếm hiện tại còn một số hạn chế như sau:

Tính liên kết chưa cao: hiện nay các hệ thống E-learning hầu hết được tổ chức ở dạng cơ sở dữ liệu quan hệ, dữ liệu được trình bày ở nhiều định dạng khác nhau như văn bản, phim, hình ảnh v.v.. mà không có một qui tắc đặc tả thống nhất để có thể dễ dàng truy xuất đến tất cả các tài nguyên có trong hệ thống để xử lý, đánh giá và đưa ra kết quả tốt nhất cho người dùng.

Sự chia sẻ tri thức kinh nghiệm chưa nhiều: chính việc tổ chức kho dữ liệu như trên đã gây khó khăn cho các chuyên gia đặc tả kinh nghiệm của họ nhằm chia sẻ cho người dùng.

Tính sàng lọc không cao: do cơ sở dữ liệu tìm kiếm lớn đặc tả ở nhiều định dạng khác nhau và yêu cầu của người dùng cũng khó trùng khớp hoàn toàn với tài liệu có sẵn nên rất khó có thể liệt kê chính xác tri thức mà người dùng muốn tìm.

Khó nhận biết: việc cơ sở dữ liệu tìm kiếm quá lớn nên kết quả trả về của các hệ thống tìm kiếm chứa tập hợp kết quả quá lớn dẫn đến rất khó cho người học có thể nhận biết chính xác đâu là kết quả mình cần, nhất là các vấn đề cần phải qua vài bước xử lý mới đến được kết quả.

2.1.2. Yêu cầu đối với hệ thống mới

Để có thể hỗ trợ một cách hiệu quả các hệ thống giải đáp thông tin cũng như hỗ trợ người dùng học sinh, sinh viên trong học tập, phát triển trong môi trường cạnh tranh và đa dạng như hiện tại, đòi hỏi hệ thống dịch vụ 1080 phải cải tiến và đổi mới. Và hệ thống mới được xây dựng phải đáp ứng được các yêu cầu sau:

- Kho dữ liệu: phải được tổ chức sao cho có thể phối hợp được các nguồn tài nguyên được mô tả ở nhiều dạng khác nhau trong hệ thống, có thể dễ dàng tiếp nhận và sử dụng hiệu quả các tri thức mới vào việc đáp ứng nhu cầu của người dùng.

- Quản lý: phải tổ chức để có thể thu nhận được các tri thức mới một cách dễ dàng đồng thời cũng đánh giá được giá trị các tri thức đó, giúp người dùng có thể tiếp cận giải quyết vấn đề chứ không đơn thuần là liệt kê tài liệu. Kết quả trả về phải minh bạch và rõ ràng.

2.2. Đề xuất giải pháp

2.2.1. Nội dung tri thức về lĩnh vực

Trong phần này chúng ta sẽ đi vào nghiên cứu cấu trúc và đặc điểm của các thành phần trong kho tri thức để từ đó chọn ra giải pháp và công nghệ để xây dựng chúng. Các thành phần của kho tri thức cần nghiên cứu để xây dựng bao gồm:

- Kho dữ liệu tri thức
- Hệ thống cập nhật và quản lý kho tri thức
- Hệ thống quản lý và khai thác kho tri thức

Để thuận tiện cho việc xây dựng hệ thống của kho tri thức và đưa vào ứng dụng trong đề tài này ta xét một số mẫu thông tin thuộc về Công nghệ thông tin sau: Tin học văn phòng, phân tích và thiết kế dự án, quản trị dự án Công nghệ thông tin

Trong các mẫu thông tin trên ta cần xác định những chủ đề, nghiệp vụ cụ thể cần tri thức để xử lý. Một số nghiệp vụ tiêu biểu cần tri thức để xử lý như:

- Soạn thảo văn bản theo mẫu
- Tìm hàm thích hợp trong Excel
- Soạn thảo bài thuyết trình
- Xây dựng hồ sơ dự án phần mềm
- Quản lý rủi ro
- ...

Tri thức cần đạt được là những thông tin trả lời cho những câu hỏi mà con người cần. Những thông tin này sẽ giúp người học nắm bắt được vấn đề hoặc hướng dẫn giải quyết những vấn đề mà họ gặp phải trong cuộc sống.

