

**BỘ GIÁO DỤC VÀ ĐÀO TẠO**  
**ĐẠI HỌC ĐÀ NẴNG**

**NGUYỄN THỊ KIM CƯƠNG**

**ỨNG DỤNG WEB NGỮ NGHĨA  
ĐỂ XÂY DỰNG HỆ THỐNG TRA CỨU  
THÔNG TIN VỀ VĂN HÓA KHMER NAM BỘ**

**Chuyên ngành : Khoa học máy tính**  
**Mã số: 60.48.01**

**TÓM TẮT LUẬN VĂN THẠC SĨ KỸ THUẬT**

**Đà Nẵng - Năm 2013**

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

**Người hướng dẫn khoa học: PGS.TS. VÕ TRUNG HÙNG**

Phản biện 1: **TS. HUỖNH HỮU HÙNG**

Phản biện 2: **PGS.TS. TRẦN CAO ĐỆ**

Luận văn được bảo vệ tại Hội đồng chấm luận văn tốt nghiệp Thạc sĩ kỹ thuật họp tại Đại học Đà Nẵng vào ngày 8 tháng 6 năm 2013.

*\* Có thể tìm hiểu luận văn tại:*

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

## MỞ ĐẦU

### 1. Tính cấp thiết của đề tài

Bản sắc văn hóa dân tộc là vấn đề trọng đại, sống còn của mỗi quốc gia. Ngày 27 tháng 7 năm 2011, Thủ tướng Chính phủ đã có quyết định số 1270/QĐ-TTg phê duyệt Đề án “*Bảo tồn, phát triển văn hóa các dân tộc thiểu số Việt Nam đến năm 2020*” gồm 06 dự án thành phần với tổng kinh phí dự kiến 1.512 tỷ đồng. Trên cơ sở đó, Bộ VH,TT&DL đã có Chỉ thị 194/CT-BVHTTDL về việc tổ chức triển khai thực hiện Đề án trên nhằm huy động sức mạnh toàn xã hội trong phát triển văn hóa dân tộc theo tinh thần Nghị quyết Đại hội Đảng toàn quốc lần thứ XI đề ra nhiệm vụ chăm lo phát triển văn hóa.

Qua quá trình cộng cư lâu đời cùng với các dân tộc Kinh - Hoa - Chăm trên mảnh đất Nam Bộ, người Khmer đã có sự giao thoa văn hóa với các dân tộc anh em; nhưng cơ bản người Khmer vẫn giữ được nét văn hóa đặc sắc, những cốt cách tinh hoa của dân tộc mình. Trên phương diện văn hóa, người Khmer Nam Bộ đã góp vào kho tàng văn hóa Nam Bộ nói riêng, văn hóa Việt Nam nói chung, vốn bản sắc làm phong phú thêm nét đặc thù văn hóa Việt Nam.

VHKMNB được bảo tồn và lưu giữ trong bảo tàng, trong sách, tài liệu, nguồn tài liệu không nhiều, lưu trữ rải rác ở nhiều nơi, thông tin không chuẩn xác và hiện chưa có hệ thống quản lý, lưu trữ chuyên biệt. Trong khi đó nhu cầu đặt ra cho việc tìm kiếm thông tin ngày càng cao, một số ít thông tin về VHKMNB được đăng tải trên các trang Web chỉ dừng lại ở việc cung cấp thông tin chứ chưa cung cấp được những tri thức cần thiết nên đã có những khó khăn nhất định trong việc tìm hiểu, tra cứu.

WWW chứa một lượng thông tin khổng lồ, người sử dụng Web có thể dễ dàng truy cập những thông tin bằng địa chỉ hoặc theo các liên kết. Web 2.0 đã đạt được những thành tựu đáng kể, nhưng nhu cầu của người dùng không dừng lại ở việc cải thiện tốc độ mà còn phải cải thiện chất lượng xử lý theo yêu cầu ngày càng nâng cao. Web 3.0 - Web ngữ nghĩa ra đời nhằm đáp ứng những yêu cầu về chất lượng đó. Thế mạnh của Web ngữ nghĩa là xử lý và tìm kiếm thông tin, cho phép chúng ta xây dựng những CSDL phục vụ tìm kiếm chính xác. Đây là một xu hướng mới còn đang được nghiên cứu và tiếp tục phát triển.

Nghiên cứu ứng dụng Web ngữ nghĩa để xây dựng hệ thống tra cứu VHKMNB là một việc làm thiết thực không chỉ đáp ứng nhu cầu tìm kiếm chính xác và cung cấp những tri thức cần thiết về VHKMNB, qua đó có thể giới thiệu những nét văn hóa đặc trưng của người KMNB nhằm phát huy những nét đẹp, những mặt tích cực, những tiềm năng, thế mạnh của dân tộc Khmer trong sự phát triển chung của cộng đồng các dân tộc VN trong thời kỳ phát triển mới của đất nước.

Xuất phát từ thực tế nêu trên cùng với sự hướng dẫn tận tình của thầy Võ Trung Hùng, tôi quyết định chọn đề tài: “*Ứng dụng Web ngữ nghĩa để xây dựng hệ thống tra cứu thông tin về văn hóa Khmer Nam Bộ*” làm luận văn tốt nghiệp.

