

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



Phan Hồng Dương

**ỨNG DỤNG CÔNG NGHỆ WEB NGỮ NGHĨA VÀ
OPENSTREETMAP PHÁT TRIỂN ỨNG DỤNG
HỖ TRỢ TÌM KIẾM THÔNG TIN NGƯỜI BỆNH COVID-19**

Chuyên ngành: Hệ thống thông tin

Mã số: 8.48.01.04

TÓM TẮT LUẬN VĂN THẠC SĨ

HÀ NỘI - 2021

Luận văn được hoàn thành tại:

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Người hướng dẫn khoa học: PGS. TS. Hoàng Hữu Hạnh

Phản biện 1: TS. Nguyễn Vĩnh An

Phản biện 2: PGS. TS. Phạm Văn Cường

Luận văn sẽ được bảo vệ trước Hội đồng chấm luận văn thạc sĩ tại Học viện Công nghệ
Bưu chính Viễn thông

Vào lúc: 9 giờ 00 ngày 30 tháng 08 năm 2021

Có thể tìm hiểu luận văn tại:

- Thư viện của Học viện Công nghệ Bưu chính Viễn thông.

PHẦN 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.

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.

OpenStreetMap (viết tắt OSM) là một dịch vụ bản đồ thế giới trực tuyến có nội dung mở. OpenStreetMap nhằm mục đích cung cấp dữ liệu địa lý do nhiều người cùng cộng tác với nhau trên hệ thống wiki. Nó thường được gọi "Wikipedia của bản đồ". Dự án OpenStreetMap được sáng lập năm 2004, chủ yếu để cạnh tranh với các công ty và cơ quan chính phủ cung cấp dữ liệu địa lý theo các điều khoản sử dụng được coi là quá chặt chẽ.

Đại dịch COVID-19 hiện nay vẫn đang diễn biến nhanh chóng với số ca nhiễm mới không ngừng gia tăng hàng ngày. Một câu hỏi đặt ra là việc truy xuất các thông tin về các ca nhiễm, nghi nhiễm và tìm kiếm các thông tin liên quan trên nền Web ngữ nghĩa. Đề tài này hướng được việc ứng dụng công nghệ Web ngữ nghĩa trong

việc phát triển các ứng dụng Web ngữ nghĩa và hệ thống dữ liệu bản đồ mở - OpenStreetMap nhằm kết nối các bộ dữ liệu để hỗ trợ hiệu quả việc tìm kiếm thông tin dịch COVID-19 và các ca nhiễm, nghi nhiễm tại Việt Nam và đặc biệt tại Hà Nội và vùng phụ cận, cung cấp thêm các thông tin liên quan đến COVID-19 để cho người sử dụng hệ thống. Từ những vấn đề nêu trên, tôi chọn đề tài luận văn là: “Ứng dụng công nghệ Web ngữ nghĩa và OpenStreetMap phát triển ứng dụng hỗ trợ tìm kiếm thông tin COVID-19”.

2. Tình hình nghiên cứu và tính mới của đề tài

Các ứng dụng của Web ngữ nghĩa trong các lĩnh vực khác nhau đã thu hút nhiều nhóm nghiên cứu. Do đó, các ứng dụng được thiết kế dựa trên các khái niệm đã đề cập và sử dụng các thông tin có thể xử lý được bởi máy tính để tạo ra động lực lớn cho việc phát triển của một thể hệ các công cụ và các ứng dụng Web mới.

Trong suốt vài năm trở lại đây, các công nghệ liên quan đến Web ngữ nghĩa đã xuất hiện hoặc đã được hiệu chỉnh nhiều. Một trong những phần quan trọng của những tiến triển này là các ngôn ngữ phát triển ontology. Tổ chức W3C đã làm việc tích cực để chuẩn hoá các chuẩn và đã phê chuẩn RDF và ngôn ngữ ontology OWL là hai chuẩn quan trọng nhất trong số các công nghệ Web ngữ nghĩa. Điều này đã cung cấp một cơ sở vững chắc để triển khai các ứng dụng Web ngữ nghĩa trên diện rộng, cũng như làm đòn bẩy cho việc đưa Web ngữ nghĩa từ nghiên cứu sang ứng dụng trong công nghiệp cũng như các chuẩn cho nó để xây dựng các thể hệ ứng dụng mới.

Điểm mới của luận văn: sử dụng OpenStreetMap – hệ thống bản đồ mở dùng để làm dữ liệu cho việc xây dựng ứng dụng Web ngữ nghĩa trong tìm kiếm thông tin COVID-19 như nêu trên.

3. Mục đích nghiên cứu

Mục đích chung: nghiên cứu về công nghệ Web ngữ nghĩa cho việc phát triển các ứng dụng Web dữ liệu (Web of Data), hệ thống OpenStreetMap (OSM) và ứng dụng của nó trong tìm kiếm thông tin COVID-19.

Các mục đích cụ thể:

+ Về mặt lý thuyết: nghiên cứu các công cụ, thư viện hỗ trợ xây dựng ứng dụng Web ngữ nghĩa. Cách thức trích xuất dữ liệu từ OpenStreetMap để làm dữ liệu cho ứng dụng Web ngữ nghĩa. Tìm hiểu và chọn những công cụ để phát triển một cách hiệu quả nhất.

+ Về mặt ứng dụng: cần xây dựng một trang Web (dữ liệu) đáp ứng yêu cầu tra cứu về thông tin COVID-19 tại Việt Nam, và thông tin liên quan về tình hình COVID-19 trên toàn thế giới hiện nay, trên nền tảng OSM.

