

BỘ GIÁO DỤC VÀ ĐÀO TẠO
ĐẠI HỌC ĐÀ NẴNG

PHẠM THANH HÙNG

**ỨNG DỤNG SEMANTIC WEB ĐỂ PHÁT TRIỂN
HỆ THỐNG TƯ VẤN VIỆC LÀM**

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. Lê Văn Sơn**

Phản biện 1: **TS. Huỳnh Công Pháp**

Phản biện 2: **PGS.TS. Đoàn Văn Ban**

Luận văn được bảo vệ trước Hội đồng chấm Luận văn tốt nghiệp thạc sĩ kỹ thuật họp tại Đại học Đà Nẵng vào ngày 15 tháng 10 năm 2011

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

➤ Lý do chọn đề tài

Web 2.0 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 (Web Semantic) 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. Web Semantic đã đi sâu vào phân tích và định hướng dữ liệu, đồng thời hỗ trợ phát triển ứng dụng Web Semantic trên nhiều lĩnh vực khác nhau. Một trong những lĩnh vực thế mạnh của Web Semantic là xử lý và tìm kiếm thông tin. Việc phân tích và định hướng nội dung lưu trữ cho phép chúng ta xây dựng những cơ sở dữ liệu phục vụ tìm kiếm chính xác hơn, tinh gọn hơn.

Nước ta đang trên đà phát triển hội nhập với thế giới, nền kinh tế thị trường nhiều thành phần mở ra nhiều cơ hội việc làm cho tất cả mọi người. Hàng ngàn cơ hội việc làm chờ đón chúng ta mỗi ngày. Nhưng làm sao những thông tin đó có thể đến với người lao động một cách nhanh nhất và chính xác nhất, người đi tìm việc không phải mất công với một đống bụi nhùi các công việc phải lựa chọn. Đó là điều trăn trở của người đi tìm việc làm cũng như đối với các công ty tuyển dụng.

Bên cạnh đó, hiện nay vấn đề việc làm là một vấn đề nhức nhối với nhiều nước trên thế giới. Nạn thất nghiệp làm ảnh hưởng không nhỏ đến quá trình phát triển của một quốc gia. Ở nước ta hàng năm

có hàng trăm ngàn sinh viên ra trường nhưng chỉ số ít trong đó kiếm được việc làm phù hợp còn lại là thất nghiệp hoặc làm những công việc không phù hợp với trình độ và năng lực của bản thân. Với những vấn đề đã nêu trên tôi đề xuất xây dựng đề tài “*Ứng dụng công nghệ semantic web để phát triển hệ thống tư vấn việc làm*”.

➤ Mục tiêu và nhiệm vụ nghiên cứu

Nghiên cứu công nghệ Semantic web, tìm hiểu hướng phát triển một ứng dụng sử dụng công nghệ Semantic web. Từ đó, phát triển ứng dụng tư vấn việc làm.

➤ Phạm vi và đối tượng nghiên cứu

❖ Đối tượng nghiên cứu

Đối tượng nghiên cứu gồm: Web ngữ nghĩa, phương pháp xây dựng và lưu trữ dữ liệu trong Web ngữ nghĩa, các công cụ dùng để xây dựng Web ngữ nghĩa, 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 và hình thức tư vấn việc làm.

❖ Phạm vi nghiên cứu

- Xây dựng tập từ vựng Ontology về tư vấn việc làm
- Xây dựng ứng dụng để tư vấn việc làm trên cơ sở tập từ vựng đã tạo

➤ Phương pháp nghiên cứu

❖ Phương pháp nghiên cứu lý thuyết

- Nghiên cứu tài liệu, công cụ và công nghệ liên quan.
- Tổng hợp các tài liệu, dữ liệu.

❖ Phương pháp khảo sát

- Tìm hiểu các hình thức tư vấn việc làm từ các trung tâm hoặc từ các hệ thống tư vấn qua mạng hiện có.

- Tìm hiểu các vấn đề liên quan các thông tin cần thiết trong quá trình tư vấn.

❖ **Phương pháp nghiên cứu thực nghiệm**

- Phân tích hệ thống, xây dựng ứng dụng.

➤ **Ý nghĩa khoa học và thực tiễn của đề tài**

- ❖ Tiếp cận công nghệ mới, phát triển ứng dụng mẫu làm bước đệm để thâm nhập sâu hơn vào công nghệ này.
- ❖ Về mặt thực tiễn: Tạo ứng dụng mẫu để dựa vào đó có thể phát triển một ứng dụng hoàn thiện góp phần cải thiện vấn đề việc làm ở nước ta hiện nay

➤ **Bố cục luận văn**

Luận văn được chia làm 3 chương

Chương 1 trình bày nội dung nghiên cứu tổng quan về web ngữ nghĩa, thực hiện nghiên cứu các lý thuyết liên quan đến công nghệ xây dựng một web ngữ nghĩa, đặc biệt khái niệm về RDF, Ontology. Ontology đóng một vai trò quan trọng do đây là phương tiện giúp cung cấp ngữ nghĩa cho các trang web.