## **2. Mục tiêu nghiên cứu**

Xây dựng và chuẩn hóa kho dữ liệu VHKMNB và ứng dụng công nghệ Web ngữ nghĩa để xây dựng hệ thống tra cứu thông tin về VHKMNB nhằm hỗ trợ việc tra cứu, đáp ứng nhu cầu tìm kiếm chính xác, cung cấp những tri thức cần thiết về VHKMNB góp phần thực hiện tốt công tác bảo tồn, phát huy bản sắc văn hóa dân tộc.

## **3. Đối tượng và phạm vi nghiên cứu**

❖ *Đối tượng nghiên cứu*

- Các vấn đề liên quan đến Web ngữ nghĩa;
- Qui trình xây dựng Web ngữ nghĩa;
- Công cụ, ngôn ngữ và công nghệ Web ngữ nghĩa;
- Các thư viện phát triển ứng dụng Web ngữ nghĩa;
- Một số ứng dụng có sẵn của Web ngữ nghĩa;
- Phân tích xử lý dữ liệu đã trích lọc trên các trang Web;
- VHKMNB, nhu cầu và hiện trạng tra cứu.

#### ❖ Phạm vi nghiên cứu

VHKMNB rất phong phú với nhiều loại hình và được lưu trữ ở nhiều nơi. Tuy nhiên trong đề tài này tôi chỉ thu thập dữ liệu và nghiên cứu về những phong tục tập quán, về văn hóa tâm linh, về lễ hội truyền thống, kiến trúc chùa chiền, về nhạc điệu, ngữ văn,... của người Khmer Nam Bộ nói chung mà cụ thể là người Khmer ở tỉnh Trà Vinh.

#### **4. Phương pháp nghiên cứu**

Phương pháp nghiên cứu tài liệu: Nghiên cứu tài liệu, công cụ, công nghệ liên quan đến Web ngữ nghĩa; Thu thập, tổng hợp thông tin về Văn hóa Khmer Nam Bộ.

Phương pháp khảo sát: Tìm hiểu về lưu trữ, quản lý các loại hình văn hóa Khmer Nam Bộ; Tìm hiểu các hệ thống tra cứu văn hóa Khmer Nam Bộ hiện có.

Phương pháp thực nghiệm: Phân tích hệ thống, tiến hành xây dựng ứng dụng.

#### **5. Bố cục đề tài**

##### Chương 1. TỔNG QUAN VỀ WEB NGỮ NGHĨA

Trong chương này giới thiệu các khái niệm tổng quan về Web ngữ nghĩa, các ngôn ngữ và công cụ xây dựng Web ngữ nghĩa tính đến

thời điểm hiện nay. Tìm hiểu kiến trúc Web ngữ nghĩa, ứng dụng và triển vọng của Web ngữ nghĩa trong tương lai.

## Chương 2. GIẢI PHÁP ĐỀ XUẤT

Chương này sẽ giới thiệu những nét đặc trưng về văn hóa Khmer Nam Bộ, phản ánh thực chất nhu cầu và hiện trạng tra cứu thông tin về văn hóa Khmer hiện nay. Tiến hành phân tích bài toán tra cứu, trình bày kết quả phân tích thiết kế hệ thống. Tìm hiểu xem để xây dựng một ứng dụng Web ngữ nghĩa cần những gì, mô hình hoạt động như thế nào, công cụ nào sẽ được lựa chọn, gói thư viện nào sẽ hỗ trợ trong quá trình triển khai ứng dụng.

## Chương 3. XÂY DỰNG HỆ THỐNG

Trong chương này sẽ xây dựng Ontology cho bài toán và xây dựng hệ thống tra cứu thông tin về Văn hoá Khmer Nam bộ đồng thời mô tả hoạt động hệ thống. Hoàn thiện trang Web ngữ nghĩa hỗ trợ việc tra cứu, tiến hành thử nghiệm và đánh giá kết quả của chương trình.

### 6. Tổng quan về vấn đề nghiên cứu

Với chủ trương đẩy mạnh ứng dụng công nghệ thông tin trong tất cả các lĩnh vực, ứng dụng công nghệ thông tin trong lĩnh vực văn hóa là điều tất yếu. Hiện nay, với lượng thông tin khổng lồ trên Internet, lượng người có nhu cầu tìm kiếm, tra cứu ngày một tăng đáng kể, công nghệ Web ngữ nghĩa ra đời đã tạo bước phát triển mới.

Web ngữ nghĩa sẽ giúp xây dựng hệ thống tra cứu thông tin nhằm hỗ trợ việc tra cứu, đáp ứng nhu cầu tìm kiếm về một lĩnh vực cụ thể và đã được nghiên cứu ứng dụng trong nhiều lĩnh vực.

Qua khảo sát thực tế tôi được biết hiện nay chưa có tổ chức, cá nhân nào ứng dụng Web ngữ nghĩa để xây dựng hệ thống tra cứu thông tin về VHKMNB. Vì vậy, tôi mạnh dạn chọn đề tài: “*Ứng*

*dụng Web ngữ nghĩa để xây dựng hệ thống tra cứu thông tin về văn hóa Khmer Nam Bộ”* làm vấn đề nghiên cứu của mình. Với mong muốn ứng dụng công nghệ mới này để xây dựng hệ thống tra cứu thông tin về VHKMNB nhằm hỗ trợ việc tra cứu, đáp ứng nhu cầu tìm kiếm nhanh chóng, chính xác và cung cấp được những tri thức cần thiết về VHKMNB góp phần thực hiện tốt công tác bảo tồn, phát huy bản sắc văn hóa dân tộc tốt hơn so với Web thường.