4. Đối tượng nghiên cứu

Trong luận văn này, tôi tập trung nghiên cứu:

- Cơ sở lý thuyết về Web ngữ nghĩa, các công nghệ mô hình hoá và công cụ lập trình ứng dụng Web ngữ nghĩa;

- OpenStreetMap và các công cụ giúp xây dựng một trang Web ngữ nghĩa. Đồng thời tìm kiếm các nguồn dữ liệu cho thông tin COVID-19 tại Việt Nam và trên thế giới (các điểm nóng).

5. Phương pháp nghiên cứu

Phương pháp nghiên cứu lý luận: nghiên cứu các tài liệu liên quan đến Web ngữ nghĩa, OpenStreetMap, tổng hợp các tài liệu, dữ liệu không gian và thời gian liên quan đến COVID-19 tại Việt Nam và trên thế giới.

Phương pháp khảo sát: Thu thập các số liệu hiện nay về COVID-19 của Việt Nam và thế giới, tìm hiểu các hệ thống tra cứu COVID-19 hiện nay để đối chiếu, so sánh.

Phương pháp thực nghiệm: Thực nghiệm trên các công cụ hỗ trợ phát triển Web ngữ nghĩa.

6. Nội dung của luận văn

Chương 1: Kiến thức cơ sở.

Chương 2: Lập trình với Web ngữ nghĩa.

Chương 3: Ứng dụng tìm kiếm thông tin bệnh nhân COVID-19 tại Việt Nam.

CHƯƠNG 1: KIẾN THỨC CƠ SỞ

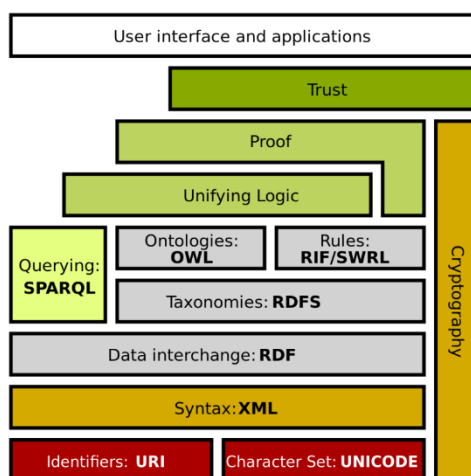
1.1 Web ngữ nghĩa

1.1.1 Khái niệm về Web ngữ nghĩa

Web ngữ nghĩa không là một phiên bản Web riêng biệt mà là một sự mở rộng của Web hiện tại, theo cách mà các thông tin được xác định ý nghĩa tốt hơn, cho phép máy tính và người cộng tác với nhau một cách tốt hơn. Web ngữ nghĩa được hình thành từ ý tưởng của Tim Berners-Lee, người phát minh ra WWW (World Wide Web), URI (Uniform Resource Identification), HTTP, và HTML. Web ngữ nghĩa là mạng lưới các nguồn thông tin được liên kết sao cho chúng có thể được xử lý và tìm kiếm dễ dàng bởi các máy tính ở phạm vi toàn cầu. Web ngữ nghĩa được xem như là một cách mô tả thông tin rất hiệu quả trên World Wide Web, và cũng được xem là một cơ sở dữ liệu có khả năng liên kết với mạng lưới toàn cầu. Web ngữ nghĩa là phương pháp cho phép người dùng định nghĩa và liên kết dữ liệu một cách có ngữ nghĩa hơn nhằm phục vụ cho máy tính có thể “hiểu” và tìm kiếm được. Ngoài ra Web ngữ nghĩa còn cung cấp môi trường được chia sẻ và xử lý dữ liệu tự động.

1.1.2 Kiến trúc Web ngữ nghĩa

Web ngữ nghĩa là tập hợp của các ngôn ngữ. Các lớp Web ngữ nghĩa sử dụng để có thể đảm bảo độ an toàn và các giá trị của thông tin được trở nên tốt nhất.



Kiến trúc Web ngữ nghĩa

- *Lớp Unicode & URI.*
- *Lớp XML cùng với những định nghĩa về không gian tên (namespace) và lược đồ (schema).*
- *Lớp RDF và RDFS.*
- *Lớp Ontology.*
- *Lớp Digital Signature.*
- *Các lớp Logic, Proof, Trust:.*

1.2 RDF, RDFS

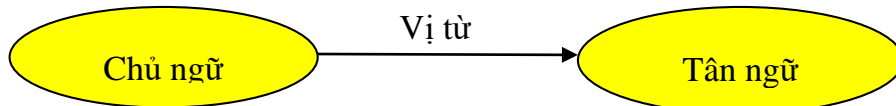
1.2.1 Giới thiệu RDF

RDF (Resource Description Framework) hay khung mô tả tài nguyên, là nền tảng cho việc biểu diễn các dữ liệu trong lĩnh vực Web có ngữ nghĩa. RDF là một tập hợp nhiều nguyên tắc dành cho ngôn ngữ đánh dấu, cung cấp mô hình dữ liệu và các cú pháp đơn giản với các hệ thống độc lập. Đồng thời, nó được thiết kế để có thể đọc được thông tin và hệ thống máy tính có thể hiểu được, chứ không đơn giản là chỉ để trình bày dữ liệu cho người dùng cuối. Biểu diễn thông tin theo mô hình RDF gồm một phát biểu (statement) ở dạng cấu trúc bộ ba (triple) với thành phần chính là: (Subject, Predicate, Object). Trong đó:

- **Subject** miêu tả đối tượng đóng vai trò là chủ thể.
- **Predicate** (property) là thuộc tính hoặc quan hệ của chủ thể.
- **Object** là giá trị đối tượng hay thuộc tính của chủ thể đã nêu. Object có thể là một giá trị nguyên thủy như số chuỗi ký tự, số nguyên,.. hoặc cũng có thể là một tài nguyên nhất định.

