

BỘ GIÁO DỤC VÀ ĐÀO TẠO
ĐẠI HỌC ĐÀ NẴNG

PHẠM KHÁNH THIÊN

**ỨNG DỤNG SEMANTIC WEB
ĐỂ XÂY DỰNG HỆ THỐNG TRA CỨU
THÔNG TIN VỀ VĂN HÓA TÂY NGUYÊN**

**Chuyên ngành: KHOA HỌC MÁY TÍNH
Mã số: 60.48.01**

TÓM TẮT LUẬN VĂN THẠC SĨ KỸ THUẬT

Đà Nẵng - Năm 2011

Công trình được hoàn thành tại
ĐẠI HỌC ĐÀ NẴNG

Người hướng dẫn khoa học: **PGS.TS Võ Trung Hùng**

Phản biện 1 :

Phản biện 2 :

Luận văn sẽ được bảo vệ trước Hội đồng chấm Luận văn tốt nghiệp thạc sĩ kỹ thuật họp tại Đại học Đà Nẵng vào ngày tháng ... năm

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

Web 2.0 là một bước đột phá về công nghệ và đã đạt được những thành tựu rất đáng kể trong việc nâng cao tính tương tác cũng như đẩy nhanh tốc độ xử lý đáp ứng yêu cầu của người dùng. Tuy nhiên trong xã hội thông tin đương đại nhu cầu của người dùng không dừng lại ở việc cải thiện tốc độ mà còn phải cải thiện chất lượng xử lý của trang web theo yêu cầu ngày càng nâng cao. Web 3.0 (Semantic web) ra đời nhằm đáp ứng những yêu cầu về chất lượng đó.

Với đặc điểm chính là nâng cao khả năng chia sẻ tài nguyên và tăng “sự hiểu biết” trong quá trình xử lý dữ liệu của máy tính. Vì vậy Semantic web đã đi sâu vào phân tích và định hướng dữ liệu đồng thời hỗ trợ phát triển ứng dụng Semantic web mở rộng trên nhiều lĩnh vực khác nhau.

Một trong những lĩnh vực thế mạnh của Semantic web là xử lý và tìm kiếm thông tin. Việc phân tích và định hướng nội dung lưu trữ cho phép ta chúng ta xây dựng những cơ sở dữ liệu phục vụ tìm kiếm chính xác hơn. Mặt khác, việc máy tính có thể “hiểu” được dữ liệu sẽ giúp tìm được nhiều thông tin có liên quan đến nội dung cần tìm.

Khu vực Tây Nguyên hiện nay có nhiều dân tộc cùng chung sống. Đây không chỉ là địa bàn chiến lược về kinh tế, chính trị, an ninh quốc phòng mà còn là khu vực có hệ thống di sản văn hóa vật thể và phi vật thể phong phú, độc đáo mang đậm bản sắc văn hóa của các tộc người thiểu số. Cùng với việc đầu tư phát triển kinh tế - xã hội, nâng cao đời sống cho đồng bào các dân tộc, công tác bảo tồn và phát huy những giá trị văn hóa của các dân tộc Tây nguyên, một trong bảy vùng văn hóa

lớn của đất nước cũng đang ngày càng được quan tâm. Song, hiện nay VHTN hầu hết được lưu giữ trong bảo tàng, trong sách, tài liệu, không được nhiều người biết đến, trong khi nhu cầu giao lưu văn hóa, kinh tế, xã hội giữa các dân tộc, vùng miền là nhu cầu thường xuyên, lâu dài.

Ngày nay, với sự phát triển của CNTT, các dịch vụ web ngày càng trở nên phổ biến thì việc xây dựng hệ thống tra cứu thông tin về VHTN là điều cần làm và cần có một hệ thống web thông minh hơn nhằm phục vụ cho việc tìm hiểu về VHTN, rút ngắn khoảng cách thông tin giữa đồng bằng và miền núi, giữa các dân tộc, đồng thời giới thiệu bản sắc văn hóa vùng Tây nguyên đến với đông đảo người dân trên mọi miền tổ quốc và cả thế giới.

Với những lí do trên tôi chọn đề tài “*Ứng dụng Semantic web để xây dựng hệ thống tra cứu thông tin về VHTN*”.

2. Mục đích nghiên cứu

Mục đích chính của đề tài là kết hợp giữa việc nghiên cứu công nghệ Semantic web với nhu cầu khai thác những thông tin về văn hóa khu vực Tây Nguyên của người dùng. Đề tài phát triển theo hướng nâng cao chất lượng tìm kiếm tra cứu thông tin về văn hóa của các dân tộc Tây Nguyên. Xác định những yêu cầu của công việc là thực hiện để xây dựng một trang Semantic web ta thấy cần phải nghiên cứu và thực hiện các công việc sau: Về mặt thuyết, trước tiên ta cần tìm hiểu Semantic web. Tìm hiểu và chọn những công cụ để phát triển một cách hiệu quả nhất. Sau đó ta cần phải đưa ra được những điểm mờ và phân tích giải quyết nó để khi tiến hành xây dựng sẽ không gặp phải sự cố không mong đợi. Về mặt phát triển ứng dụng ta cần xây dựng một trang web để ứng dụng tra cứu về văn hóa của các dân tộc Tây Nguyên. Chương trình phải đáp ứng đúng hai yêu cầu cốt lõi đó là tìm kiếm chính xác hơn và tùy chọn tìm kiếm nhiều dữ liệu liên quan hơn.

3. Đối tượng và phạm vi nghiên cứu

- Tìm hiểu các lý thuyết về Semantic Web và VHTN.
- Tìm hiểu, nghiên cứu các phương pháp xây dựng và lưu trữ dữ liệu trong Semantic Web, các công cụ dùng để thực hiện triển khai hệ thống thành chương trình như Visual Studio, các ngôn ngữ lập trình, các gói thư viện mở rộng.
- Đề tài tập trung vào nghiên cứu và ứng dụng Semantic web vào lĩnh vực quản lý các thông tin về VHTN. Đồng thời xây dựng một chương trình minh họa dùng để tìm kiếm các thông tin về VHTN.