## CHƯƠNG 1

### TỔNG QUAN VỀ WEB NGŨ NGHĨA

#### 1.1. WEB NGŨ NGHĨA

##### 1.1.1. World Wide Web và những hạn chế

Người ta nói rằng WWW là linh hồn của Internet. Thật vậy, Internet chỉ trở nên thực sự hấp dẫn khi WWW xuất hiện. Dữ liệu trong HTML đúng trong ngữ cảnh nhưng có khi lại không đúng với những ngữ cảnh khác. Vì vậy, nếu như các thành phần chính yếu của dữ liệu trong Web trình bày theo dạng thức thông thường, thì thật khó sử dụng dữ liệu này một cách phổ biến. Do đó cần phải mở rộng Web để máy có thể hiểu, có thể tích hợp dữ liệu, cũng như tái sử dụng dữ liệu thông qua các ứng dụng khác nhau.

##### 1.1.2. Sự ra đời của Web ngữ nghĩa

WWW chứa một lượng thông tin khổng lồ, người sử dụng Web có thể dễ dàng truy cập thông tin bằng địa chỉ hoặc theo các liên kết. Tuy nhiên, các trang Web hiện nay người sử dụng phải tự suy luận, tổng hợp và trích rút thông tin mình cần. Điều đó đã đặt ra thách thức là làm sao để khai thác thông tin trên Web hiệu quả hay nói cách khác là làm thế nào để máy tính có thể xử lý tự động được chúng. Muốn vậy thì Web phải có khả năng mô tả thông tin theo cách mà máy tính có thể hiểu được.

##### 1.1.3. Định nghĩa Web ngữ nghĩa

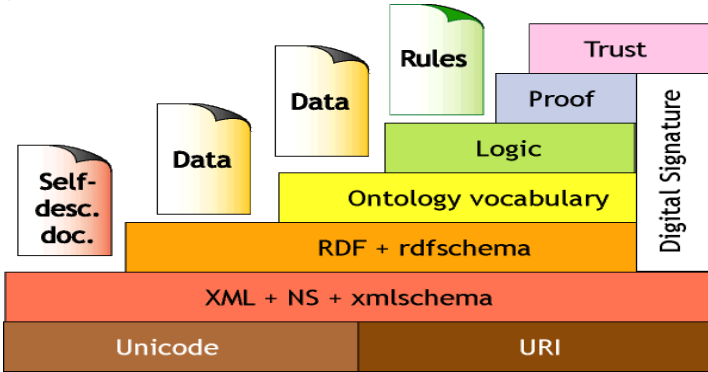
Web ngữ nghĩa được hình thành từ ý tưởng của Tim Berners Lee, Web ngữ nghĩa là sự mở rộng của Web hiện tại mà trong đó thông tin được định nghĩa rõ ràng sao cho con người và máy tính có thể làm việc với nhau một cách hiệu quả hơn.



## 1.2. KIẾN TRÚC WEB NGỮ NGHĨA

### 1.2.1. Giới thiệu

Web ngữ nghĩa được xây dựng trên nền hệ thống Web hiện tại và được coi là sự mở rộng của Web hiện tại có bổ sung thêm ngữ nghĩa vào dữ liệu trên Web.



Hình 1.5. Kiến trúc Web ngữ nghĩa đề xuất của Tim Berners Lee

### 1.2.2. Vai trò các tầng trong kiến trúc Web ngữ nghĩa

## 1.3. NỘI DUNG XÂY DỰNG WEB NGỮ NGHĨA

### 1.3.1. XML và RDF trên Web

Web ngữ nghĩa được phát triển trên nền Web hiện tại, trên nền XML, chúng ta có thể thấy rất rõ trong sơ đồ kiến trúc của Web ngữ nghĩa. Việc xây dựng ngôn ngữ chuẩn cho XML và RDF là thiết yếu để biểu diễn dữ liệu cho Web ngữ nghĩa.

### 1.3.2. Các ngôn ngữ biểu diễn Ontology

Một số ngôn ngữ được đề xuất với các khả năng biểu diễn tăng dần như: RDFS, DAML+OIL, OWL, ... và tiếp tục được mở rộng trong tương lai.

### 1.3.3. Phát triển nâng cao Web ngữ nghĩa

Các công việc chuẩn hoá các ngôn ngữ biểu diễn dữ liệu XML, siêu dữ liệu RDF hay ngôn ngữ biểu diễn Ontology là các công việc nền tảng, cơ sở tạo ra chuẩn chung cơ sở để phát triển hệ thống và các ứng dụng trên Web ngữ nghĩa.

## 1.4. NGÔN NGỮ VÀ CÔNG CỤ XÂY DỰNG WEB NGỮ NGHĨA

Có nhiều ngôn ngữ cho Web ngữ nghĩa, hầu hết những ngôn ngữ này dựa trên XML hay sử dụng XML làm cú pháp. Một số ngôn ngữ sử dụng RDF và RDFschema.