Ví dụ: Tri thức cần đạt được trong chủ đề “Tìm hiểu hồ sơ dự án phần mềm” trong lĩnh vực quản lý dự án để trả lời cho các câu hỏi sau:

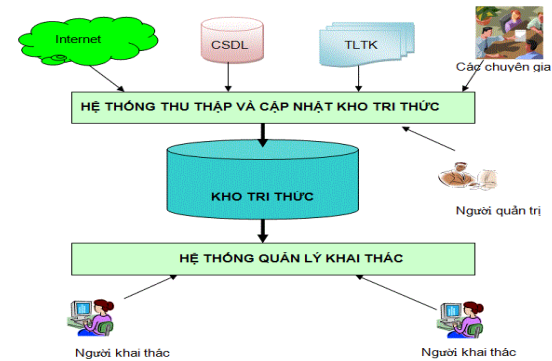
- Hồ sơ dự phần mềm tổng quát bao gồm những hồ sơ nào?
- Hồ sơ dự án phần mềm được phân loại theo các tiêu chí nào?
- Các bước thực hiện hồ sơ của các phần ra sao?
- Mẫu của từng loại hồ sơ đó ra sao?
- Hướng dẫn để điền vào hồ sơ đó như thế nào?
- Làm thế nào để tìm được các thông tin liên quan từng hồ sơ?
- Làm thế nào để xác định được bảng qui định hồ sơ dự án phần mềm cho phù hợp với dự án đang quan tâm?

Xây dựng kho tri thức chúng ta phải chọn chủ đề có cơ sở tri thức rõ ràng và các vấn đề giải quyết cần đến kinh nghiệm. Trong đề tài này ta chọn chủ đề “Tìm hiểu hồ sơ dự án phần mềm” để khảo sát với cơ sở tri thức được xây dựng dựa trên các tiêu chuẩn CMMI về

hệ thống sản xuất phần mềm và qui trình sản xuất phần mềm của hãng Rational (RUP).

2.2.2. *Phác thảo mô hình hệ thống*

Từ những yêu cầu ở trên ta nhận thấy hệ thống mới mang dáng dấp của một hệ chuyên gia và cần có ba thành phần cơ bản là kho dữ liệu, hệ thống khai thác và hệ thống tổ chức quản lý. Xác định đối tượng người dùng và các hệ thống bên ngoài tác động vào hệ thống. Ta có thể mô hình hóa hệ thống như sau:



Hình 2-1. Mô hình hệ thống kho tri thức

Trong đó:

- Kho tri thức: nơi tập trung dữ liệu và tri thức của hệ thống.
- Hệ thống thu thập và cập nhật kho tri thức: là hệ thống phần mềm có khả năng thu nhận và cập nhật dữ liệu để làm giàu kho tri thức.
- Hệ thống khai thác: là hệ thống phần mềm được xây dựng để tạo ra môi trường khai thác giúp người sử dụng dễ dàng khai thác hệ thống.

- Người khai thác: là các khai thác viên, những người cần tìm tri thức từ hệ thống.
- Các chuyên gia: là những người cập nhật các tri thức vào hệ thống.
- Người quản trị: là người chịu trách nhiệm quản lý hệ thống tri thức, bao gồm công tác giám sát tính đúng đắn của tri thức cũng như cách thức đặc tả tri thức cho hệ thống.

2.2.3. Cấu trúc của kho dữ liệu

Để kho tri thức có thể giải quyết được những yêu cầu đã nêu ở trên thì ngoài hệ thống dữ liệu, kho tri thức cần có cơ chế thông minh để hướng dẫn người dùng. Từ đó, có thể khái quát kho tri thức gồm ba phần chính là:

Kho dữ liệu: là nơi lưu trữ các dữ liệu vật lý như bài giảng, câu hỏi, hình ảnh, các mẫu liên quan đến hồ sơ, tài liệu hướng dẫn.v.v.. nó là một phần cơ bản của cơ sở tri thức.

Cơ sở tri thức: bao gồm các luật và các sự kiện được thu thập từ các chuyên gia về các lĩnh vực mà người học quan tâm và được kỹ sư tri thức đặc tả lại vào hệ thống.