4. Phương pháp nghiên cứu

Để thực hiện đề tài này chúng tôi kết hợp hai nhóm phương pháp nghiên cứu chính, đó là:

- Phương pháp nghiên cứu lý thuyết.
- Phương pháp nghiên cứu thực tế.

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

- Về mặt khoa học, đề tài nghiên cứu tiếp cận với công nghệ mới về xử lý thông tin của Semantic Web. Nghiên cứu đề xuất hướng phát triển một ứng dụng đảm bảo tính chính xác thông tin.
- Về mặt thực tiễn, ứng dụng thành quả của công nghệ thông tin vào lĩnh vực văn hóa. Việc xây dựng hệ thống tra cứu thông tin về VHTN có ý nghĩa hết sức thiết thực trong việc tìm hiểu, nghiên cứu và gìn giữ bản sắc văn hóa các dân tộc vùng Tây Nguyên.

6. Cấu trúc của luận văn

Nội dung của luận văn được chia thành 3 chương:

Chương 1: Tổng quan về Semantic Web.

Chương 2: Giải pháp xây dựng ứng dụng Semantic Web.

Chương 3: Phát triển ứng dụng.

CHƯƠNG 1: TỔNG QUAN VỀ SEMANTIC WEB

Trong chương này chúng tôi giới thiệu tổng quan nhất về Web và Semantic web. Quy trình xây dựng một trang Semantic web, trong chương này chúng tôi cũng đề cập đến các công cụ hỗ trợ tốt nhất cho lĩnh vực Semantic web hiện nay cũng như các khái niệm về văn hóa, thực trạng và nhu cầu tra cứu về VHTN.

1.1. Semantic web

1.1.1. Giới thiệu về Semantic web

1.1.1.1. World wide web và những hạn chế của nó

1.1.1.2. Sự ra đời của Semantic web

1.1.2. Quy trình xây dựng Semantic web

Phần này tôi trình bày mô hình cải tiến của mô hình Noy và McGuinness :

- (1) Xác định mục đích phát triển ontology.
- (2) Kỹ thuật nắm bắt ontology (ontology capture).
- (3) Xem xét sử dụng lại các ontology đang tồn tại.
- (4) Mã hoá ontology (ontology coding).
- (5) Cải tiến ontology.
- (6) Tích hợp các ontology thành phần.
- (7) Kiểm thử.
- (8) Duy trì.

1.1.3. Một số ứng dụng có sẵn của Semantic web

1.1.3.1. Công cụ tìm kiếm Kngine

Công cụ tìm kiếm Kngine dùng để tìm kiếm theo yêu cầu với việc phân tích ngữ nghĩa. Trang web này được chia thành các mục để người dùng có thể tìm kiếm trong miền nhỏ hơn, cho kết quả chính xác hơn.

1.1.3.2. Công cụ tìm kiếm Swoogle

Đây là công cụ tìm kiếm kết hợp giữa việc thu thập văn bản theo ngữ nghĩa (SWDs). Ứng dụng này khai thác thuật toán PageRank nên mặc dù phải truy xuất trên nhiều miền dữ liệu khác nhau nhưng vẫn có tốc độ tương đối nhanh.

1.1.3.3. Công cụ tìm kiếm Hakia

Hakia là công ty công nghệ tìm kiếm ngữ nghĩa. Nhiệm vụ của Hakia là triển khai các giải pháp tìm kiếm ngữ nghĩa để đáp ứng những thách thức của người dùng với hiệu quả kinh doanh cao và chi phí thấp.

1.1.3.4. Trang web microformats.

Microformats tập hợp dữ liệu trên các định dạng sẵn có hiện nay để phân tích và tái sử dụng.

1.2. Công cụ và công nghệ liên quan đến Semantic web

Trong phần này, chúng tôi trình bày sơ lược về các công cụ, công nghệ liên quan đến Semantic web.

1.2.1. Công cụ xây dựng ontology Protégé

1.2.1.1. Bộ công cụ WebODE

1.2.1.2. Bộ công cụ OilED

1.2.1.3. Bộ công cụ Protégé

1.2.2. Truy vấn trên dữ liệu Semantic web

1.2.3. Thư viện phát triển ứng dụng

1.2.3.1. Jena – Một Framework của Semantic web

1.2.3.2. Công cụ hỗ trợ lập trình phát triển Semantic web trên nền tảng .NET

1.3. Văn hóa Tây Nguyên

1.3.1. Văn hóa và đặc trưng VHTN

1.3.1.1. Khái niệm văn hóa

Văn hóa là từ thường được hiểu theo nghĩa rộng hẹp khác nhau. Trong hoạt động giao tiếp hàng ngày, văn hóa được dùng theo nghĩa hẹp để chỉ học thức, lối sống. Văn hóa trong đề tài này được hiểu theo nghĩa rộng, đó là toàn bộ những gì do con người sáng tạo ra có tính giá trị, có thể là văn hoá vật thể, phi vật thể.

1.3.1.2. Đặc trưng VHTN

Đặc trưng văn hóa của một dân tộc là những giá trị tiêu biểu về tinh thần và vật chất mà dân tộc đó đã tích lũy trong quá trình lịch sử, nó có tính bền vững, có ý nghĩa lâu dài, có giá trị khu biệt.

Tây nguyên có nhiều thành phần dân tộc đang sinh sống, nguồn gốc là các dân tộc thiểu số gồm: Ba Na, Xê đăng, Gia Rai, Êđê, Giê-Triêng, BRâu, Rơ Mâm, K'Ho..., đời sống của các dân tộc Tây Nguyên gắn với nương rẫy, tự cấp tự túc, phụ thuộc nhiều vào thiên nhiên cho nên họ sùng bái các lực lượng thiên nhiên, họ quan niệm những gì liên quan cuộc sống cũng có thần. Từ đặc điểm đó mà VHTN rất đa dạng và phong phú, với những đặc trưng được chia từng loại như sau: Văn hóa lễ hội, luật tục, kiến trúc, nhạc cụ dân tộc, trang phục, ẩm thực, ngôn ngữ ...

