

TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI

**ĐỒ ÁN TỐT NGHIỆP**

**Xây dựng cơ sở tri thức ngữ nghĩa và chatbot về  
du lịch Hà Nội**

**DƯƠNG HUỆ LINH**

linh.dh183942@sis.hust.edu.vn

**Ngành: Công nghệ thông tin**

**Giảng viên hướng dẫn:** TS. Đỗ Bá Lâm

Chữ ký GVHD

**Khoa:** Khoa học máy tính

**Trường:** Công nghệ thông tin và Truyền thông

**HÀ NỘI, 08/2022**

# LỜI CAM KẾT

Họ và tên sinh viên: Dương Huệ Linh  
Điện thoại liên lạc: 0868372366  
Email: linh.dh183942@sis.hust.edu.vn  
Lớp: IT3-03  
Hệ đào tạo: Đại học chính quy

Tôi – *Dương Huệ Linh* – cam kết Đồ án Tốt nghiệp (ĐATN) là công trình nghiên cứu của bản thân tôi dưới sự hướng dẫn của *Tiến sĩ Đỗ Bá Lâm*. Các kết quả nêu trong ĐATN là trung thực, là thành quả của riêng tôi, không sao chép theo bất kỳ công trình nào khác. Tất cả những tham khảo trong ĐATN – bao gồm hình ảnh, bảng biểu, số liệu, và các câu trích dẫn – đều được ghi rõ ràng và đầy đủ nguồn gốc trong danh mục tài liệu tham khảo. Tôi xin hoàn toàn chịu trách nhiệm với dù chỉ một sao chép vi phạm quy chế của nhà trường.

*Hà Nội, ngày 07 tháng 08 năm 2022*

Tác giả ĐATN

*Dương Huệ Linh*

# LỜI CẢM ƠN

Đề tài “Xây dựng cơ sở tri thức ngữ nghĩa và chatbot về du lịch Hà Nội” là nội dung tôi chọn để nghiên cứu và làm đồ án tốt nghiệp sau bốn năm theo học đại học chuyên ngành Công nghệ thông tin tại Đại học Bách Khoa Hà Nội.

Lời đầu tiên, tôi xin chân thành gửi lời cảm ơn đến các thầy giáo, cô giáo trong Trường Công nghệ Thông tin và Truyền thông đã tận tình giảng dạy, truyền đạt những kiến thức cơ sở và chuyên ngành, quan tâm, giúp đỡ tôi trong suốt quá trình học tập ở trường.

Đặc biệt, tôi xin được bày tỏ lòng biết ơn sâu sắc tới thầy giáo TS. Đỗ Bá Lâm thuộc bộ môn Hệ thống thông tin. Thầy đã trực tiếp chỉ bảo, định hướng, góp ý và giúp đỡ tận tình trong suốt quá trình nghiên cứu để tôi có thể hoàn thiện đồ án này.

Cuối cùng, tôi xin được gửi lời cảm ơn chân thành tới gia đình, bạn bè đã động viên, đóng góp ý kiến và giúp đỡ trong quá trình học tập, nghiên cứu và hoàn thành đồ án tốt nghiệp.

## TÓM TẮT NỘI DUNG ĐỒ ÁN

Đồ án ”Xây dựng cơ sở tri thức ngữ nghĩa và chatbot về du lịch Hà Nội” tập trung nghiên cứu và thực hiện xây dựng cơ sở tri thức ngữ nghĩa về du lịch lễ hội ở Hà Nội, từ đó phát triển hệ thống tra cứu và hỏi đáp sử dụng kho tri thức đã phát triển ở trên. Trong quá trình hoàn thành đồ án, tôi đã thực hiện các công việc: thu thập dữ liệu, xây dựng cơ sở tri thức về du lịch Hà Nội, xây dựng chatbot, xây dựng website tra cứu thông tin. Tôi đã sử dụng Python Selenium để crawl dữ liệu, Protégé để lưu trữ cơ sở tri thức và dịch vụ xử lý ngôn ngữ tự nhiên LUIS để huấn luyện chatbot. Hệ thống sau khi hoàn thành hoạt động tương đối ổn định, đáp ứng được các yêu cầu đề ra, thể hiện được ưu điểm của web ngữ nghĩa như khả năng suy diễn và tính linh hoạt trong sử dụng dữ liệu.

## MỤC LỤC

<b>CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI.....</b>	<b>1</b>
1.1 Đặt vấn đề.....	1
1.2 Mục tiêu và định hướng giải pháp .....	2
1.3 Đối tượng và phạm vi nghiên cứu.....	2
1.4 Bố cục đồ án .....	2
<b>CHƯƠNG 2. CƠ SỞ LÝ THUYẾT .....</b>	<b>4</b>
2.1 Web ngữ nghĩa .....	4
2.1.1 Giới thiệu Web ngữ nghĩa .....	4
2.1.2 Kiến trúc Web ngữ nghĩa .....	5
2.1.3 RDF.....	7
2.1.4 RDFS.....	7
2.1.5 Ontology .....	8
2.1.6 Ngôn ngữ truy vấn SPARQL .....	10
2.2 Chatbot.....	12
2.2.1 Giới thiệu chatbot.....	12
2.2.2 Các thành phần của chatbot.....	12
2.2.3 Xây dựng Chatbot .....	14
2.2.4 Chatbot ứng dụng công nghệ Semantic Web .....	15
<b>CHƯƠNG 3. PHÂN TÍCH VÀ THIẾT KẾ HỆ THỐNG .....</b>	<b>17</b>
3.1 Tổng quan giải pháp.....	17
3.2 Kiến trúc hệ thống .....	17
3.2.1 Các thành phần xây dựng cơ sở tri thức .....	17
3.2.2 Các thành phần xây dựng chatbot .....	18
3.2.3 Các thành phần xây dựng website .....	18

3.3 Xây dựng cơ sở tri thức về du lịch ở Hà Nội .....	19
3.3.1 Thu thập dữ liệu.....	19
3.3.2 Xử lý dữ liệu.....	19
3.3.3 Xây dựng Ontology .....	20
3.3.4 Tạo cơ sở tri thức .....	25
3.3.5 Tích hợp Ontology với Linked Open Data .....	26
3.4 Xây dựng hệ thống chatbot .....	28
3.4.1 Đặc tả usecase .....	28
3.4.2 Mô tả kịch bản giữa Chatbot và người dùng .....	29
3.4.3 Cài đặt và xử lý ngôn ngữ tự nhiên.....	30
3.4.4 Truy vấn dữ liệu bằng SPARQL .....	32
3.5 Xây dựng trang web tra cứu thông tin về du lịch Hà Nội .....	35
3.5.1 Mô hình hóa chức năng .....	35
3.5.2 Mô hình hóa hành vi .....	38
<b>CHƯƠNG 4. TRIỂN KHAI HỆ THỐNG.....</b>	<b>40</b>
4.1 Công nghệ sử dụng .....	40
4.2 Kết quả thực nghiệm .....	44
4.2.1 Kết quả thu thập dữ liệu.....	44
4.2.2 Kết quả xây dựng cơ sở tri thức .....	45
4.2.3 Kết quả cài đặt Chatbot hỗ trợ du lịch .....	48
4.2.4 Kết quả cài đặt Website tra cứu thông tin.....	62
<b>CHƯƠNG 5. KẾT LUẬN .....</b>	<b>68</b>
5.1 Kết luận .....	68
5.2 Hướng phát triển trong tương lai .....	69
<b>TÀI LIỆU THAM KHẢO.....</b>	<b>70</b>

## DANH MỤC HÌNH VẼ

Hình 2.1	Google Knowledge Graph . . . . .	4
Hình 2.2	Kiến trúc phân tầng Web ngữ nghĩa [4] . . . . .	5
Hình 2.3	Mô hình bộ ba RDF . . . . .	7
Hình 2.4	Các thành phần cơ bản của chatbot [10] . . . . .	12
Hình 3.1	Quá trình xây dựng cơ sở tri thức . . . . .	17
Hình 3.2	Sơ đồ kiến trúc hệ thống chatbot . . . . .	18
Hình 3.3	Mô hình MVC [12] . . . . .	18
Hình 3.4	Một số dữ liệu quán ăn thu thập được . . . . .	19
Hình 3.5	Geocoding by SmartMonkey . . . . .	20
Hình 3.6	Sơ đồ các lớp trong Ontology . . . . .	22
Hình 3.7	Dữ liệu khách sạn trong trang bản tính . . . . .	26
Hình 3.8	Giao diện Cellfie của Protégé . . . . .	26
Hình 3.9	Truy vấn dữ liệu DBpedia bằng Sparql Endpoint . . . . .	27
Hình 3.10	Kết quả truy vấn DBpedia . . . . .	28
Hình 3.11	Kết quả tạo liên kết với DBpedia . . . . .	28
Hình 3.12	Sơ đồ kịch bản giữa Chatbot và người dùng . . . . .	30
Hình 3.13	Giao diện quản lý intent của LUIS . . . . .	31
Hình 3.14	Giao diện quản lý entity của LUIS . . . . .	32
Hình 3.15	Kết quả truy vấn SPARQL . . . . .	33
Hình 3.16	Use case tổng quan . . . . .	36
Hình 3.17	Biểu đồ tuần tự Xem thông tin . . . . .	38
Hình 3.18	Biểu đồ tuần tự Truy vấn SPARQL . . . . .	39
Hình 4.1	WebProtégé . . . . .	40
Hình 4.2	Protégé Desktop . . . . .	40
Hình 4.3	Lưu trữ Ontology dưới nhiều định dạng . . . . .	41
Hình 4.4	Minh họa tab chức năng trong Protégé . . . . .	41
Hình 4.5	Minh họa giao diện Cellfie plugin . . . . .	42
Hình 4.6	Minh họa tab SWRL . . . . .	42
Hình 4.7	Minh họa chức năng tạo intents, entities và utterances . . . . .	43
Hình 4.8	Minh họa chức năng Test của LUIS . . . . .	43
Hình 4.9	Minh họa vòng đời phát triển ứng dụng . . . . .	43
Hình 4.10	Ví dụ một số khách sạn trên Google Travel . . . . .	44
Hình 4.11	Một số dữ liệu khách sạn được crawl . . . . .	45
Hình 4.12	Một số dữ liệu khách sạn sau khi được xử lý và lưu trữ . . . . .	45

Hình 4.13 Các lớp của Ontology . . . . .	45
Hình 4.14 Object property của Ontology . . . . .	46
Hình 4.15 Data property của Ontology . . . . .	46
Hình 4.16 Các thể hiện của lớp điểm du lịch . . . . .	47
Hình 4.17 Các thuộc tính và quan hệ của một thể hiện trong lớp điểm du lịch . . . . .	47
Hình 4.18 Lời chào hệ thống . . . . .	48
Hình 4.19 Hướng dẫn của Chatbot . . . . .	48
Hình 4.20 Minh họa chức năng xem thông tin địa điểm du lịch (1) . . . . .	49
Hình 4.21 Minh họa chức năng xem thông tin địa điểm du lịch (2) . . . . .	49
Hình 4.22 Minh họa chức năng xem thông tin lễ hội . . . . .	50
Hình 4.23 Minh họa chức năng xem thông tin đặc sản . . . . .	50
Hình 4.24 Minh họa chức năng xem thông tin chợ . . . . .	51
Hình 4.25 Minh họa chức năng xem thông tin siêu thị . . . . .	51
Hình 4.26 Minh họa chức năng xem thông tin khách sạn . . . . .	52
Hình 4.27 Minh họa chức năng xem thông tin quán ăn . . . . .	52
Hình 4.28 Minh họa chức năng xem thông tin quán bar . . . . .	53
Hình 4.29 Minh họa chức năng xem thông tin quán cafe . . . . .	53
Hình 4.30 Minh họa chức năng tìm kiếm thông tin địa điểm du lịch thông qua quận huyện . . . . .	54
Hình 4.31 Minh họa chức năng tìm kiếm thông tin địa điểm du lịch thông qua phường xã . . . . .	55
Hình 4.32 Minh họa chức năng tìm kiếm thông tin đặc sản thông qua phường xã hoặc quận huyện . . . . .	55
Hình 4.33 Minh họa chức năng tìm kiếm thông tin lễ hội thông qua phường xã hoặc quận huyện . . . . .	56
Hình 4.34 Minh họa chức năng tìm kiếm thông tin khách sạn thông qua phường xã . . . . .	56
Hình 4.35 Minh họa chức năng tìm kiếm thông tin khách sạn thông qua quận huyện . . . . .	57
Hình 4.36 Minh họa chức năng tìm kiếm thông tin chợ thông qua phường xã hoặc quận huyện . . . . .	57
Hình 4.37 Minh họa chức năng tìm kiếm thông tin siêu thị thông qua phường xã hoặc quận huyện . . . . .	58
Hình 4.38 Minh họa chức năng tìm kiếm thông tin quán ăn thông qua phường xã . . . . .	58

Hình 4.39 Minh họa chức năng tìm kiếm thông tin quán ăn thông qua quận huyện . . . . .	59
Hình 4.40 Minh họa chức năng tìm kiếm thông tin quán bar thông qua phường xã . . . . .	59
Hình 4.41 Minh họa chức năng tìm kiếm thông tin quán bar thông qua quận huyện . . . . .	60
Hình 4.42 Minh họa chức năng tìm kiếm thông tin quán cafe thông qua phường xã . . . . .	60
Hình 4.43 Minh họa chức năng tìm kiếm thông tin lễ hội thông qua địa điểm du lịch . . . . .	61
Hình 4.44 Minh họa chức năng tìm kiếm thông tin đặc sản thông qua địa điểm du lịch . . . . .	61
Hình 4.45 Minh họa chức năng tìm kiếm thông tin địa điểm du lịch thông qua lễ hội hoặc đặc sản . . . . .	62
Hình 4.46 Minh họa trang chủ của website . . . . .	63
Hình 4.47 Minh họa trang danh sách chợ . . . . .	64
Hình 4.48 Minh họa trang thông tin chi tiết chợ . . . . .	65
Hình 4.49 Minh họa trang nhập truy vấn SPARQL . . . . .	66
Hình 4.50 Minh họa trang kết quả truy vấn . . . . .	67