### 1.4.1. RDF - Nền tảng của Web ngữ nghĩa

#### a. Giới thiệu RDF

Có thể nói, RDF chính là nền tảng của Web ngữ nghĩa, là linh hồn của Web ngữ nghĩa. RDF mô tả siêu dữ liệu về các tài nguyên trên Web. RDF dựa trên cú pháp XML.

#### b. Mô hình RDF cơ bản

Mô hình cơ bản của RDF gồm ba đối tượng: Tài nguyên, thuộc tính và phát biểu.

### 1.4.2. RDF Schema

#### a. Giới thiệu

RDF schema cung cấp một khung để mô tả các lớp, thuộc tính của ứng dụng cụ thể.

#### b. Định nghĩa Class

Các resource trên Web có thể chia thành các nhóm gọi là class.

### ***c. Định nghĩa Property***

RDF Schema cũng cung cấp một bộ từ vựng để mô tả làm thế nào mà các thuộc tính (Property) và lớp (Class) có thể được sử dụng cùng với nhau trong dữ liệu RDF.

### **1.4.3. Truy vấn dữ liệu**

#### ***a. Giới thiệu***

Tổ chức W3C đã phát triển một ngôn ngữ chuẩn để truy vấn dữ liệu RDF với việc cung cấp nhiều chức năng cho các developer và end users cách để viết và thực thi một câu truy vấn. SPARQL là một ngôn ngữ để truy cập thông tin từ các đồ thị RDF.

#### ***b. Tạo một câu truy vấn đơn giản***

Một câu truy vấn bao gồm 2 mệnh đề, mệnh đề SELECT và mệnh đề WHERE. Mệnh đề SELECT định danh các biến mà ứng dụng quan tâm và mệnh đề WHERE bao gồm các khuôn mẫu bộ ba (Triple pattern).

### **1.4.4. Ontology**

#### ***a. Khái niệm Ontology***

Ontology là tập từ vựng để mô hình hóa thế giới bên ngoài, nó đưa ra các khái niệm cơ bản và định nghĩa quan hệ giữa các khái niệm đó trong một miền lĩnh vực. Đồng thời Ontology còn cung cấp các ràng buộc, là các giá định cơ sở về ý nghĩa mong muốn của bộ từ vựng.

#### ***b. Mục đích xây dựng Ontology***

#### ***c. Các thành phần Ontology***

Ontology thường miêu tả: Cá thể (Individuals), lớp (Classes), thuộc tính (Propertises) và quan hệ (Relation).

#### ***d. Ngôn ngữ OWL***

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”.

#### ***e. Công cụ phát triển Ontology***

Trong số các công cụ hiệu chỉnh Ontology hiện nay thì Protégé là công cụ được sử dụng rộng rãi và hiệu quả nhất.

### **1.5. ỨNG DỤNG VÀ TRIỂN VỌNG CỦA WEB NGỮ NGHĨA**

#### **1.5.1. Một số ứng dụng có sẵn của Web ngữ nghĩa**

#### **1.5.2. Triển vọng phát triển của Web ngữ nghĩa**

### **1.6. NHẬN XÉT**

Việc tìm kiếm dựa trên Web ngữ nghĩa sẽ thu được kết quả tìm kiếm chính xác hơn là điều tất yếu. Với mục tiêu là máy tính có thể hiểu được thông tin trên Web, các kết quả trả về sẽ đáp ứng nhu cầu ngày càng cao của người dùng. Hơn nữa tìm kiếm dựa trên Web ngữ nghĩa người dùng không cần phải nhớ chính xác từ khóa.

Bên cạnh đó, vẫn còn một vấn đề đáng ngại khi xây dựng hệ thống tìm kiếm với công nghệ này. Trong khi hầu hết tài nguyên trên Internet hiện nay đều có dạng phi cấu trúc mà việc tìm kiếm dựa trên Web ngữ nghĩa chỉ có thể thực hiện trên các tài liệu có cấu trúc. Do đó, khi muốn thực hiện tìm kiếm dựa trên Web ngữ nghĩa để tìm toàn bộ tài nguyên trên Internet là điều không mấy khả thi và chỉ có thể áp dụng các ứng dụng này trong phạm vi nhất định với các tài liệu tuân theo chuẩn của Web ngữ nghĩa.

## **CHƯƠNG 2**

### **GIẢI PHÁP ĐỀ XUẤT**

#### **2.1. VÀI NÉT VỀ VHKMNB**

##### **2.1.1. Đặc trưng của VHKMNB**

Qua quá trình cộng cư lâu đời cùng với các dân tộc Kinh - Hoa - Chăm... trên mảnh đất Nam Bộ, người Khmer đã có sự giao thoa văn hóa với các dân tộc anh em; nhưng cơ bản người Khmer vẫn giữ được nét văn hóa đặc sắc, những cốt cách tinh hoa của dân tộc mình. Thể hiện rõ nét nhất qua các ngôi chùa Khmer và sinh hoạt ở mỗi phum sóc, gắn liền với Phật giáo Nam tông Tiểu thừa, qua tiếng nói, chữ viết, các lễ hội truyền thống, các hình thức nghệ thuật như kiến trúc, điêu khắc, hội họa, âm nhạc, sân khấu, ca múa,...