1.3.2. Nhu cầu việc tra cứu VHTN

1.3.3. Hiện trạng hệ thống phục vụ tra cứu VHTN

Để có cơ sở xây dựng hệ thống mới, tôi tiến hành khảo sát hiện trạng hệ thống phục vụ tra cứu hiện tại theo 2 nguồn thông tin, đó là nguồn thông tin trực tiếp các sở VH-TT&DL và qua mạng Google. Kết quả sau khi tiến hành khảo sát thực tế có thể nói chưa có 1 hệ thống phục vụ tra cứu thông tin VHTN hoàn hảo và đầy đủ thông tin để giúp người sử dụng có thể tra cứu mọi lúc mọi nơi một cách nhanh chóng, tiết kiệm chi phí.

CHƯƠNG 2: GIẢI PHÁP XÂY DỰNG ỨNG DỤNG SEMANTIC WEB

Trong chương này chúng tôi trình bày các bước tìm hiểu về các yêu cầu của bài toán, phân tích và chọn giải pháp cũng như công nghệ để thực hiện bài toán từ đó định hướng xây dựng một ứng dụng Semantic web tiện lợi nhất.

2.1. Giới thiệu bài toán

Việc xây dựng một chương trình ứng dụng giải quyết một bài toán thực tế cần thực hiện rất nhiều công việc khác nhau, ta đi vào công việc đầu tiên là xác định các yêu cầu của bài toán.

2.1.1. Yêu cầu bài toán

2.1.1.1. Đặt vấn đề

❖ Đối tượng sử dụng

Với nội dung của bài toán này ta thấy đối tượng sử dụng là những người nghiên cứu, sinh viên các trường đại học, cao đẳng, học sinh các trường phổ thông và du khách...Trang web này mong muốn phát triển để liên kết với website của bộ VH-TT & DL cũng như website của các tỉnh, Sở VH-TT & DL của các tỉnh trong khu vực Tây Nguyên.

❖ Yêu cầu bài toán

Bài toán đặt ra những yêu cầu xây dựng một trang web có chức năng hỗ trợ tìm kiếm thông tin về VHTN đáp ứng các yêu cầu như:

- Cho phép người dùng nhập dữ liệu động.
- Cho phép người dùng có thể tùy chọn hướng truy xuất dữ liệu từ dân tộc, lễ hội, nhạc cụ,...
- Trang tìm kiếm phải đáp ứng hai chức năng chính đó là tìm kiếm theo tên dân tộc và tìm kiếm theo tùy biến.
- Tạo các trang web thông tin về các dân tộc, kiến trúc, lễ hội, nhạc cụ,...

2.1.1.2. Phân tích vấn đề

Nếu theo cách lập trình với web 2.0 thì các thông tin liên quan đến một dân tộc có thể bao gồm những thông tin như “lễ hội”, “trang phục”, “kiến trúc”,...Nhìn tổng quát ta thấy cách này tương đối ổn nhưng xét chi tiết hơn ta thấy rằng dữ liệu của các bảng thông tin liên quan rất khó có thể thay đổi.

Dữ liệu quan hệ giữa các dân tộc cũng là một vấn đề. Với yêu cầu phép người dùng thêm bớt các quan hệ, điều này cũng lại rất khó thực thi với các hệ quản trị cơ sở dữ liệu trước đây.

Từ những khó khăn trên tôi đưa ra giải pháp đó là sử dụng công nghệ Semantic web để giải quyết bài toán.

Công nghệ Semantic web với đặc điểm lưu trữ dữ liệu dưới định dạng XML và mô hình dữ liệu thông minh nên việc lưu trữ dữ liệu có tính tùy biến cao và hỗ trợ tìm kiếm nhanh, thông tin chất lượng hơn.

Với công việc xác định là phát triển một trang Semantic web tôi xây dựng ứng dụng gồm 2 phần chính đó là xây dựng ontology và thiết kế xây dựng trình duyệt thông minh để hiển thị kết quả.

❖ Ontology

Trong phần này chúng ta xây dựng các lớp, các thuộc tính tạo những mối quan hệ đồng cấp và phân cấp theo những chuẩn đã được định nghĩa bởi tổ chức W3C và cả những định nghĩa mới riêng biệt cho chương trình.

❖ Trình duyệt web

Phần trình duyệt ta không xây dựng một trình duyệt mới hoàn toàn đáp ứng đầy đủ các yêu cầu truy cập dữ liệu ở bất kỳ ontology nào mà ta xây dựng trình duyệt tương tự các ứng dụng web nhằm hiển thị thông tin cơ sở dữ liệu đã xây dựng.

2.1.2. Giải pháp

Cấu trúc của một máy tìm kiếm theo công nghệ Semantic web, về cơ bản cũng có cấu trúc tương tự với một máy tìm kiếm thông thường, bao gồm 2 thành phần chính là giao diện truy vấn và kiến trúc bên trong.

Giao diện truy vấn:

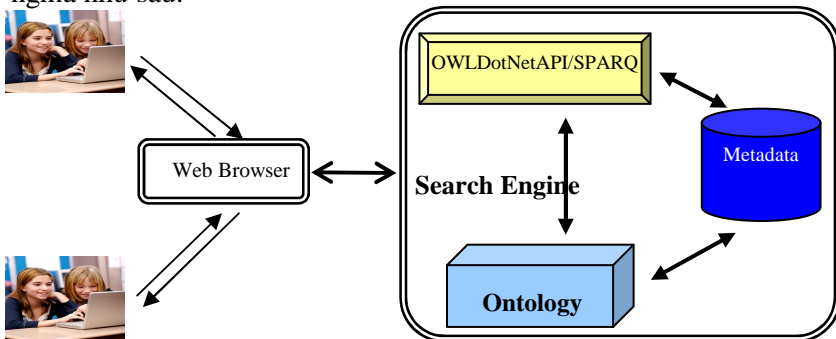
- Cho phép người dùng nhập yêu cầu tìm kiếm
- Hiển thị kết quả tìm kiếm

Phần kiến trúc bên trong: là phần cốt lõi của máy tìm kiếm bao gồm các thành phần:

- Phân tích yêu cầu
- Tìm kiếm kết quả cho yêu cầu
- Dữ liệu tìm kiếm, mạng ngữ nghĩa

Sự khác biệt trong cấu trúc của máy tìm kiếm ngữ nghĩa so với tìm kiếm thông thường nằm ở phần kiến trúc bên trong, cụ thể ở 2 phần: phân tích câu hỏi và tập dữ liệu tìm kiếm.