1.2.1.1 Đồ thị RDF

Đồ thị RDF là tập hợp của các bộ ba hợp lại tạo thành. Các chủ ngữ các là các nút trong đồ thị, trong cung đồ thị là vị từ và tân ngữ trong bộ ba. Bộ ba được mô tả dạng nút - cung - nút. Hướng của các đồ thị rất quan trọng và luôn bắt đầu từ chủ ngữ qua vị từ đến tân ngữ.



Mối quan hệ giữa các thành phần trong bộ ba RDF

1.2.1.2 Cú pháp RDF/XML

Để biểu diễn những phát biểu RDF trong một cách thức để máy có khả năng xử lý được, RDF sử dụng rất nhiều ngôn ngữ trong đó có ngôn ngữ XML. XML (Extensible Markup Language) là ngôn ngữ đánh dấu mở rộng. RDF định nghĩa ra một loại ngôn ngữ đánh dấu XML đặc biệt, được biết giống như RDF/XML, để sử dụng trong việc biểu diễn các thông tin của bộ ba trong RDF. Như phần trước đã giới thiệu một đồ thị được xem là một tập những đường đi của các nút, cung vị từ, nút, cung vị từ, nút, cung vị từ,... chúng bao bọc toàn đồ thị. Trong RDF/XML điều này trở thành sự nối tiếp liên tục của các phần tử bên trong các phần tử mà xen kẽ giữa các phần tử đối với những nút và cung vị từ. Điều này được gọi là một loạt những nút/cung. Nút ở vị trí bắt đầu của chuỗi nối tiếp trở thành phần tử ở ngoài cùng, cung vị từ tiếp theo trở thành một phần tử con.

1.2.2 Giới thiệu về RDFS

RDF sử dụng XML để có thể trao đổi những mô tả tài nguyên web nhưng những tài nguyên mô tả có thể thuộc nhiều loại khác nhau, bao gồm những tài nguyên XML và không XML. RDF dựa vào những tiện ích để cho phép các tài nguyên web được xử lý tự động. RDF được dùng trong nhiều ứng dụng khác nhau, ví dụ: trong khám phá tài nguyên để giúp cho việc tìm kiếm cho kết quả tốt hơn, hay trong việc Các lớp và các thuộc tính RDF chia thành loại để mô tả nội dung và những mối quan

hệ sẵn có tại một trang web đặc biệt bằng các phần mềm thông minh để làm cho việc chia sẻ và trao đổi tri thức được dễ dàng hơn. Trong khi RDF là ngôn ngữ không thiết lập các giả định về bất kỳ lĩnh vực ứng dụng cụ thể nào, cũng không định nghĩa ngữ nghĩa của lĩnh vực chuyên biệt nào thì RDFS cung cấp phương pháp xây dựng một mô hình đối tượng từ dữ liệu thực tế và cho chúng ta ngữ nghĩa của dữ liệu.

1.3 Các ngôn ngữ truy vấn của Web ngữ nghĩa

1.3.1 Simple Protocol And RDF Query Language (SPARQL)

Như chúng ta đã biết RDF là một cách để mô tả thông tin về các tài nguyên Web một cách linh động. Nó được sử dụng để mô tả nhiều loại thông tin ví dụ như thông tin về cá nhân, về hệ thống mạng toàn cầu, các siêu dữ liệu (meta-data) về đa phương tiện như âm thanh và hình ảnh cũng như là cung cấp một phương tiện để tích hợp các nguồn tài nguyên khác. Với một lượng thông tin quá lớn như vậy, để dữ liệu có thể được truy vấn nhanh chóng và chính xác, 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 nhà phát triển web và người dùng cuối cách để viết và thực thi một câu truy vấn. Phần này mô tả ngôn ngữ truy vấn Protocol And RDF Query Language (SPARQL) để truy cập dữ liệu RDF được phát triển bởi nhóm RDF Data Access Working Group.

1.3.1.1 Viết một truy vấn đơn giản

1.3.1.2. Cú pháp các biến.

1.4 LINKED DATA

1.4.1 Khái niệm

Trong các hoạt động tính toán của máy tính, Linked data được mô tả như phương thức tạo ra dữ liệu có cấu trúc để có thể liên kết được với nhau và trở nên có ích. Dữ liệu Liên kết (Linked Data) là một trong những khái niệm và trụ cột chính của Web ngữ nghĩa (Semantic Web), còn được biết tới như là Web của Dữ liệu. Web

ngữ nghĩa tất cả là về các liên kết giữa các tập hợp dữ liệu có thể hiểu được không chỉ cho con người mà còn cho máy tính, và Dữ liệu Liên kết đưa ra các thực hành tốt nhất để tạo ra các liên kết đó. Dữ liệu Liên kết là tập hợp các nguyên tắc thiết kế cho việc chia sẻ dữ liệu được liên kết với nhau trên Web mà máy tính đọc được.

1.4.2 Các quy tắc triển khai Linked Data

Tim Berners-Lee liệt kê bốn quy tắc triển khai Linked data trong các thảo luận của mình được tóm gọn trong các dòng sau:

1. Dùng URI có thể định danh các tài nguyên.
2. Sử dụng HTTP URI để các tài nguyên được tham chiếu và tìm kiếm chúng ("tham chiếu lại") bởi mọi người hay các ứng dụng.
3. Cung cấp các thông tin hữu ích về tài nguyên khi các URI của nó được tham chiếu lại, sử dụng định dạng tiêu chuẩn như RDF/XML.
4. Gồm các liên kết đến liên kết khác, các URL có thể có quan hệ bên trong dữ liệu tìm thấy để đẩy mạnh khả năng tìm kiếm các thông tin liên quan khác trên Web.