Chương 2 chúng tôi giới thiệu một số công cụ hỗ trợ xây dựng web ngữ nghĩa và giới thiệu về môi trường, ngôn ngữ để phát triển ứng dụng.

Chương 3 Đề xuất giải pháp, giới thiệu quá trình phát triển, cài đặt ứng dụng và đưa ra một số kết quả thực hiện của ứng dụng.

Chương 1 TỔNG QUAN VỀ SEMANTIC WEB

Trong chương này, tôi dành để giới thiệu những kết quả nghiên cứu tổng quan về web ngữ nghĩa, các vấn đề liên quan đến web ngữ nghĩa

1.1 Web ngữ nghĩa (Semantic Web – SW)

1.1.1 Giới thiệu

Web ngữ nghĩa là sự mở rộng của Web hiện tại mà trong đó thông tin được định nghĩa rõ ràng sao cho con người và máy tính có thể cùng làm việc với nhau một cách hiệu quả hơn. Mục tiêu của Web có ngữ nghĩa là để phát triển các chuẩn chung và công nghệ cho phép máy tính có thể hiểu được nhiều hơn thông tin trên Web, sao cho chúng có thể hỗ trợ tốt hơn việc khám phá thông tin, tích hợp dữ liệu (*dữ liệu liên kết động*), và tự động hóa các công việc.

1.1.2 Web ngữ nghĩa là gì?

Web ngữ nghĩa được phát triển bởi Tim- Berners Lee, cha đẻ của WWW, URIs, HTTP và HTML. Theo Ông, “ *Web ngữ nghĩa là sự mở rộng của Web hiện tại, cho phép người dùng có thể truy tìm, phối hợp, sử dụng lại và trích lọc thông tin một cách dễ dàng và chính xác* ”. (Tim- Berners Lee, XML-2000).

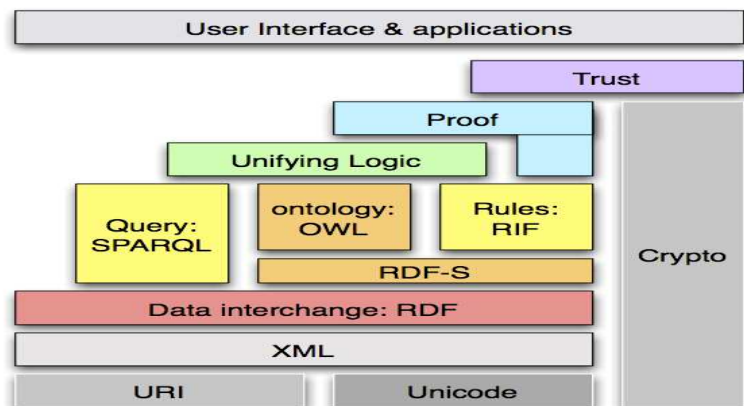
1.1.2.1 Nội dung xây dựng Web ngữ nghĩa

Để xây dựng hệ thống Web ngữ nghĩa thay thế cho World Wide Web hiện tại, các nhà nghiên cứu đang nỗ lực và tập trung nghiên cứu với ba hướng chính sau:

- Chuẩn hoá các ngôn ngữ biểu diễn dữ liệu (XML) và siêu dữ liệu (RDF) trên Web.
- Chuẩn hoá các ngôn ngữ biểu diễn Ontology cho Web có ngữ nghĩa.

- Phát triển nâng cao Web có ngữ nghĩa (Semantic Web Advanced Development - SWAD).

1.1.2.2 Kiến trúc phân tầng của Semantic Web



Hình 1.1 Kiến trúc phân tầng của web ngữ nghĩa năm 2006

Trong cấu trúc trên mỗi tầng có một vai trò nhất định:

➤ Tầng URI, UNICODE

- ❖ **Unicode:** là một bảng mã chuẩn chung có đủ các ký tự để thống nhất sự giao tiếp trên tất cả các quốc gia.
- ❖ **URI (Uniform Resource Identifier):** là kí hiệu nhận dạng Web đơn giản.

➤ Tầng XML

XML – (eXtensible Markup Language) là ngôn ngữ đánh dấu mở rộng, cho phép người dùng có thể tùy ý thêm vào những thẻ theo yêu cầu của mình.

➤ Tầng RDF

RDF (Resource Description Framework): Khung mô tả tài nguyên - RDF được W3C giới thiệu để cung cấp một cú pháp chuẩn để tạo, thay đổi và sử dụng các chú thích trong Web ngữ nghĩa.

➤ Tầng Ontology

Ontology Vocabulary Bộ từ vựng ontology được xây dựng trên cơ sở tầng RDF và RDFS, cung cấp biểu diễn ngữ nghĩa mềm dẻo cho tài nguyên web và có khả năng hỗ trợ lập luận.

➤ Tầng Logic

Việc biểu diễn các tài nguyên dưới dạng các bộ từ vựng ontology có mục đích là để máy có thể lập luận được. Mà cơ sở lập luận chủ yếu dựa vào logic.