Mô hình được đề xuất trong luận văn cho ứng dụng tìm kiếm ngữ nghĩa như sau:



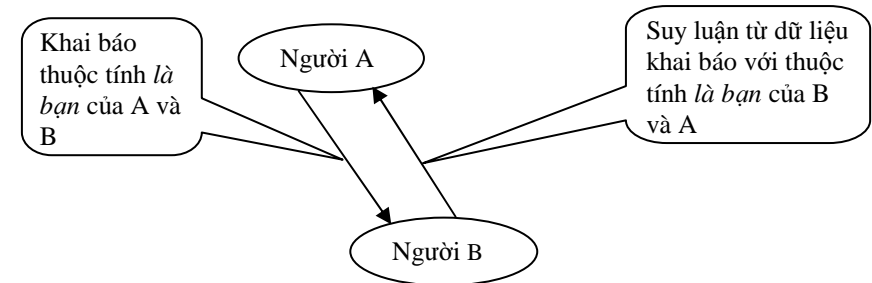
Hình 2.1. Mô hình ứng dụng tìm kiếm thông tin

2.1.3. Giải quyết yêu cầu

2.1.3.1. Suy luận dữ liệu

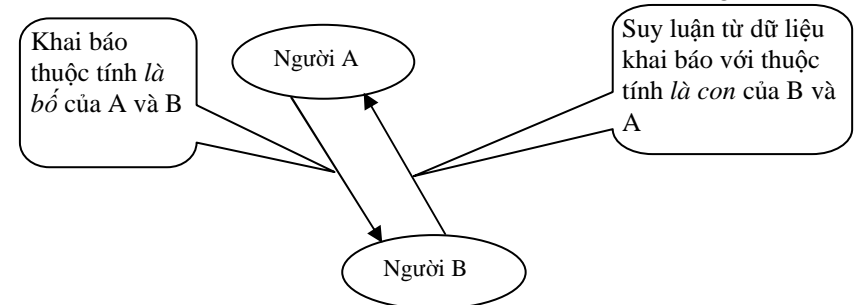
Semantic web hỗ trợ suy luận dữ liệu thông minh nhằm hỗ trợ việc suy luận dữ liệu theo quy tắc của ontology. Ngôn ngữ Ontology hỗ trợ rất nhiều các suy luận khác nhau. Tuy nhiên, trong giới hạn của đề án này chúng tôi chỉ đưa ra hai loại hình suy luận dữ liệu sau:

- ✓ Suy luận ngang cấp là suy luận dựa trên cùng một thuộc tính của hai đối tượng quan hệ. Ví dụ A là bạn của B thì B cũng là bạn của A. Thuộc tính là *bạn* là thuộc tính hai chiều.



Hình 2.2. Mô tả suy luận ngang cấp trong Semantic web

- ✓ Suy luận ngược là cách suy luận dựa trên các thuộc tính khác nhau của hai đối tượng. Ví dụ A là bố của B thì B là con của A. Thuộc tính là *bố* và thuộc tính là *con* là hai thuộc tính ngược nhau.



Hình 2.3. Mô tả suy luận ngược trong Semantic web

Chương trình hỗ trợ hai Properties cơ bản là Properties đảo ngược và Properties tương đương. Thuật toán Filling sẽ tìm kiếm tất cả các

thông tin chưa được khai báo để điền đầy đủ thông tin phục vụ yêu cầu tìm kiếm.

2.1.3.2. Các chức năng chính

Chức năng cơ bản của ứng dụng là tìm kiếm, việc áp dụng kỹ thuật tìm kiếm trên Ontology sẽ trả về những kết quả chính xác hơn nhờ những ưu điểm về lưu trữ dữ liệu RDF so với dữ liệu truyền thống, có cách tổ chức dữ liệu đơn giản, đồng nhất, cấu trúc bộ ba giúp dễ truy xuất thông tin bởi các hệ thống suy luận.

❖ **Duyệt theo ngữ nghĩa**

Duyệt cây phân cấp, theo loại hình dịch vụ: cung cấp cách tìm kiếm địa điểm theo phân cấp trên cây Ontology. Người dùng không cần nhập thông tin tìm kiếm mà vẫn có thể tìm kiếm được tài liệu cần thiết bằng cách duyệt cây theo loại hình dịch vụ.

❖ **Tìm kiếm theo từ khóa**

Cách tìm kiếm phổ biến và có nhu cầu nhiều nhất là tìm kiếm theo từ khóa. Từ khóa chính là từ gợi ý để so sánh nó với các thông tin của dữ liệu qua đó sẽ tìm ra kết quả cần thiết.

❖ **Tìm kiếm nâng cao**

Hệ thống đưa ra một số tùy chọn làm điều kiện người dùng dựa vào một số gợi ý đó để đưa ra điều kiện phù hợp với yêu cầu của mình. Sử dụng cách tìm kiếm nâng cao sẽ có kết quả chính xác hơn.

2.2. Các công cụ hỗ trợ

2.2.1. Công cụ xây dựng ứng dụng Protégé

2.2.1.1. Đặc điểm của Protégé

Đây là phần mềm miễn phí và được phát triển bởi trường Đại học Stanford và Mark Musen. Chức năng nổi bật nhất của phần mềm này là cho phép người dùng sử dụng tạo ra các ontology để phát triển Semantic web theo đúng chuẩn của ngôn ngữ W3C OWL.

