

BỘ GIÁO DỤC VÀ ĐÀO TẠO
ĐẠI HỌC ĐÀ NẴNG

LÊ THỊ TUYẾT

ỨNG DỤNG SEMANTIC WEB
ĐỂ XÂY DỰNG HỆ THỐNG TRA CỨU
BIÊN ĐẢO VIỆT NAM

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. NGUYỄN THANH BÌNH

Phản biện 2: PGS.TS. TRƯƠNG CÔNG TUẤN

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 18 tháng 5 năm 2013.

Có thể tìm hiểu luận văn tại:

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

MỞ ĐẦU

1. Lý do chọn đề tài

Sự bùng nổ các thông tin trực tuyến trên Internet và World Wide Web tạo ra một lượng thông tin khổng lồ đưa ra thách thức là làm thế nào để có thể khai phá hết được lượng thông tin này một cách hiệu quả nhằm phục vụ đời sống con người. Sự ra đời của thế hệ web thứ ba – semantic web (hay web ngữ nghĩa) do W3C khởi xướng đã mở ra một bước tiến của công nghệ web giúp giải quyết được thách thức trên.

Một trang web thông thường được biểu diễn bằng HTML, các chương trình tìm kiếm hiện tại sẽ không hiểu được thông tin trên file HTML đó. Ví dụ ta có một thẻ trong trang html là `<b> Ông là Shakespeare </b>`. Một chương trình tìm kiếm thông thường đọc được đoạn đó, nó sẽ tách thành các từ: ông, là, Shakespeare, coi chúng là các từ khóa rồi đưa vào CSDL, mà không hiểu được nghĩa của mỗi từ và các thông tin liên quan.

Web ngữ nghĩa thì khác, đoạn văn bản trên sẽ được trình bày trong những mối liên hệ, như "Shakespeare" là thể hiện của một thuộc tính có tên là "tên", thuộc tính này nằm trong một lớp có tên là "Tác giả". Bằng những định nghĩa và thuộc tính này, thay vì phải tìm kiếm theo các từ khóa, các hệ thống tìm kiếm sẽ tìm trong không gian các mối liên hệ phụ thuộc, và sẽ cho ra những kết quả liên quan đến các định nghĩa, liên hệ đó. Như vậy, các hệ thống tìm kiếm sẽ cho ra kết quả chính xác hơn so với tìm kiếm dựa theo từ khóa.

Dựa trên nền tảng XML, các thông tin trong web ngữ nghĩa sẽ dễ dàng được trao đổi với nhau, bởi các hệ thống đều biết cách "đọc" các cấu trúc dữ liệu dựa trên XML. Các tầng phía trên của web ngữ nghĩa cho phép định nghĩa, chú giải các thẻ XML. Như vậy, ở mức cao hơn

XML, web ngữ nghĩa không những cho phép trao đổi thông tin mà nó còn cho phép máy tính có thể hiểu được một phần thông tin trong đó và tự động thực hiện một số công đoạn thay con người.

Dựa trên nền tảng lô-gic, các đối tượng, thuộc tính sẽ được trình bày bằng các luật lô-gic, điều này sẽ cho phép các chương trình suy luận, lập luận, và chứng minh dựa trên các biểu diễn bằng lô-gic trong web ngữ nghĩa. Đây chính là điều mà web hiện tại không thể làm được.

Việt Nam là một quốc gia ven biển có bờ biển dài trên 3.200 km, có các vùng biển và thềm lục địa khoảng một triệu km², gần 3.000 đảo nằm rải rác trên biển Đông từ Bắc chí Nam, bao gồm các đảo ven bờ và hai quần đảo Hoàng Sa, Trường Sa nằm giữa biển. Biển và đảo ngày càng có vai trò quan trọng về nhiều mặt kinh tế, quân sự, chính trị... Vì vậy, lịch sử phát triển của đất nước ta luôn gắn chặt với việc bảo vệ các vùng biển và hải đảo thuộc chủ quyền của đất nước.

Thời gian qua, sự biến đổi khôn lường của tình hình thế giới, khu vực và trên biển Đông khiến nhiệm vụ phòng thủ, bảo vệ đất nước, nhất là an ninh trên biển trở thành nhiệm vụ nhiều khó khăn, thách thức. Muốn bảo vệ vững chắc chủ quyền biển đảo thì mỗi người cần phải có kiến thức nhất định về biển đảo nước mình nhưng trên thực tế không phải ai cũng biết nhiều về biển đảo Việt Nam. Việt Nam có bao nhiêu hòn đảo, các tỉnh ven biển Việt Nam gồm những tỉnh nào chắc không phải học sinh, sinh viên nào cũng nắm được. Những kiến thức cơ bản về lãnh thổ được trang bị ở chương trình địa lý ở bậc phổ thông, thế nhưng để hiểu rõ về biển đảo Việt Nam thì sách giáo khoa ở bậc phổ thông chưa có điều kiện đề cập tới. Ngay cả sinh viên chuyên ngành Địa ở các trường Cao đẳng cũng chỉ được học 2 trình về biển đảo. Các trang web

về biển đảo chưa có nhiều, việc tìm kiếm thông tin về biển đảo gặp rất nhiều khó khăn.

Hiện nay Đảng và Nhà nước đã tổ chức rất nhiều cuộc thi, các hoạt động liên quan đến biển đảo như “Tuần lễ biển đảo Việt Nam”, “Thiên lương biển đảo quê hương”, “Vì biển đảo thân yêu”, hay “Trường Sa, Hoàng Sa là của Việt Nam”... Để hoà chung khí thế sôi động đó cũng như muốn đóng góp phần nhỏ giúp mọi người có thể tra cứu nhanh, thông minh hơn để có các thông tin quý báu về lịch sử cũng như các sự kiện liên quan về biển đảo Việt Nam, tôi đăng ký thực hiện đề tài **“ Ứng dụng Semantic web để xây dựng hệ thống tra cứu biển đảo Việt Nam”**.

2. Mục tiêu và nhiệm vụ

Đề tài: **“ Ứng dụng Semantic web để xây dựng hệ thống tra cứu biển đảo Việt Nam”** là sự kết hợp giữa việc nghiên cứu công nghệ web ngữ nghĩa với nhu cầu khai thác những thông tin về biển, đảo Việt Nam của người dùng. Mục tiêu của tôi đề ra là tổng quan được về web ngữ nghĩa và các công cụ liên quan, xây dựng kho Ontology về biển đảo Việt Nam và ứng dụng thành công web ngữ nghĩa vào một ứng dụng cụ thể trong thực tế đó là xây dựng website tìm kiếm thông minh giúp mọi người tìm thấy nhanh, đầy đủ nhất thông tin về biển đảo Việt Nam, nhằm nâng cao tình yêu biển đảo và tạo thế trận toàn dân bảo vệ Tổ quốc.