Mô tơ suy diễn: để phát huy hiệu quả sử dụng, kho tri thức cần được đặc tả phù hợp và có mô tơ suy diễn hợp lý để tạo sự thông minh cho kho tri thức. Mô tơ suy diễn phải được xây dựng dựa trên sự đặc tả của cơ sở tri thức và kho dữ liệu.

2.2.4. Hệ thống cập nhật và quản lý tri thức

Để kho tri thức được ứng dụng một cách hiệu quả cần có một hệ thống phần mềm để thu nhận, quản lý và đặc tả kho tri thức. Hệ thống phần mềm này phải đảm bảo cho người dùng dễ dàng tiếp cận được với kho tri thức.

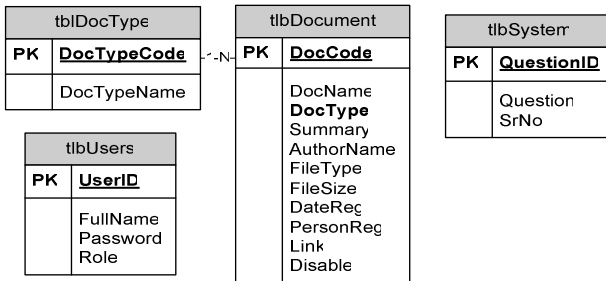
2.2.5. Hệ thống khai thác kho tri thức

Hệ thống này tạo ra môi trường học tập theo dạng ứng dụng của hệ chuyên gia. Hệ thống sẽ đưa ra các câu hỏi để hướng dẫn người học và dựa vào các câu trả lời của người sử dụng để đưa ra kết quả hợp lý.

2.3. Phân tích hệ thống

2.3.1. Mô hình cơ sở dữ liệu cho hệ thống

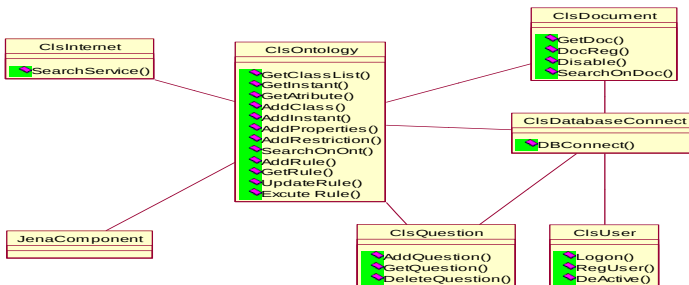
Cơ sở dữ liệu cho hệ thống nhằm tạo ra kho dữ liệu làm cơ sở tri thức cho hệ thống đồng thời chứa các thông tin cho hệ thống.



Hình 2. 6. Mô hình cơ sở dữ liệu cục bộ

2.3.2. Sơ đồ lớp để xây dựng hệ thống

Đây là sơ đồ mô tả cấu trúc viết mã lệnh của hệ thống theo mô hình hướng đối tượng



Hình 2.7. Sơ đồ lớp của hệ thống thu phát tri thức

2.4. Xây dựng Ontology cho kho tri thức

2.4.1. Thiết kế Ontology

Cụ thể ở đây ta xây dựng Ontology để dạy cho chủ đề “Tìm hiểu hồ sơ dự án phần mềm” trong quản lý dự án phần mềm.

Đối với chủ đề “Tìm hiểu hồ sơ dự án phần mềm” những câu hỏi đặt ra với người tìm hiểu là:

- Hồ sơ một dự án phần mềm bao gồm những phần nào?
- Làm thế nào để xác định được hồ sơ nào cần cho dự án, hồ sơ nào không? Phân loại chúng như thế nào?
- Đầu ra của tiến trình này là gì?
- Làm thế nào để hoàn thành hồ sơ đó?
- Làm thế nào để có được các mẫu hồ sơ đó?
- Muốn tìm tài liệu tham khảo, bài giảng về những vấn đề liên quan thì phải làm như thế nào?

Ontology được xây dựng phải có đủ tri thức để trả lời được các câu hỏi trên một cách tốt nhất.