2.2.1.2. Protégé sử dụng giao diện đồ họa

Phần mềm hỗ trợ người sử dụng lập trình tương tác bằng giao diện một cách nhanh chóng, phù hợp giữa suy nghĩ của người lập trình và nội dung mà các lớp Protégé tạo ra.

2.2.1.3. Protégé phát triển để tích hợp các công cụ

Protégé cung cấp một số điểm mở rộng nơi các nhà phát triển có thể chủ động thêm các thành phần mà ta thường gọi là plug-ins.

2.2.2. Bộ Visual Studio.Net

Visual Studio.NET là một môi trường tích hợp triển khai phần mềm. Nó được thiết kế để lập ra một tiến trình viết mã, gỡ rối và biên dịch thành một assembly dễ dàng. Visual Studio.NET có một ứng dụng multiple-document-interface rất tinh vi, qua đó người dùng có thể liên kết mọi thứ để phát triển đoạn mã của mình, bao gồm:

Text Editor: cho phép viết đoạn mã C#.

Design view editor: nó cho phép đặt giao diện người dùng và các control dữ liệu truy cập trong dự án.

Các cửa sổ hỗ trợ: cho phép xem và sửa đổi những khía cạnh khác nhau trên dự án.

Biên dịch trong lòng môi trường: để thay cho việc chạy trình biên dịch C# từ dòng lệnh.

Intergated MSDN help: Visual studio.NET có thể gọi tài liệu MSDN.

2.2.3. Thư viện phát triển ứng dụng

Trong phần này sẽ trình bày về hai công cụ hỗ trợ người lập trình phát triển Semantic web dựa trên nền tảng .NET mà tôi đã lựa chọn để thực hiện đề tài.

2.2.3.1. SemWeb

SemWeb với các tính năng cốt lõi như đọc/ghi dữ liệu XML với bộ ba RDF, liên tục lưu trữ dữ liệu với nền tảng SQL và các truy vấn SPARQL cơ bản đã được kiểm nghiệm nhiều lần.

❖ Giấy phép

Đây là bản mã nguồn mở, được sử dụng miễn phí.

❖ Đặc điểm của Semweb

Một số đặc điểm nổi trội đó là tốc độ xử lý và khả năng giao tiếp với các cơ sở dữ liệu liên kết để truy cập dữ liệu.

2.2.3.2. OwlDotNetApi

OwlDotNetApi là một OWL API với bộ phân tích cú pháp viết bằng C# theo công nghệ .NET dựa trên phân tích cú pháp RDF Drive. Hoàn toàn phù hợp với đặc điểm kỹ thuật của W3C.

❖ Chức năng

Mục tiêu của OwlDotNetApi là đọc/ghi dữ liệu của XML dựa trên đồ thị với các cạnh tương ứng với thuộc tính liên kết và các đỉnh tương ứng với các nút hay còn gọi là các lớp.

Xuất phát từ việc đồ thị hoá nội dung của dữ liệu nên OwlDotNetApi đáp ứng được hầu hết tất cả các chuẩn mà W3C đưa ra. Tuy nhiên việc truy cập dữ liệu không thông qua câu lệnh truy vấn nên việc lập trình với thư viện này chưa thuận lợi về thời gian xử lý.

OwlDotNetApi có hai phần chính là lớp chức năng và lớp giao tiếp.

2.2.3.3. Ví dụ

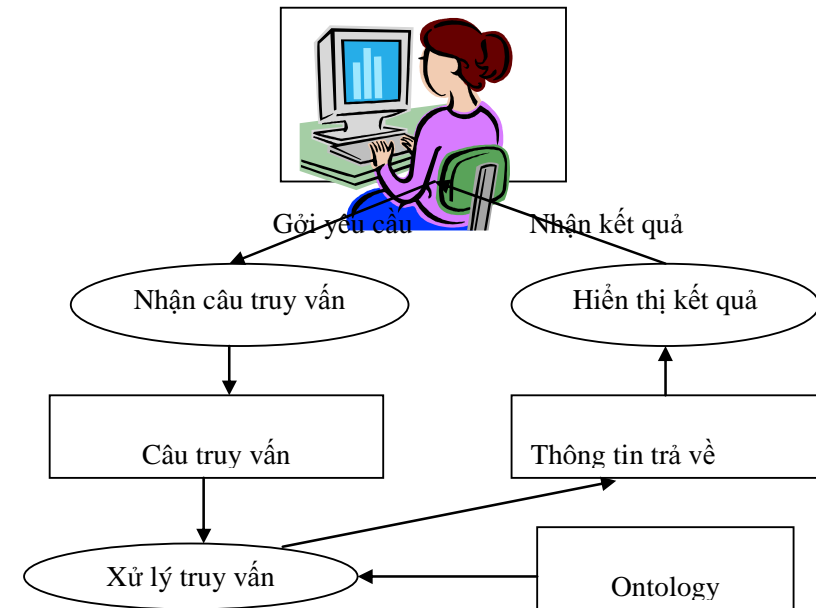
Phần này sẽ giới thiệu số lệnh cơ bản trong thư viện này.

CHƯƠNG 3: PHÁT TRIỂN ỨNG DỤNG

Đây là chương tổng hợp và là câu trả lời chính xác nhất đầy đủ nhất cho những gì chúng tôi đã trình bày ở 2 chương trước.

3.1. Mô hình hoạt động của hệ thống

Ứng dụng được viết trên nền web nên việc quản lý cập nhật, hiệu chỉnh thông tin của người quản trị cũng như việc truy xuất và tìm kiếm dữ liệu của người dùng đầu cuối đều thông qua mạng internet. Ở đây người duyệt web đóng vai trò trung gian theo sơ đồ xử lý sau:



Hình 3.1. Quy trình truy xuất dữ liệu

3.2. Quy trình xây dựng ứng dụng

Để thiết kế công cụ tìm kiếm ngữ nghĩa ứng dụng trên, luận văn đề xuất mô hình hỗ trợ việc tìm gồm các công đoạn sau:

- Giai đoạn 1 : Thiết kế Ontology
- Giai đoạn 2: Xây dựng ứng dụng

3.3. Xây dựng ontology

Ontology là nền tảng để xây dựng dữ liệu cho ứng dụng. Dữ liệu trên ontology có thể là nội dung thông tin cũng có thể là liên kết đến các ontology khác có chung kết nối. Trong nội dung của đề tài này chúng tôi chỉ đề cập đến vấn đề xây dựng thông tin và sử dụng lại thông tin đó theo mô hình Semantic web mà chưa mở rộng ra các thông tin trên những ontology có sẵn trên internet.