## **DANH MỤC BẢNG BIỂU**

Bảng 3.1	Danh sách tác nhân của hệ thống . . . . .	35
Bảng 3.2	Danh sách ca sử dụng . . . . .	35
Bảng 3.3	Bảng đặc tả usecase "Xem thông tin" . . . . .	36
Bảng 3.4	Bảng đặc tả usecase "Truy vấn SPARQL" . . . . .	37
Bảng 3.5	Danh sách dữ liệu đầu vào . . . . .	37
Bảng 4.1	Bảng thống kê dữ liệu thu thập được . . . . .	44

# CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI

## 1.1 Đặt vấn đề

Hà Nội có vị trí và địa thế đẹp, thuận lợi, là Thủ đô, trung tâm chính trị, kinh tế, văn hoá, khoa học và đầu mối giao thông quan trọng của cả nước. Trải qua hơn 2.000 năm lịch sử, Hà Nội là nơi diễn ra, chứng kiến nhiều sự kiện lớn, nhiều biến cố, thăng trầm của riêng mảnh đất này và của chung đất nước, vì thế Hà Nội thực sự trở thành nơi lăng hồn núi sông ngàn năm, nơi hội tụ, kết tinh và lan tỏa các giá trị văn hóa của Việt Nam. Được thừa hưởng truyền thống cùng những giá trị văn hiến ấy, đồng thời luôn có ý thức bảo tồn và phát huy những giá trị văn hóa tốt đẹp, Hà Nội vững bước trên con đường phát triển và hội nhập như hiện nay, vừa là mảnh đất đậm đà bản sắc văn hóa, vừa là Thủ đô năng động, hiện đại. Tính đến cuối năm 2015, Hà Nội có 5.922 di tích lịch sử văn hóa, trong đó có 2.396 di tích đã được xếp hạng là di tích quốc gia, di tích quốc gia đặc biệt, di sản thế giới. Hà Nội có 1.793 di sản văn hóa phi vật thể được nhận diện, kiểm kê, đưa vào danh sách để bảo vệ; trong đó, có 1.206 lễ hội; 215 tập quán xã hội và tín ngưỡng; 175 nghệ thủ công truyền thống; 106 di sản về tri thức dân gian; 79 nghệ thuật trình diễn dân gian; 14 di sản về ngữ văn dân gian. Không chỉ nổi tiếng với lịch sử lâu đời, giàu bản sắc, văn hóa truyền thống dân tộc, vùng đất này còn được biết đến là một thành phố năng động và hiện đại, một đô thị sôi động từ sáng tới khuya, là trung tâm kinh tế hiện đại, phát triển lớn của cả nước và khu vực. Với những lợi thế trên, ngành du lịch Thủ đô trong những năm qua đã liên tục phát triển, tạo dấu ấn. Năm 2019, Hà Nội đã đón gần 29 triệu lượt khách, tăng 10,1% so với năm 2018. Trong đó có hơn 7 triệu lượt khách quốc tế (tăng 17%), khách nội địa đạt gần 22 triệu lượt (tăng 8%), tổng thu từ khách du lịch đạt 103.807 tỷ đồng (tăng 34%).

Tiềm năng phát triển du lịch lớn cũng đồng nghĩa với nhu cầu tìm kiếm thông tin của người dùng ngày càng tăng cao. Tuy nhiên với lượng thông tin khổng lồ như hiện nay trên web thì việc tìm kiếm tri thức mất khá nhiều thời gian và công sức. Web ngữ nghĩa đã ra đời nhằm giải quyết vấn đề trên. Web ngữ nghĩa là hệ thống các thông tin được định nghĩa một cách rõ ràng nhằm mục đích giúp máy tính có thể hiểu được ngữ nghĩa, từ đó việc tìm kiếm thông tin cũng dễ dàng và chính xác hơn.

Ngoài ra, Chatbot cũng là một công cụ hữu ích hỗ trợ tra cứu thông tin dễ dàng và thuận tiện. Hiện nay Chatbot đang ngày càng trở nên phổ biến trong tất cả các lĩnh vực và đặc biệt là trong lĩnh vực du lịch, nhà hàng và ăn uống. Chatbot đem đến nhiều lợi ích cho cả doanh nghiệp du lịch và khách hàng. Với khả năng truy

cập thông tin nhanh chóng, kết nối tức thì với hệ sinh thái địa phương, chatbot góp phần mang đến chuyến du lịch an toàn và trải nghiệm vượt trội.

Do vậy, trong đồ án lần này em đã chọn đề tài ”Xây dựng cơ sở tri thức ngữ nghĩa và chatbot về du lịch Hà Nội”.

## **1.2 Mục tiêu và định hướng giải pháp**

### **Mục tiêu:**

Đề tài hướng đến xây dựng một Ontology về các địa điểm du lịch, từ đó xây dựng hệ thống tra cứu và hỏi đáp về các thông tin liên quan đến du lịch ở Hà Nội.

### **Định hướng giải pháp:**

Phương pháp lý thuyết: Nghiên cứu tìm hiểu về web ngữ nghĩa cũng như các công cụ để có thể xây dựng và truy vấn cơ sở tri thức ngữ nghĩa. Tìm hiểu các công nghệ để xây dựng chatbot.

Phương pháp thực nghiệm: Thu thập và xử lý các dữ liệu liên quan đến du lịch Hà Nội. Tiến hành cài đặt hệ thống.

## **1.3 Đối tượng và phạm vi nghiên cứu**

### **Đối tượng nghiên cứu:**

Đồ án tập trung nghiên cứu thông tin các địa điểm du lịch ở Hà Nội, cơ sở lý thuyết về web ngữ nghĩa, cơ sở lý thuyết chatbot và cách xây dựng chatbot.

### **Phạm vi nghiên cứu:**

Xây dựng Ontology lưu trữ tập dữ liệu về các địa điểm du lịch ở Hà Nội và các thông tin liên quan như lễ hội, âm thực, khách sạn, quán ăn, quán cafe, siêu thị,... Khai thác các tính năng về truy xuất dữ liệu trong Ontology và xây dựng hệ thống cho phép người dùng tra cứu và hỏi đáp các thông tin liên quan đến du lịch Hà Nội.

## **1.4 Bố cục đồ án**

Chương 2 trình bày 2 phần chính là cơ sở lý thuyết của Web ngữ nghĩa và giới thiệu về chatbot. Đây là những kiến thức nền tảng giúp tôi có cái nhìn tổng quan về Web ngữ nghĩa và chatbot cũng là bước đầu tiên trong quá trình tìm hiểu để xây dựng hệ thống.

Sau khi đã trang bị được những kiến thức cần thiết từ Chương 2. Chương 3 trình bày quá trình phân tích thiết kế, từng bước xây dựng hệ thống. Từ giai đoạn thu thập dữ liệu đến tạo Ontology, cài đặt chatbot và website được nêu rõ các bước tiến hành cùng với công nghệ sử dụng.

Từ những phân tích thiết kế trong chương 3, chương 4 trình bày những kết quả đạt được trong từng giai đoạn, tổng kết cả quá trình tìm hiểu và thực hành.

Qua kết quả nghiên cứu, phân tích và thực nghiệm, chương 5 kết luận rút ra những gì đã đạt được trong quá trình làm đồ án, kết quả hệ thống và hướng phát triển trong tương lai.

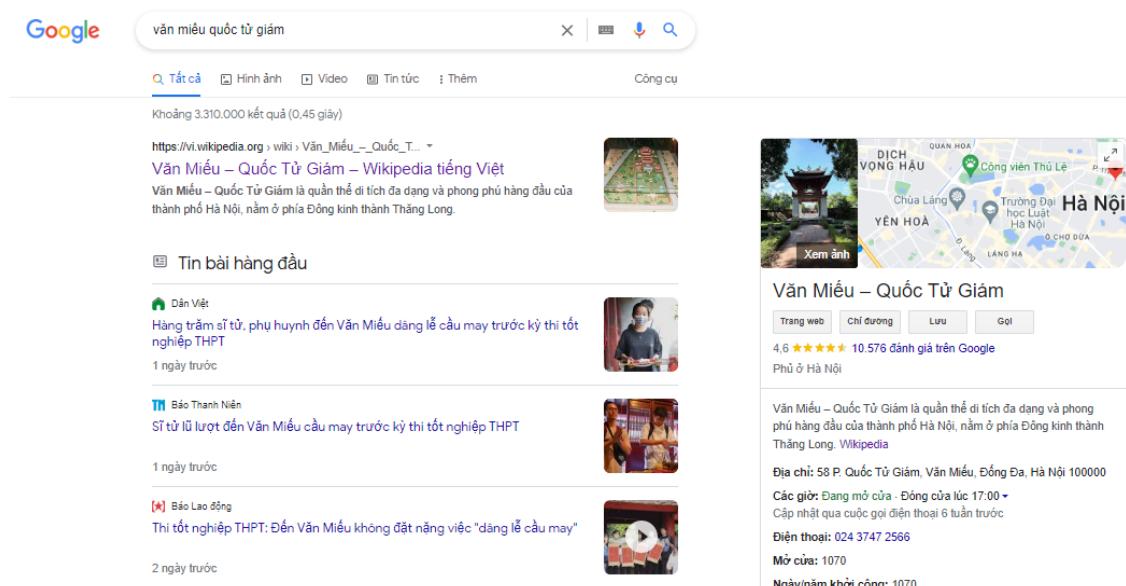
## CHƯƠNG 2. CƠ SỞ LÝ THUYẾT

### 2.1 Web ngữ nghĩa

#### 2.1.1 Giới thiệu Web ngữ nghĩa

Theo Tim Berners-Lee (cha đẻ của WWW, URI, HTML và HTTP), Web ngữ nghĩa là một mở rộng của Web hiện tại trong đó ý nghĩa của thông tin được thiết lập tường minh, mở ra khả năng cộng tác tốt hơn giữa người và máy tính [1]. Semantic Web là một phương pháp cho phép định nghĩa và liên kết dữ liệu một cách có ngữ nghĩa hơn nhằm phục vụ cho máy tính có thể “hiểu” được. Ngoài ra, Semantic Web còn cung cấp một môi trường chia sẻ và xử lý dữ liệu tự động bằng máy tính [2].

Một ví dụ về ứng dụng tìm kiếm của Semantic Web là Google Knowledge Graph - cơ sở tri thức được Google và các dịch vụ của nó sử dụng để nâng cao kết quả của công cụ tìm kiếm với thông tin được thu thập từ nhiều nguồn khác nhau.



Hình 2.1: Google Knowledge Graph

Khi tìm kiếm trên google từ khóa "văn miếu quốc tử giám", dữ liệu (mô tả, địa chỉ, thời gian mở cửa, ...) hiển thị ngay tại trang tìm kiếm mà không cần truy cập vào bất kỳ đường dẫn nào. Để làm được điều này thì Search Engine của Google phải áp dụng công cụ tìm kiếm ngữ nghĩa và Website mang thông tin phải được mô hình hóa thành dữ liệu ngữ nghĩa.

Một số ưu điểm của Web ngữ nghĩa:

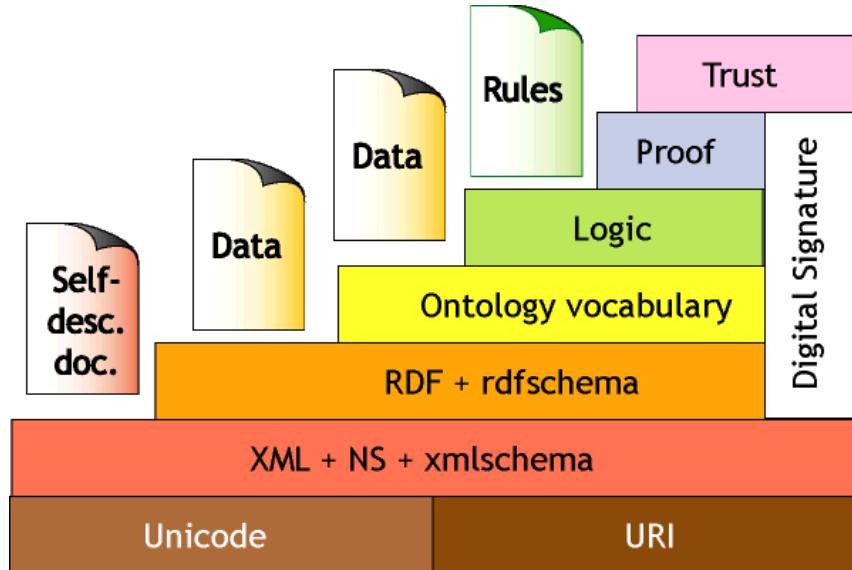
- **Máy có thể hiểu được thông tin trên Web:** Web hiện hành chỉ cho con người đọc chứ không dành cho máy hiểu. Semantic Web tạo ra Web mà cả người và

máy đều có thể hiểu [3].