Các bước xây dựng một Ontology cho một vấn đề nào đó thông thường thực hiện theo các bước sau:

Bước 1: Xác định các tri thức mà người dùng cần ở Ontology

Để xác định được các tri thức mà người dùng cần ta liệt kê tất cả các câu hỏi mà Ontology cần trả lời xung quanh miền vấn đề đang xây dựng. Sau đó trả lời điển hình các câu hỏi đó. Dựa vào yêu cầu ở phần đặt vấn đề đối với Ontology cho chủ đề “Tìm hiểu hồ sơ dự án phần mềm” ta nhận thấy Ontology cần trả lời các câu hỏi sau:

- Hồ sơ dự phần mềm tổng quát bao gồm những hồ sơ nào?
- Các thủ tục nào qui định quá trình Xây dựng hồ sơ dự án phần mềm?

- Làm thế nào để xác định được hồ sơ cho một dự án? Chúng được xác định theo tiêu chí nào?
- Các bước để lập nên các hồ sơ đó như thế nào?
- Mẫu của từng loại hồ sơ đó ra sao?
- Danh mục các hồ sơ qui định cho một dự án cụ thể như thế nào?
- Làm thế nào để tìm được các thông tin liên quan đến từng hồ sơ?

Dựa vào cơ sở tri thức để trả lời cho các câu hỏi ở trên.

Bước 2: Liệt kê các thuật ngữ quan trọng trong Ontology

Viết ra danh sách tất cả các thuật ngữ mà chúng ta sử dụng hay muốn giải thích cho người dùng. Đối với Ontology đang xét dựa vào kết quả trả lời của các câu hỏi ở bước 1 chúng ta có thể xác định các thuật ngữ quan trọng sau:

- Dự án: chứa các thông tin về các loại dự án phần mềm.
- Các tiêu chí: Các tiêu chí phân loại dự án.
- Quy trình: Các giai đoạn xây dựng hồ sơ dự án.
- Tài liệu: Các tài liệu cần thiết để xây dựng nên hồ sơ
- Thủ tục: Thủ tục để thực hiện các bước trong quy trình.
- Chuỗi tìm kiếm: để tìm tài liệu liên quan từ cơ sở dữ liệu

hoặc Internet.

- Tài liệu liên quan: xác định các tài liệu liên quan.
- Nguồn tài liệu: chỉ rõ lấy nguồn tài liệu từ đâu? Internet

hay cơ sở dữ liệu cục bộ.

Bước 3: Định nghĩa các lớp và sơ đồ lớp cho Ontology

Từ các từ khóa quan trọng ở trên ta xác định những từ đại diện cho nhóm các đối tượng để tạo thành các lớp. Tiếp theo xác

định mối quan hệ giữa các lớp với nhau để tạo ra Ontology. Các lớp là các thuật ngữ mô tả cho một nhóm các đối tượng trong thực tế.

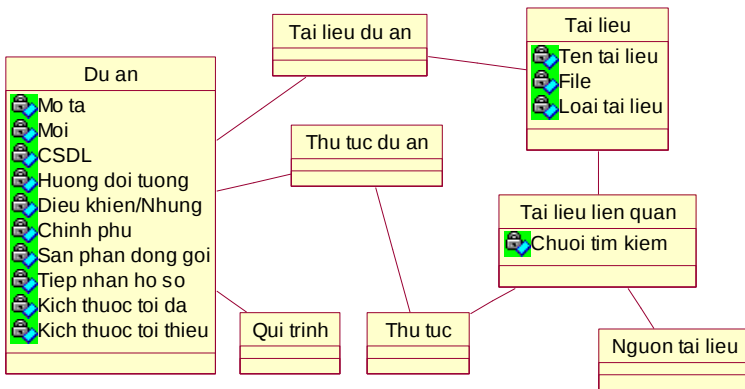
Đầu tiên chúng ta xem xét các thuật ngữ đã liệt kê ở trên và nhận thấy các thuật ngữ có thể chuyển thành lớp bao gồm:

Dự án	Quy trình	Thủ tục
Tài liệu	Tài liệu liên quan	Nguồn tài liệu

Bước 4: Định nghĩa các thuộc tính của lớp

Nếu chỉ có tên lớp thì sẽ không đủ thông tin để trả lời cho các câu hỏi đã nêu ở bước 1. Do vậy, một khi đã định nghĩa các lớp, chúng ta phải mô tả cấu trúc bên trong của các lớp đó. Sau đó phải kiểm tra lại khả năng trả lời ở các câu hỏi ở bước 1 để quyết định thêm hay bỏ đi một lớp nào đó.