Với những mục tiêu đề ra thì ta phải nghiên cứu và thực hiện những nhiệm vụ sau: Về lý thuyết phải nghiên cứu về web ngữ nghĩa, các công cụ, thư viện hỗ trợ xây dựng web ngữ nghĩa. Tìm hiểu và chọn những công cụ để phát triển một cách hiệu quả nhất. Về mặt phát triển ứng dụng ta cần xây dựng một trang web đáp ứng yêu cầu tra cứu về biển

đảo Việt Nam. Chương trình này có tính chất minh họa cho lý thuyết nghiên cứu.

3. Đối tượng và phạm vi nghiên cứu

3.1. Đối tượng nghiên cứu:

Trong luận văn này, tôi tập trung nghiên cứu thông tin biển đảo Việt Nam, cơ sở lý thuyết về web ngữ nghĩa và các công cụ giúp xây dựng một trang web ngữ nghĩa.

3.2. Phạm vi nghiên cứu:

Phát triển trang web đồng thời cũng yêu cầu trích lọc và sử dụng lại dữ liệu đã có hiện nay ở các trang web 2.0. Tuy nhiên, giả thuyết rằng dữ liệu trích lọc đã có và chúng ta chỉ quan tâm đến việc làm thế nào để phân tích, xử lý dữ liệu.

Hệ thống biển đảo Việt Nam vô cùng phong phú, nhiều bãi biển, đảo, quần đảo, vịnh...tuy nhiên tôi chỉ tiến hành xây dựng hệ thống phục vụ nhu cầu tìm kiếm về các bãi biển chính và các đảo chính của nước ta.

4. Phương pháp nghiên cứu

- Phương pháp nghiên cứu lý luận: nghiên cứu các tài liệu liên quan đến web ngữ nghĩa, hệ thống biển đảo Việt Nam, tổng hợp các tài liệu, dữ liệu.
- Phương pháp khảo sát: Điều tra các số liệu, thông tin về biển đảo, tìm hiểu các hệ thống tra cứu biển đảo hiện có.
- Phương pháp thực nghiệm: thực nghiệm trên các công cụ hỗ trợ phát triển web ngữ nghĩa.

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

Ý nghĩa khoa học: tổng hợp, trình bày, phân tích những vấn đề liên quan đến web ngữ nghĩa và sẽ tạo ra một tài liệu khoa học đáng tin cậy

để tham khảo bởi những người muốn nghiên cứu và ứng dụng web ngữ nghĩa.

Ý nghĩa thực tiễn: đề xuất một qui trình và mô tả chi tiết các bước để phát triển một web ngữ nghĩa. Minh họa cụ thể qua ứng dụng web ngữ nghĩa xây dựng hệ thống tra cứu thông tin về biển đảo Việt Nam, có ý nghĩa hết sức thiết thực như cung cấp những thông tin chính thống về biển đảo Việt Nam để mọi người hiểu biết một cách cặn kẽ hơn. Lòng tự hào dân tộc sẽ làm cho người Việt Nam yêu quý vùng biển quê mình hơn. Thế hệ trẻ cũng được tuyên truyền, giáo dục và khơi gợi ý thức chủ quyền dân tộc về biển đảo của Tổ quốc.

6. Bố cục của luận văn

Chương 1: TỔNG QUAN VỀ WEB NGỮ NGHĨA. Tập trung vào nghiên cứu về web ngữ nghĩa từ khái niệm, ý nghĩa, điểm khác biệt so với web hiện tại, những ngôn ngữ và công cụ liên quan.

Chương 2 PHÂN TÍCH ỨNG DỤNG. Phân tích và thiết kế hệ thống, để làm được 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 chúng ta trong quá trình triển khai thành ứng dụng.

Chương 3: THIẾT KẾ VÀ CÀI ĐẶT ỨNG DỤNG. Nội dung của chương này tiến hành xây dựng chương trình dựa trên những phân tích và thiết kế hệ thống mà ta đã có, các kết quả đạt được sau khi chúng ta triển khai ứng dụng thành công.

Ngoài ra, để đánh giá lại kết quả đạt được trong quá trình nghiên cứu ta xây dựng phần tổng kết để xem xét những kết quả đã đạt được, những khúc mắc còn tồn đọng đồng thời xác định những mục tiêu và hướng phát triển để chất lượng đề tài này ngày càng nâng cao.

CHƯƠNG 1: TỔNG QUAN VỀ WEB NGŨ NGHĨA

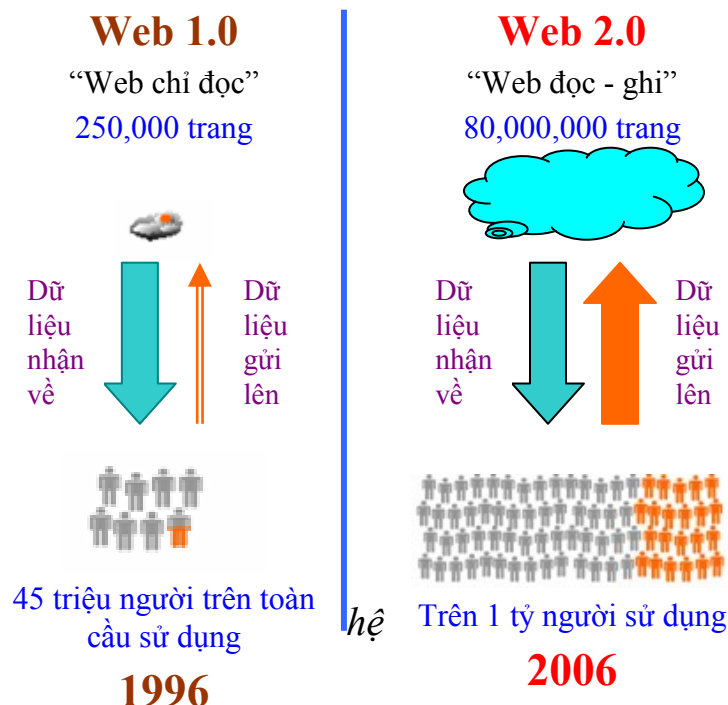
1.1. SEMANTIC WEB – WEB NGŨ NGHĨA

1.1.1. Thế hệ Web 1.0 và 2.0

Web 1.0 là thời kỳ của Netscape. Trong đó người dùng chỉ sử dụng thông tin. Thông tin trên website có nội dung tĩnh. Mức tương tác giữa những người cung cấp và người sử dụng thông tin bị hạn chế. Các nhà cung cấp thông tin tập trung và công khai thông tin là chính.