- **Thông tin được tìm kiếm nhanh chóng và chính xác hơn:** Với Semantic Web, việc tìm kiếm thông tin trở nên dễ dàng hơn khi mọi thứ được đặt trong ngữ cảnh. Mục tiêu của Semantic Web chính là phát triển các tiêu chuẩn và kĩ thuật giúp máy tính hiểu nhiều thông tin trên Web, tìm ra các thông tin dồi dào hơn, tích hợp, duyệt dữ liệu, và tự động hóa các thao tác. Semantic Web không chỉ giúp chúng ta tìm kiếm thông tin chính xác hơn mà còn khiến cho máy tính có thể tích hợp thông tin từ nhiều nguồn khác nhau, biết so sánh các thông tin với nhau [3].
- **Dữ liệu liên kết động:** Với Semantic Web, chúng ta có thể kết hợp các thông tin đã được mô tả và giàu ngữ nghĩa với bất kì nguồn dữ liệu nào. Không chỉ cung cấp các URI cho tài liệu như đã làm trong quá khứ mà còn cho con người, các khái niệm, các mối liên hệ [3].

### 2.1.2 Kiến trúc Web ngữ nghĩa

Semantic Web là một tập hợp/ một ngăn xếp các chuẩn. Các lớp trong kiến trúc của Semantic Web phối hợp với nhau để đảm bảo độ an toàn và tối ưu giá trị của thông tin.



Hình 2.2: Kiến trúc phân tầng Web ngữ nghĩa [4]

### Lớp Unicode và URI

Unicode và URI là lớp đầu tiên trong kiến trúc Web ngữ nghĩa, giữ vai trò nền tảng, cung cấp phương tiện nhằm định danh các tài nguyên trong Semantic Web. Tài nguyên ở đây có thể là bất cứ thứ gì có danh tính chẵng hạn như một trang web, một tài liệu, một hình ảnh hay một người.

- Unicode được coi là tiêu chuẩn chung trong hệ thống mã hóa, đại diện cho tất cả các ngôn ngữ khác nhau trên thế giới.
- URI (Uniform Resource Identifier) là những chuỗi ký tự tham chiếu đến để nhận dạng và định danh tài nguyên.

### **Lớp XML, NS và XMLSchema**

Lớp này cung cấp khả năng tích hợp các định nghĩa Semantic Web với các chuẩn XML. XML (Extensible Markup Language) là ngôn ngữ đánh dấu mở rộng, có chức năng truyền dữ liệu và mô tả nhiều loại dữ liệu khác nhau. XML cho phép người dùng tự định nghĩa cấu trúc, có thể được sử dụng để biên soạn các tài liệu Web, phù hợp để chia sẻ trong môi trường Internet. NS (Name Space) là cơ chế đặt tên các đối tượng để tránh vấn đề trùng lặp. XSD (XML Schema Definition) được sử dụng để định nghĩa kiểu phần tử, thuộc tính và dữ liệu, mô tả cấu trúc và nội dung của dữ liệu XML.

### **Lớp RDF và RDFS**

RDF (Resource Description Framework) là một mô hình dữ liệu cho phép biểu diễn các phát biểu ngữ nghĩa về dữ liệu. RDFS (Resource Description Framework Schema) cung cấp bộ từ vựng mô tả một cách có ngữ nghĩa về các lớp và thuộc tính của tài nguyên RDF, tổng quát hóa phân cấp các thuộc tính và lớp đó, thêm vào khả năng bổ sung một số dữ liệu ở mức lược đồ.

### **Lớp Ontology**

Ontology là tập hợp khái niệm và quan hệ giữa chúng trong một miền tri thức. Ontology được xây dựng dựa trên RDFS, và cung cấp một bộ từ vựng lớn hơn, cũng như có cú pháp đa dạng hơn RDFS [5].

### **Lớp Logic**

Lớp Logic cung cấp các biểu diễn tri thức, luật suy diễn, đưa ra kết luận dựa trên các điều kiện đầu vào.

### **Lớp Proof**

Lớp Proof hỗ trợ kiểm tra tính đúng đắn và hợp lệ của các suy diễn.

### **Lớp Digital Signature**

Lớp Digital Signature được sử dụng để định danh tài liệu.

### **Lớp Trust**

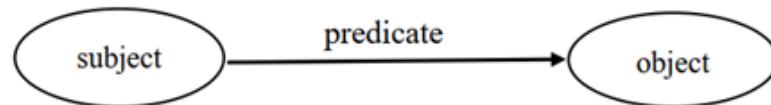
Lớp Trust cung cấp một cơ chế để đánh giá mức độ tin tưởng, đáng tin cậy của thông tin.

### 2.1.3 RDF

RDF là một mô hình dữ liệu được sử dụng để mô tả khái niệm hoặc mô hình hóa thông tin được diễn dịch trong các tài nguyên web, sử dụng trong các định dạng cú pháp khác nhau [6].

Các khái niệm cơ bản trong RDF:

- **Tài nguyên (Resources):** Tài nguyên là đối tượng mà chúng ta muốn mô tả, mọi tài nguyên đều có URI để định danh.
- **Thuộc tính (Properties):** Thuộc tính là một loại tài nguyên đặc biệt dùng để mô tả mối quan hệ giữa các tài nguyên. Thuộc tính trong RDF cũng được xác định bởi các URI.
- **Phát biểu (Statements):** Phát biểu là các mô tả về dữ kiện. Mỗi phát biểu là một bộ ba chủ thể - thuộc tính - giá trị.
  - **Chủ thể (Subject):** Chủ thể là thành phần thông tin đang được mô tả.
  - **Thuộc tính (Predicate):** Thuộc tính của chủ thể đang được mô tả.
  - **Giá trị (Object):** Xác định giá trị của thuộc tính của chủ thể đang được mô tả.



**Hình 2.3:** Mô hình bộ ba RDF

Ví dụ:

```

<http://www.example.org/index.html>
<http://www.example.org/terms/creation-date> "August 8, 2022" .
  
```

Trong ví dụ này:

- Chủ thể: là URL <http://www.example.org/index.html>
- Thuộc tính: là URL <http://www.example.org/terms/creation-date>
- Giá trị: chuỗi "August 8, 2022"

### 2.1.4 RDFS

RDF Schema cung cấp các nguyên tắc mô hình hóa để tổ chức các đối tượng Web thành hệ thống phân cấp. RDFS làm giàu thêm ngữ nghĩa bằng cách bổ sung các từ vựng để mô tả các lớp và thuộc tính, các mối quan hệ của lớp con và lớp cha.

cũng như các giới hạn về miền và phạm vi. RDF Schema được xây dựng dựa trên RDF.

Một số từ khóa RDFS đưa ra, đồng thời quy ước ý nghĩa tương thích với mô hình dữ liệu RDF.

- Từ khóa cho lớp

**rdfs:Class:** Tập hợp các tài nguyên có những thuộc tính chung.

**rdfs:subClassOf:** Các lớp được liên kết bởi quan hệ subClassOf có thể tạo thành một cấu trúc phân cấp.

- Từ khóa cho thuộc tính

**rdfs:domain** và **rdfs:range:** Các giới hạn miền và phạm vi cho phép hạn chế các tài nguyên có thể áp dụng thuộc tính và giá trị của thuộc tính.

**rdfs:subPropertyOf:** Tạo quan hệ phân cấp cho các thuộc tính.

- Các lớp đặc biệt

**rdfs:Resource:** Lớp của tất cả các tài nguyên.

**rdfs:Literal:** Lớp của tất cả các hằng giá trị (chuỗi).

**rdfs:Datatype:** Gắn tài nguyên với kiểu dữ liệu của nó.

### 2.1.5 Ontology

Ontology cung cấp một bộ từ vựng chung bao gồm các khái niệm, các thuộc tính quan trọng và mối quan hệ giữa các khái niệm và các thuộc tính này [7]. Ngoài bộ từ vựng, ontology còn cung cấp các ràng buộc, đôi khi các ràng buộc này được coi như các giả định cơ sở về ý nghĩa mong muốn của bộ từ vựng, nó được sử dụng trong một miền mà có thể được giao tiếp giữa người và các hệ thống ứng dụng phân tán hỗn tạp khác.

Các thành phần của Ontology:

- **Lớp (Classes):** Tập hợp các đối tượng có chung thuộc tính.
- **Thuộc tính (Properties):** Thuộc tính mô tả tính chất, đặc điểm của các đối tượng hay mối liên hệ giữa các đối tượng.
- **Thể hiện (Individuals):** Các thể hiện là các thành phần cơ bản nền tảng của một Ontology, là đối tượng đại diện cho lớp.
- **Mệnh đề (Axioms):** Mệnh đề là một biểu diễn về mối quan hệ giữa các khái niệm và các vai trò.
- **Ràng buộc (Restrictions):** Cung cấp mô tả về những ràng buộc của các thuộc

tính trong Ontology.

- **Quan hệ (Relations):** Cung cấp mô tả về mối liên hệ giữa các đối tượng trong Ontology.

Khả năng diễn đạt của RDF và RDF Schema vẫn còn hạn chế, việc giao tiếp trên Web ngữ nghĩa sẽ đòi hỏi nhiều mô tả phức tạp hơn RDF và RDF Schema cung cấp. Điều này dẫn đến nhu cầu tìm ra một ngôn ngữ chung, phong phú hơn để mô tả Ontology. Năm 2001, OWL (Ontology Web Language) được DAML+OIL phát triển và xây dựng dựa trên cơ sở RDF để biểu diễn những tri thức đa dạng và phức tạp về các đối tượng, các lớp, và các mối quan hệ. OWL trở thành quy chuẩn W3C năm 2004, phiên bản tiếp theo là OWL 2 được phát hành vào năm 2009 và cập nhật phiên bản 2 năm 2012 [8]. W3C chia OWL thành ba tầng/ngôn ngữ, mỗi ngôn ngữ nhỏ hướng tới đáp ứng các nhu cầu sử dụng khác nhau.

- **OWL Full:** Sử dụng tất cả các thành phần của OWL, cho phép các thành phần này kết hợp tùy ý với RDF và lược đồ RDF. OWL Full hoàn toàn tương thích với RDF, về cả cú pháp và ngữ nghĩa, khả năng diễn đạt mạnh, phạm vi suy diễn vô cùng lớn.
- **OWL DL (Description Logic):** Là một tập con của OWL Full, giới hạn ứng dụng các thành phần từ OWL và RDF, hỗ trợ suy diễn hiệu quả nhưng mất tính tương thích với RDF.
- **OWL Lite:** Tập con của OWL DL, với nhiều giới hạn hơn nữa, đơn giản đối với người dùng nhưng khả năng diễn đạt hạn chế.

OWL sử dụng cú pháp RDF/XML và có ngữ nghĩa hình thức được xây dựng trên cơ sở lô-gic bậc nhất. Một số thành phần cơ bản trong tài liệu OWL:

- **Header**

Phần tử gốc của tài liệu OWL là *rdf:RDF*, chứa các khai báo về các thuật ngữ sẽ sử dụng.

```
<rdf:RDF
    xmlns:owl = "http://www.w3.org/2002/07/owl#"
    xmlns:rdf = "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:rdfs = "http://www.w3.org/2000/01/rdf-schema#"
    xmlns:xsd = "http://www.w3.org/2001/XMLSchema#">
```

Ontology header *owl:Ontology* mô tả trạng thái của Ontology hiện tại, bao gồm các thông tin như nhãn, phiên bản, liên kết với Ontology khác,...

```
<owl:Ontology rdf:about="">
    <owl:versionInfo> ... </owl:versionInfo>
```

```

<rdfs :comment>... </ rdfs :comment>
<owl :imports rdf :resource = "... " />
<rdfs :label>... </ rdfs :label>
</owl :Ontology>

```

- **Class**

Các lớp được định nghĩa bằng cú pháp *owl:Class*. Chúng ta có thể tạo các quan hệ cha con giữa các lớp thông qua thuộc tính *rdfs:subClassOf*. Tất cả các lớp owl đều là *rdfs:subClassOf* của *owl:Thing*, *owl:Nothing* là *rdfs:subClassOf* của tất cả các lớp owl.

Ví dụ:

```

<owl :Class rdf :about = "#quanCafe">
    <rdfs :subClassOf rdf :resource = "#diemAnUong" />
</owl :Class>

```

- **Property**

OWL có hai loại thuộc tính:

- Object properties: Thuộc tính đối tượng gắn các đối tượng với các đối tượng khác. Ví dụ:

```

<owl :ObjectProperty rdf :about = "#thuocQuanHuyen">
    <rdfs :domain rdf :resource = "#diemDuLich" />
    <rdfs :range rdf :resource = "#quanHuyen" />
</owl :ObjectProperty>

```

- Datatype properties: Thuộc tính kiểu dữ liệu gắn các đối tượng với kiểu dữ liệu. Ví dụ:

```

<owl :DatatypeProperty rdf :about = "#tenQuanHuyen">
    <rdfs :domain rdf :resource = "#quanHuyen" />
    <rdfs :range rdf :resource =
        " http ://www.w3.org/2001/XMLSchema#string " />
</owl :DatatypeProperty>

```

### 2.1.6 Ngôn ngữ truy vấn SPARQL

SPARQL (SPARQL Protocol And RDF Query Language) là một ngôn ngữ truy vấn RDF, tên của nó là viết tắt của giao thức SPARQL và ngôn ngữ truy vấn RDF. SPARQL được tạo ra là một chuẩn để truy cập dữ liệu RDF và được coi là một trong những công nghệ chủ chốt của Semantic Web [9].

Ngôn ngữ SPARQL có bốn hình thức truy vấn khác nhau cho các mục đích khác nhau:

Truy vấn **SELECT**: Sử dụng để trích xuất dữ liệu từ cơ sở tri thức, các kết quả được trả về trong định dạng bảng.

Truy vấn **CONSTRUCT**: Sử dụng để tổng hợp bộ ba (rdf triples), kết quả trả về là đồ thị RDF.

Truy vấn **ASK**: Sử dụng để kiểm tra xem có hay không bộ ba theo mẫu, kết quả dạng trả về là True/False.