##### **2.1.2. Nhu cầu và hiện trạng việc tra cứu VHKMNB**

Nhu cầu tra cứu VHKMNB là rất lớn. Hiện nay VHKMNB được bảo tồn và lưu giữ rải rác ở nhiều nơi, thông tin không chuẩn xác và hiện chưa có hệ thống quản lý, lưu trữ chuyên biệt.

Trong khi đó nhu cầu đặt ra cho việc tìm kiếm thông tin ngày càng cao, một số ít thông tin về VHKMNB được đăng tải trên các trang Web chỉ dừng lại ở việc cung cấp thông tin chứ chưa cung cấp được những tri thức cần thiết nên đã có những khó khăn nhất định trong việc tìm hiểu, tra cứu. Vì vậy việc nghiên cứu ứng dụng Web ngữ nghĩa để xây dựng hệ thống tra cứu VHKMNB là một việc làm thiết thực nhằm đáp ứng nhu cầu tìm kiếm chính xác và cung cấp những tri thức cần thiết về VHKMNB.

## 2.2. PHÂN TÍCH BÀI TOÁN

### 2.2.1. Giới thiệu bài toán

#### a. Yêu cầu bài toán

##### ❖ *Đối tượng sử dụng*

Đối tượng sử dụng là các nhà khoa học, sinh viên các trường đại học, cao đẳng, học sinh các trường phổ thông và du khách.

##### ❖ *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 VHKMNB đáp ứng các yêu cầu như: Cho phép người dùng nhập dữ liệu động, tùy chọn được hướng tiếp cận dữ liệu, người dùng có thể tùy chọn hướng truy xuất dữ liệu từ tên tỉnh thành trong khu vực hay tên của các loại hình văn hóa.

#### b. Phân tích vấn đề

Công nghệ Web Semantic với đặc điểm lưu trữ dữ liệu dưới định dạng XML nên tính tùy biến rất cao. Sự tùy biến này cho phép lưu trữ dữ liệu, thêm và gỡ bỏ các định dạng dữ liệu dễ dàng. Công nghệ Web Semantic sử dụng mô hình dữ liệu thông minh, lưu trữ dữ liệu dưới dạng thông tin mà máy có thể hiểu được. Điều đó không chỉ giúp việc tìm kiếm nhanh mà còn hỗ trợ sử dụng truy xuất thông tin chất lượng hơn.

Phát triển một trang Web semantic ta cần xây dựng ứng dụng gồm 2 phần chính đó là xây dựng Ontology và xây dựng trình duyệt thông minh để hiển thị kết quả.

##### ❖ *Xây dựng Ontology*

Trong phần này chúng ta xây dựng các lớp, các thuộc tính tạo những mối quan hệ đẳng 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.

### **2.2.2. Giải pháp**

#### ***a. Phát thảo kiến trúc tổng thể hệ thống***

Mô hình được đề xuất trong luận văn cho ứng dụng tìm kiếm ngữ nghĩa phải thể hiện được hai thành phần cơ bản của hệ thống là giao diện Web (Web Browser) và máy tìm kiếm (Search Engine).

#### ❖ *Giao diện Web (Web Browser)*

Giao diện Web đóng vai trò cầu nối giữa người dùng với hệ thống, nó thực hiện vai trò tiếp nhận câu truy vấn của người dùng và hiển thị kết quả câu truy vấn.

#### ❖ *Máy tìm kiếm (Search Engine)*

Đây là chức năng chính của chương trình thực hiện các thao tác:

- Tổ chức và lưu trữ Ontology.
- Thực hiện truy vấn yêu cầu của người dùng trên Ontology và trả về kết quả cho Web Browser theo yêu cầu của người dùng.

#### ***b. Giải quyết yêu cầu tìm kiếm***

Điểm mạnh của Web ngữ nghĩa so với Web 2.0 là cho phép khả năng suy luận dữ liệu thông minh nhằm hỗ trợ việc suy luận dữ liệu theo quy tắc của Ontology.

### 2.2.3. Các chức năng cơ bản của ứng dụng

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.

#### ❖ *Tìm kiếm theo 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. Tuy nhiên, nếu ta tìm kiếm với từ khóa như thông thường thì sẽ không có độ chính xác cao.

#### ❖ *Tìm kiếm nâng cao*

Các kết quả tìm kiếm với từ khóa thường có kết quả trả về quá nhiều khi dữ liệu lớn. Sử dụng cách tìm kiếm nâng cao sẽ có kết quả chính xác hơn. Với cách tìm kiếm này, hệ thống đưa ra một số tùy chọn làm điều kiện tìm kiếm. Người dùng dựa vào một số gợi ý tùy chọn đó để đưa ra điều kiện phù hợp với yêu cầu của mình.