Web 2.0 là thời kỳ của Google và weblogs đóng vai trò quan trọng. Ở thời điểm này, người sử dụng thông tin đóng vai trò tích cực. Họ không chỉ sử dụng thông tin, mà còn thêm và chia sẻ thông tin. Trang HTML được tạo tự động. Web 2.0 không phải là một ứng dụng độc lập mà là ứng dụng chạy trên nền Web Server.

Dưới đây là hình ảnh minh họa web 1.0 và web 2.0.



1.1.2. Hạn chế web 2.0 và sự ra đời của Semantic Web (SW)

World Wide Web (gọi tắt là Web) đã trở thành một kho tàng thông tin khổng lồ của nhân loại và một môi trường chuyển tải thông tin không thể thiếu được trong thời đại công nghệ thông tin ngày nay.

Điều đó đã đặt ra thách thức làm sao để khai thác thông tin trên web một cách hiệu quả, mà cụ thể là làm thế nào để máy tính có thể trợ giúp xử lý tự động được chúng. Muốn vậy, web phải có khả năng mô tả các sự vật theo cách mà máy tính có thể “hiểu” được. Hầu hết các trang web được lấp đầy các thông tin và các tag liên quan. Các tag này chủ yếu thể hiện về mặt định dạng như <H1>... Một số trang có gắn các tag để tạo ngữ nghĩa như <META> tuy nhiên chúng thiếu gắn kết với nhau để cung cấp ngữ nghĩa cho trang web.

Chính những vấn đề này đã thúc đẩy sự ra đời của ý tưởng web ngữ nghĩa (Semantic Web), một thế hệ mới của web, mà chính cha đẻ của World Wide Web là Tim Berners-Lee đề xuất vào năm 1998.

1.1.3. Khái niệm web ngữ nghĩa.

Theo 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 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 (thông tin được tìm kiếm nhanh chóng và chính xác hơn), tích hợp dữ liệu (dữ liệu liên kết động), và tự động hóa các công việc.

1.1.4. Ích lợi web ngữ nghĩa

Máy có thể hiểu được thông tin trên web

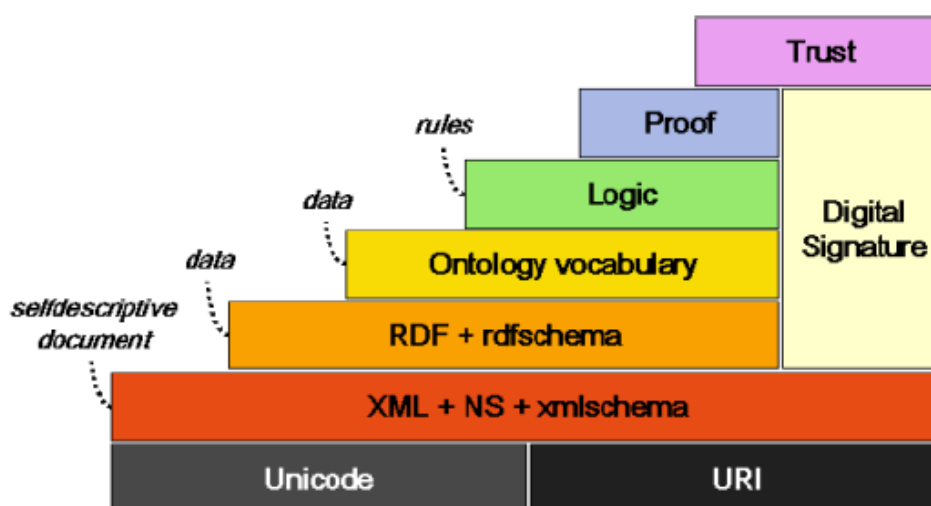
Thông tin được tìm kiếm nhanh chóng và chính xác hơn

Dữ liệu liên kết động

Hỗ trợ công cụ tự động hóa

1.1.5. Kiến trúc của web ngữ nghĩa

Web ngữ nghĩa là một tập hợp một chồng (stack) các ngôn ngữ. Tất cả các lớp của web ngữ nghĩa được sử dụng để đảm bảo độ an toàn và giá trị thông tin trở nên tốt nhất.



Hình 1.2. Kiến trúc của web ngữ nghĩa

Chúng ta sẽ tìm hiểu các lớp trong kiến trúc web ngữ nghĩa và vai trò của các lớp đó.

Lớp Unicode & URI

Lớp XML cùng với các định nghĩa về namespace và schema

Lớp RDF [RDF] và RDFSyntax [RDFS]

Đó là cấu trúc dữ liệu biểu diễn ngữ nghĩa. Được phát triển dựa trên kỹ thuật lưu trữ dữ liệu của XML và kiểu cấu trúc dữ liệu thông minh. RDF sau này được định nghĩa mở rộng nhằm tinh gọn dữ liệu và được định nghĩa bằng RDFs – RDF Schema.

Lớp Ontology

Ontology là cấu trúc dữ liệu biểu diễn ngữ nghĩa nâng cao. Được phát triển trên nền tảng RDF có phát triển thêm những định nghĩa về từ vựng ngữ nghĩa.

Lớp 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 trong khi cơ sở lập luận chủ yếu dựa vào lô-gic. Chính vì vậy mà các ontology được ánh xạ sang lô-gic, cụ thể là lô-gic mô tả để có thể hỗ trợ lập luận.

Lớp Proof

Lớp này đưa ra các luật để suy luận.

Lớp Trust

Để đảm bảo tính tin cậy của các ứng dụng trên web ngữ nghĩa.

1.1.6. URI – Bộ định danh tài nguyên duy nhất

URI - Uniform Resource Identifier, là một định danh web giống như các chuỗi bắt đầu bằng “http” hay “ftp”. Một dạng thức quen thuộc của URI là **URL - Uniform Resource Locator** hoặc **URN -Uniform Resource Name**.

1.1.7. Lập trình web ngữ nghĩa

Một ứng dụng SW gồm một số các thành phần riêng biệt, được chia thành 2 nhóm chính là các thành phần SW và công cụ SW liên quan.

1.1.8. 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)

1.1.9. Semantic Search Engine (Tìm kiếm ngữ nghĩa)

Vấn đề hiện nay là đa số các Search Engines hiện có đều thuộc loại Keyword Search Engine. Cơ chế của chúng là định kì duyệt web để phát hiện ra những sự thay đổi, rồi lập chỉ mục những thay đổi này. Tuy nhiên, phương pháp này gặp hai vấn đề chính sau đây:

- Một từ khóa có thể có một hay nhiều ý nghĩa tùy theo từng ngữ cảnh và Search Engine không thể hiện mối quan hệ giữa các từ khóa với nhau.
- Các trang web có cùng ý nghĩa với câu truy vấn của người sử dụng sẽ không tồn tại trong kết quả trả về.