3.3.1. Mô hình biểu diễn Ontology

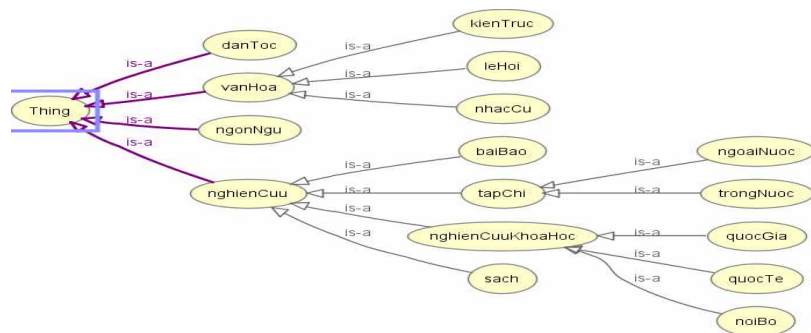
Với mục đích chia nhỏ thông tin để quản lý chúng tôi tiến hành xây dựng các đối tượng được phân chia thành các lớp, các cá thể và các mối quan hệ. Trong ontology này tất cả các thông tin về VHTN sẽ được phân rã và lưu trữ dưới các dạng khác nhau.

3.3.2. Thực hiện xây dựng HignOnt

Việc xây dựng ontology tương tự như cách nhập dữ liệu vào cơ sở dữ liệu.

Phân tích các đối tượng của ứng dụng tôi chia thành các đối tượng:

- Lớp : DanToc, VanHoa, NgonNgu, NghienCuu
- Thuộc tính: CuaDanToc, ...
- Cá thể: Kinh, BaNa...



Hình 3.2. Các phân cấp lớp trong HighlandResource Ontology

3.3.2.1. Lớp

- Lớp *DanToc* trong *HignOnt* dùng để lưu các cá thể là các dân tộc thuộc VHTN. Các cá thể dân tộc đều là con của lớp này. Lớp này là lớp nhỏ nhất và không được chia thành các lớp con.
- Lớp *VanHoa*: thể hiện thông tin các loại hình văn hóa của các dân tộc tây nguyên. Lớp này có thể chứa các cá thể hoặc các lớp con tương ứng với dữ liệu thành phần của VHTN.

Trong đề tài này chúng tôi chia thành các loại hình văn hóa tương ứng với các lớp con như sau: Lớp *KienTruc*, Lớp *NhacCu*, Lớp *LeHoi*.

- Lớp *NgonNgu* thể hiện thông tin các loại hình ngôn ngữ mà các dân tộc sử dụng hoặc ngôn ngữ của các công trình nghiên cứu về VHTN. Lớp này không được chia nhỏ thành các lớp con.
- Lớp *ngghienCuu* mô tả thông tin của các công trình nghiên cứu về VHTN.

3.3.2.2. Thuộc tính

Các thuộc tính thể hiện mối quan hệ giữa các đối tượng dữ liệu (individual) với nhau hoặc quan hệ với dữ liệu. Ngôn ngữ ontology chia thuộc tính ra thành hai loại khác nhau là thuộc tính quan hệ và thuộc tính dữ liệu. Trong ứng dụng này ta phát triển ứng dụng với các thuộc tính sau:

❖ Thuộc tính dữ liệu

Đây là một số thuộc tính dữ liệu cơ bản có trong chương trình.

- Thuộc tính *coTieuDe* thể hiện một công trình nghiên cứu tiêu đề cụ thể là gì và được quy định là kiểu dữ liệu chuỗi.
- Thuộc tính *coNoiDung* thể hiện một công trình có thông tin nội dung xác định và được quy định kiểu dữ liệu văn bản.
- Thuộc tính *namXuatBan* thể hiện năm xuất bản của công trình nghiên cứu cụ thể. Thuộc tính này kiểu số.

- Thuộc tính *coTen* thể hiện thông tin tên của một dân tộc, loại nhạc cụ, loại hình kiến trúc... và tên có kiểu dữ liệu là chuỗi.
- Thuộc tính *diaDiem* thể hiện nơi lưu trữ của một công trình nghiên cứu hoặc thông tin cư trú của một dân tộc.



Hình 3.3. Các thuộc tính dữ liệu trong ứng dụng

❖ Thuộc tính quan hệ

Đây là một số thuộc tính quan hệ cơ bản có trong chương trình.

- Thuộc tính *veDanToc* khai báo công trình nghiên cứu về tây nguyên cụ thể đang xét có liên quan đến dân tộc nào. Thuộc tính này thể hiện mối quan hệ giữa một nghiên cứu cụ thể và các dân tộc.
- Thuộc tính *duaTrenNgonNgu* thể hiện mối liên quan giữa công trình nghiên cứu đó với các ngôn ngữ thể hiện trên các báo cáo, các bài báo liên quan...
- Thuộc tính *veLoaiHinhVanHoa* thể hiện mối quan hệ giữa công trình nghiên cứu đó về loại hình văn hóa nào.



Hình 3.4. Các thuộc tính quan hệ trong ứng dụng

3.3.2.3. Cá thể

Các cá thể là các thông tin chính thể hiện nội dung dữ liệu của ứng dụng. Trong ứng dụng này tôi chỉ đưa ra một số cá thể tiêu biểu cho từng nhóm lớp và thể hiện được mối quan hệ giữa các đối tượng.