## 2.3. QUY TRÌNH XÂY DỰNG WEB NGỮ NGHĨA

Bước 1: Xác định lĩnh vực và phạm vi của Ontology

Bước 2: Xem xét việc sử dụng lại các Ontology có sẵn

Bước 3: Liệt kê các thuật ngữ quan trọng

Bước 4: Xác định các lớp và phân cấp của các lớp



Bước 5: Xác định các thuộc tính

Bước 6: Xác định ràng buộc của các thuộc tính

Bước 7: Tạo các thực thể

## 2.4. CÁC CÔNG CỤ HỖ TRỢ

### 2.4.1. Công cụ xây dựng ứng dụng Protégé

Hiện nay có rất nhiều phần mềm hỗ trợ xây dựng Ontology. Một trong những phần mềm được sử dụng nhiều nhất là Protégé - OWL.

#### *a. Đặc điểm của Protégé*

Đây là phần mềm miễn phí dùng để tạo ra các mô hình và các ứng dụng bằng cách sử dụng các Ontology. 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.

#### *b. Protégé sử dụng giao diện đồ họa*

Ngay từ phiên bản Protégé - API, thì phần mềm Protégé đã không chỉ cho phép tạo mô hình bằng cách thủ công mà nó còn cho phép người sử dụng giao diện đồ họa để phát triển. Một trong những cơ sở của việc lập trình bằng giao diện đồ họa với Protégé là cơ chế tạo lớp, nó cho phép người lập trình giao tiếp rõ ràng với những sự kiện thay đổi.

#### *c. Protégé phát triển để tích hợp các công cụ*

Mục đích của Protégé là hỗ trợ người phát triển tạo ra được các Ontology một cách dễ dàng nhất. Ban đầu đó là vấn đề giao diện đồ họa, tiếp nữa đó là những công cụ thêm vào để tạo ra các chức năng đặc biệt khác.

### 2.4.2. Bộ Visual Studio.NET

Visual Studio.NET là một môi trường tích hợp triển khai phần mềm (Integrated Development Environment - IDE). Nó được thiết kế để lập ra một tiến trình viết mã, gỡ rối và biên dịch thành một Assembly

một cách dễ dàng. Visual Studio.NET cho chúng ta một ứng dụng Multiple - Document - Interface rất tinh vi, trong đó người sử dụng có thể liên kết mọi thứ để phát triển đoạn mã của mình.

### **2.4.3. Thư viện phát triển ứng dụng**

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

#### **a. SemWeb**

SemWeb có 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.

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

## **2.5. NHẬN XÉT**

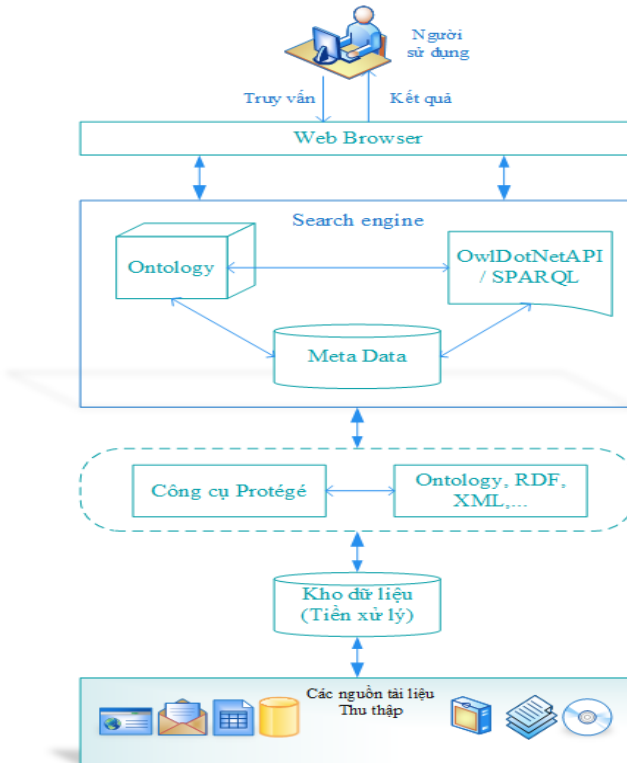
Qua những vấn đề đã trình bày trong chương 2 này cho chúng ta thấy được nhu cầu và hiện trạng của việc tra cứu thông tin về Văn hoá KMNB. Từ đó đã đề ra những giải pháp cụ thể góp phần quyết thực trạng đó. Tôi đã tiến hành phân tích bài toán tra cứu thông tin về văn hoá KMNB, phác thảo kiến trúc tổng thể của hệ thống cũng như quy trình xây dựng Web ngữ nghĩa. Tiến hành lựa chọn công cụ thích hợp để xây dựng và triển khai ứng dụng.

## CHƯƠNG 3

### XÂY DỰNG HỆ THỐNG

Nhằm cụ thể hoá những phân tích ở chương 2, trong chương 3 sẽ xây dựng Ontology cho bài toán và xây dựng hệ thống tra cứu thông tin về Văn hoá Khmer Nam bộ đồng thời mô tả hoạt động hệ thống. Hoàn thiện trang Web ngữ nghĩa hỗ trợ việc tra cứu, tiến hành thử nghiệm và đánh giá kết quả chương trình.