1.4.3 Truy vấn Web ngữ nghĩa

Những đồ thị RDF được truy xuất thông qua URI trên Web ngữ nghĩa là một phần của một tập dữ liệu liên kết. Tập dữ liệu liên kết này là một đồ thị RDF chứa dữ liệu về nhiều thực thể hay đối tượng. Để tạo ra tập dữ liệu liên kết này, chúng ta có thể dùng một số cách phổ biến như: truy vấn dữ liệu dựa vào giao diện Linked Data, trên các kho lưu trữ trực tuyến hay trên các giao diện Web. Để xuất bản tập dữ liệu liên kết theo công thức Linked Data, người ta đưa ra khái niệm kết xuất RDF hay một điểm đầu cuối SPARQL để truy vấn vào tập dữ liệu liên kết này. Một kết xuất RDF là một tài liệu RDF lớn chứa đồ thị RDF cấu thành tập dữ liệu liên kết. Một điểm đầu cuối SPARQL là một dịch vụ truy vấn dựa trên giao thức HTTP, xử lý các truy vấn SPARQL trên tập dữ liệu liên kết. SPARQL – ngôn ngữ truy vấn cho đồ thị RDF, dựa trên các mẫu đồ thị RDF và đối sánh đồ thị con.

Các tiếp cận về xử lý truy vấn có thể kể đến như xếp kho dữ liệu (Data Warehousing), các máy tìm kiếm (Search Engines), truy vấn liên hiệp (Query Federation), truy vấn liên hiệp khám phá (Discovery Query Federation) hay quét liên kết (Linh traversal), dựa trên một số thuộc tính như không gian dữ liệu vào, giao diện yêu cầu, truy xuất đến dữ liệu nguồn, hỗ trợ cấu trúc dữ liệu, tính đáp ứng, mức độ truy vấn dữ liệu, độ chính xác, tính cập nhật.

CHƯƠNG 2: LẬP TRÌNH VỚI WEB NGỮ NGHĨA

2.1 Môi trường lập trình Web ngữ nghĩa

2.1.1 Môi trường lập trình IntelliJ

2.1.1.1 Giới thiệu.

IntelliJ IDEA là một trình soạn thảo (IDE) để phát triển các phần mềm máy tính. Nó được phát triển bởi nhà sản xuất JetBrains (trước đây gọi là IntelliJ), và được cấp phép Apache. IntelliJ có 2 phiên bản: IntelliJ Ultimate (thương mại) và IntelliJ Community (cộng đồng). Cả hai có thể được sử dụng cho phát triển phần mềm.

2.1.1.2 Các thành phần cơ bản

- *Kiểm soát phiên bản*
- *Xây dựng công cụ*
- *Kiểm thử*
- *Trình biên dịch*
- *Thiết bị đầu cuối*
- *Công cụ cơ sở dữ liệu*
- *Máy chủ ứng dụng*

2.1.2 Jena

2.1.2.1 Giới thiệu

Jena là một thư viện của Java framework dùng cho việc xây dựng các ứng dụng Web ngữ nghĩa. Cung cấp một môi trường lập trình cho RDF, RDFS and OWL, SPARQL. Bao gồm các công cụ suy diễn từ các luật cơ sở (rule-based inference engine). Phiên bản đầu tiên ra đời do HP Lab xây dựng Jena 2 ra đời vào 2003 – phiên bản hiện tại là Jena 3.0.

2.1.2.2 Jena API

Kiến trúc của Jena được thiết kế để cho phép tích hợp một cách dễ dàng các thành phần xử lý có thể thay thế như bộ phân tích dữ liệu (parser), xuất bản (writer), lưu trữ và truy vấn.

Jena API bao gồm một tập các giao diện Java mô tả các nguồn tài nguyên (resources), các thuộc tính (properties), các phát biểu (statements) và các mô hình (model) dưới dạng khung mô tả tài nguyên RDF (Resource Description Framework).

2.1.2.3 Kiến trúc Jena 3

RDFNode interface: Cung cấp các thành phần của các bộ ba RDF {predicate, subject, object}.

Literal interface: Dùng để mô tả các chuỗi và chuyển đổi một số kiểu của Java như String, Int, và Double. Các đối tượng hiện thực giao diện Property có thể là một predicate trong bộ ba {predicate, subject, object}.

Statement interface: mô tả bộ ba {predicate, subject, object}. Đồng thời nó cũng có thể được dùng như một đối tượng .

2.2 Mô hình hóa thông tin với Jena

2.2.1 Chương trình Hello Word! trong Jena

2.2.2 Tạo mô hình RDF

2.2.2.1 Tạo tài nguyên và thêm thuộc tính

2.2.2.2 Thêm mối quan hệ, xuất ra tập tin RDF

2.2.3 Đọc mô hình RDF

Phần này tôi sử dụng truy vấn SPARQL để xuất ra *?s ?p ?o* của tất cả các tài nguyên trong tập tin *family.rdf*.