➤ Tầng Proof

Tầng này đưa ra các luật để suy luận. Cụ thể từ các thông tin đã có ta có thể suy ra các thông tin mới.

➤ Tầng Trust

Đảm bảo tính tin cậy của các ứng dụng trên Web ngữ nghĩa.

1.2 RDF Nền tảng của Semantic Web

1.2.1 Giới thiệu về RDF

RDF là một thành phần quan trọng của Semantic Web, được đặt trên XML, RDF sử dụng cú pháp của XML để biểu diễn thông tin. Ngôn ngữ XML dùng để biểu diễn thông tin trong RDF được gọi là RDF/XML. Thông qua định dạng này, các thông tin trong RDF có thể được trao đổi dễ dàng giữa các hệ thống máy tính cũng như các hệ điều hành hay các ngôn ngữ lập trình ứng dụng khác nhau.

1.2.2 RDF là gì?

RDF (Resource Description Framework) là một “bộ khung” được sử dụng để mô tả các nguồn tài nguyên trên Internet

1.2.3 Mô hình RDF

Mô hình cơ bản của RDF gồm ba đối tượng sau:

- **Tài nguyên (Resources):** là tất cả những gì được mô tả bằng biểu thức RDF.

- **Thuộc tính** (Properties): thuộc tính, đặc tính, hoặc quan hệ dùng để mô tả tính chất của tài nguyên.
- **Phát biểu** (Statements): mỗi phát biểu gồm ba thành phần sau:
 - ❖ Subject (Tài nguyên): địa chỉ hay vị trí tài nguyên muốn mô tả
 - ❖ Predicate (Vị ngữ): xác định tính chất của tài nguyên.
 - ❖ Object (Bổ ngữ): có thể là một giá trị nguyên thủy hoặc **cũng** có thể là một tài nguyên.

Mỗi một phát biểu (subject, predicate, object) còn gọi là một bộ ba (triple).

Ví dụ: Xét phát biểu sau

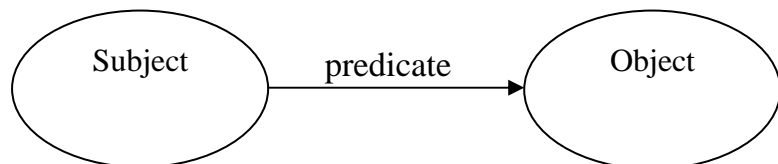
“Son có anh là Minh”

Phát biểu trên được phân ra thành các phần sau:

Subject	son
Predicate	hasBrother (có anh)
Object	Minh

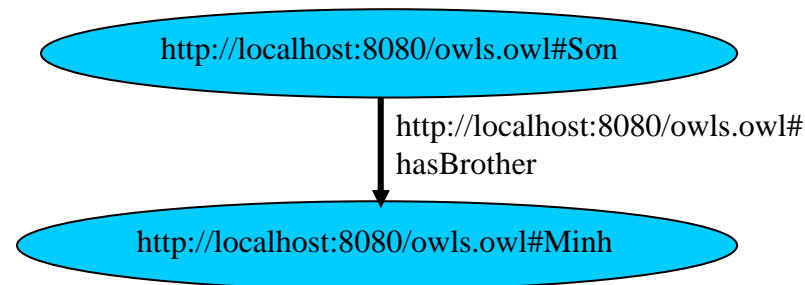
1.2.4 Đồ thị RDF

Một tập hợp các RDF Triple được gọi là một đồ thị RDF (RDF Graph).



Hình 1.2 Mối quan hệ giữa các thành phần trong triple

Vi dụ 1: Mô hình hóa cho phát biểu trên



Hình 1.3 Mô hình bộ ba Triple

1.2.5 Namespace

Namespace là một tập các tên (*name*), được định danh bởi các URI, được sử dụng trong các tài liệu XML như các element type và attribute name

1.2.6 Literal

Literal được sử dụng để biểu diễn các giá trị như con số, ngày tháng, chuỗi... Bất cứ cái gì có thể biểu diễn bởi một giá trị Literal cũng có thể được biểu diễn dưới dạng một URI.

1.2.7 Kiểu dữ liệu có cấu trúc

1.2.7.1 RDF Container

RDF sử dụng một số phần tử đặc biệt để xây dựng các danh sách, gọi là các “bộ chứa” (RDF Container). Ba phần tử chính được dùng để mô tả các nhóm là `rdf:Bag`, `rdf:Seq` và `rdf:Alt`.

1.2.7.2 RDF Collection

RDF collection cho phép khai báo một tập hợp đóng. Cấu trúc của RDF Collection tương tự như một danh sách, có phần tử đầu (`rdf:first`), phần tử kế (`rdf:rest`) và phần tử cuối (`rdf:nil`).

1.2.8 RDFS (RDF Schema)

RDFS được sử dụng để định nghĩa các nguồn tài nguyên và các lớp. Trong RDF, tất cả mọi thứ đều được coi là các nguồn tài nguyên, bản thân các lớp cũng là các nguồn tài nguyên, nhưng bên trong nó cũng có thể là tập hợp các nguồn tài nguyên khác.