#### 3.1. MÔ HÌNH KIẾN TRÚC TỔNG THỂ CỦA HỆ THỐNG



Hình 3.1. Kiến trúc tổng thể của hệ thống

## 3.2. PHÁT TRIỂN ỨNG DỤNG

### 3.2.1. Quy trình phát triển ứng dụng

Để thiết kế công cụ tìm kiếm ngữ nghĩa cho bài toán này sau khi thu thập dữ liệu cần thực hiện các giai đoạn sau:

Giai đoạn 1: Xây dựng Ontology

Giai đoạn 2: Xây dựng giao diện

Giai đoạn 3: Thử nghiệm

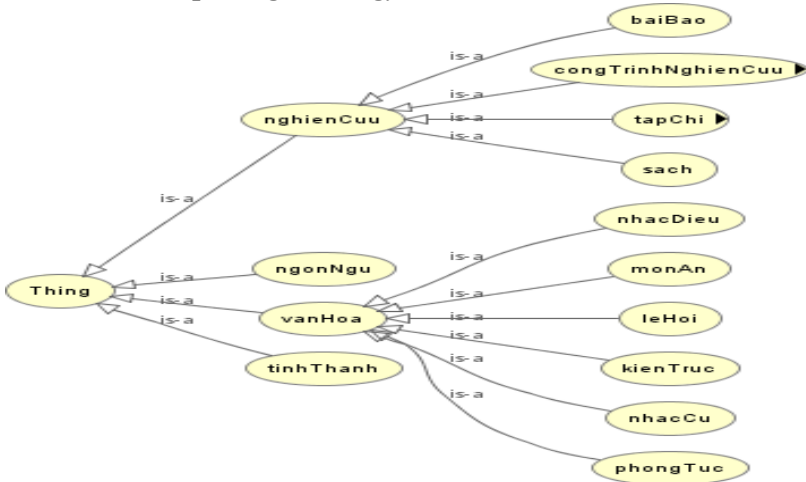
Giai đoạn 4: Đánh giá kết quả của chương trình.

### 3.2.2. Môi trường phát triển ứng dụng

Phần mềm Protégé 4.1 để thiết kế Ontology VHKMNB; Bộ Visual Studio 2010; Ngôn ngữ lập trình C# trên nền ASP.Net MVC; Ngôn ngữ truy vấn dữ liệu RDF: SPARQL; Thư viện mã nguồn OwlDotNetApi; Ứng dụng được chạy trên trình duyệt web.

### 3.2.3. Xây dựng Ontology

#### a. Các lớp trong Ontology



Hình 3.3. Các lớp trong ứng dụng

### ***b. Các thuộc tính trong Ontology***

Trong ứng dụng này ta xác định thuộc tính cho các lớp như sau: coDiaChi, coTieuDe, coHinhAnh, namXuatBan, CoNoiDung, laTacGiaCua, VeLeHoi, laBaiBaoCua, laNghienCuuCua, duaTrenNgonNgu, veNghienCuu, coTen, cuaTacGia, VeKientruc, VeTinhThanh, VeNhacCu, VePhongTuc, VeTonGiao, VeNhacDieu, VeMonAn, ...

### ***c. Các cá thể trong Ontology***

Đối với ứng dụng này, bước đầu tôi chỉ thu thập dữ liệu từ các nguồn khác nhau như các bài báo, các bài viết trên Internet, cơ sở dữ liệu về VHKMNB và chỉ đưa ra một vài cá thể tiêu biểu đại diện cho từng nhóm và đồng thời cũng thể hiện được các mối quan hệ giữa các đối tượng trong ứng dụng bao gồm: CaMau, CanTho, VinhLong, TraVinh, SocTrang, NN\_Viet, VH\_LeHoi, VH\_KT\_ChuaChien, VH\_NC\_NhacNguAm, VH\_ND\_Mua, NC\_Sach, NC\_BB\_0001, NC\_BB\_0002, ...

#### **3.2.4. Xây dựng giao diện**



*Hình 3.7. Giao diện chính của hệ thống*

### ***a. Thực hiện chức năng tìm kiếm***

Chức năng của tìm kiếm đơn giản là dựa vào từ khóa chỉ tên hoặc các thông tin về Văn hóa KMNB hiện có trong Ontology. 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 hệ thống cung cấp.

Đoạn code sau đây thực hiện chức năng tìm kiếm

```
[HttpPost]
public ActionResult SearchResult(FormCollection
collection){
    string strSearch =
collection["startsWith"].ToLower();
    string Category = collection["Category"];
    ViewBag.Message = string.IsNullOrEmpty(strSearch)
? "Từ khóa tìm kiếm" : strSearch;
    string file =
Path.Combine(Server.MapPath("~/App_Data"),
VHKMNB.owl");
    IOwlParser parser = new OwlXmlParser();
    IOwlGraph graph = parser.ParseOwl(file);
    var _list = CategoryRepository.SearchList(graph,
strSearch);
    return View(_list);
}
```

Kết quả của việc tìm kiếm đơn giản bằng từ khóa thường quá nhiều khi dữ liệu lớn. Để kết quả tìm kiếm được thỏa đáng hơn thì việc cung cấp thêm thông tin ngữ nghĩa cho quá trình tìm kiếm được chính xác rất cần thiết. Hệ thống hỗ trợ thêm chức năng tìm theo tính thành và

các thông tin liên quan như tác giả, ngôn ngữ,... Mục đích của việc tìm kiếm này cho phép người dùng có thể tùy chọn nhiều hơn một điều kiện dựa vào tính phân cấp trong Ontology và các thuộc tính quan hệ.