Truy vấn **DESCRIBE**: Sử dụng để trích xuất một đồ thị RDF mô tả các tài nguyên khớp với biến được cho.

Mỗi dạng truy vấn đều dùng khối lệnh bên trong từ khóa WHERE để hạn chế truy vấn mặc dù trường hợp truy vấn DESCRIBE từ khóa WHERE là tùy chọn.

Cấu trúc của câu truy vấn:

Cú pháp	Mô tả
PREFIX abc: <...> PREFIX xyz: <...>	Khai báo tiền tố (không bắt buộc)
SELECT CONSTRUCT ASK DESCRIBE	Biểu thức kết quả
FROM <...> FROM NAMED <...>	Khai báo tập dữ liệu (không bắt buộc)
WHERE { ... }	Mẫu truy vấn
GROUP BY ... HAVING ... ORDER BY ... LIMIT ... OFFSET ... VALUES ...	Hiệu chỉnh kết quả (không bắt buộc)

Ví dụ câu truy vấn tên, địa chỉ và tọa độ của chợ:

```
SELECT DISTINCT ?name ?location ?lat ?lng
WHERE
{
?x ex:tenCho ?name;
ex:diaChi ?location ;
ex:lat ?lat ;
ex:lng ?lng .
```

}

## 2.2 Chatbot

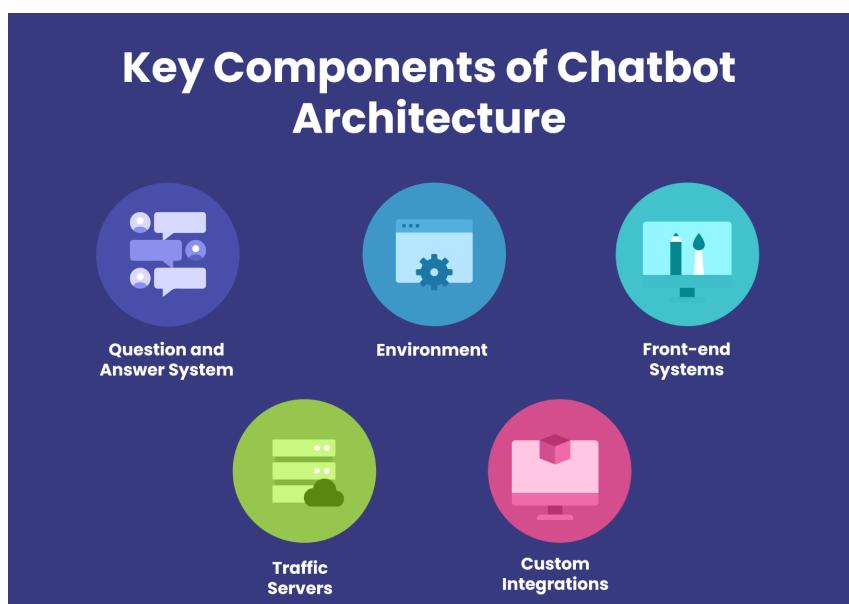
### 2.2.1 Giới thiệu chatbot

Chatbot là một chương trình máy tính mà con người dùng có thể giao tiếp với máy thông qua các ứng dụng nhắn tin. Chatbot sẽ sử dụng trí tuệ nhân tạo và xử lý ngôn ngữ tự nhiên để phân tích những gì con người nói hoặc nhập vào, xác định yêu cầu được đưa ra và phản hồi lại.

Ngày nay, công nghệ chatbot hầu như có mặt ở khắp mọi nơi, một số ứng dụng phổ biến của chatbot có thể kể đến như trợ lý ảo (Siri của Apple, Trợ lý Google, Amazon Alexa,...), bán hàng, chăm sóc khách hàng, cập nhật tin tức, tìm kiếm thông tin, thanh toán trực tuyến,... Nhờ khả năng hoạt động 24/7, tự động hóa các quy trình, công suất xử lý lớn, chatbot giúp cho con người tiết kiệm đáng kể thời gian, chi phí và nâng cao năng suất lao động.

### 2.2.2 Các thành phần của chatbot

Kiến trúc của chatbot phụ thuộc vào nhiều yếu tố khác nhau như trường hợp sử dụng, miền, loại chatbot,... Tuy nhiên, quy trình hội thoại cơ bản vẫn giống nhau. Dưới đây là một số thành phần quan trọng của chatbot:



**Hình 2.4:** Các thành phần cơ bản của chatbot [10]

- **Question and Answer System**

Hệ thống hỏi đáp có nhiệm vụ trả lời các câu hỏi của người dùng. Hệ thống này bao gồm các thành phần: huấn luyện thủ công (manual training) và huấn luyện tự động (automated training).

Huấn luyện thủ công là biên soạn danh sách câu hỏi và câu trả lời tương ứng giúp chatbot nhanh chóng xác định câu trả lời của các câu hỏi có liên quan nhất.

Huấn luyện tự động là tải lên các tài liệu nghiệp vụ, chẳng hạn như tài liệu về chính sách và các loại hỏi đáp khác, đồng thời hướng dẫn bot tự huấn luyện. Từ những tài liệu này, hệ thống tạo ra một danh sách các câu hỏi và câu trả lời. Nhờ vậy mà chatbot có thể trả lời chúng một cách trôi chảy và nhanh chóng [10].

#### • Environment

Môi trường chịu trách nhiệm chính trong việc diễn giải ngữ cảnh trong thông điệp của người dùng bằng cách sử dụng xử lý ngôn ngữ tự nhiên (NLP).

NLP Engine là thành phần cốt lõi của kiến trúc chatbot. Nó diễn giải những gì người dùng nói và chuyển đổi thành các đầu vào có cấu trúc mà hệ thống có thể xử lý. Công cụ NLP sử dụng các thuật toán học máy nâng cao để xác định ý định của người dùng và sau đó đối sánh nó với danh sách ý định được hỗ trợ của bot.

NLP Engine có hai thành phần:

- Phân loại ý định (intent classifier): Công cụ phân loại ý định ánh xạ thông tin người dùng yêu cầu và loại hành động được thực hiện bởi phần mềm.
- Trích xuất thực thể (entity extractor): Công cụ trích xuất thực thể chịu trách nhiệm xác định các từ khóa quan trọng từ truy vấn của người dùng nhằm xác định những gì người dùng đang tìm kiếm.

Ngoài ra, NLP cũng có thể được mở rộng để bao gồm cơ chế phản hồi và học có nguyên tắc giúp chatbot giao tiếp hiệu quả theo nhiều ngữ cảnh khác nhau.

- Cơ chế phản hồi (feedback mechanism): Hệ thống nhận phản hồi từ người dùng để đánh giá xem bot có hoạt động hiệu quả trong cuộc trò chuyện hay không và người dùng có hài lòng với phản hồi của bot hay không. Phần học này có thể được tích hợp vào chính chatbot. Điều này thúc đẩy bot học hỏi từ những sai lầm và tự sửa chữa trong các cuộc trò chuyện trong tương lai.
- Học có nguyên tắc (policy learning): Nguyên tắc giúp chatbot đem lại trải nghiệm hài lòng tổng thể cho người dùng cuối thông qua việc tạo ra các chuỗi hành động hợp lý để điều hướng cuộc trò chuyện một cách mạch lạc và thỏa mãn người dùng. Chatbot cố gắng học hỏi từ sự tương tác trước

đó của người dùng và theo dõi luồng trò chuyện mà nó đã có với những người dùng tương tự trong quá khứ. Từ đó chatbot lưu trữ thông tin, hiểu rõ về hành vi và phương thức giao tiếp của người dùng để hoạt động hiệu quả.

- **Front-End Systems**

Hệ thống mà người dùng tương tác với chatbot. Một số hệ thống phổ biến như: Facebook Messenger, WhatsApp Business, Slack, Google Hangouts, trang web hoặc ứng dụng di động,...

- **Traffic Server**

Máy chủ xử lý các yêu cầu lưu lượng truy cập của người dùng và định tuyến chúng đến các thành phần thích hợp. Phản hồi từ các thành phần bên trong thường được chuyển qua máy chủ lưu lượng đến Front-End Systems.

- **Custom Integrations**

Chatbot có thể được tích hợp với các hệ thống phụ trợ như CRM, cơ sở dữ liệu, ứng dụng thanh toán, lịch và nhiều công cụ khác, để nâng cao khả năng của chatbot, tăng trải nghiệm người dùng.

### **2.2.3 Xây dựng Chatbot**

Quá trình xây dựng Chatbot gồm các bước: **Xác định ý định người dùng:** Xác định ý định người dùng (Intent) khi tương tác với hệ thống thông qua câu hỏi hay câu truy vấn của người dùng.

**Trích xuất thông tin:** Trích xuất các thông tin liên quan đến đối tượng, bổ nghĩa cho một ý định qua việc nhận biết các Entity. Danh sách các entity đã được khai báo và huấn luyện từ trước.

Ví dụ: Người dùng có ý định hỏi "thông tin của danh lam thăng cảnh", intent này có thể thể hiện qua một số câu hỏi như :"thông tin Thành Cổ Loa?", "Thành Cổ Loa là gì?", "Thành Cổ Loa ở đâu?",... Trong câu trên "Thành Cổ Loa" đóng vai trò là thể hiện của entity "địa điểm du lịch", từ đó hệ thống có thể truy xuất ra thông tin liên quan đến địa điểm

**Quản lý hội thoại:** Quản lý hội thoại để đảm bảo việc trao đổi giữa người và máy được thông suốt. Dựa trên đầu vào của người dùng (câu nói hoặc đoạn chat), mô đun NLU tạo ra biểu diễn ngữ nghĩa cho đầu vào đó. Sau đó mô đun quản lý hội thoại chịu trách nhiệm điều chỉnh toàn bộ cấu trúc của cuộc đối thoại, nhận đầu vào từ mô đun NLU, duy trì một số loại trạng thái, xử lý logic và quyết định hành động tiếp theo cần phản hồi lại cho người dùng là gì, sau đó chuyển đầu ra cho mô

đun NLG. Thông thường, các kiến trúc dựa trên frame-based hoặc finite state được sử dụng bởi hầu hết các mô đun quản lý hội thoại.

#### 2.2.4 Chatbot ứng dụng công nghệ Semantic Web

Một số ưu điểm khi sử dụng Ontology ứng dụng vào Chatbot

- **Khả năng tìm kiếm:** Ontology biểu diễn các khái niệm trong các lĩnh vực khác nhau và tập trung vào mối quan hệ giữa chúng. Những khái niệm này có thể được sử dụng trực tiếp trong chatbot, để tìm ra các từ ghép, từ đồng nghĩa và các quan hệ khác giữa các khái niệm. Dựa vào nguồn dữ liệu phong phú, các bot tìm kiếm có thể cung cấp nội dung phù hợp nhất cho người. Semantic web không chỉ cải thiện tìm kiếm truyền thống mà còn tạo điều kiện cho hành trình trải nghiệm người dùng thông minh và liền mạch hơn.
- **Khả năng suy diễn:** Ontology có lợi thế là các khái niệm được kết nối với nhau thành một biểu đồ, cho phép máy tính tìm kiếm và sử dụng các quy tắc lập luận hay thậm chí là suy diễn để tìm ra các lập luận mới. Nhờ vậy, chatbot có thể khám phá các đường dẫn hội thoại một cách linh hoạt hơn, tăng khả năng suy diễn ngữ nghĩa cho Chatbot.
- **Dữ liệu phân cấp:** Trong Ontology, các khái niệm được nhóm lại với nhau thông qua các nhãn cho phép phân biệt các cấp độ tri thức khác nhau, từ đó chatbot có thể dễ dàng tham vấn vào từng mức độ tùy theo yêu cầu sử dụng.
- **Cú pháp đơn giản:** Cơ sở tri thức có tính linh hoạt, mềm dẻo, dễ dàng thực hiện các thao tác tìm kiếm, thêm, sửa, xóa.

Ví dụ: Khi người dùng muốn tìm các địa điểm du lịch và khách sạn ở cùng một huyện, thay vì phải liên kết các bảng với nhau như trong cơ sở dữ liệu quan hệ, với SPARQL ta chỉ cần truy vấn đơn giản:

```

SELECT DISTINCT ?name ?nameHo
WHERE
{
  ?x :tenQuanHuyen ?nameDi .
  ?y :tenDiemDuLich ?name ;
  :thuocQuanHuyen ?x .
  ?z :tenKhachSan ?nameHo ;
  :thuocQuanHuyen ?x .

  FILTER( regex( lcase(?nameDi) , "huyen Quoc Oai" ) )
}

```

Ngoài ra khi người dùng muốn xóa một bảng trong cơ sở dữ liệu quan hệ, nếu bảng đó có liên kết khóa ngoại từ một bảng khác thì cần thực hiện xóa khóa

ngoại trước hoặc cần lần lượt xóa các bảng theo một thứ tự nhất định. Trong Ontology, thao tác này được thực hiện dễ dàng hơn nhiều, khi người dùng xóa một thể hiện, các thể hiện có quan hệ với nó sẽ tự động xóa các thuộc tính quan hệ mà không gặp bất kỳ lỗi, xung đột nào.

- **Khả năng chia sẻ:** Dễ dàng chia sẻ thông tin giữa các cơ sở tri thức không đồng nhất nhờ cấu trúc thông tin độc lập không có các ràng buộc như cơ sở dữ liệu quan hệ.

## CHƯƠNG 3. PHÂN TÍCH VÀ THIẾT KẾ HỆ THỐNG

### 3.1 Tổng quan giải pháp

Các bước xây dựng hệ thống:

Bước 1: Tìm kiếm dữ liệu về du lịch Hà Nội từ các trang web như Wikipedia, Google Travel hay các trang web du lịch Hà Nội, tiến hành Crawl dữ liệu và lưu lại dưới dạng tệp excel.