- Cá thể *DT_Kinh* lưu thông tin về dân tộc kinh mà ứng dụng cần có liên quan đến dân tộc này. Người quản trị có thể thêm vào các thông tin chưa được khai báo mà người dùng yêu cầu.

```
<owl:DanToc rdf:about="#DT_Kinh">
  <rdf:type rdf:resource="#owl:Thing"/>
  <rdfs:label>Dân tộc kinh</rdfs:label>
  <rdfs:comment>luu thông tin về dân tộc
  kinh như tên gọi là kinh...</rdfs:comment>
  <diaDiem>Tất cả vùng miền thuộc đất nước
  Việt Nam</diaDiem>
</owl: DanToc >
```

- Cá thể *NN_Viet* lưu thông tin về ngôn ngữ Tiếng Việt. Đây là cá thể chứa thông tin và thể hiện mối quan hệ với các cá thể khác sẽ được giới thiệu bên dưới.
- Cá thể *VH_KT_NhaRong* lưu thông tin về loại hình văn hóa nhà rông của tây nguyên. Cá thể này là con của lớp Kiến trúc thuộc lớp Văn Hóa.

- Cá thể *VH_NC_CongChieng* lưu thông tin về loại hình văn hóa Công Chiêng tây nguyên. Cá thể này là con của lớp Nhạc cụ thuộc lớp Văn Hóa.
- Cá thể *NC_BB_0001* thể hiện bài báo viết về VHTN của tác giả Bích Hồng. Bài viết này nghiên cứu sâu về nhạc cụ công chiêng tây nguyên của dân tộc BaNa và được viết bằng ngôn ngữ tiếng Việt với địa chỉ tìm kiếm là trên mạng.
- Cá thể *NC_CTNC_0002* thể hiện công trình nghiên cứu cấp nhà nước về tây nguyên có tên Sử thi Ba Na được Nguyễn Quang Tuyên nghiên cứu năm 2010 và đã được công nhận.

3.4. Thiết kế chương trình

Từ nguồn dữ liệu tổng hợp như trên, chúng tôi tiến hành xây dựng ứng dụng để khai thác một cách có hiệu quả nhất. Đầu tiên để công việc tiến hành thuận lợi hơn chúng tôi đã khai thác thư viện mã nguồn mở OwlDotNetApi. Trong quá trình xây dựng chúng tôi sử dụng một số thuật toán sau đây để khai thác dữ liệu.

3.4.1. Điền dữ liệu

Thuật toán này dùng để điền đầy các quan hệ của ứng dụng và tạo cho ứng dụng có thông tin hai chiều. Ví dụ khi người dùng khai báo công trình có mã số *NC_CTNC_0002 của tác giả Nguyễn Quang Tuyên* thì ứng dụng tự động hiểu tác giả Nguyễn Quang Tuyên có nghiên cứu đề tài đó.

Thuật toán được tiến hành như sau:

- Mở tệp tin chứa ontology
- Đọc tất cả các Properties có khai báo SymmetricProperty hoặc đưa vào danh sách đối chiếu.
- Duyệt qua tất cả các đỉnh của ontology

- o Nếu một đỉnh có chứa quan hệ cần điền đầy theo danh sách đối chiếu ở trên (B1)
 - Điền thông tin quan hệ ngược lại đồng
 - Quay lại xét cho đỉnh vừa điền như B1
- o Ngược lại bỏ qua bước này

- Đóng truy cập vào ontology

3.4.2. Duyệt theo ngữ nghĩa

Duyệt theo cây phân cấp của thông tin: chức năng này thực hiện đọc dữ liệu từ Ontology bằng cách duyệt theo phân cấp của các lớp cao nhất như ngôn ngữ, dân tộc... để truy cập đến thông tin về một công trình nghiên cứu nào đó. Ví dụ, người dùng cần tìm những người thuộc ngôn ngữ Tiếng Anh (NN_Anh) thì người dùng có thể thực hiện bằng cách chọn Ngôn Ngữ (tất cả các công trình nghiên cứu viết bằng ngôn ngữ đang chọn). Tiếp theo chọn Tiếng Anh với ý nghĩa là chọn tất cả các công trình nghiên cứu viết bằng tiếng Anh. Chức năng này được thực hiện dựa trên tính phân cấp cha - con giữa các lớp. Các bước được thực hiện như sau:

B1: Thực hiện đọc, hiển thị nội dung từng lớp dựa trên Properties phân cấp trong lớp *NgonNgu* để người dùng lựa chọn.

B2: Liệt kê tất cả các individual thuộc lớp *NgonNgu* mà người dùng đã lựa chọn.

B3: Truy cập thông tin chi tiết (dựa trên Properties) của các individual và hiển thị kết quả.

3.4.3. Tìm kiếm cơ bản

Chức năng của tìm kiếm cơ bản là dựa vào từ khóa chỉ tên hoặc các thông tin của một công trình nghiên cứu về tây nguyên. Việc tìm kiếm dựa trên sự đối chiếu về thông tin mà người dùng nhập vào tùy thuộc vào các thuộc tính mà người quản trị cung cấp.

Các bước thực hiện

B1: Duyệt tìm tất cả các địa chỉ liên quan đến từ khóa địa điểm được yêu cầu truy vấn.

B2: Nếu có ít nhất 1 địa điểm thỏa mãn ở B1 thì thực hiện:

- Đọc các Object Properties thỏa B1.

B3: Xác định các lớp có Properties nằm trong B2,

- Duyệt qua các individual của lớp vừa xác định.
- Kiểm tra các tính quan hệ giữa các các thể để đưa ra các giá trị liên quan.

B4: Hiển thị kết quả.

3.4.4. Tìm kiếm nâng cao

Dù muốn hay không thì các kết quả của tìm kiếm đơn giản (với từ khóa) thường quá nhiều khi dữ liệu lớn. Để tìm kiếm được thỏa đáng hơn thì việc cung cấp thêm thông tin ngữ nghĩa cho quá trình tìm kiếm được chính xác là hết sức cần thiết. Nhưng khi phải cung cấp thêm thông tin thì tìm kiếm lại ‘chật’ và có thể không trả ra kết quả tìm kiếm dù yêu cầu mà người dùng cần tìm là có thật. Cũng bởi chính người sử dụng cũng không biết chắc thông tin mình cung cấp là chính xác hoặc họ chưa thực sự biết về địa điểm mình cần tìm có thể tìm thấy ở đâu.