Tìm kiếm ngữ nghĩa (semantic search): là tìm kiếm thông tin không dựa trên sự hiện diện của từ khóa hay cụm từ, mà dựa vào nghĩa của từ..

1.1.10. Một số thành tựu của web ngữ nghĩa trên thế giới và ở Việt Nam

- a. *ABS (Activity Based Search) tại ĐH Stanford*
- b. *Công cụ tìm kiếm Swoogle*
- c. *Công cụ tìm kiếm Kngine*
- d. *VN-KIM (Viet Nam Knowledge and Information*

Management)

1.2. CÔNG CỤ VÀ CÔNG NGHỆ LIÊN QUAN ĐẾN WEB NGỮ NGHĨA

1.2.1. Ontology

- a. *Khái niệm*

Ontology là một thuật ngữ có nguồn gốc từ Triết học diễn tả các thực thể tồn tại trong tự nhiên và các mối quan hệ giữa chúng. Ontology xuất hiện trong Công nghệ Thông tin trong lĩnh vực Trí Tuệ Nhân Tạo nhằm giải quyết vấn đề về chia sẻ và tái sử dụng tri thức.

“Ontology là một tập các khái niệm và quan hệ giữa các khái niệm được định nghĩa cho một lĩnh vực nào đó nhằm vào việc biểu diễn và trao đổi thông tin để con người và máy tính có thể hiểu được”.

b. Các thành phần của ontology.

Các thành phần chính của Ontology là: Lớp (Class), thuộc tính (Property), thực thể (Individual).

Lớp (class): là một bộ những thực thể, các thực thể được mô tả lô-gic để định nghĩa các đối tượng của lớp; lớp được xây dựng theo cấu trúc phân cấp cha con như là một sự phân loại các đối tượng. Ví dụ như lớp Person.

Thực thể (Individual): được xem là thể hiện của một lớp, làm rõ hơn về lớp đó và có thể được hiểu là một đối tượng nào đó trong tự nhiên (Thực thể của lớp Person là An, Hoa,...).

Thuộc tính (Property):

- thể hiện quan hệ nhị phân của các thực thể (quan hệ giữa hai thực thể) như liên kết hai thực thể với nhau. Ví dụ thuộc tính “Là cha” liên kết hai thực thể ‘An’ và ‘Hoa’ với nhau.
- Thuộc tính có 3 kiểu:
 - Object Property: Liên kết thực thể này với thực thể khác.
 - DataType Property: Liên kết thực thể với kiểu dữ liệu XML Schema, RDF literal

- Annotation Property: Thêm các thông tin metadata về lớp, thuộc tính hay thực thể khác thuộc 2 kiểu trên.

c. Phương pháp xây dựng Ontology

Quy trình phát triển Ontology phổ biến là quy trình phát triển gồm 7 bước do Stanford Center for Biomedical Informatics Research đưa ra.

d. Ngôn ngữ xây dựng Ontology

Hiện tại, các ngôn ngữ xây dựng ontology (ngôn ngữ ontology) điển hình bao gồm LOOM, LISP, Ontolingua, XML, SHOE, OIL, DAML+OIL và OWL.

e. Công cụ xây dựng Ontology

Bộ công cụ xây dựng và phát triển Ontology bao gồm các tool hỗ trợ và môi trường giúp người dùng có thể xây dựng một Ontology mới từ bản thiết kế mới hoặc sử dụng lại những Ontology có sẵn. Những bộ công cụ được sử dụng rộng rãi gần đây bao gồm OntoEdit, Oiled, webODE, OWL S-Editor và Protégé.

1.2.2. RDF – Nền tảng của web ngữ nghĩa

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 tuy nhiên XML chỉ mô tả dữ liệu, RDF còn có khả năng biểu diễn ngữ nghĩa giữa chúng thông qua các tài nguyên được định danh bằng URI.

1.2.3. Truy vấn trên dữ liệu web ngữ nghĩa

Các ngôn ngữ truy vấn có thể được phân nhóm thành ba dòng khác biệt theo các khía cạnh như mô hình dữ liệu, tính biểu trưng, hỗ trợ thông tin lướt đồ, và các kiểu truy vấn gồm SPARQL, “dòng RQL” và một số loại ngôn ngữ truy vấn RDF khác sử dụng các mô hình khác, như là sử dụng các luật, hoặc như các ngôn ngữ suy diễn như TRIPLE và Xcerpt.

CHƯƠNG 2: PHÂN TÍCH ỨNG DỤNG

Trong chương này tôi trình bày về nhu cầu tra cứu biển đảo, hiện trạng của các hệ thống tra cứu biển đảo hiện tại, từ đó đưa ra các yêu cầu của bài toán, phân tích và chọn giải pháp cũng như công nghệ để thực hiện bài toán.

2.1. BIỂN ĐẢO VIỆT NAM

2.1.1. Tổng quan

Việt Nam là một quốc gia ven biển có bờ biển dài trên 3.200 km, có các vùng biển và thềm lục địa khoảng một triệu km², gần 3.000 đảo nằm rải rác trên biển Đông từ Bắc chí Nam, bao gồm các đảo ven bờ và hai quần đảo Hoàng Sa, Trường Sa nằm giữa biển. Biển và đảo ngày càng có vai trò quan trọng về nhiều mặt kinh tế, quân sự, chính trị... Vì vậy, lịch sử phát triển của đất nước ta luôn gắn chặt với việc bảo vệ các vùng biển và hải đảo thuộc chủ quyền của đất nước.

2.1.2. Vai trò biển đảo

Về phát triển kinh tế

Về quốc phòng - an ninh

Về tư tưởng, văn hóa, giáo dục

2.2. NHU CẦU VIỆC TRA CỨU BIỂN ĐẢO

Hiện nay Đảng và Nhà nước đã tổ chức rất nhiều cuộc thi, các hoạt động liên quan đến biển đảo như “Tuần lễ biển đảo Việt Nam”, “Thiên lương biển đảo quê hương”, “Vì biển đảo thân yêu, hay “Trường Sa, Hoàng Sa là của Việt Nam”... Mục đích để mỗi người bổ sung thêm những kiến thức nhất định để bảo vệ biển đảo, chủ quyền của dân tộc. Kiến thức về biển đảo không được trang bị nhiều ở trường học, cũng

không có nhiều sách đề cập tới. Chính vì vậy nhu cầu tra cứu về biển đảo rất là lớn, nhất là trong tình hình hiện nay.