RDF/RDFS trước đây được coi là một thành phần cốt yếu để biểu diễn nội dung trong Semantic Web. Tuy nhiên, chúng không đủ mạnh để mô tả thông tin một cách chi tiết. Cụ thể hơn, chúng không có các ràng buộc cục bộ giữa domain và range, các ràng buộc tồn tại, tập hợp, các thuộc tính bắc cầu nghịch đảo, đối xứng... (các thuộc tính liên quan đến logic), và đặc biệt là chúng không có chuẩn ngữ nghĩa nên khó hỗ trợ cho suy diễn, lập luận.

1.3 Ontology và ngôn ngữ Web OWL

1.3.1 Khái niệm Ontology

Ontology cung cấp một bộ từ vựng chung dùng để mô tả một lĩnh vực nghĩa là một loại đối tượng hay khái niệm hiện hữu, cùng với các thuộc tính và quan hệ giữa chúng và lời đặc tả cho nghĩa của những từ trong bộ từ vựng.

1.3.2 Các kiểu kiến trúc của Ontology

Trong môi trường mở như Web, các ontology được phát triển và bảo trì một cách độc lập trong môi trường phân tán. Do đó hai hệ thống có thể sử dụng hai ontology khác nhau để mô tả cho hai domain tương tự nhau, vấn đề này được gọi là không thống nhất ontology. Có ba kiến trúc cơ bản nhằm giải quyết vấn đề này, đó là: ontology đơn, đa ontology, và ontology phức hợp:

1.3.3 Vai trò của Ontology

Danh sách dưới đây sẽ phân tích vai trò của Ontology trong ngữ cảnh ứng dụng Web có ngữ nghĩa.

- Chia sẻ sự hiểu biết chung giữa các ứng dụng và con người.
- Cho phép sử dụng lại tri thức.
- Đưa ra các giả thiết rõ ràng về miền.
- Phân tách tri thức lĩnh vực với tri thức thao tác.
- Phân tích tri thức lĩnh vực. Phân tích hình thức của các khái niệm, cần thiết cho việc tái sử dụng và mở rộng Ontology.

1.3.4 Các thành phần của Ontology

- **Các cá thể (Individuals):** Các cá thể là các thành phần cơ bản, nền tảng của một Ontology.
- **Các lớp (Classes):** các lớp là các nhóm, tập hợp các đối tượng **trừu** tượng. Chúng có thể chứa các cá thể, các lớp khác, hay là sự phối hợp của cả hai.
- **Các thuộc tính (Properties):** các đối tượng trong **Ontology** có thể được mô tả thông qua việc khai báo các thuộc tính của chúng. Mỗi một thuộc tính đều có tên và giá trị của thuộc tính đó. Các thuộc tính được sử dụng để lưu trữ các thông tin mà đối tượng có thể có.
- **Các mối quan hệ (Relation):** Một mối quan hệ là một thuộc tính có giá trị là một đối tượng nào đó trong Ontology.

1.3.5 Ngôn ngữ OWL

OWL (The Web Ontology Language) là một ngôn ngữ gần như XML dùng để mô tả các hệ cơ sở tri thức. OWL là một ngôn ngữ đánh dấu dùng để xuất bản và chia sẻ dữ liệu trên Internet thông qua những mô hình dữ liệu gọi là "Ontology". OWL biểu diễn ý nghĩa của các thuật ngữ trong các từ vựng và mối liên hệ giữa các thuật ngữ này để đảm bảo phù hợp với quá trình xử lý bởi các phần mềm.

Chương 2 CHƯƠNG PHƯƠNG PHÁP VÀ CÔNG CỤ XÂY DỰNG WEB SEMANTIC

Hiện có rất nhiều công cụ có khả năng hỗ trợ người thiết kế giải quyết những bài toán liên quan. Có thể kể ra một số như: Sesame, Protégé, Ontolingua, Chimaera, OntoEdit, OidEd... Ngoài ra trong nội dung chương này còn giới thiệu sơ lược một số ứng dụng được xây dựng trên nền Semantic web.

2.1 Một số công cụ xây dựng ontology

2.1.1 Phần mềm Chimaera

Chimaera cũng là một ứng dụng được phát triển bởi đại học Stanford, với mục đích ban đầu nhằm giải quyết hai vấn đề là: trộn các Ontology và chuẩn đoán lỗi, phân tích tính nhất quán giữa các Ontology phân tán.

2.1.2 Phần mềm Sesame

Sesame là một phần mềm mã nguồn mở trong dự án European IST On-To-Knowledge của công ty Hà Lan Aduna giúp cho việc lưu trữ và truy vấn dữ liệu RDF và RDFS.

Ngày càng nhiều ứng dụng trong lĩnh vực Web ngữ nghĩa sử dụng Sesame như một ứng dụng chủ cho phép lưu trữ và truy vấn dữ liệu RDF và RDFS. Lý do Sesame trở nên phổ biến chính là ở những ưu điểm về kiến trúc của nó, không chỉ về mặt thiết kế mà còn về tính ứng dụng, như sau:

- Sesame được thiết kế như một phần mềm trung gian giữa người dùng và kho lưu trữ dữ liệu.
- Sesame cung cấp một cơ chế cho phép lưu trữ ổn định và truy vấn hiệu quả dữ liệu RDF và RDFS.
- Người phát triển phần mềm có thể sử dụng Sesame như một thư viện khi xây dựng các ứng dụng thao tác với Sesame.