Kết quả xuất ra dưới định dạng Turtle:

s	p	o
<http://family/Son>	<http://purl.org/vocab/relationship/vochongcua>	<http://family/Tung>
<http://family/Son>	<http://purl.org/vocab/relationship/chamecua>	<http://family/Tuyet>
<http://family/Tung>	<http://purl.org/vocab/relationship/concua>	<http://family/hai>
<http://family/Tung>	<http://purl.org/vocab/relationship/concua>	<http://family/hoa>
<http://family/Tung>	<http://purl.org/vocab/relationship/anhemcua>	<http://family/Nam>
<http://family/Tung>	<http://purl.org/vocab/relationship/vochongcua>	<http://family/Son>
<http://family/Tung>	<http://purl.org/vocab/relationship/chamecua>	<http://family/Tuyet>
<http://family/hoa>	<http://purl.org/vocab/relationship/vochongcua>	<http://family/hai>
<http://family/hoa>	<http://purl.org/vocab/relationship/chamecua>	<http://family/Nam>
<http://family/hoa>	<http://purl.org/vocab/relationship/chamecua>	<http://family/Tung>
<http://family/Tuyet>	<http://purl.org/vocab/relationship/concua>	<http://family/Tung>
<http://family/Tuyet>	<http://purl.org/vocab/relationship/concua>	<http://family/Son>
<http://family/trung>	<http://purl.org/vocab/relationship/vochongcua>	<http://family/mai>
<http://family/mai>	<http://purl.org/vocab/relationship/anhemcua>	<http://family/hai>
<http://family/mai>	<http://purl.org/vocab/relationship/vochongcua>	<http://family/trung>
<http://family/hai>	<http://purl.org/vocab/relationship/anhemcua>	<http://family/mai>
<http://family/hai>	<http://purl.org/vocab/relationship/vochongcua>	<http://family/hoa>
<http://family/hai>	<http://purl.org/vocab/relationship/chamecua>	<http://family/Nam>
<http://family/hai>	<http://purl.org/vocab/relationship/chamecua>	<http://family/Tung>
<http://family/Nam>	<http://purl.org/vocab/relationship/concua>	<http://family/hai>
<http://family/Nam>	<http://purl.org/vocab/relationship/concua>	<http://family/hoa>
<http://family/Nam>	<http://purl.org/vocab/relationship/anhemcua>	<http://family/Tung>

Dữ liệu xuất dưới dạng Turtle

2.3 Chuyển đổi dữ liệu Web ngữ nghĩa

Trong phần này tôi sử dụng thư viện *jxl.jar* để thao tác với tập tin Excel. Sau đó sử dụng Jena để tạo mô hình dữ liệu RDF. Thêm các tài nguyên và thuộc tính với dữ liệu đọc từ file Excel. Cuối cùng xuất dữ liệu ra tập tin RDF với định dạng XML.

2.4 Framework Spring MVC

2.4.1 Giới thiệu về Framework Spring MVC

Spring là một trong những framework để phát triển các ứng dụng Java được sử dụng bởi hàng triệu lập trình viên trên thế giới. Spring giúp tạo ra các ứng dụng có hiệu năng cao, dễ dàng kiểm thử, sử dụng lại code một cách hợp lý. Spring gọn nhẹ và trong suốt (nhẹ: với kích thước nhỏ, version cơ bản chỉ khoảng 2 megabytes; trong suốt: hoạt động một cách trong suốt với đối lập trình viên), Spring là một mã nguồn mở, được phát triển, chia sẻ và có cộng đồng người dùng rất lớn trên thế giới. Spring Framework xây dựng dựa trên 2 nguyên tắc thiết kế chính là: Dependencies Injection và Aspect Orient Programming.

CHƯƠNG 3: ỨNG DỤNG TÌM KIẾM THÔNG TIN BỆNH NHÂN COVID-19 TẠI VIỆT NAM

3.1 Giới thiệu bài toán tìm kiếm thông tin bệnh nhân COVID-19

Từ dữ liệu hồ sơ bệnh án (trích xuất biểu đồ) có thể cho chúng ta có thể biết thêm về bệnh nhân COVID-19 và quá trình bệnh của họ. Những dữ liệu này có thể bao gồm thông tin về nhân khẩu học (tuổi, chủng tộc/dân tộc, giới tính), cũng như các triệu chứng, phương pháp điều trị và kết quả. Các nhà khoa học có thể sử dụng trích xuất biểu đồ để tìm hiểu những người có khả năng mắc bệnh nặng hơn, những dịch vụ chăm sóc y tế bệnh nhân đã nhận được và bệnh nhân đã hồi phục chưa. Với những nội dung đã tìm hiểu ở các phần trên, luận văn sẽ sử dụng công nghệ Web ngữ nghĩa để xây dựng một ứng dụng minh họa nhằm phục vụ cho việc lưu trữ, truy vấn, tìm kiếm các thông tin bệnh nhân COVID-19 tại Việt Nam.

3.1.1 Kiến trúc hệ thống

3.1.2 Chuẩn bị dữ liệu cho bài toán

3.1.2.1 Dữ liệu của người bệnh COVID-19 tại Việt Nam

Dữ liệu về thông tin người bệnh được lấy từ trang tin về dịch bệnh viêm đường hô hấp cấp COVID-19 của Bộ Y tế (Website: <https://ncov.moh.gov.vn/>). Dữ liệu sẽ được copy vào file Excel gồm các thông tin về bệnh nhân, địa điểm di chuyển của người bệnh.