2.3. HIỆN TRẠNG HỆ THỐNG PHỤC VỤ TRA CỨU BIỂN ĐẢO

Để có cơ sở xây dựng hệ thống mới, tôi tiến hành khảo sát hiện trạng hệ thống phục vụ tra cứu hiện nay thì thấy việc tra cứu hiện tại gặp rất nhiều khó khăn đặc biệt là ở Tây Nguyên nơi tôi đang sinh sống. Hầu hết các tỉnh chưa có hệ thống quản lý biển đảo của tỉnh. Muốn tìm hiểu biển đảo ở tỉnh khác thì còn khó khăn hơn.

Có thể nói chưa có một hệ thống phục vụ tra cứu biển đảo Việt Nam chuyên nghiệp, 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.

2.4. MÔ TẢ HỆ THỐNG

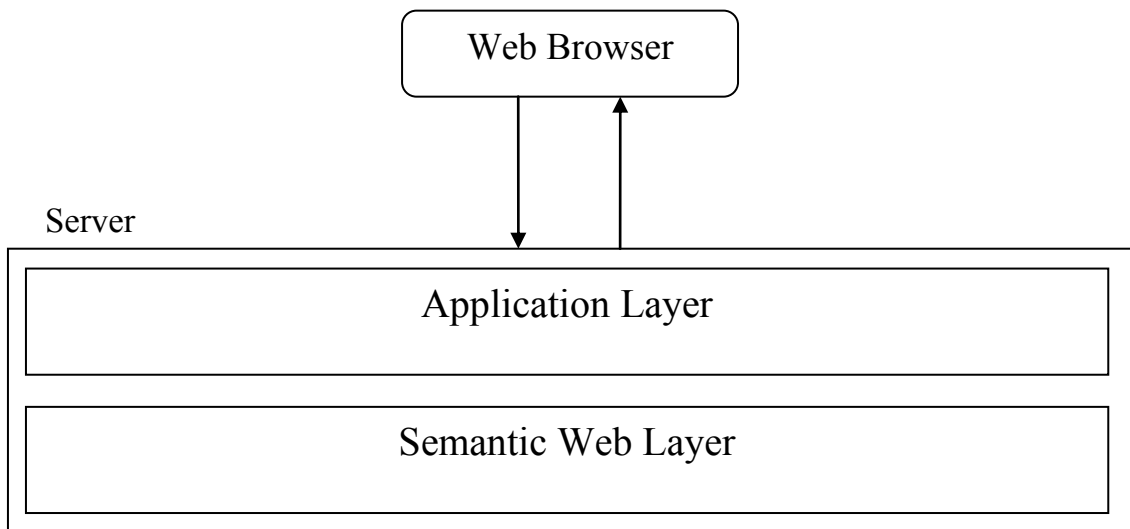
2.4.1. Yêu cầu đối với hệ thống

Khi xây dựng hệ thống, cần đáp ứng được các yêu cầu sau:

- Xây dựng hệ thống tra cứu thông tin về biển đảo Việt Nam. Hệ thống cũng lưu trữ thông tin như tài nguyên, di sản văn hóa liên quan tới biển đảo.
- Hệ thống xây dựng trên các công cụ mở, đảm bảo tính phổ dụng và có chất lượng tốt.
- Cung cấp SPARQL endpoints nhằm giúp các công cụ web ngữ nghĩa khác có thể tiếp cận nguồn tài nguyên.
- Tìm kiếm cơ bản và mở rộng. Trong đó việc tìm kiếm đảm bảo có suy luận ngược tương ứng với một số thuộc tính nghịch đảo, thuộc tính bắc cầu, ... trong các đối tượng.

2.4.2. Kiến trúc tổng thể

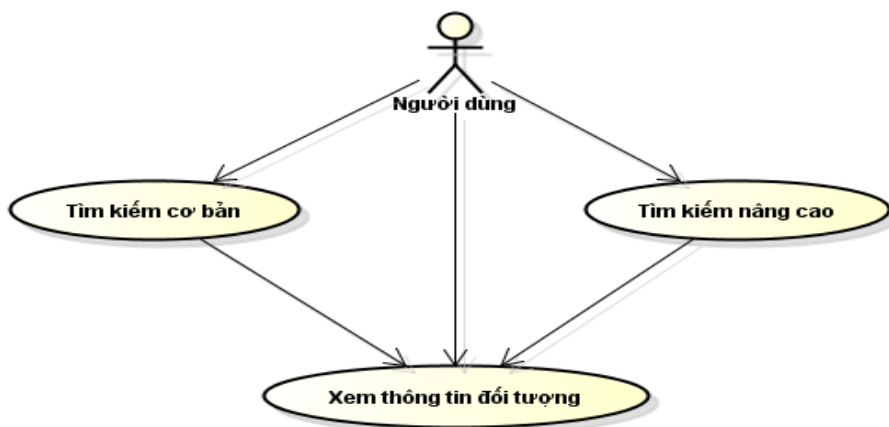
Theo yêu cầu bài toán và đặc điểm của web ngữ nghĩa, kiến trúc tổng thể của hệ thống cần xây dựng sẽ như mô hình sau. Web Browser: là trình duyệt phía người sử dụng, sẽ có các yêu cầu lên máy chủ (Server). Mô hình trên Server sẽ có hai tầng: tầng ứng dụng (Application Layer) và tầng web ngữ nghĩa (Semantic Web Layer).



2.5. BIỂU ĐỒ CA SỬ DỤNG CỦA ỨNG DỤNG

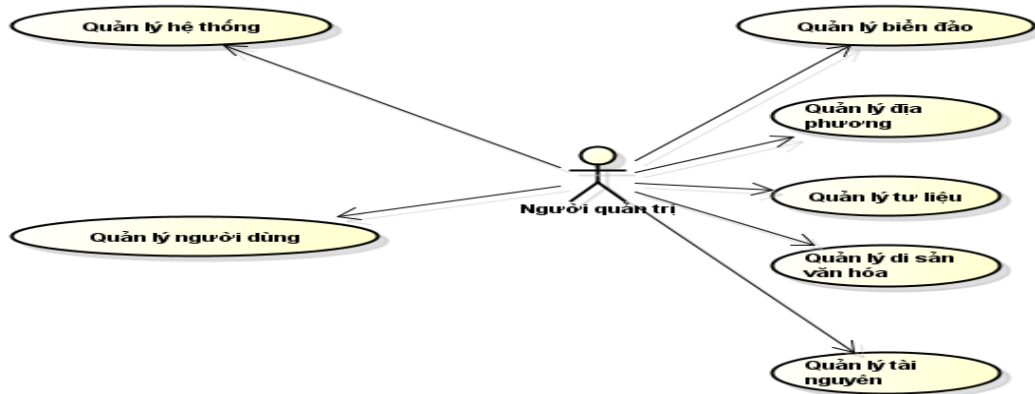
Biểu đồ ca sử dụng của ứng dụng sẽ chia làm hai đối tượng sử dụng chính: đó là người dùng và người quản trị.