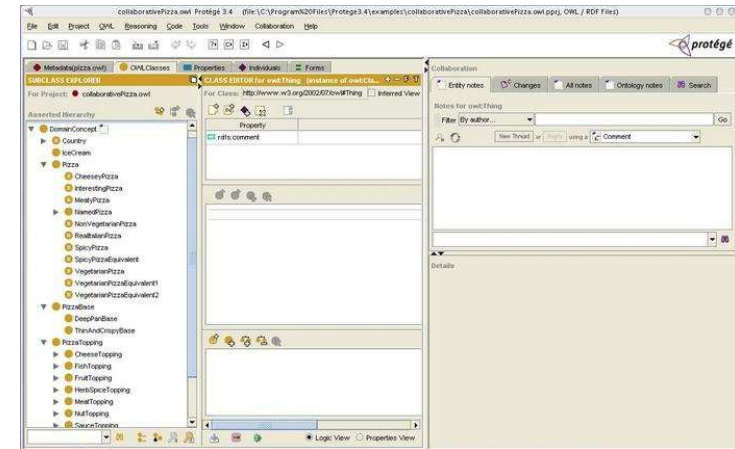
2.1.3 Công cụ Jena

Jena là một Java Framework dùng để xây dựng các ứng dụng Web ngữ nghĩa. Jena cung cấp môi trường lập trình cho RDF, RDFS, OWL và SPARQL - ngôn ngữ truy vấn cho RDF. Jena bao gồm các thành phần và tính năng sau:

2.1.4 Phần mềm Protégé

Protégé là bộ phần mềm mã nguồn mở Java nổi tiếng. Protégé được nghiên cứu và phát triển từ năm 1998 bởi nhóm nghiên cứu của Mark Musen, ĐH. Stanford nhằm quản lý các thông tin trong lĩnh vực sinh y học. Mã nguồn Protégé có thể được tìm thấy tại website:

<http://smi-protege.stanford.edu/repos/protege/owl/trunk>



Hình 2.1 Giao diện phần mềm Protégé

2.1.4.1 Đặc điểm của Protégé

Chức năng nổi bật nhất của phần mềm này là cho phép người dùng sử dụng tạo ra các ontology để phát triển Web Semantic theo đúng chuẩn của ngôn ngữ W3C OWL.

Protégé có hai phiên bản OWL và API.

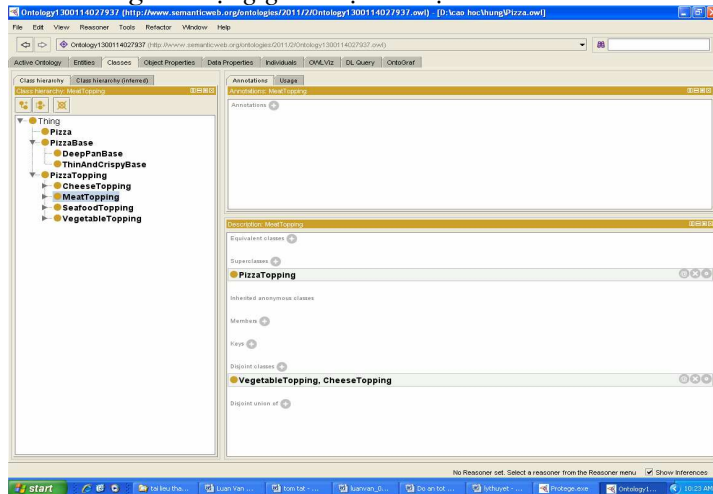
Phiên bản Protégé-API có nền tảng từ OKBC (Open Knowledge Base Connectivity). OKBC là một ứng dụng lập trình giao tiếp thực hiện truy xuất dữ liệu thông minh.

Phiên bản Protégé-OWL được phát triển dựa trên hai yêu cầu chính. Đầu tiên là yêu cầu định nghĩa các đối tượng và quan hệ tồn tại giữa chúng. Sau đó là yêu cầu xây dựng các đặc điểm kỹ thuật phục vụ ý tưởng chia sẻ thông tin.