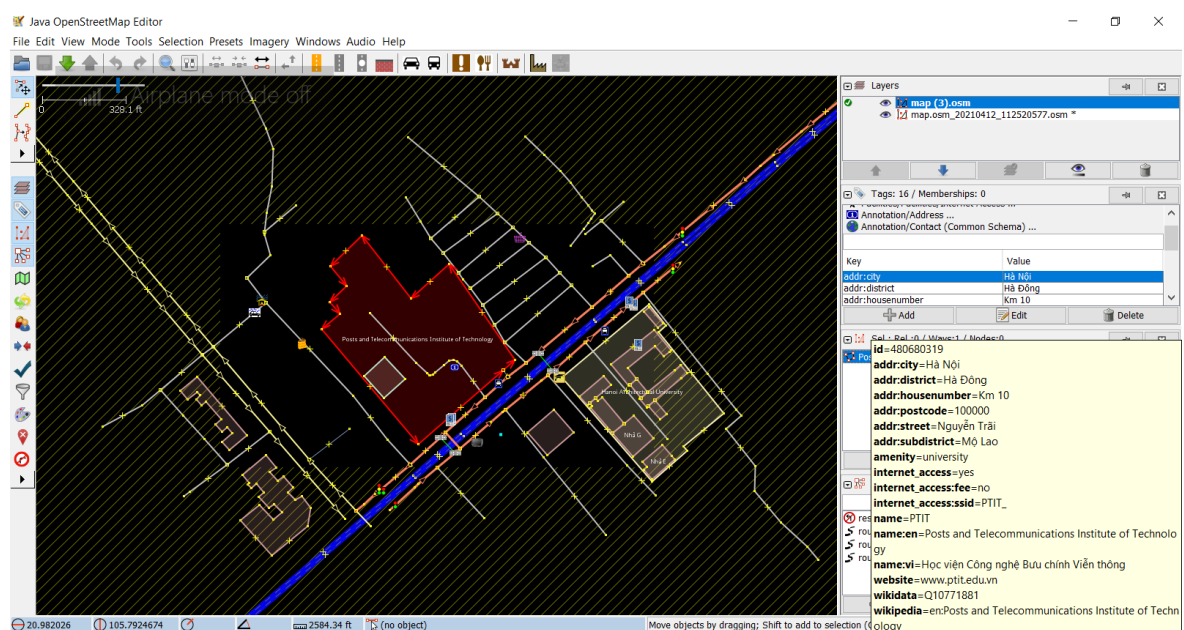
3.1.2.2 Dữ liệu từ OpenStreetMap

OpenStreetMap, hoặc OSM, giống như Wikipedia cho bản đồ. Đó là nguồn mở, người dùng đã đóng góp, và miễn phí để sử dụng, và, giống như Wikipedia, nó được chứng minh là đáng tin cậy đáng kinh ngạc. Ngay cả khi bạn chưa từng nghe về nó trước đây, rất có thể là bạn đã sử dụng dữ liệu của nó. Nó được sử dụng bởi FourSquare, Pokemon Go, Craigslist, Tesla, và một danh sách dài các ứng dụng và dịch vụ khác. Trong khi nó không phải là hữu ích cho chuyển hướng hàng ngày, bộ

dữ liệu không gian địa lý rộng lớn của nó đã nhiều lần chứng minh hữu ích không chỉ cho các ứng dụng, nhưng cho công việc nhân đạo và cứu trợ khẩn cấp .

OpenStreetMap ban đầu được tạo ra ở Anh bởi Steve Coast, và kể từ khi thành lập năm 2004, cơ sở dữ liệu thông tin bản đồ của nó đã phát triển đáng kể. Dữ liệu của họ đến từ việc đóng góp hàng loạt từ các thực thể chứa rất nhiều dữ liệu bản đồ và từ những người dùng tặng thời gian và năng lượng của họ để giúp xây dựng và duy trì bản đồ.

Hệ thống tìm kiếm khi nhập vào một địa điểm sẽ được tìm kiếm bởi các tập tin RDF trên hệ thống. Sau khi trích được thông tin dựa trên OSM_ID sẽ được nhúng vào OpenStreetMap để hiển thị lên bản đồ.



Dữ liệu Học viện Công nghệ Bưu chính Viễn thông trên OpenStreetMap

3.1.2.3 Dữ liệu từ *LinkedGeoData*

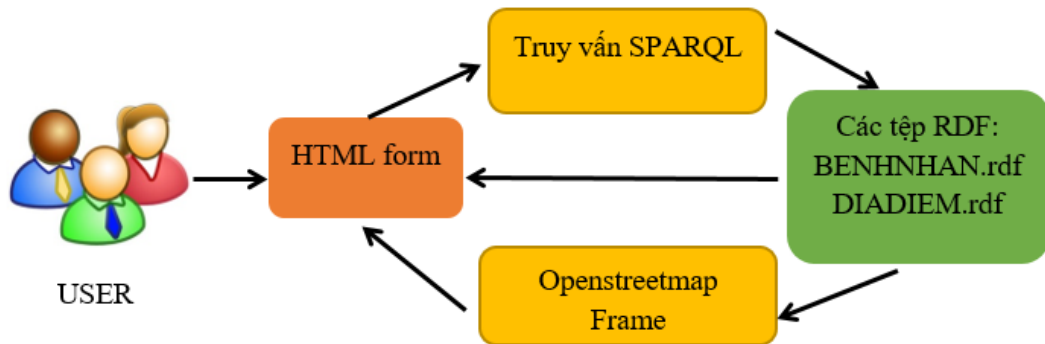
LinkedGeoData là dự án của nhóm nghiên cứu AKSW với hơn 1,2 tỉ lượt truy cập dựa thông tin trên tập tin hành tinh OpenStreetMap từ năm 2015.

LinkedGeoData là một dự án để thêm một không gian vào dữ liệu của Web ngữ nghĩa. LinkedGeoData sử dụng thông tin thu thập được bởi dự án OpenStreetMap và làm cho nó có sẵn dưới dạng cơ sở tri thức RDF theo nguyên tắc dữ liệu có liên

kết. Nó liên kết dữ liệu này với các cơ sở tri thức khác trong sáng kiến liên kết dữ liệu mở.