Biểu đồ ca sử dụng của người dùng



Hình 2.2. Biểu đồ ca sử dụng của người dùng

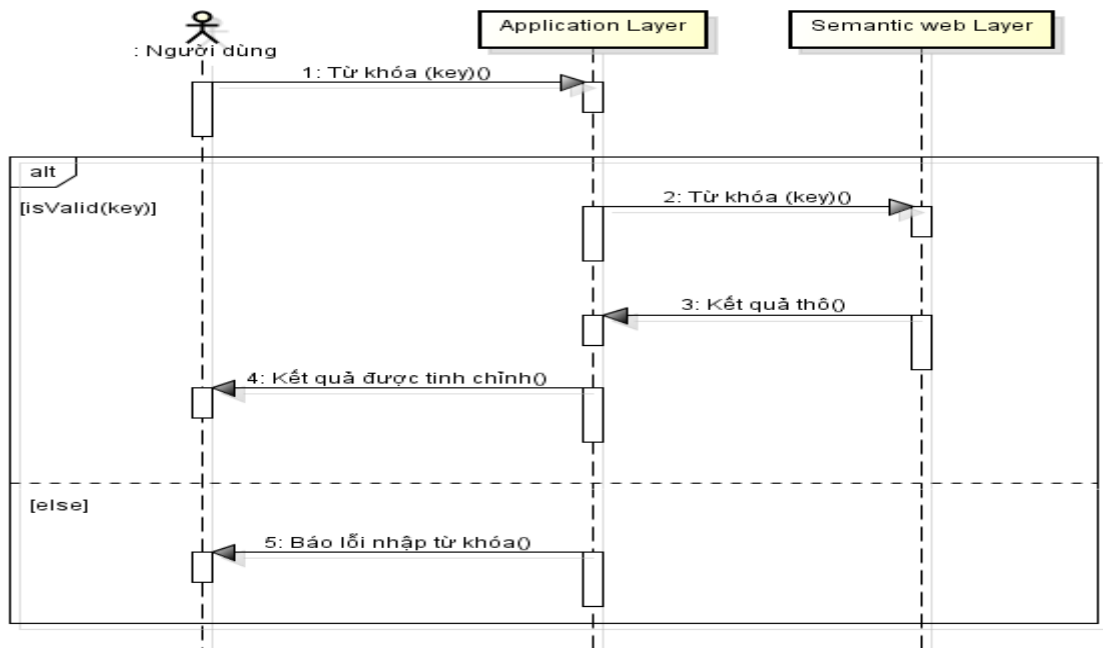
Biểu đồ ca sử dụng của người quản trị



Hình 2.3. Biểu đồ ca sử dụng của người quản trị

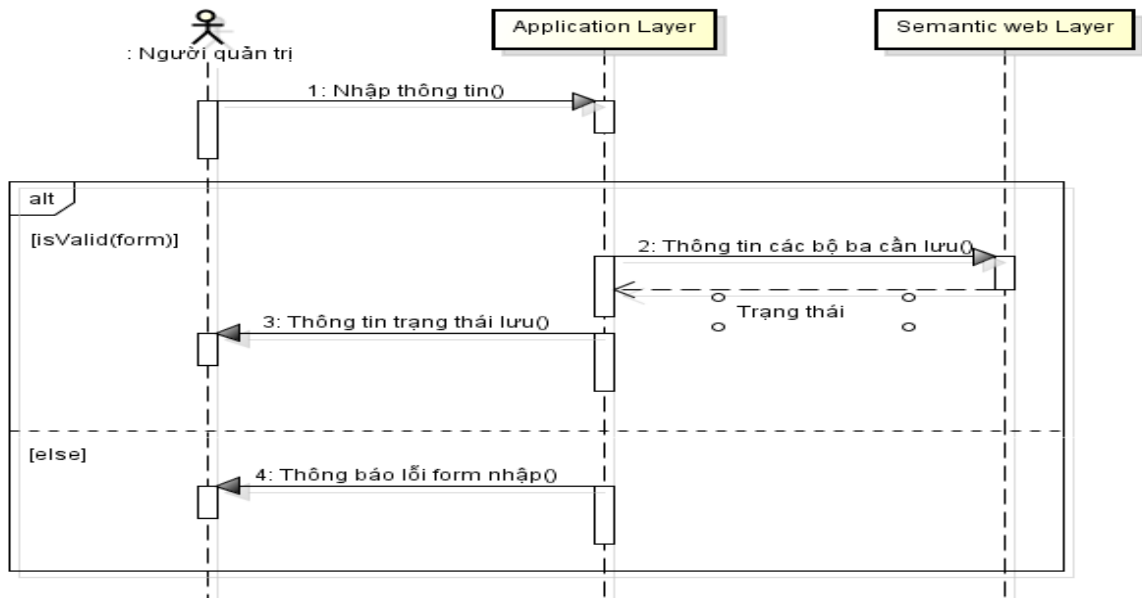
2.6. LƯỢC ĐỒ TUẦN TỰ CÁC CHỨC NĂNG CHÍNH CỦA ỨNG DỤNG

2.6.1. Chức năng tìm kiếm



Hình 2.4. Biểu đồ tuần tự chức năng tìm kiếm

2.6.2. Chức năng thêm một đối tượng (đảo, quần đảo...)



Hình 2.5. Biểu đồ tuần tự chức năng thêm đối tượng

2.7. CÁC CÔNG CỤ ĐỀ XUẤT

2.7.1. Zend Framework

Zend Framework là một nền tảng cho phép xây dựng website nhanh hơn và mạnh hơn, xây dựng trên ngôn ngữ lập trình PHP. Đây là một framework nguồn mở và có sự hỗ trợ lớn từ cộng đồng.

2.7.2. ARC2

ARC là một hệ thống web ngữ nghĩa linh hoạt sử dụng ngôn ngữ PHP. Đây là công cụ miễn phí, nguồn mở, dễ sử dụng và chạy được trên hầu hết các web server (PHP 5.3 trở lên).

Các đặc tính của ARC2:

- Dễ triển khai: sử dụng ngôn ngữ PHP và cơ sở dữ liệu MySQL phổ biến nên rất đơn giản cho việc cài đặt sử dụng.
- Hỗ trợ phân tích nhiều dạng: RDF/XML, N-Triples, Turtle, SPARQL + SPOG, Legacy XML, HTML tag soup, RSS 2.0, Google Social Graph API JSON...