Các đối tượng xây dựng chính của Protégé là:

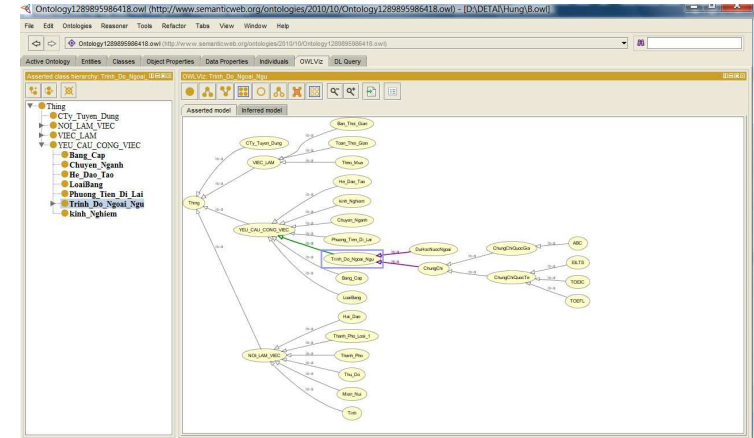
- Classes – tổ chức các quan hệ tham chiếu và các kiểu thực thi
- Axioms – mô hình câu lệnh đúng
- Instances – các thể hiện, các thành phần của đối tượng
- Domain – giới hạn của ontology
- Vocabulary – các lớp và khai báo

2.1.4.2 Protégé sử dụng giao diện đồ họa



Hình 2.2 Giao tiếp bằng đồ họa của Protégé

2.1.4.3 Protégé phát triển để tích hợp các công cụ

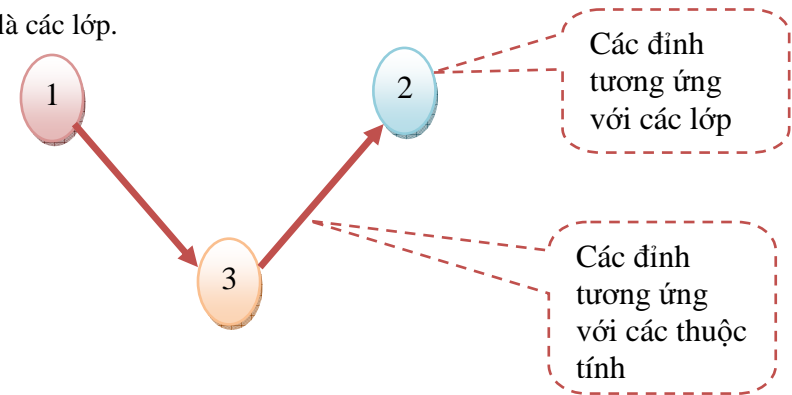


Hình 2.3 Protégé tích hợp công cụ OWL Viz

2.2 Thư viện phát triển ứng dụng

2.2.1 Thư viện mã nguồn mở OWLDotNetAPI

Nhiệm vụ của OwlDotNetApi là kết hợp C# và OWL, thực hiện 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.



Hình 2.4 Mô hình quan hệ giữa các nút và các cạnh

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 có hiệu quả không cao. OwlDotNetApi có hai phần chính là lớp chức năng và lớp giao tiếp:

2.3 Phương pháp xây dựng ontology

Dựa trên các bước xây dựng ontology của Noy và McGuinness ta có sự tinh gọn công việc trong mỗi bước như sau:

- **Bước 1.** Xác định mục đích phát triển ontology.
- **Bước 2.** Nắm bắt kỹ thuật xây dựng ontology:
Bước này gồm ba giai đoạn như sau:
 - ❖ Xác định phạm vi của ontology
 - ❖ Chọn phương thức nắm bắt ontology
 - ❖ Định nghĩa các khái niệm trong ontology
- **Bước 3.** Xem xét sử dụng lại các ontology đang tồn tại.
- **Bước 4.** Mã hoá ontology

Lựa chọn trình biên tập ontology dựa trên các yêu cầu của lĩnh vực và chức năng của ontology. Mã hóa ontology là tiến trình lập, gồm các bước con sau:

- ❖ Mã hóa ontology tổng quát.
- ❖ Định nghĩa lớp.
- ❖ Sắp xếp các lớp theo cây phân cấp
- ❖ Định nghĩa thuộc tính và mô tả giá trị của thuộc tính.
- **Bước 5.** Cải tiến ontology

Bao gồm hai giai đoạn:

Cải tiến mã hóa bên trong (intra-coding): cải tiến trong quá trình mã hóa. Trong khi mã hóa, nếu phát hiện ra lỗi hoặc yêu cầu mới, mã cần được cải tiến để hiệu chỉnh hoặc thực hiện yêu cầu mới đó.

Cải tiến mã hóa bên ngoài (extra-coding): hiệu chỉnh lỗi phát hiện được trong quá trình kiểm thử, và những mở rộng của ontology theo các phân hệ ứng dụng.

- **Bước 6:** Kiểm thử
Phát hiện nhược điểm của ontology
- **Bước 7:** Duy trì

Thực hiện các việc hiệu chỉnh, thích ứng hoặc hoàn tất ontology. Hiệu chỉnh là xem xét vấn đề mắc phải khi truy vấn ontology và hiệu chỉnh ontology để khắc phục các vấn đề này. Thích ứng bao gồm việc điều chỉnh ontology theo các yêu cầu mới phát sinh. Hoàn tất ontology là phát triển cải tiến ontology trong tương lai.