3.5. Kết quả thử nghiệm

Sau khi chạy chương trình ứng dụng ta có những kết quả như sau

3.5.1. Kết quả tìm kiếm cơ bản

Tây	Kết quả tìm kiếm:
	A. Bài báo:
	1. Tiềm năng du lịch tây nguyên
	Tác giả ngôn ngữ
	2. Lễ hội cúng bến nước của người Ê đê
	dân tộc
Văn hóa	B. Công trình nghiên cứu
	1. Sử thi BaNa
	dân tộc tác giả ngôn ngữ
Ngôn ngữ	
Nghiên cứu	

Hình 3.10. Kết quả tìm kiếm cơ bản

3.5.2. Kết quả tìm kiếm nâng cao

Tây	Kết quả tìm kiếm nâng cao:
	A. Bài báo:
	1. Tiềm năng du lịch tây nguyên
	Thông tin liên quan: Tác giả ngôn ngữ
	B. Công trình nghiên cứu
	1. Sử thi BaNa
	Thông tin liên quan: dân tộc tác giả ngôn ngữ
	2. Sử thi Ê đê
	Thông tin liên quan: dân tộc
Văn hóa	
Ngôn ngữ	
Nghiên cứu	

Hình 3.11. Kết quả tìm kiếm nâng cao

3.5.3. Chi tiết thông tin một cá thể

Dân tộc	NỘI DUNG CHI TIẾT	
Kinh	Tên công trình	Sử thi Ba Na
Bana	Dân tộc	BANA
Êđê	Nội dung	Nói rằng ở các tộc người Tây nguyên có một kho tàng sử thi đồ sộ, gọi chung là kho tàng sử thi Tây nguyên, là một sự thật. Bằng chứng là trong một số các tộc người bản địa ở đây như Êđê, M'Nông, Ba Na, Xê Đăng, Gia Rai, Raglai..., chúng tôi đã sưu tầm được rất nhiều tác phẩm văn học truyền miệng, trong đó phần lớn là sử thi. Bấy giờ khẳng định con số cuối cùng là bao nhiêu thì rất khó vì việc sưu tầm chưa kết thúc, hơn nữa phần lớn các tác phẩm đó chưa được phiên âm và phiên dịch nên chưa xác định thật chính xác cái nào là sử thi đích thực...
Hrê	Loại hình VH	Sử thi
Tây	Tác giả	Nguyễn Quang Tuyên
	Năm xuất bản	2010
Văn hóa		
Ngôn ngữ		
Nghiên cứu		

Hình 3.12. Kết quả xem chi tiết các cá thể

KẾT LUẬN

Với việc phát triển ngày càng nhanh của dữ liệu văn bản thì xây dựng những ontology theo từng lĩnh vực là vấn đề tất yếu. Nghiên cứu xây dựng ứng dụng Semantic web trong việc tìm kiếm về VHTN đã có được những kết quả ban đầu.

Trước tiên đó là kết quả về phần lý thuyết, nghiên cứu này đã nêu ra được những nét đặc trưng của Semantic web. Hơn nữa, việc đi sâu vào nghiên cứu RDF và ontology – những thành phần quan trọng nhất của Semantic web, đã cho thấy được khả năng và hiệu quả sử dụng cao của thế hệ web này.

Tiếp theo, nghiên cứu còn đưa ra được những công cụ nào là tối cần thiết để phát triển một ứng dụng Semantic web hiệu quả nhất. Song song với nó là việc giải quyết vấn đề giao tiếp giữa người và máy nhất là vấn đề đa ngôn ngữ trong thế hệ web này.

Sau cùng là kết quả thực nghiệm với hệ thống tra cứu về VHTN, đề tài đã khai thác hai thuộc tính đảo ngược và thuộc tính tương đương để xây dựng hai loại hình suy luận ngược và suy luận ngang cấp bằng cách dùng thuật toán Filling sẽ tìm kiếm tất cả các thông tin chưa được khai báo để điền đầy đủ thông tin phục vụ yêu cầu tìm kiếm. Trang web đã chứng minh với nền tảng lý thuyết đã nghiên cứu kết hợp với mô hình phát triển và những công cụ hỗ trợ phát triển với công nghệ .NET, hoàn toàn có thể xây dựng thành công một ứng dụng web 3.0. Ứng dụng này còn chứng minh tính vượt trội của công nghệ web 3.0 với những web đã xây dựng trước đây.

Bên cạnh những thành công đã đạt được thì nghiên cứu này còn hạn chế đó là ngôn ngữ Ontology hỗ trợ rất nhiều các suy luận khác nhau nhưng việc khai thác các suy luận vào ứng dụng còn ít mới chỉ có hai suy luận.

Về mặt ứng dụng cũng mới đưa ra được những chức năng có tính chất chứng minh cho lý thuyết mà chưa có sự đầu tư nhiều về chất lượng hình ảnh giao tiếp với người sử dụng.

Tuy nhiên, trong tương lai đề tài này có thể phát triển tiếp tục ứng dụng được vào các vấn đề thực tiễn. Để đạt được điều đó cần phát triển theo các hướng như: Trước tiên đó là nghiên cứu về mặt lý thuyết và xây dựng thêm nhiều suy luận mới. Sau đó cần nghiên cứu mở rộng trình duyệt web sao cho có thể đáp ứng vấn đề truy xuất dữ liệu từ những tài nguyên có sẵn trên mạng.

Trên đây là những nghiên cứu về lý thuyết và ứng dụng về Semantic web. Nghiên cứu này dựa trên lý thuyết về Semantic web do tổ chức W3C đưa ra và đã phát triển thành công ứng dụng cho lĩnh vực tra cứu thông tin về VHTN.