Xác định thuộc tính cho các lớp và bổ sung vào sơ đồ ta được Ontology đầy đủ như hình bên dưới:



Hình 2. 9. Ontology sau khi đã bổ sung thuộc tính

Bước 5: Tạo các đối tượng

Bước sau cùng là việc tạo ra các đối tượng của các lớp trên sơ đồ lớp. Việc tạo đối tượng trên các lớp giúp chúng ta mô tả được tất cả các trường hợp của các lớp trong hệ thống.

Tiến hành tạo tất cả các đối tượng cần thiết cho các lớp trên Ontology ta sẽ hoàn tất việc xây dựng Ontology.

2.4.2. Soạn thảo Ontology

Để có thể sử dụng được Ontology vào ứng dụng thì sau khi thiết kế xong chúng ta phải đặc tả nó ở dạng cấu trúc dữ liệu tương ứng. Trong đề tài này chúng ta sẽ đặc tả Ontology theo định dạng XML và sử dụng cấu trúc OWL.

2.4.3. Xây dựng các luật và câu truy vấn trên Ontology

Trong Ontology mà chúng ta đang xét, trước tiên chúng ta phải xây dựng các luật để phân loại các dự án. Chúng ta tiến hành xây dựng các luật như sau:

Các luật cho các qui định về phân loại qui mô dự án như sau:

Nếu dự án có nhân công thực hiện dưới 5 Manmonth thì dự án được xem là rất nhỏ.

$$Du_an \quad (?x) \quad \wedge \quad haskichthuoc \quad (?x, \quad ?y) \quad \wedge \quad swrlb:lessThan(?y, 5) \quad \rightarrow \quad Rat_nho(?x)$$

Nếu dự án có nhân công thực hiện từ 5 đến 10 Manmonth thì dự án được xem là dự án nhỏ.

$$Du_an \quad (?x) \quad \wedge \quad haskichthuoc \quad (?x, \quad ?y) \wedge swrlb:lessThan(?y, 10) \wedge swrlb:greaterThan(?y, 5) \rightarrow nho(?x)$$

Tương tự như vậy chúng ta xây dựng tất cả các luật cần thiết để điều khiển quá trình cập nhật Ontology và tiến hành bổ sung khi cần thiết.

Các luật sẽ thực hiện việc điều khiển khi cập nhật Ontology. Các câu truy vấn trên Ontology sẽ đảm trách việc truy xuất Ontology để trả lời các câu hỏi cho người sử dụng. Chúng ta xây dựng một số câu truy vấn để trả lời các câu hỏi cho Ontology đang xét như sau:

- ✓ Hồ sơ dự án phần mềm tổng quát bao gồm những gì?

Tai_lieu(?x) -> query:select(?x)

- ✓ Các thủ tục nào qui định quá trình Xây dựng hồ sơ dự án phần mềm?

Qui_trinh(?x) -> query:select(?x)

- ✓ Làm thế nào để xác định được hồ sơ cho một dự án?

Du_an(?x)^swrlb:equal(?x, "Ten du an") ^

Tai_lieu_du_an(?x,?y)-> query:select(?y)

Tương tự như vậy, chúng ta tạo ra các câu truy vấn cần thiết để truy vấn tri thức từ Ontology. Thông thường, trong hệ thống có người quản trị tri thức chịu trách nhiệm chuyển các yêu cầu theo ngôn ngữ tự nhiên thành dạng luật hoặc truy vấn để sử dụng về sau.

CHƯƠNG 3 - CÀI ĐẶT HỆ THỐNG THỬ NGHIỆM

3.1. Môi trường và công cụ cài đặt

3.2. Thiết kế hệ thống

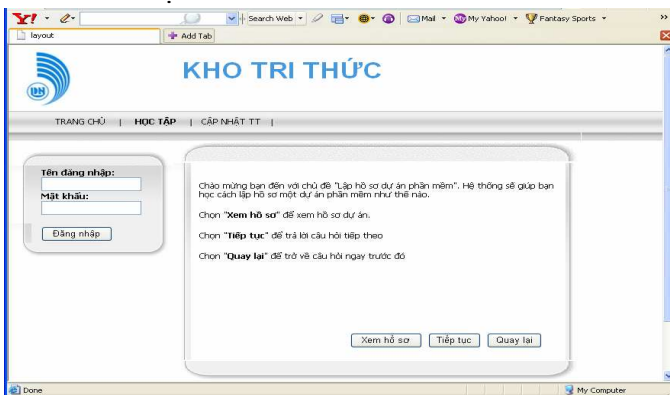
3.3. Cài đặt các chức năng hệ thống

3.4. Thử nghiệm

Hệ thống được thiết kế theo cơ chế của hệ chuyên gia, người sử dụng sẽ trả lời các câu hỏi từ hệ thống và dựa vào các câu trả lời đó hệ thống sẽ đưa ra những kết quả hoặc hướng dẫn hợp lý.