Bước 2: Dựa trên dữ liệu thu thập được ở bước 1, thiết kế và xây dựng Ontology về du lịch Hà Nội.

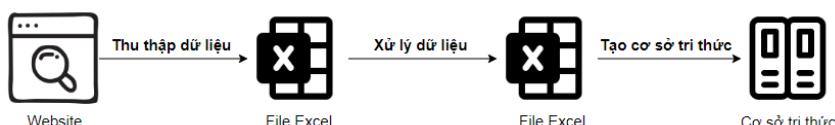
Bước 3: Tiến hành chuyển đổi dữ liệu đã thu thập được ở bước 1 sang dạng dữ liệu ngữ nghĩa theo Ontology đã xây dựng ở bước 2.

Bước 4: Thiết kế và xây dựng Chatbot hỏi đáp về du lịch Hà Nội. Tiến hành xác định ý định, trích rút thông tin từ truy vấn của người dùng, sử dụng SPARQL để truy vấn Ontology đã xây dựng ở bước 3 và phản hồi lại người dùng.

Bước 5: Xây dựng trang web biểu diễn ontology.

### 3.2 Kiến trúc hệ thống

#### 3.2.1 Các thành phần xây dựng cơ sở tri thức



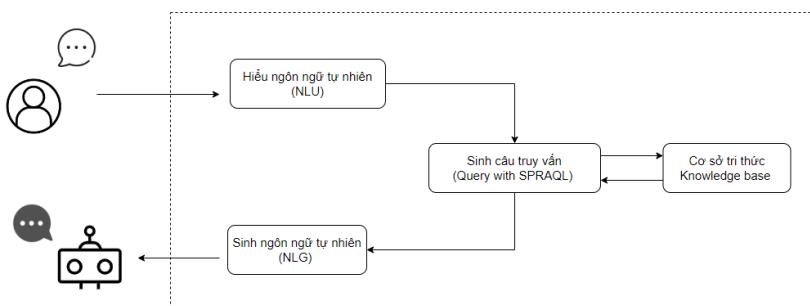
Hình 3.1: Quá trình xây dựng cơ sở tri thức

Thu thập dữ liệu: Tìm kiếm các thông tin, bài viết trên các trang web như Wikipedia, Google Travel hay các trang web du lịch Hà Nội sau đó lưu trữ dữ liệu dưới dạng bảng trong excel sử dụng thư viện Python Selenium.

Xử lý dữ liệu: Sau khi thu thập dữ liệu từ các trang web, tiến hành chỉnh sửa các thông tin sai, trích xuất ra các đối tượng cần thiết.

Tạo cơ sở tri thức: Các đối tượng sau khi trích xuất được chuyển đổi sang dạng bộ ba RDF dựa trên Ontology đã thiết kế và tạo ra các thể hiện của Ontology, tập hợp tất cả các bộ ba RDF tạo thành cơ sở tri thức.

### 3.2.2 Các thành phần xây dựng chatbot



Hình 3.2: Sơ đồ kiến trúc hệ thống chatbot

**Hiểu ngôn ngữ tự nhiên (NLU):** Sử dụng LUIS (Language Understanding Intelligent Service) huấn luyện bot để nhận diện ý định (intent) và xác định các thực thể (entity).

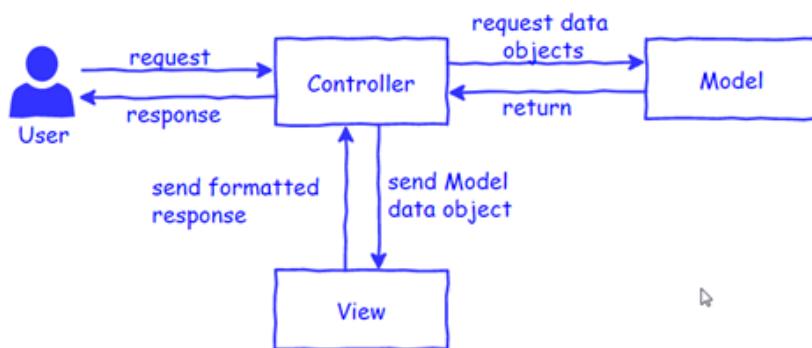
**Sinh ngôn ngữ tự nhiên (NLG):** NLG (Natural Language Generation) chuyển đổi dữ liệu có cấu trúc thành ngôn ngữ tự nhiên, tự động tạo phản hồi cho người dùng dựa vào các lớp, thuộc tính và thể hiện trong Ontology.

**Truy vấn SPARQL:** Sử dụng thư viện RDFLib, một thư viện Python hỗ trợ làm việc với RDF để truy vấn SPARQL.

**Cơ sở tri thức:** Ontology sẽ được xây dựng theo các bước mô tả ở phần 3.2.1.

### 3.2.3 Các thành phần xây dựng website

Website thiết kế theo kiến trúc MVC (Model – View – Controller). MVC là một mẫu thiết kế (design pattern) được sử dụng phổ biến để tạo cấu trúc cho nhiều trang web, ứng dụng. MVC chia một ứng dụng thành ba phần tương tác được với nhau để tách biệt giữa cách thức mà thông tin được xử lý nội hàm và phần thông tin được trình bày và tiếp nhận từ phía người dùng [11].



Hình 3.3: Mô hình MVC [12]

**Controller:** Quản lý các tương tác từ người dùng để xử lý các thao tác logic liên

quan.

**Model:** Thiết lập logic cho dữ liệu của ứng dụng, chứa các phương thức xử lý dữ liệu, truy xuất dữ liệu.

**View:** Các thành phần dùng để hiển thị giao diện người dùng, nhận tương tác trực tiếp từ người dùng, view thường được tạo dựa vào thông tin dữ liệu model.

### 3.3 Xây dựng cơ sở tri thức về du lịch ở Hà Nội

#### 3.3.1 Thu thập dữ liệu

Thu thập dữ liệu là một giai đoạn có ý nghĩa vô cùng quan trọng đối với quá trình nghiên cứu khoa học. Việc thu thập dữ liệu thường tốn nhiều thời gian, công sức và chi phí. Dữ liệu được thu thập cần đảm bảo tính chính xác và trung thực, số lượng và chất lượng để duy trì tính toàn vẹn của nghiên cứu.

Để đảm bảo các yêu cầu trên, tôi đã lựa chọn thu thập khoảng 75 bài viết về các địa điểm du lịch được đăng trên cổng thông tin điện tử Wikipedia cùng với dữ liệu về lễ hội, đặc sản, khách sạn, siêu thị, quán ăn,...trên địa bàn thành phố Hà Nội. Ngoài ra thực hiện thu thập bài viết từ một vài trang web khác trên Internet. Các nguồn dữ liệu chính được sử dụng:

[https://vi.wikipedia.org/wiki/Du\\_lịch\\_Hà\\_Nội](https://vi.wikipedia.org/wiki/Du_lịch_Hà_Nội)

<https://www.google.com/travel/>

Dữ liệu sau khi thu thập được lưu trữ dưới dạng bảng trong excel.

A	B	C	D	E	F	G
1	tên quán ăn	địa chỉ	sdt	mở cửa		
2	0 Bánh Dày Ngon	28 Ng. 580 Đ. Trường C		Thứ Sáu 06:15–22:45		
3	1 Pao Quán Hà Trì Hà Đô Ngã 5 Hà Trì, Hà Cầu, H 098 255 89 46			Thứ Sáu 10:00–22:30		
4	2 Quán Ăn Ngon	18 Phan Bội Châu, Cửa 090 212 69 63		Thứ Sáu 06:45–22:00		
5	3 Quán Kiến 100	100 P. Tuệ Tĩnh, Nguyễn 097 839 99 83		Thứ Sáu 10:00–23:00		
6	4 Hẻm Quán - Hoàng Cầu 161 P. Hoàng Cầu, Chợ 024 3538 1636			Thứ Sáu 10:30–22:30		
7	5 Phủi Quán	17A P. Nguyễn Đình C 0888 775 555				
8	6 Nhà Hàng Lương Sơn Q.1 P. Quan Nhân, Nhân 093 771 66 88			Thứ Sáu 10:00–23:00		
9	7 Ngoại Ô Quán - Ấm thư 19 P. Nguyễn Gia Thiều 094 286 81 89			Thứ Sáu 07:00–21:00		
10	8 Cơm Niêu Tố Uyên	101C2 P. Phạm Ngọc Th 024 3852 6662		Thứ Sáu 10:30–21:00		
11	9 Chic Restaurant	2RHP+GV6, Phố Nguyễn 098 358 86 86		Thứ Sáu 06:00–22:00		
12	10 Nhà hàng quán ăn ngon 14 P. Đặng Tiến Đông, 094 550 21 25			Thứ Sáu 09:00–23:00		
13	11 Hẻm Quán - 360 Xã Đà 360 P. Xã Đà, Phương 024 3559 1616			Thứ Sáu 10:30–22:30		
14	12 Tiêu Dao Quán	51 P. Hoàng Cầu, Chợ 1090 490 90 80		Thứ Sáu 09:30–22:00		
15	13 Bánh Cuốn Bà Hoành	66 P. Tô Hiến Thành, B 098 908 35 70		Thứ Sáu 06:00–20:00		
16	14 Phủi quán	08 Ng. Hàng Bột, Cát Li 0888 775 555		Thứ Sáu 10:00–23:00		

**Hình 3.4:** Một số dữ liệu quán ăn thu thập được

#### 3.3.2 Xử lý dữ liệu

Dữ liệu sau khi thu thập cần trích rút để lấy các thực thể cần thiết. Cột địa chỉ cần tách ra để lấy thông tin phường/ xã và quận/ huyện/ thị xã.

Sử dụng tiện ích mở rộng của Excel - Geocoding by SmartMonkey để trích xuất kinh độ và vĩ độ từ địa chỉ ban đầu.

	A Address	B Country	C Latitude	D Longitude	E Address found
1					
2	52 phố Đặng Thai Mai, phường Quảng An, Tây Hồ, thành phố Hà Nội		21,0639483	105,8257028	52 P. Đặng Thai Mai, Quảng An, Tây Hồ, Hà Nội, Vietnam
3	Xã Phù Đổng, huyện Gia Lâm, thành phố Hà Nội		21,060339	105,9641124	Phù Đổng, Gia Lâm, Hà Nội, Vietnam
4	Xã Phù Linh, huyện Sóc Sơn, Hà Nội		21,2802077	105,8466437	Phù Linh, Sóc Sơn, Hà Nội, Vietnam
5	2RW6+GH8, Hoàng Hoa Thám, Bưởi, Tây Hồ, Hà Nội		21,0419911	105,8205621	Đ. Hoàng Hoa Thám, Tây Hồ, Hà Nội, Vietnam
6	Làng Ngãi Cầu, xã An Khánh, huyện Hoài Đức, Hà Nội		20,9889673	105,7194391	ĐT72, An Khánh, Hoài Đức, Hà Nội, Vietnam
7	xã Cổ Loa, Đông Anh, Hà Nội		21,1156865	105,8701316	Cổ Loa, Đông Anh, Hà Nội, Vietnam
8	phường Láng Thượng, quận Đống Đa, thành phố Hà Nội		21,0220022	105,8040796	Láng Thượng, Đống Đa, Hanoi, Vietnam
9	Xã Hương Sơn, huyện Mỹ Đức, Hà Nội		20,6041258	105,7644596	Hương Sơn, Mỹ Đức, Hà Nội, Vietnam
10	Đèn Voi Phục, làng Thủ Lệ, phường Ngọc Khánh, quận Ba Đình, Hà Nội.		21,0292133	105,8046769	2RH3+WV4, 306B P. Kim Mã, Ngọc Khánh, Ba Đình, Hà Nội, Vietnam
11	Đèn - Đinh Kim Liên (phường Phương Liên, Đống Đa, Hà Nội)		21,0106528	105,8382111	148 P. Kim Hoa, Phương Liên, Đống Đa, Hà Nội, Vietnam
12	Gò Đống Đa, Phường Quang Trung, Quận Đống Đa, Hà Nội		21,011981	105,824451	Đống Đa Mound, Quang Trung, Đống Đa, Hanoi, Vietnam
13	Xã Sài Sơn, huyện Quốc Oai, Hà Nội		21,032456	105,6471178	Sài Sơn, Quốc Oai, Hà Nội, Vietnam
14	Định Bát Tràng, xã Bát Tràng, Gia Lâm, Hà Nội		20,9765119	105,9097294	XWG5+JV5, Bát Tràng, Gia Lâm, Hanoi, Vietnam
15	Thôn Hạ Lôi, xã Mê Linh, huyện Mê Linh, Hà Nội		21,1559316	105,7365841	Hạ Lôi, Văn Khê, Mê Linh, Hà Nội, Vietnam
16	Định Đại Đô, làng Đại Đô, xã Võng La, huyện Đông Anh, Hà Nội		21,1025608	105,7644596	Võng La, Đông Anh, Hanoi, Vietnam

+    Geocoding Details    Trang tính1    

**Hình 3.5:** Geocoding by SmartMonkey

### 3.3.3 Xây dựng Ontology

#### a, Đặc tả Ontology

Ontology mô tả một cách có ngữ nghĩa các thông tin liên quan đến các du lịch ở Hà Nội. Ontology hướng đến đối tượng là tất cả những ai có mong muốn tìm hiểu về du lịch ở Hà Nội, hay muốn đi du lịch ở Hà Nội. Phạm vi của Ontology sẽ được xây dựng bao gồm thông tin của địa điểm du lịch, lễ hội, đặc sản, khách sạn, chợ, siêu thị, quán ăn, quán bar, quán cafe... Trong tương lai, Ontology có khả năng mở rộng với nhiều dữ liệu hơn, tăng phạm vi tìm kiếm và phục vụ được nhiều đối tượng hơn.