3.2 Giải pháp hệ thống

3.2.1 Mô hình hệ thống



Mô hình hệ thống tra cứu thông tin bệnh nhân

Khi người dùng chọn người bệnh, quá trình tìm kiếm từ khóa sẽ được thực hiện tại truy vấn SPARQL trên dữ liệu RDF từ các tệp BENHNNHAN.rdf, DIADIEM.rdf và LICHSU.rdf. Các thông tin từ file RDF sẽ được hiển thị trực tiếp lên HTML form, đồng thời dựa vào thông tin từ OpenStreetMap để hiển thị lên trình duyệt thông tin chi tiết địa điểm cần tìm kiếm.

3.2.2 Tạo dữ liệu RDF

Với các dữ liệu tải về từ OpenStreetMap, tôi sẽ dùng Jena API để chuyển đổi sang dữ liệu RDF và được lưu với tên DIADIEM.rdf. Đồng thời để tạo dữ liệu ngữ nghĩa minh họa tôi tạo thêm tập tin BENHNNHAN.rdf, LICHSU.rdf từ dữ liệu file Excel để biết thông tin bệnh nhân và lịch sử di chuyển của họ. Tiếp đó, tôi dùng bộ từ vựng FOAF(<http://xmlns.com/foaf/0.1/>) để mô tả các thông tin trên lần lượt tương ứng với các thuộc tính sau: FOAF.name, FOAF.age, FOAF.openid, FOAF.status, cùng các bộ từ vựng của Dbpedia(<https://dbpedia.org/ontology/>) và DC (<http://purl.org/dc/elements/1.1/>)

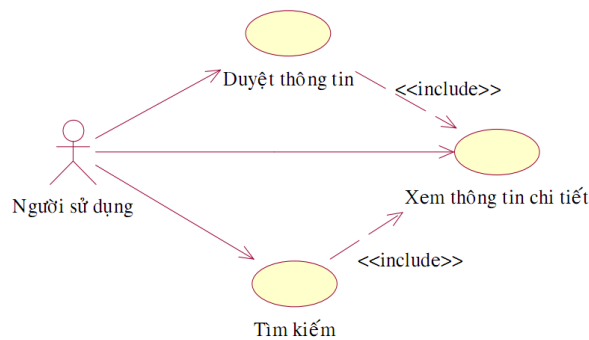
3.2.3 Truy vấn dữ liệu

Để thực hiện truy vấn trên tập tin RDF, tôi sử dụng ngôn ngữ truy vấn Web ngữ nghĩa SPARQL với nội dung nhập từ HTML form. Sau đó xuất kết quả truy vấn dưới định dạng HTML. Sử dụng ngôn ngữ lập trình Java.

3.3 Xây dựng hệ thống

3.3.1 Mô tả hệ thống

Phần này sẽ xây dựng mô hình hệ thống với mục đích cung cấp một cách chi tiết về các chức năng cơ bản như tìm kiếm thông tin, xem thông tin bệnh nhân (duyet thông tin) và xem các thông tin chi tiết của địa điểm. Mô hình được thể hiện như hình sau:



Use case hệ thống tìm kiếm thông tin bệnh nhân COVID-19

3.3.2 Đặc tả chức năng

- Chức năng tìm kiếm: cho phép người sử dụng truy xuất thông tin trong hệ thống bằng cách chọn người bệnh trên form. Hệ thống duyệt file DIADIEM.rdf tìm tất cả các tài nguyên liên quan đến từ khóa muốn tìm và trả về kết quả. Khi lấy được thông tin phù hợp sẽ hiển thị tại khung OpenStreetMap.

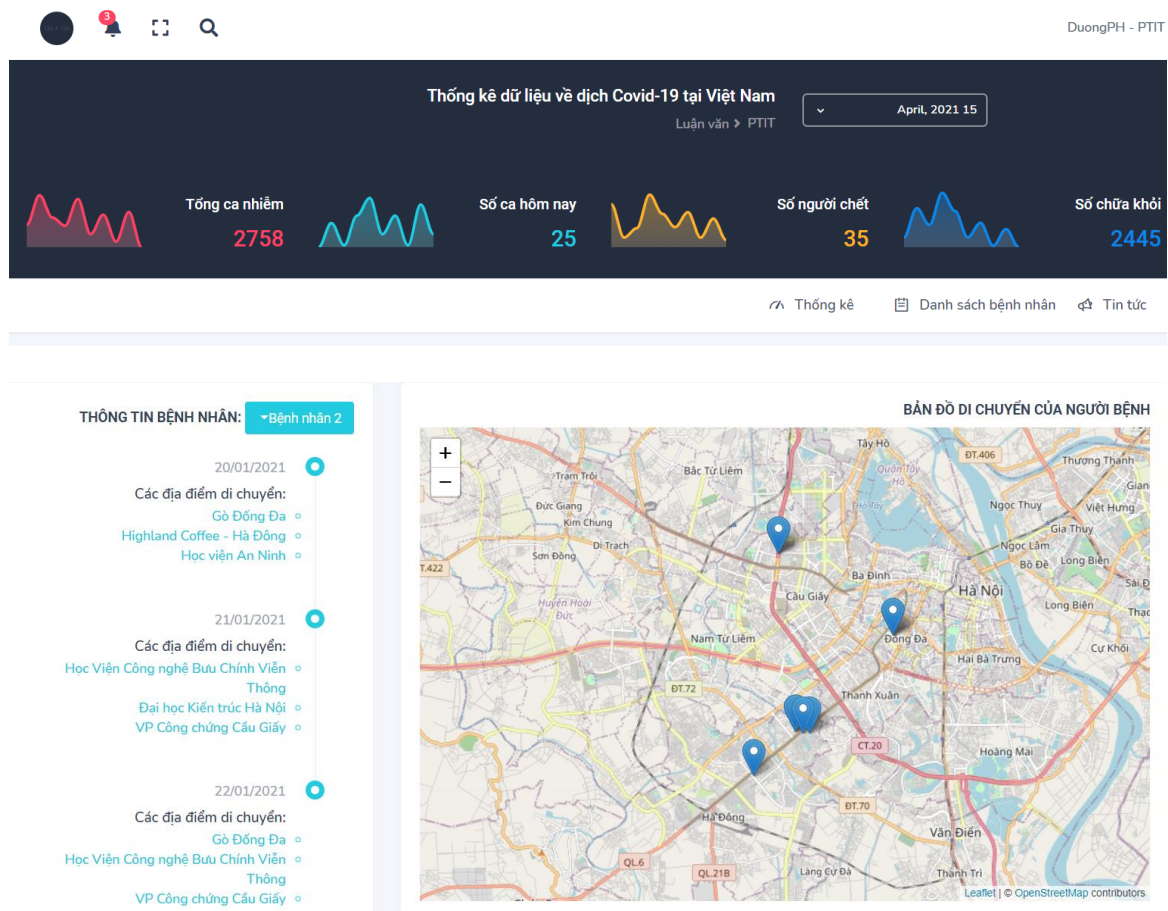
- Chức năng duyệt thông tin: Chức năng này cho phép người sử dụng xem thông tin trong danh sách các tài nguyên liên quan đến từ khóa bằng cách chọn mục thông tin cần xem. Hệ thống tìm các đối tượng liên quan đến thông tin được chọn