Trước tiên chúng ta sử dụng mô đun dành cho người sử dụng. Hệ thống tiến hành theo các bước sau:

- Hệ thống chào mừng và yêu cầu người sử dụng chọn.
- Hệ thống sẽ đưa ra câu hỏi.
- Người sử dụng trả lời các câu hỏi.
- Hệ thống căn cứ vào câu trả lời của người sử dụng và đưa ra mẫu hồ sơ của dự án theo những tiêu chí đã chọn.
- Căn cứ vào yêu cầu của người sử dụng để xuất ra mẫu hoặc tìm kiếm tài liệu tham khảo.



Hình 3. 5. Màn hình hiển thị chức năng tìm kiếm

Hình trên, là màn hình học tập, đầu tiên hệ thống sẽ thông tin cho người sử dụng những thông tin chào mừng và một số hướng dẫn ban đầu. Người sử dụng có thể thực hiện các tùy chọn sau:

- Xem hồ sơ: hệ thống sẽ hiển thị hồ sơ của dự án phần mềm tổng quát.
- Tiếp tục: hệ thống sẽ đưa ra các câu hỏi tương tác với người sử dụng.

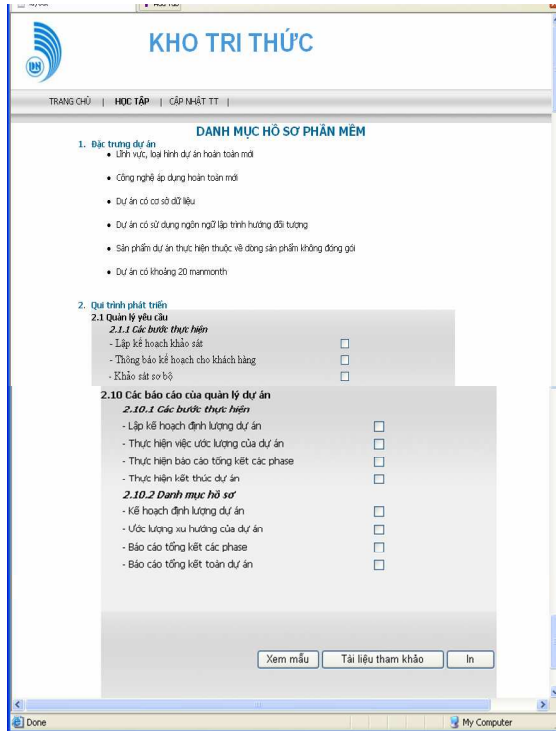


Hình 3. 7. Tương tác giữa khai thác viên và hệ thống

Hệ thống đưa ra câu hỏi và người sử dụng trả lời các câu hỏi đó. Lúc này những nút lệnh sẽ được xử lý như sau:

- Xem hồ sơ: hệ thống sẽ hiển thị hồ sơ của dự án phần mềm theo kết quả trả lời các câu hỏi của người học.
- Tiếp tục: hệ thống sẽ đưa ra câu hỏi tiếp theo cho người học.
- Quay lại: Hệ thống sẽ quay lại câu hỏi trước đó.

Tương tự như vậy hệ thống tiếp tục đưa ra các câu hỏi và người sử dụng trả lời cho đến hết hoặc khi người sử dụng kích chọn xem hồ sơ.



Hình 3. 8. Danh mục hồ sơ phần mềm

Ngoài ra, hệ thống còn có những chức năng khác như cập nhật tri thức, cập nhật câu hỏi, cập nhật dữ liệu lên cơ sở dữ liệu cục bộ.v.v...