#### b, Lớp trong Ontology

Ontology về du lịch được xây dựng, định nghĩa thông qua các lớp sau:

- Lớp **diemDuLich**:

- Lớp **diemDuLich** là lớp lưu trữ dữ liệu về các địa điểm du lịch.
- Lớp được gán nhãn: <rdfs:label>điểm du lịch</rdfs:label>

- Lớp **leHoi**:

- Lớp **leHoi** là lớp lưu trữ dữ liệu về các lễ hội ở Hà Nội.
- Lớp được gán nhãn: <rdfs:label>lễ hội</rdfs:label>

- Lớp **dacSan**:

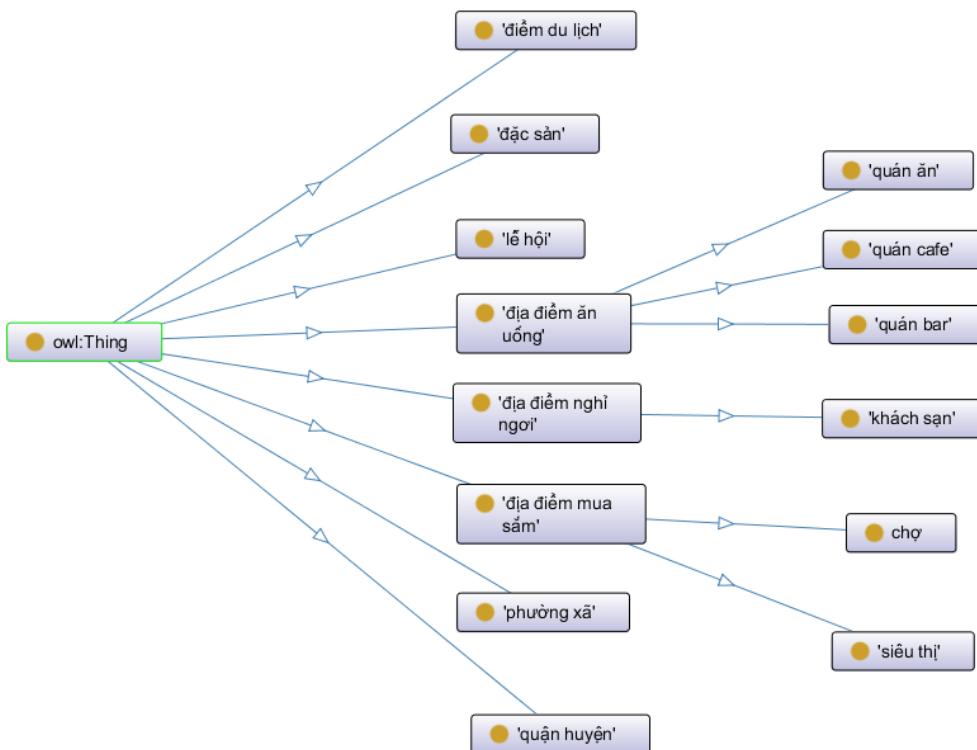
- Lớp **dacSan** là lớp lưu trữ dữ liệu về các đặc sản ở Hà Nội.
- Lớp được gán nhãn: <rdfs:label>đặc sản</rdfs:label>

- Lớp **phuongXa**:

- Lớp **phuongXa** là lớp lưu trữ dữ liệu về các phường, xã ở Hà Nội.

- Lớp được gán nhãn: <rdfs:label>phường xã</rdfs:label>
- Lớp **quanHuyen**:
  - Lớp **quanHuyen** là lớp lưu trữ dữ liệu về các quận, huyện, thị xã ở Hà Nội.
  - Lớp được gán nhãn: <rdfs:label>quận huyện</rdfs:label>
- Lớp **diemMuasam**:
  - Lớp **diemMuasam** là lớp lưu trữ dữ liệu về các địa điểm phục vụ cho việc mua sắm.
  - Lớp được gán nhãn: <rdfs:label>địa điểm mua sắm</rdfs:label>
- Lớp **diemAnUong**:
  - Lớp **diemAnUong** là lớp lưu trữ dữ liệu về các địa điểm phục vụ cho việc ăn uống.
  - Lớp được gán nhãn: <rdfs:label>địa điểm ăn uống</rdfs:label>
- Lớp **diemNghiNgoi**:
  - Lớp **diemNghiNgoi** là lớp lưu trữ dữ liệu về các địa điểm phục vụ cho việc nghỉ ngơi.
  - Lớp được gán nhãn: <rdfs:label>địa điểm nghỉ ngơi</rdfs:label>
- Lớp **cho**:
  - Lớp **cho** là lớp con của lớp diemMuasam và lưu trữ thông tin về các chợ ở Hà Nội.
  - Lớp được gán nhãn: <rdfs:label> chợ</rdfs:label>
- Lớp **sieuThi**:
  - Lớp **sieuThi** là lớp con của lớp diemMuasam và lưu trữ thông tin về các siêu thị ở Hà Nội.
  - Lớp được gán nhãn: <rdfs:label> siêu thi</rdfs:label>
- Lớp **quanAn**:
  - Lớp **quanAn** là lớp con của lớp diemAnUong và lưu trữ thông tin về các nhà hàng, quán ăn ở Hà Nội.
  - Lớp được gán nhãn: <rdfs:label>quán ăn</rdfs:label>
- Lớp **quanBar**:

- Lớp **quanBar** là lớp con của lớp diemAnUong và lưu trữ thông tin về các quán bar ở Hà Nội.
- Lớp được gán nhãn: <rdfs:label>quán bar</rdfs:label>
- Lớp **quanCafe**:
  - Lớp **quanCafe** là lớp con của lớp diemAnUong và lưu trữ thông tin về các quán cafe ở Hà Nội.
  - Lớp được gán nhãn: <rdfs:label>quán cafe</rdfs:label>
- Lớp **khachSan**:
  - Lớp **khachSan** là lớp con của lớp diemNghiNgoi và lưu trữ thông tin về các khách sạn ở Hà Nội.
  - Lớp được gán nhãn: <rdfs:label>khách sạn</rdfs:label>



**Hình 3.6:** Sơ đồ các lớp trong Ontology

### c, Thuộc tính trong Ontology

#### *Object properties:*

- Thuộc tính **thuocPhuongXa**:
  - Domain: diemDuLich, leHoi, dacSan, khachSan, cho, sieuThi, quanAn, quanBar, quanCafe

- Range: phuongXa
- Thuộc tính **thuocQuanHuyen**:
  - Domain: diemDuLich, leHoi, dacSan, khachSan, cho, sieuThi, quanAn, quanBar, quanCafe, phuongXa
  - Range: quanHuyen
- Thuộc tính **coLeHoi**:
  - Domain: diemDuLich
  - Range: leHoi
- Thuộc tính **coDacSan**:
  - Domain: diemDuLich
  - Range: dacSan
- Thuộc tính **taiDiemDuLich**:
  - Domain: leHoi, dacSan
  - Range: diemDuLich

*Data properties:*

- Domain: diemDuLich - Các thuộc tính mô tả thông tin về địa điểm gồm
  - Thuộc tính **tenDiemDuLich**: tên địa điểm du lịch.
  - Thuộc tính **tenKhac**: tên gọi khác.
  - Thuộc tính **moTa**: giới thiệu về địa điểm du lịch.
  - Thuộc tính **diaChi**: địa chỉ của địa điểm du lịch.
  - Thuộc tính **loaiDiaDiem**: loại địa điểm du lịch (bảo tàng, di tích lịch sử, hồ,...).
  - Thuộc tính **thoiGian**: thời gian đóng mở cửa.
  - Thuộc tính **gia**: giá vé tham quan.
  - Thuộc tính **lat**: vĩ độ của địa điểm du lịch.
  - Thuộc tính **lng**: kinh độ của địa điểm du lịch.
  - Thuộc tính **sdt**: số điện thoại liên hệ.
  - Thuộc tính **mail**: địa chỉ mail liên hệ.
  - Thuộc tính **website** : địa chỉ website của địa điểm du lịch.

- Domain: leHoi - Các thuộc tính mô tả thông tin về lễ hội gồm
  - Thuộc tính **tenLeHoi**: tên lễ hội.
  - Thuộc tính **moTa**: giới thiệu về lễ hội.
  - Thuộc tính **diaChi**: địa chỉ diễn ra lễ hội.
  - Thuộc tính **thoiGian**: thời gian diễn ra lễ hội.
  - Thuộc tính **lat**: vĩ độ của địa điểm diễn ra lễ hội.
  - Thuộc tính **Ing**: kinh độ của địa điểm diễn ra lễ hội.
- Domain: dacSan - Các thuộc tính mô tả thông tin về đặc sản gồm
  - Thuộc tính **tenDacSan**: tên đặc sản.
  - Thuộc tính **tenKhac**: tên khác của đặc sản.
  - Thuộc tính **moTa**: giới thiệu về đặc sản.
  - Thuộc tính **diaChi**: địa chỉ của đặc sản.
  - Thuộc tính **lat**: vĩ độ của địa điểm có đặc sản.
  - Thuộc tính **Ing**: kinh độ của địa điểm có đặc sản.
- Domain: khachSan - Các thuộc tính mô tả thông tin về khách sạn gồm
  - Thuộc tính **tenKhachSan**: tên khách sạn.
  - Thuộc tính **diaChi**: địa chỉ của khách sạn.
  - Thuộc tính **gia**: giá phòng trung bình trong 1 đêm.
  - Thuộc tính **lat**: vĩ độ của khách sạn.
  - Thuộc tính **Ing**: kinh độ của khách sạn.
  - Thuộc tính **sdt**: số điện thoại liên hệ.
- Domain: cho - Các thuộc tính mô tả thông tin về chợ gồm
  - Thuộc tính **tenCho**: tên chợ.
  - Thuộc tính **diaChi**: địa chỉ của chợ.
  - Thuộc tính **lat**: vĩ độ của chợ.
  - Thuộc tính **Ing**: kinh độ của chợ.
- Domain: sieuThi - Các thuộc tính mô tả thông tin về siêu thị gồm
  - Thuộc tính **tenSieuThi**: tên siêu thị.
  - Thuộc tính **diaChi**: địa chỉ của siêu thị.

- Thuộc tính **lat**: vĩ độ của siêu thị.
- Thuộc tính **lng**: kinh độ của siêu thị.
- Thuộc tính **thoiGian**: thời gian đóng mở cửa.
- Thuộc tính **sdt**: số điện thoại liên hệ.
- Domain: quanAn - Các thuộc tính mô tả thông tin về quán ăn gồm
  - Thuộc tính **tenQuanAn**: tên nhà hàng, quán ăn.
  - Thuộc tính **diaChi**: địa chỉ của quán ăn.
  - Thuộc tính **lat**: vĩ độ của quán ăn.
  - Thuộc tính **lng**: kinh độ của quán ăn.
  - Thuộc tính **thoiGian**: thời gian đóng mở cửa.
  - Thuộc tính **sdt**: số điện thoại liên hệ.
- Domain: quanBar - Các thuộc tính mô tả thông tin về quán bar gồm
  - Thuộc tính **tenQuanBar**: tên quán bar.
  - Thuộc tính **diaChi**: địa chỉ của quán bar.
  - Thuộc tính **lat**: vĩ độ của quán bar.
  - Thuộc tính **lng**: kinh độ của quán bar.
  - Thuộc tính **thoiGian**: thời gian đóng mở cửa.
  - Thuộc tính **sdt**: số điện thoại liên hệ.
- Domain: quanCafe - Các thuộc tính mô tả thông tin về quán cafe gồm
  - Thuộc tính **tenCafe**: tên quán cafe.
  - Thuộc tính **diaChi**: địa chỉ của quán cafe.
  - Thuộc tính **lat**: vĩ độ của quán cafe.
  - Thuộc tính **lng**: kinh độ của quán cafe.
  - Thuộc tính **thoiGian**: thời gian đóng mở cửa.
  - Thuộc tính **sdt**: số điện thoại liên hệ.

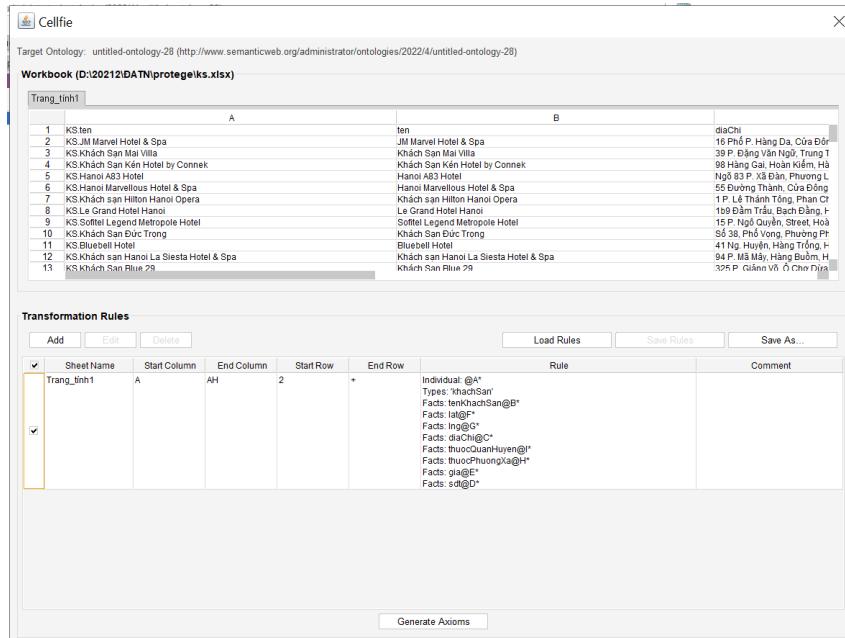
### 3.3.4 Tạo cơ sở tri thức

Dữ liệu về địa điểm du lịch, lễ hội, đặc sản, khách sạn, chợ, siêu thị, quán ăn, quán bar, quán cafe đã được lưu vào các trang bảng tính (các file excel định dạng .xlxs) sau bước xử lý dữ liệu.