CHƯƠNG 3: THIẾT KẾ VÀ CÀI ĐẶT ỨNG DỤNG

3.1. 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 : Thu thập dữ liệu về biển đảo Việt Nam
- Giai đoạn 2: Thiết kế Ontology

3.2. XÂY DỰNG KHO DỮ LIỆU

Dữ liệu về biển đảo chủ yếu được lấy từ các nguồn thông tin sau:

- http://vi.wikipedia.org/wiki/Danh_sách_đảo_ở_Việt_Nam
- <http://dangcongsan.vn/cpv/Modules/News/ListObjectNews.aspx?coid=10184>
- <http://www.biendao.org/>

3.3. XÂY DỰNG ONTOLOGY

Một số lớp quan trọng:

Lớp *Biển đảo*: là lớp cha của các lớp về các đối tượng biển đảo:

Biển, Bãi biển, Đảo, Quần đảo, Vịnh.

Lớp *Địa phương*: là lớp chỉ đối tượng là các địa phương. Trong đó, một địa phương này có thể thuộc một địa phương khác.

Lớp *di sản văn hóa*: lớp mô tả các di sản văn hóa. Các di sản có thể là vật thể hoặc phi vật thể.

Lớp *tài nguyên*: mô tả tài nguyên nào đó. Một biển, đảo hoặc quần đảo có thể có nhiều loại tài nguyên khác nhau: khoáng vật, du lịch, ...

Lớp *tư liệu*:

Một số thuộc tính dữ liệu có trong Ontology:

- Thuộc tính Thuộc Địa phương
- Thuộc tính Thuộc Quần đảo: Mô tả một Đảo thuộc một Quần đảo.
- Thuộc tính Thuộc Biển: Mô tả một Đảo thuộc một Biển.

- Thuộc tính Có Biển: mô tả một Địa phương có Biển.
- Thuộc tính Có Tư liệu: Mô tả một đối tượng biển đảo có tư liệu nào đó.

3.4. PHÁT TRIỂN HỆ THỐNG TRUY VẤN

Truy vấn lấy thông tin về một loại đối tượng, truy vấn đơn giản để lấy các đối tượng có thông tin liên quan tới từ khóa cần tìm, truy vấn của tìm kiếm nâng cao, chính xác tới các thông tin được người sử dụng đưa vào.

3.5. PHÁT TRIỂN HỆ THỐNG WEB

Hệ thống website được xây dựng trên Zend 1.12, một framework bằng ngôn ngữ PHP mạnh mẽ cho việc tạo các ứng dụng web hiện đại. Trong đó cài đặt thư viện ARC2 vào như là một công cụ cho web ngữ nghĩa.

3.6. KẾT QUẢ CHƯƠNG TRÌNH

Thêm đối tượng vào danh sách.

Quản lý Đảo
Thêm

Tên:

Thuộc khu vực:

Vĩ độ:

Kinh độ:

Có di sản văn hóa:

- ☐ Di tích lịch sử Nhà tù Côn Đảo
- ☐ Di sản thế giới: giá trị thẩm mỹ
- ☐ Di sản thế giới: giá trị địa chất

Có di tài nguyên:

- ☐ Du lịch
- ☐ Sinh học

Mô tả:

Hòn Tằm(đảo Thủy Kim Sơn) là một đảo rộng hơn 110ha nằm giữa vùng biển trong xanh bốn mùa đầy nắng và gió. Hòn Tằm cách thành phố Nha

Trở lại danh sách

Google Map Satellite

Laos, Thailand, Cambodia, Vietnam, Hainan, Macau, Hong Kong, Guang, Spr, Paracel Islands, South China

Hình 3.1. Giao diện chức năng thêm đảo

Tìm kiếm cơ bản:



IreSea

Tìm kiếm cơ bản

cù lao

Tìm

Tìm kiếm nâng cao

Số kết quả: 2 (0.21203804016113 giây)

► **Cù Lao Xanh (Đảo)**

Có tên **Cù Lao Xanh**

Mô tả Đảo **Cù Lao Xanh** (xã Nhơn Châu, TP Quy Nhơn - Bình Định) bao đời nay đẹp mà nghèo, thừa dân người trẻ. Công cuộc "làm mới" đảo để xốc lại đời sống vật chất, tinh thần người dân đảo, với những công trình mang tính căn cơ, dù quyết tâm nhưng cũng vô vàn cam go ...

► **Cù Lao Chàm (Đảo)**

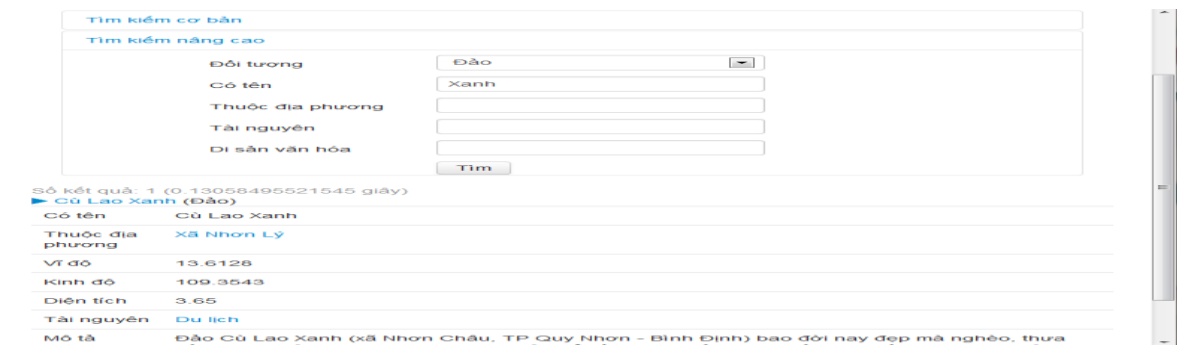
Có tên **Cù Lao Chàm**

Mô tả **Cù lao** Chàm là một cụm đảo, về mặt hành chính trực thuộc xã đảo Tân Hiệp, thành phố Hội An, tỉnh Quảng Nam, nằm cách bờ biển Cửa Đại 15 km và đã được UNESCO công nhận là khu dự trữ sinh quyển thế giới. **Cù lao** Chàm bao gồm 8 đảo: Hòn **Lao**, Hòn Dài, Hòn Mỏ, Hòn Khô me, Hòn Khô con, Hòn Lã, Hòn Tai, Hòn Ông. Dân số trên các hòn đảo này gồm khoảng 3.000 người.