và trả về các kết quả cho người sử dụng. Kết quả được hiển thị tại khung OpenStreetMap.

- Chức năng xem thông tin chi tiết: Chức năng này cho phép người sử dụng xem thông tin chi tiết của một tài nguyên trong hệ thống dựa trên khung OpenStreetMap.

3.3.3 Thiết kế giao diện

3.3.3.1 Giao diện trang thống kê

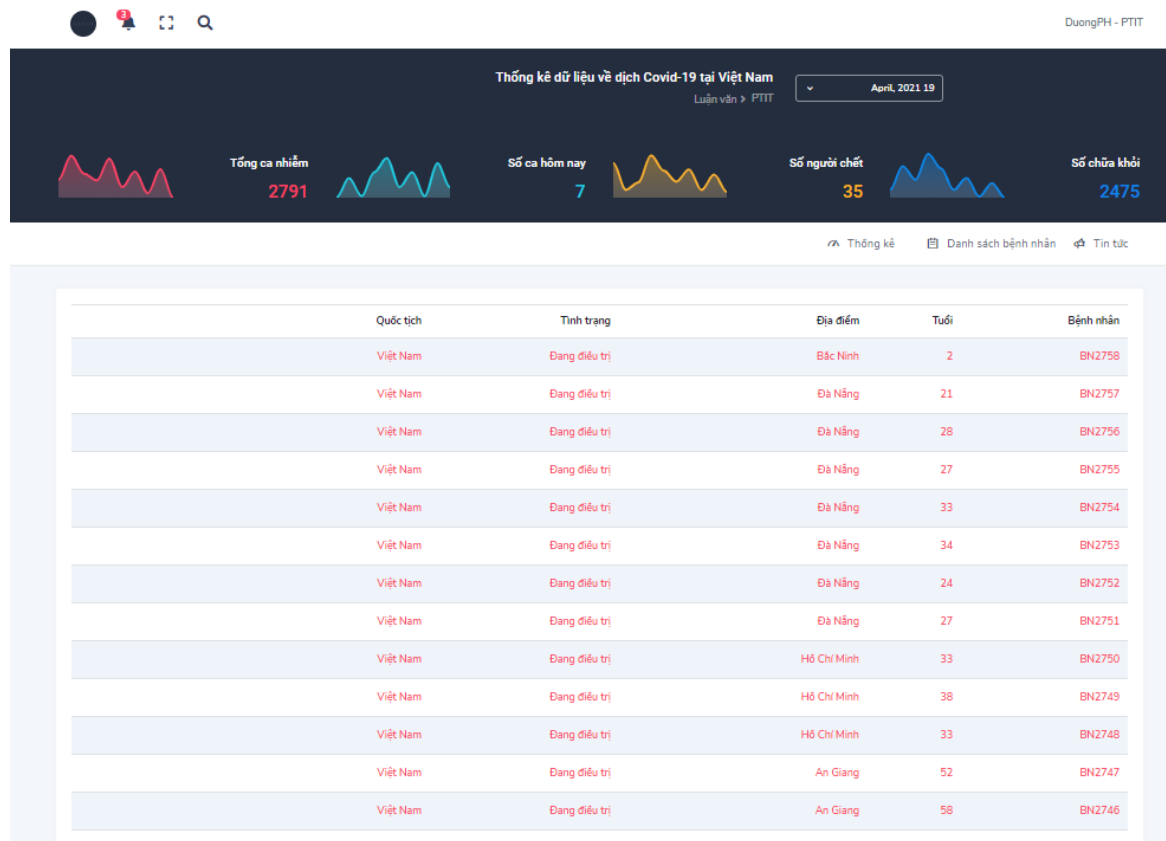


Giao diện trang chủ tra cứu thông tin người bệnh COVID-19

Khi người dùng chọn bệnh nhân hệ thống sẽ hiển thị tất cả các sự kiện, địa điểm chính xác của người bệnh tương ứng di chuyển trong thời gian trước theo

ngày cụ thể. Ngoài ra các thông số như Tổng ca nhiễm, Số ca hôm nay, Số người chết, Số chữa khỏi được cập nhật mới nhất từ Website của bộ y tế <https://ncov.moh.gov.vn/>

3.3.3.2 Giao diện trang chi tiết thông tin người bệnh



Giao diện chi tiết thông tin người bệnh COVID-19