A	B	C	D	E	F	G	H	I	J
1	K.ten	ten	diaChi	sdt	gia	lat	Ing	PX.phuongXa	QH.quanHuyen
2	K.SJM	Marvel Hotel	16 Phố P. Hàng Da, Cửa Đông, Hoàn Kiếm, Hà Nội	024 3823 8855	1.318.665 đ	21.0313856	105.8467474	PX.Hàng Da	QH.Hoàn Kiếm
3	K.S.Khach	Khách San Mai	1/39 P. Đặng Văn Ngữ, Trung Tự, Đống Đa, Hà Nội	050.400.500	510.400 đ	21.0082251	105.8320227	PX.Trung Tự	QH.Đống Đa
4	K.S.Khach	Khách San Kén	1/9 Hàng Gai, Hoàn Kiếm, Hà Nội 100000	096 995 34 43	495.001 đ	21.0312773	105.8494042	PX.Hàng Gai	QH.Hoàn Kiếm
5	K.S.Hanoi	Hanoi A83 Hotel	Ngõ 83 P. Xã Đàn, Phường Liễn, Đống Đa, Hà Nội 10000	024 3742 5555	240.000 đ	21.0126934	105.8352371	PX.Phường Liễn	QH.Đống Đa
6	K.S.Hanoi	Hanoi Marvel	55 Đường Thành, Cửa Đông, Hoàn Kiếm, Hà Nội 100000	024 3266 9240	1.091.469 đ	21.0312774	105.8471416	PX.Hàng Bông	QH.Hoàn Kiếm
7	K.S.Khach	Khách sạn Hilton	1 P. Lê Thánh Tôn, Phan Chu Trinh, Hoàn Kiếm, Hà Nội	024 3933 0594	2.114.844 đ	21.0231954	105.8577909	PX.Phản Chu Trinh	QH.Hoàn Kiếm
8	K.S.Le	Le Grand Hotel	1/99 Hàng Trống, Hai Bà Trưng, Hà Nội	024 3984 5350	289.710 đ	21.0119554	105.8652486	PX.Bạch Đằng	QH.Hai Bà Trưng
9	K.S.Sofitel	Sofitel Legend Metropole	15 P. Ngô Quyền, Street, Hoàn Kiếm, Hà Nội 100000	024 3826 6919	5.231.819 đ	21.02559	105.8564147	PX.Tràng Tiền	QH.Hoàn Kiếm
10	K.S.Khach	Khách San Đức	T5 38, Phố Vọng, Phường Phượng Mai, Quận Đống Đa, Hà Nội	098 276 68 23	32.875 đ	20.0291582	105.8494237	PX.Phường Định	QH.Hai Bà Trưng
11	K.S.Bluebell	Bluebell Hotel	41 P. Hàng Bông, Hoàn Kiếm, Hà Nội	097 999 51.23	150.000 đ	21.0291582	105.8494237	PX.Hàng Trống	QH.Hoàn Kiếm
12	K.S.Khach	Khách sạn Handa	9A P. Hàng Mành, Hoàn Kiếm, Hà Nội	024 3926 5441	88.500 đ	21.0084409	105.8529512	PX.Hàng Buồm	QH.Hoàn Kiếm
13	K.S.Khach	Khách San Biển	2335 P. Giải Phóng, Ô Chợ Dừa, Đống Đa, Hà Nội	024 3512 4013	246.628 đ	21.0238363	105.8206156	PX.O Chợ Dừa	QH.Đống Đa
14	K.S.Khach	Khách San Royal	95 P. Hàng Da, Hàng Bông, Hoàn Kiếm, Hà Nội	024 3824 7443	164.826 đ	21.0292317	105.8467342	PX.Hàng Bông	QH.Hoàn Kiếm
15	K.S.A25	Ho A25 Hotel & Spa	68/4 P. Minh Khai, Vinh Phú, Hai Bà Trưng, Hà Nội	024 3987 4720	657.505 đ	21.000662	105.8710033	PX.Vinh Phú	QH.Hai Bà Trưng
16	K.S.ANATO	ANATO HOTEL	26-28-30 P. Nhà Chung, Street, Hoàn Kiếm, Hà Nội	096 157 55 99	1.079.208 đ	21.0282151	105.8497072	PX.Hàng Trống	QH.Hoàn Kiếm

Hình 3.7: Dữ liệu khách sạn trong trang bản tính

Để chuyển đổi dữ liệu từ bảng tính vào Ontology, ta sử dụng Cellfie - Protégé Desktop plugin hỗ trợ đọc dữ liệu từ các trang bảng tính vào cơ sở tri thức dựa trên luật được định nghĩa trước. Ngoài ra, Protégé còn hỗ trợ lưu lại các luật đã tạo dưới dạng file JSON để tái sử dụng trong những lần thêm dữ liệu tiếp theo.



Hình 3.8: Giao diện Cellfie của Protégé

### 3.3.5 Tích hợp Ontology với Linked Open Data

Dữ liệu Liên kết (Linked Data) là tập hợp các nguyên tắc thiết kế cho việc chia sẻ dữ liệu được liên kết với nhau trên Web, và phải đảm bảo máy tính hiểu được. Dữ liệu Mở (Open Data) là dữ liệu mà bất kỳ ai đều có thể sử dụng và phân phối. Các tập hợp dữ liệu vừa mở và vừa có liên kết là Dữ liệu Mở Liên kết (Linked Open Data). Năm 2010, Tim Berners-Lee đã gợi ý mô hình 5 sao cho Dữ liệu Mở Liên kết như sau:

- One-star (\*): Dữ liệu sẵn sàng trên Web, ở bất kỳ định dạng nào, nhưng với một giấy phép mở. Người dùng có thể xem, tìm kiếm, lưu trữ, thay đổi dữ liệu và chia sẻ dữ liệu đó với bất kỳ ai.

- Two-star (\*\*): Dữ liệu mở phải là dữ liệu có cấu trúc mà máy đọc được, ví dụ, định dạng CSV, Json, Excel. Người sử dụng dữ liệu mở 2 sao có thể làm bất kỳ điều gì họ muốn như với dữ liệu mở 1. Hơn nữa, họ có thể xử lý dữ liệu bằng phần mềm độc quyền (Excel) và chuyển đổi nó sang định dạng có cấu trúc khác.
- Three-star (\*\*\*) : Dữ liệu sử dụng các định dạng mở (CSV, JSON) thay vì định dạng độc quyền (Excel).
- Four-star (\*\*\*\*): Dữ liệu mở chỉ dùng các chuẩn của W3C như RDF, SPARQL. Mỗi mục dữ liệu có URI tương ứng.
- Five-star (\*\*\*\*\*): Dữ liệu được liên kết với các nguồn dữ liệu khác. Dữ liệu Mở Liên kết (Linked Open Data) 5 sao là dữ liệu mở sẵn sàng trên Web được kết nối với các dữ liệu khác.

Trong đồ án này, tôi đã lựa chọn tích hợp dữ liệu trong Ontology với dữ liệu mở liên kết 5 sao trên DBpedia - một dự án tận dụng nguồn lực cộng đồng để trích xuất thông tin có cấu trúc từ Wikipedia và làm cho các thông tin này sẵn sàng trên Web để cung cấp cho các ứng dụng khác [13]. Việc liên kết với nguồn dữ liệu mở phong phú như DBpedia sẽ giúp bổ sung và nâng cao chất lượng của cơ sở tri thức.

Đầu tiên, thực hiện truy vấn trên Sparql Endpoint của DBpedia để trích xuất ra các URI cần thiết. Ví dụ câu truy vấn để lấy ra các di tích lịch sử ở Hà Nội.

The screenshot shows the SPARQL Query Editor interface. At the top, there are tabs for 'SPARQL Query Editor', 'About', and 'Tables'. Below that, a 'Default Data Set Name (Graph IRI)' field contains 'http://dbpedia.org'. In the 'Query Text' section, a query is entered:

```
select distinct ?Site where {?Site dbo:wikiPageWikiLink dbc:Historical_sites_in_Hanoi .} LIMIT 100
```

**Hình 3.9:** Truy vấn dữ liệu DBpedia bằng Sparql Endpoint

DBpedia cho phép người dùng lưu kết quả dưới nhiều định dạng như HTML, XML, JSON, Turtle, RDF/XML, N-Triples,... Kết quả của câu truy vấn dưới dạng HTML:

SPARQL   HTML5 table
Site
<a href="http://dbpedia.org/resource/Quán_Thánh_Temple">http://dbpedia.org/resource/Quán_Thánh_Temple</a>
<a href="http://dbpedia.org/resource/Presidential_Palace,_Hanoi">http://dbpedia.org/resource/Presidential_Palace,_Hanoi</a>
<a href="http://dbpedia.org/resource/Hanoi_Ceramic_Mosaic_Mural">http://dbpedia.org/resource/Hanoi_Ceramic_Mosaic_Mural</a>
<a href="http://dbpedia.org/resource/Hanoi_Opera_House">http://dbpedia.org/resource/Hanoi_Opera_House</a>
<a href="http://dbpedia.org/resource/Láng_Pagoda">http://dbpedia.org/resource/Láng_Pagoda</a>
<a href="http://dbpedia.org/resource/Ho_Chí_Minh_Mausoleum">http://dbpedia.org/resource/Ho_Chí_Minh_Mausoleum</a>

**Hình 3.10:** Kết quả truy vấn DBpedia

Sau khi so khớp kết quả, tạo liên kết giữa các mục dữ liệu có liên quan trong Ontology và DBpedia thông qua thuộc tính *owl:sameAs*. Ví dụ địa điểm du lịch "Đền Quán Thánh" sau khi tạo liên kết.

**Hình 3.11:** Kết quả tạo liên kết với DBpedia

## 3.4 Xây dựng hệ thống chatbot

### 3.4.1 Đặc tả usecase

#### a, Chức năng "Xem thông tin"

Mô tả: Chức năng cung cấp thông tin liên quan đến một địa điểm (địa điểm du lịch, khách sạn, chợ, siêu thị, quán ăn, quán bar, quán cafe), lễ hội hay đặc sản.

Các thông tin trên sẽ được hiển thị khi người dùng nhập tên của đối tượng. Đối với những đối tượng có nhiều tên gọi, trong quá trình thu thập thông tin cũng sẽ cố gắng thu thập các tên gọi khác để tăng khả năng tìm kiếm.

Ví dụ: Khi nhập "thông tin **tenDiemDuLich**", kết quả sẽ trả về các thông tin:

- Tên địa điểm du lịch.
- Địa chỉ của địa điểm du lịch.
- Mô tả về địa điểm du lịch.

#### **b, Chức năng "Tìm kiếm thông tin"**

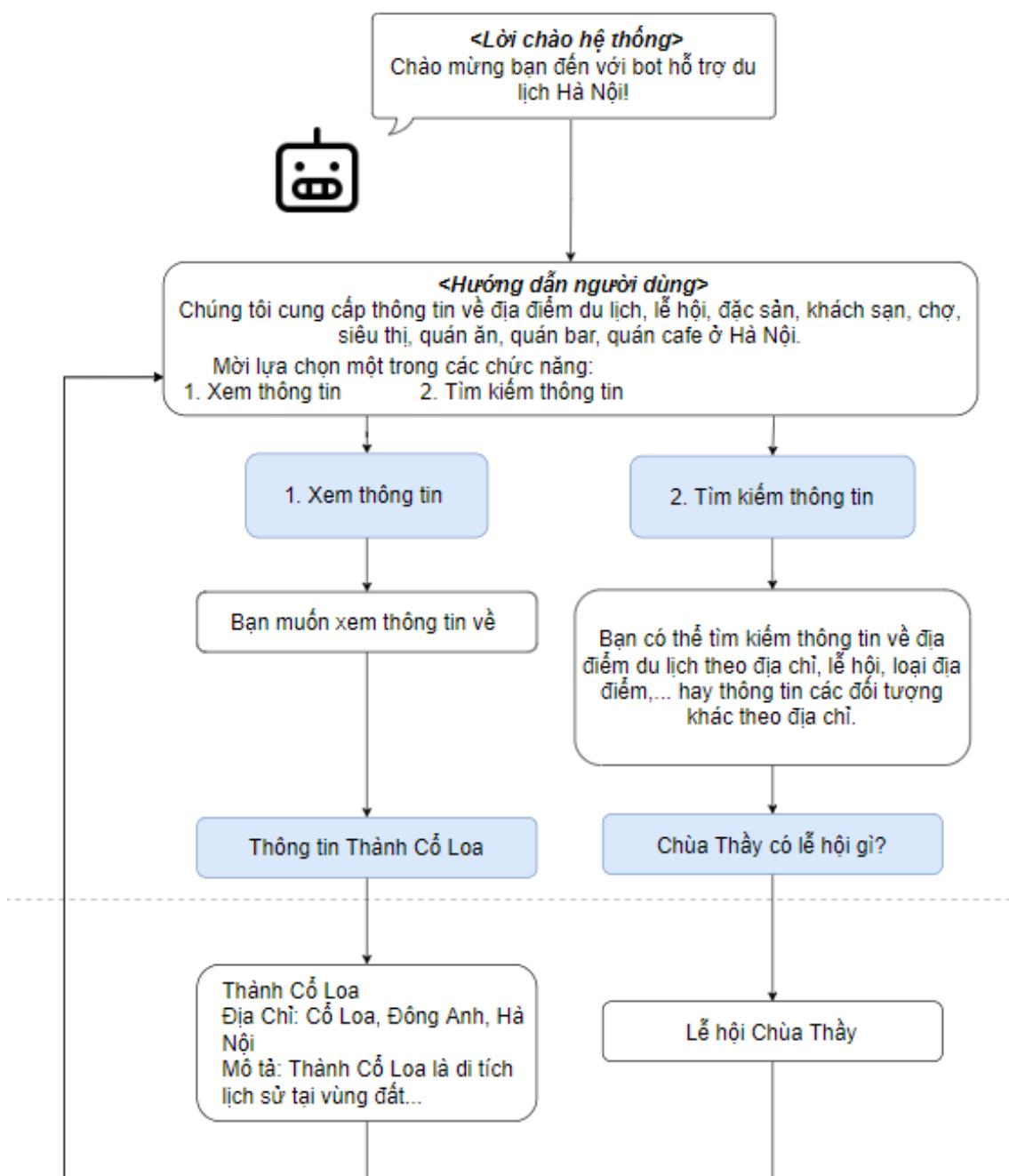
Mô tả: Tìm kiếm đối tượng mà người dùng quan tâm dựa trên thông tin liên quan.