### ***b. Duyệt theo ngữ nghĩa***

Duyệt theo cây phân cấp của thông tin là quá trình đọc dữ liệu từ Ontology bằng cách duyệt theo cấp của các lớp để truy cập đến thông tin về một bài báo hay một nghiên cứu nào đó.

### **3.3. ĐÁNH GIÁ KẾT QUẢ CỦA HỆ THỐNG**

Luận văn sử dụng phần mềm Protégé xây dựng Ontology văn hóa KMNB và thống tra cứu văn hóa KMNB. Sau một thời gian thử nghiệm người sử dụng đã có những nhận xét thông qua phiếu đánh giá được tính trên tỉ lệ phần trăm như bảng 3.1.

*Bảng 3.1 Bảng đánh giá kết quả thử nghiệm hệ thống tra cứu*

Hệ thống Đối tượng	Số NSD	Đáp ứng nhu cầu			Hình thức	
		Tốt	Khá	Hạn chế	Phù hợp	Hạn chế
SV Lớp SPNVKM	20	70%	20%	10%	85%	15%
HS Trường PTDTNT	32	66%	25%	9%	87%	13%
Du Khách	14	72%	21%	7%	86%	14%
Sở VH TT&DL Tỉnh	17	65%	24%	11%	82%	18%
Ban Dân tộc Tỉnh	11	64%	27%	9%	82%	18%

Việc triển khai ứng dụng đã bước đầu ghi nhận được những kết quả khả quan. Hệ thống đã cơ bản giải quyết được vấn đề đặt ra, đã đáp ứng được nhu cầu tìm kiếm một cách nhanh chóng, chính xác, cung cấp một cách có hiệu quả những thông tin cần thiết về Văn hóa KMNB.



## KẾT LUẬN

### 1. Kết luận

Với lượng thông tin khổng lồ trên Internet, lượng người có nhu cầu tìm kiếm, tra cứu ngày một tăng đáng kể, công nghệ Web ngữ nghĩa ra đời đã tạo bước phát triển mới cho thế hệ Web hiện tại. Web ngữ nghĩa thật sự mang lại nhiều thuận lợi nhưng để thật sự hiểu rõ và nắm vững Web ngữ nghĩa là điều không dễ dàng. Trong quá trình thực hiện luận văn tôi đã tham khảo những thông tin trên các bài báo, sách, tạp chí và những nguồn khác trên Internet, cùng với sự hướng dẫn tận tình của Thầy Võ Trung Hùng và đã cơ bản hoàn thành những yêu cầu đặt ra.

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. Trong luận văn này tôi đã xây dựng thành công Ontology VHKMNB bằng phần mềm protégé 4.1, đã ứng dụng công nghệ Web ngữ nghĩa xây dựng hệ thống tra cứu thông tin về VHKMNB đạt được những kết quả ban đầu.

Về lý thuyết, luận văn đã nghiên cứu những nét đặc trưng về cơ sở lý thuyết của Web ngữ nghĩa. Luận văn còn đưa ra được những công cụ cần thiết để phát triển một ứng dụng Web ngữ nghĩa hiệu quả nhất. Đặc biệt, với xu hướng đưa các ứng dụng lên nền Web như hiện nay thì Web ngữ nghĩa ngày càng trở nên thực tiễn. Nghiên cứu về ngôn ngữ lập trình C#, sử dụng các gói thư viện hỗ trợ và nhúng vào chương trình. Nắm rõ được cách xây dựng, lưu trữ thông tin về một đối tượng trong Web ngữ nghĩa.

Về ứng dụng, phát triển hệ thống tra cứu thông tin về VHKMNB đã chứng minh được nền tảng lý thuyết nghiên cứu kết hợp giữa 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 công nghệ Web đã xây dựng trước đây.

Hệ thống tra cứu đã xây dựng chức năng cho phép người dùng nhập dữ liệu mới tại trang Web, đồng thời xem và tìm kiếm thông tin. Hệ thống cho phép truy xuất dữ liệu từ file và những tài nguyên có sẵn trên Internet nhằm cung cấp dữ liệu phong phú hơn.

Ứng dụng chỉ đưa ra những chức năng có tính chất minh họa cho phân lý thuyết, chưa có sự đầu tư nhiều về giao diện.

## **2. Hướng phát triển của đề tài**

Qua quá trình thực hiện đề tài nhận thấy vẫn còn nhiều vấn đề khác liên quan đến lĩnh vực này cần được quan tâm như:

Hệ thống chỉ hỗ trợ tra cứu thông tin về VHKMNB và sẽ là nền tảng để phát triển hệ thống tra cứu thông tin về các lĩnh vực .

Triển khai ứng dụng trên Server thực tế, cũng như môi trường Internet để hệ thống có điều kiện hỗ trợ tốt cho nhiều người dùng như mong muốn ban đầu.

Ứng dụng cần có một bộ từ điển từ đồng nghĩa phong phú về các lĩnh vực mà ứng dụng hỗ trợ giúp cho việc tìm kiếm theo ngữ nghĩa đạt kết quả tốt hơn.