3.4.1. *Đánh giá kết quả của hệ thống*

Qua quá trình thực hiện của hệ thống, có thể nhận thấy hệ thống đã giúp người sử dụng xác định được cấu trúc của hồ sơ phần mềm theo những đặc trưng của dự án. Bên cạnh đó hệ thống cũng cung cấp cho người sử dụng các bước thực hiện để hoàn thành hồ sơ phần mềm thông qua cơ sở tri thức có sẵn. Ngoài ra, hệ thống cũng cho phép cập nhật mới các luật, các tri thức một cách dễ dàng tạo ra quá trình phát triển cho hệ thống về sau.

So sánh với các hệ thống hiện tại hệ thống này đã khắc phục được những điểm sau:

- Về việc liên kết các tri thức: các tri thức trong hệ thống dù được định dạng ở nhiều dạng khác nhau nhưng được được mô tả trên Ontology được định dạng theo cấu trúc chuẩn XML và được liên kết với nhau theo phương pháp đặc tả cây tri thức nên dễ dàng truy xuất và xử lý.

- Chia sẻ được kinh nghiệm từ các chuyên gia: hệ thống đã cho phép bổ sung tri thức và tri thức được sắp xếp, liên kết với nhau trên Ontology và chịu sự chi phối của các luật trên Ontology nên các chuyên gia có thể bổ sung dần dần những kinh nghiệm của mình mỗi khi cần thiết và hệ thống vẫn đảm bảo sự gắn kết giữa chúng với nhau. Điều này tạo điều kiện cho việc mở rộng kho tri thức được dễ dàng hơn.

- Về tính minh bạch của kết quả trả về: do đã giới hạn kết quả tìm kiếm cũng như sử dụng các chuỗi tìm kiếm thông qua kinh nghiệm của các chuyên gia nên kết quả trả về cũng dễ nhận diện hơn.

- Hướng dẫn người sử dụng: hệ thống hướng dẫn người sử dụng giải quyết vấn đề theo cách của các chuyên gia bằng cách đưa ra các câu hỏi để tương tác với người sử dụng (cách vận hành của hệ thống).

Như vậy hệ thống kho tri thức mà đề tài xây dựng đã giải quyết được những vấn đề cơ bản mà các hệ thống tìm kiếm hiện tại đang gặp phải

KẾT LUẬN

Trong khuôn khổ một luận văn thạc sĩ, sau khi tiến hành nghiên cứu đề tài, chúng tôi đã đạt được một số kết quả nhất định sau đây:

Những kết quả đạt được:

Về nghiên cứu lý thuyết : Đề tài đã xây dựng được một hệ thống hỗ trợ tìm kiếm vừa có tính hướng dẫn vừa có tính hỗ trợ giải quyết vấn đề thông qua việc chia sẻ tri thức từ các chuyên gia, giúp giải quyết được những khó khăn cơ bản của các hệ thống tìm kiếm hiện thời. Bên cạnh đó, luận văn cũng trình bày được một cách tổng quát các khái niệm về tri thức, quản lý tri thức, đặc tả tri thức, xử lý tri thức v.v.. Đặc biệt, luận văn đã đi sâu vào phân tích và xây dựng ứng dụng trên Ontology.

Về mặt ứng dụng: Có thể khẳng định đề tài đã đáp ứng được các mục tiêu đề ra là tạo ra kho tri thức để hỗ trợ công tác đào tạo nhằm góp phần nâng cao chất lượng của các hệ thống hỗ trợ tìm kiếm. Kết quả của đề tài có thể được áp dụng vào các hệ thống tìm kiếm như một chức năng nâng cao để tăng năng lực của các công cụ tìm kiếm hoặc có thể hỗ trợ cho các hệ thống tra cứu và giải đáp thông tin.v.v...

Hạn chế và hướng phát triển:

Đề tài mới chỉ giải quyết những vấn đề lý thuyết về cách thức xây dựng kho tri thức và xây dựng ứng dụng cho một chủ đề nhỏ. Để đề tài thực sự đi vào cuộc sống cần bổ sung thêm tri thức liên quan đến nhiều chủ đề, đáp ứng cơ bản các thông tin trên mọi lĩnh vực của người dân, từng bước mở rộng đưa vào phục vụ khai thác dịch vụ Bưu điện 1080.

Ngoài ra, để có thể xây dựng ứng dụng một cách hiệu quả và thuận tiện hơn cho người sử dụng