Chương 3 PHÁT TRIỂN ỨNG DỤNG

3.1 Mô tả bài toán

Tìm kiếm việc làm luôn là một ứng dụng rất quan trọng và thực tế cho thấy có rất nhiều công cụ hỗ trợ tư vấn và tìm kiếm việc làm đã và đang được sử dụng. Nhưng hiệu quả của nó mang lại chưa cao. Dựa vào thực tế đó và những nghiên cứu về công nghệ web ngữ nghĩa, tôi xây dựng một ứng dụng hỗ trợ tư vấn việc làm sử dụng công nghệ web ngữ nghĩa. Ứng dụng được xây dựng nhằm phục cho đối tượng là tất cả mọi người lao động trong xã hội có nhu cầu tìm kiếm một việc làm phù hợp.

Đối với người đi tìm việc làm, sau khi nhập vào các thông tin cá nhân cần thiết hệ thống sẽ đưa ra danh sách các việc làm phù hợp với năng lực của họ và đồng thời đưa ra một số thông tin tư vấn liên quan đến công ty có tuyển dụng việc làm đó. Ngoài ra ứng dụng còn cho phép chúng ta tìm việc làm bằng cách nhập vào các từ khóa. Điều này cho phép chúng ta tự tra cứu các danh mục việc làm theo chủ đề.

3.2 Kiến trúc chung của ứng dụng Semantic web

Cấu trúc của một máy tìm kiếm theo công nghệ web ngữ nghĩa, về cơ bản cũng có cấu trúc tương tự với một máy tìm kiếm thông thường, bao gồm 2 thành phần chính

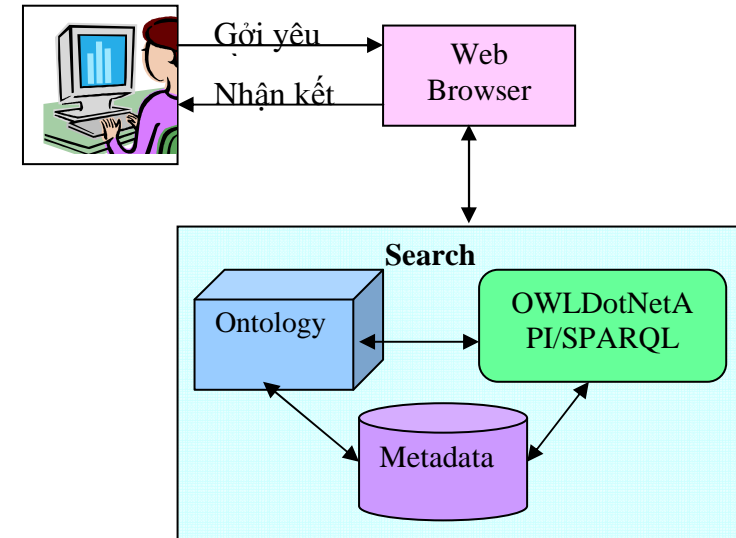
- Giao diện truy vấn: cho phép người dùng nhập yêu cầu tìm kiếm
- Hiển thị kết quả tìm kiếm

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

- Phân tích yêu cầu
- Tìm kiếm kết quả cho yêu cầu

- Dữ liệu tìm kiếm, mạng ngữ nghĩa

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



Hình 3.1 Mô hình ứng dụng tìm kiếm ngữ nghĩa thông tin

- Web Browser : đóng vai trò giao tiếp với người dùng.
- Search Engine: Đây là chức năng chính của chương trình thực hiện các thao tác sau:
 - ❖ Tổ chức lưu trữ Ontology và theo các quan hệ, thuộc tính của Ontology để hiển thị các thông tin theo dạng phân cấp.
 - ❖ Thực hiện truy vấn yêu cầu của người dùng trên Ontology và trả về kết quả cho Web Browser địa điểm theo yêu cầu của người dùng.

3.3 Quy trình xây dựng ứng dụng

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

➤ Giai đoạn 1 : Thiết kế Ontology

Thiết kế Ontology nhằm cung cấp một bộ từ vựng chung bao gồm các khái niệm, các thuộc tính quan trọng và các định nghĩa về các khái niệm. Các thuộc tính này bao quanh chủ đề về việc làm.

Ngoài bộ từ vựng, Ontology còn cung cấp các ràng buộc.

➤ Giai đoạn 2: Xây dựng ứng dụng

3.4 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.

3.4.1 Mô hình phân cấp lớp trong 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 tư vấn việc làm sẽ được phân rã và lưu trữ dưới các dạng khác nhau. Sau đây là mô hình phân cấp lớp trong ứng dụng:



Hình 3.3 Phân cấp lớp trong Ontology

3.4.2 Thực hiện xây dựng HignOnt

Vấn đề xây dựng ontology tương tự như cách người dùng nhập dữ liệu vào cơ sở dữ liệu. Ở đây, tất cả dữ liệu và cấu trúc phân tách nó sẽ được khai báo thông qua các công cụ. Các đối tượng chính của ontology như class, properties, individuals sẽ được xem xét và điền đầy thông tin một cách càng cụ thể thì khả năng khai thác về sau sẽ càng thuận tiện.