Liên <http://www.culaocham.com.vn>

Hình 3.2. Giao diện chức năng tìm kiếm cơ bản

Tìm kiếm nâng cao: Ta đưa vào các điều kiện tìm kiếm chi tiết hơn khi đó kết quả trả về sẽ chính xác hơn như thuộc loại đối tượng nào, thuộc khu vực nào...



Tìm kiếm cơ bản

Tìm kiếm nâng cao

Đối tượng

Có tên

Thuộc địa phương

Tài nguyên

Di sản văn hóa

Tìm

Số kết quả: 1 (0.13058495521545 giây)

► **Cù Lao Xanh (Đảo)**

Có tên **Cù Lao Xanh**

Thuộc địa phương **Xã Nhơn Lý**

Vĩ độ 13.6128

Kinh độ 109.3543

Diện tích 3.65

Tài nguyên **Du lịch**

Mô tả Đảo **Cù Lao Xanh** (xã Nhơn Châu, TP Quy Nhơn - Bình Định) bao đời nay đẹp mà nghèo, thừa

Hình 3.3. Giao diện chức năng tìm kiếm nâng cao

3.7. ĐÁNH GIÁ

Dựa vào mô hình tổng quát của hệ thống và sản phẩm sau khi cài đặt, chúng ta có thể thấy một phần của hệ thống này tương tự như các web 2.0. Tuy nhiên, sự khác biệt ở đây hệ thống tra cứu thông tin biển đảo xây dựng dữ liệu trên nền tảng của web ngữ nghĩa.

KẾT LUẬN

Ngày nay với một lượng thông tin khổng lồ trên mạng, đồng thời số lượng người dùng truy cập trên mạng rất nhiều đã làm cho việc tìm kiếm, trình bày và duy trì thông tin trên web ngày càng trở nên khó khăn hơn. Đúng lúc này công nghệ web ngữ nghĩa ra đời đã thật sự đem lại niềm hy vọng, một bước phát triển mới cho thế hệ web hiện tại. Công nghệ web ngữ nghĩa đã giúp cho con người có thể thêm ngữ nghĩa vào các tài liệu bằng ngôn ngữ mà máy tính có thể hiểu được. Điều này làm cho máy tính hiểu được thông tin trên web, từ đó giúp cho việc tìm kiếm được nhanh hơn và chính xác hơn. Với công nghệ web ngữ nghĩa dữ liệu trên web đã được định nghĩa và liên kết theo cách mà máy tính có thể hiểu được không chỉ cho những mục đích hiển thị mà còn cho mục đích tự động, tích hợp và tái sử dụng dữ liệu qua các ứng dụng khác nhau. 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. Qua quá trình nghiên cứu, tìm hiểu và thực hiện đề tài, kết quả đề tài đã đạt được cơ bản các mục tiêu đề ra.

Kết quả về mặt nghiên cứu:

- Nắm được hạn chế của các thế hệ web trước và sự ra đời cũng như những thuận lợi mà web ngữ nghĩa mang lại.
- Nắm được kiến trúc web ngữ nghĩa và vai trò các lớp trong kiến trúc web ngữ nghĩa.
- Tìm hiểu về RDF, một ngôn ngữ nền tảng và giữ vai trò khá quan trọng trong kiến trúc web ngữ nghĩa bao gồm: lược đồ RDF dùng để mô tả các tài nguyên dưới dạng đồ thị, cú pháp RDF/XML dùng để tạo ra một tài liệu RDF, RDF Schema dùng

để định nghĩa các lớp và các thuộc tính.

- Tìm hiểu về Ontology, như khái niệm, các thành phần, ngôn ngữ và công cụ xây dựng ontology.
- Tìm hiểu về ngôn ngữ truy vấn trong RDF đó là ngôn ngữ SPARQL, cách tạo ra một câu truy vấn cũng như việc truy vấn dữ liệu trong tài liệu RDF.
- Hiểu được cách sử dụng thư viện ARC2, mô hình thiết kế web MVC (Model – View – Controller, framework ZEND và nhiều công cụ khác.

Kết quả về chương trình ứng dụng:

- Ứng dụng minh họa đã hoàn tất được mục tiêu của luận văn đề ra là nắm vững các kỹ thuật xây dựng ontology, lưu trữ và truy vấn trên dữ liệu RDF bằng cách sử dụng thư viện ARC2 trên nền Zend và ngôn ngữ PHP.
- Đóng góp thêm vào kho ontology một ontology đầy đủ về biển, đảo, vịnh, bãi biển. Sau đó thiết lập Setup EndPoint để các ứng dụng bên ngoài có thể sử dụng ontology đó.
- Qua việc thực hiện đề tài tôi cũng đã nắm được cách sử dụng một ontology có sẵn vào ứng dụng của mình như geo/wgs84-post là hệ tọa độ toàn cầu của Mỹ, hệ tọa độ mà các thiết bị GPS hiện nay đều sử dụng.
- Về mặt chương trình kết quả thu được là hai phân hệ chương trình: một phân hệ dành cho người sử dụng và một phân hệ dành cho người quản trị. Đối phân hệ dành cho người dùng có hỗ trợ việc tìm kiếm thông tin của biển đảo theo tìm kiếm cơ bản và nâng cao. Đối với phân hệ dành cho người quản trị hỗ trợ việc cập nhật thông tin về các đối tượng trong hệ thống như thêm,

sửa, xóa.

Hướng phát triển

Luận văn đã được khép lại với những kết quả đạt được theo yêu cầu của đề tài đưa ra. Tuy nhiên qua quá trình thực hiện đề tài tôi nhận thấy vẫn còn nhiều vấn đề khác liên quan đến lĩnh vực này cần được quan tâm. Do vậy những vấn đề này tôi xem như hướng phát triển của đề tài để nghiên cứu tiếp sau này.

- Đối với phân lý thuyết: Còn phải nghiên cứu sâu hơn về web ngữ nghĩa, đặc biệt là một số tầng trong kiến trúc web ngữ nghĩa (như tầng Lô-gic hay là Proof), nghiên cứu các luật, các giải thuật, trích lọc dữ liệu từ web hiện tại từ đó mới có thể mở rộng ứng dụng thông qua việc xử lý ngôn ngữ tự nhiên để đưa tính ngữ nghĩa (semantic) vào trong ứng dụng của mình.
- Đối với ứng dụng đã xây dựng: Vì dữ liệu của ứng dụng chỉ có tính chất minh họa nên để đạt được những kết quả mà công nghệ Semantic mang lại cần tổ chức dữ liệu phong phú hơn. Cần phải đi sâu vào các thuật toán để xây dựng ứng dụng thành bộ máy tìm kiếm thông minh hơn.