3.4.2.1 Các lớp chính trong ontology tư vấn việc làm

- Lớp Cty_TuyenDung
- Lớp Noi_Lam_Viec
- Lớp Viec_Lam
- Lớp Yeu_Cau_Cong_Viec

3.4.2.2 Thuộc tính

Các thuộc tính thể hiện mối quan hệ giữa các lớp, 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.

- Thuộc tính dữ liệu: Thuộc tính *coTenGoi*, *coPhuCap*, *coBaoHiem*, *coDiaChi*, *coMucLuong*, *coNgayTuyenDung*

Ngoài ra còn một số thuộc tính khác như: *coMucDiem*, *coSoNgayPhep*, ...

- Thuộc tính quan hệ: Thuộc tính *coYeuCauBangCap*, *coHeDaoTao*, *coChuyenNganh*, *thuocCongTy*, *coNoiLamViec*, *coPhuongTienDiLai*

3.4.2.3 Cá thể

Một ứng dụng bất kỳ về semantic đều chứa rất nhiều cá thể. Mỗi cá thể là một đối tượng không thể chia nhỏ hơn.

Sau đây là một số cá thể tiêu biểu: *Cá thể CTTD_CTy_CNSG*, *CVV_KSCK*, *YCC_BangDaiHoc*, *HDT_ChinhQuy*, *YCCN_CoKhi*

3.5 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. Những phương thức trong thư viện này đã hỗ trợ cho chúng tôi rất nhiều về việc phân loại và xử lý dữ liệu. Trong quá trình xây dựng chúng tôi sử dụng một số thuật toán sau đây để khai thác dữ liệu.

3.5.1 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.

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
 - Nếu một đỉnh có chứa quan hệ cân đối đầy theo danh sách đối chiếu ở trên (B1)
 - Điền thông tin quan hệ ngược lại
 - Quay lại xét cho đỉnh vừa điền như B1
 - Ngược lại bỏ qua bước này
- Đóng truy cập vào ontology

3.5.2 Duyệt theo ngữ nghĩa

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 và hiển thị nội dung từng lớp dựa trên Properties phân cấp trong lớp *Noi_Lam_Viec* để người dùng lựa chọn.

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

B3: Truy cập thông tin chi tiết của các individual và hiển thị kết quả.

3.5.3 Tìm kiếm cơ bản

Chức năng của tìm kiếm cơ bản là dựa vào từ khóa chỉ tên hoặc một thông tin nào đó về việc làm.

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.

3.5.4 Tìm kiếm nâng cao

Sau đây là giao diện tìm kiếm nâng cao

TÌM KIẾM NÂNG CAO	
Khu vực	
Miền núi	Bằng Đại học
Hải đảo	Ngành học Cơ khí
Tỉnh	Hệ đào tạo Chính quy
Thành Phố	Kinh 3 năm
Thủ đô	Hình thức Toàn thời gian
Chuyên Ngành	Khu vực Thành Phố
Công ty	Ngoại ngữ Bảng C tiếng Anh
Hình thức làm	Tim

Hình 3.9 Giao diện tìm kiếm nâng cao

KẾT LUẬN

Qua luận văn này chúng tôi đã nghiên cứu và trình bày những vấn đề then chốt của lĩnh vực Web ngữ nghĩa, trong đó đi sâu vào nghiên cứu thành phần quan trọng của công nghệ này đó là RDF và Ontology, nó cung cấp một hệ thống mã hóa đơn giản và nhất quán hỗ trợ người sử dụng tìm kiếm và truy hồi thông tin một cách hiệu quả. Nghiên cứu một số phương pháp, công cụ để xây dựng một ứng dụng web ngữ nghĩa đồng thời qua đó tiến hành xây dựng một ứng dụng thử nghiệm để minh họa cho công việc nghiên cứu trên. Cụ thể chúng tôi đã tiến hành xây dựng một Ontology về lĩnh vực việc làm. Ontology này mô tả tổng quát được các thực thể cơ bản trong dữ liệu về việc làm với mục đích đưa ra sự lựa chọn chính xác phù hợp với năng lực của người lao động.

Bên cạnh những thành công đã đạt được thì nghiên cứu này còn chưa có sự kết nối giữa dữ liệu được trích lọc từ web 2.0 và dữ liệu dùng trong ontology. Bản thân trích lọc và sử dụng dữ liệu từ web 2.0 là một vấn đề lớn còn có nhiều điểm chưa thống nhất trong nhiều nghiên cứu khác nhau trên thế giới.

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. Thêm nữa chương trình chưa khai thác được những dữ liệu liên quan đã được xây dựng trong các ontology trên mạng.

Trong thời gian đến, để hoàn thiện ứng dụng đã phát triển, chúng tôi xây dựng một Ontology hoàn thiện hơn về lĩnh vực việc làm, nhằm tư vấn giúp người lao động có thể tìm một việc làm phù hợp với bản thân. Tiếp đến, chúng tôi sẽ tích hợp với Ontology của những lĩnh vực khác nhau để mở rộng và có thể tận dụng hết những ưu điểm của công nghệ tìm kiếm trên web ngữ nghĩa.