- Địa điểm du lịch: tìm theo quận/ huyện, phường/xã, loại địa điểm, lễ hội, đặc sản.
- Lễ hội: tìm theo quận/ huyện, phường/xã, địa điểm du lịch.
- Đặc sản: tìm theo quận/ huyện, phường/xã, địa điểm du lịch.
- Khách sạn: tìm theo quận/ huyện, phường/xã.
- Chợ: tìm theo quận/ huyện, phường/xã.
- Siêu thị: tìm theo quận/ huyện, phường/xã.
- Quán ăn: tìm theo quận/ huyện, phường/xã.
- Quán bar: tìm theo quận/ huyện, phường/xã.
- Quán cafe: tìm theo quận/ huyện, phường/xã.

Ví dụ: Khi tìm kiếm địa điểm du lịch theo quận huyện - "địa điểm du lịch ở quận Hoàn Kiếm", chatbot trả về danh sách các địa điểm theo yêu cầu.

#### **3.4.2 Mô tả kịch bản giữa Chatbot và người dùng**

Kịch bản Chatbot với người dùng được mô tả qua sơ đồ dưới đây:



**Hình 3.12:** Sơ đồ kịch bản giữa Chatbot và người dùng

### 3.4.3 Cài đặt và xử lý ngôn ngữ tự nhiên

LUIS là một dịch vụ AI, lưu trữ các đoạn hội thoại trên nền tảng đám mây và áp dụng trí thông minh của máy để tùy chỉnh vào văn bản, ngôn ngữ tự nhiên, cách trò chuyện của người dùng để dự đoán ý nghĩa tổng thể, và đưa ra các thông tin chi tiết, liên quan, giúp quá trình trò chuyện trở nên dễ dàng, tự nhiên, linh hoạt hơn. Sử dụng dịch vụ LUIS (<https://www.luis.ai>) để định nghĩa và huấn luyện các ý định và thuộc tính phục vụ cho việc nhận dạng các thông tin cần thiết từ người dùng.

Ví dụ: Tìm kiếm các địa điểm du lịch thông qua loại địa điểm

timDiaDiemTheoLoai 

Machine learning features 

+ Add feature

Examples 

✓ Confirm all entities  Move to  Delete ...

Example user input  
Type an example of what a user might say and hit Enter.

trung tâm mua sắm ở hà nội  
loaiDiaDiem

tôi muốn tham quan di tích lịch sử  
loaiDiaDiem

**Hình 3.13:** Giao diện quản lý intent của LUIS

Để trích xuất được thông tin từ người dùng, ta cần tiến hành huấn luyện, xác định danh sách các thực thể có thể xuất hiện trong câu hỏi thoại.

Ví dụ với entity: "loaiDiaDiem"

The screenshot shows the LUIS entity management interface for the entity 'loaiDiaDiem'. At the top, there's a navigation bar with 'List items' (underlined), 'Examples', and 'Roles'. Below the navigation, a note says: 'List entities represent a fixed, closed set of related words along with their synonyms. List entities are extracted from the training data.' There are two buttons: 'Import values' and 'Delete'. A section titled 'Normalized values ↑' has a dropdown menu for 'Synonyms'. Below this, there's a text input field 'Type in a list item ...'. A list of entities follows:

- bảo tàng
- di tích lịch sử
- hồ, công viên, khu sinh ...
- kiến trúc tôn giáo, tâm l...
- làng nghề
- trung tâm mua sắm

The last item, 'trung tâm mua sắm', is selected, and its value 'trung tâm mua sắm' is shown in a text input field with a clear button 'X'.

**Hình 3.14:** Giao diện quản lý entity của LUIS

#### 3.4.4 Truy vấn dữ liệu bằng SPARQL

Sử dụng thư viện Python RDFLib (Dan Krech) để truy vấn SPARQL. RDFLib là thư viện Python dành cho làm việc với RDF, một ngôn ngữ đơn giản nhưng mạnh mẽ để biểu diễn thông tin. Thông qua thư viện này, Python là một trong những ngôn ngữ thao tác RDF chính, ngôn ngữ còn lại là Java. Thư viện này chứa các trình phân tích cú pháp / tuần tự hóa cho hầu hết tất cả các tuần tự hóa RDF đã biết, chẳng hạn như RDF / XML, Turtle, N-Triples và JSON-LD,... Thư viện cũng chứa cả phần kết thúc trong bộ nhớ và phần cuối Graph trong bộ nhớ để lưu trữ thông tin RDF và nhiều hàm tiện lợi để khai báo không gian tên đồ thị, truy vấn SPARQL,...

Ví dụ về câu truy vấn vào file data.owl để lấy thông tin địa điểm du lịch "thành cổ loa".

```
import rdflib
rdflib.registerplugins()
import rdflib

g = rdflib.Graph()
```

```

filename = "data/data.owl"
g.parse(filename, format='xml')
value = "thanh co loa"

query = """PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX xml: <http://www.w3.org/XML/1998/namespace>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX ex:
<http://www.semanticweb.org/administrator/ontologies/2022/4/untitled-ontology-28#>
SELECT DISTINCT ?name ?nameEx ?location ?info
WHERE
{
?x ex:tenDiemDuLich ?name ;
ex:diaChi ?location ;
ex:moTa ?info .

OPTIONAL{?x ex:tenKhac ?nameEx}
FILTER(lcase(?name) = """+value+""" || lcase(?nameEx) = """+value+""")}
} LIMIT 1
"""

rows = g.query(query)
for row in rows:
    print(row.name)
    print(row.location)
    print(row.info)

```

Kết quả của câu truy vấn:

Thành Cố Loa Cố Loa, Đông Anh, Hà Nội Thành Cố Loa là di tích lịch sử tại vùng đất xã Cố Loa, huyện Đông Anh, Hà Nội. Cố Loa là kinh đô của nhà nước Âu Lạc dưới thời An Dương Vương vào khoảng thế kỷ thứ 3 trước Công nguyên và cửa nhà nước phong kiến dưới thời Ngô Quyền thế kỷ 10.
--

**Hình 3.15:** Kết quả truy vấn SPARQL

#### a, Chức năng "Xem thông tin"

Dựa tên đối tượng mà Chatbot trích xuất được từ thông tin do người dùng cung cấp, hệ thống thực hiện truy vấn trong cơ sở dữ liệu tìm ra đối tượng tương ứng. Ví dụ mẫu câu truy vấn SPARQL lấy thông tin lễ hội trong đó *value* là tên lễ hội mà người dùng nhập vào.

```
query = """PREFIX owl: <http://www.w3.org/2002/07/owl#>
```

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX xml: <http://www.w3.org/XML/1998/namespace>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX ex:
<http://www.semanticweb.org/administrator/ontologies/2022/4/untitled-ontology-28#>
SELECT DISTINCT ?name ?location ?time ?info
WHERE
{
?x ex:tenLeHoi ?name;
ex:diaChi ?location;
ex:thoiGian ?time;
ex:moTa ?info.

FILTER(1case(?name) = """+value+"""")
}
LIMIT 1"""

```

### b, Chức năng "Tìm kiếm thông tin"

Chatbot thực hiện truy vấn tìm đối tượng mà người dùng yêu cầu dựa vào các thông tin liên quan xoay quanh. Ví dụ mẫu câu truy vấn SPARQL tìm địa điểm du lịch thông qua tên lễ hội trong đó *value* là tên lễ hội mà người dùng nhập vào.

```

query = """PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX xml: <http://www.w3.org/XML/1998/namespace>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX ex:
<http://www.semanticweb.org/administrator/ontologies/2022/4/untitled-ontology-28#>
SELECT DISTINCT ?name
WHERE
{
?y ex:tenLeHoi ?nameFe;
ex:taiDiemDuLich ?x.
?x ex:tenDiemDuLich ?name.

FILTER(regex(1case(?nameFe),""" + value + """))}
"""

```

### 3.5 Xây dựng trang web tra cứu thông tin về du lịch Hà Nội

Trang web xây dựng dựa trên Ontology đã tạo nhằm trực quan hóa cơ sở tri thức giúp người dùng tìm kiếm thông tin nhanh chóng, dễ dàng.

#### 3.5.1 Mô hình hóa chức năng

##### a, Giới thiệu chung

Tác nhân của hệ thống:

**Bảng 3.1:** Danh sách tác nhân của hệ thống

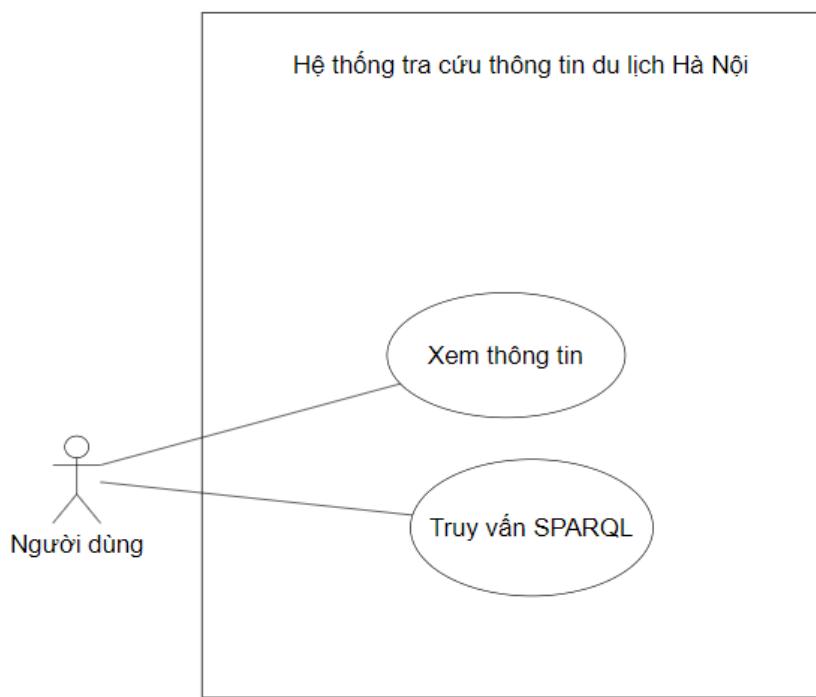
STT	Tên tác nhân	Mô tả
1	Người dùng	Có thể xem danh sách các đối tượng (địa điểm du lịch, lễ hội, đặc sản,...), xem thông tin chi tiết từng đối tượng, truy vấn SPARQL.

Các ca sử dụng:

**Bảng 3.2:** Danh sách ca sử dụng

STT	Mã usecase	Tên usecase	Mô tả usecase	Tác nhân tương tác
1	UC001	Xem thông tin	Cho phép người dùng xem danh sách tài nguyên và thông tin từng tài nguyên	Người dùng
2	UC002	Truy vấn SPARQL	Cho phép người dùng truy vấn Ontology bằng ngôn ngữ SPARQL	Người dùng

Usecase tổng quan:

**Hình 3.16:** Usecase tổng quan**b, Đặc tả các usecase**

- Chức năng "Xem thông tin"

Bảng đặc tả:

**Bảng 3.3:** Bảng đặc tả usecase "Xem thông tin"

Mã Usecase	UC001	Tên Usecase	Xem thông tin
Tác nhân	Người dùng		
Mô tả	Cho phép người dùng xem thông tin các tài nguyên trong hệ thống.		
Tiền điều kiện	Không		
Luồng sự kiện chính	STT	Thực hiện bởi	Hành động
	1	Người dùng	Chọn loại tài nguyên cần xem
	2	Hệ thống	Hiển thị danh sách tài nguyên
	3	Người dùng	Chọn tài nguyên cần xem
	4	Hệ thống	Hiển thị thông tin chi tiết của tài nguyên

Luồng sự kiện thay thế	STT	Thực hiện bởi	Hành động
Hậu điều kiện	Không		

- Chức năng "Truy vấn SPARQL"

Bảng đặc tả:

**Bảng 3.4:** Bảng đặc tả usecase "Truy vấn SPARQL"

Mã Usecase	UC002	Tên Usecase	Truy vấn SPARQL
Tác nhân	Người dùng		
Mô tả	Cho phép người truy vấn các quan hệ và thuộc tính liên quan đến tài nguyên trong hệ thống.		
Tiền điều kiện	Không		
Luồng sự kiện chính	STT	Thực hiện bởi	Hành động
1	Người dùng	Chọn chức năng truy vấn SPARQL	
2	Hệ thống	Hiển thị giao diện trang truy vấn	
3	Người dùng	Nhập vào câu truy vấn và xác nhận	
4	Hệ thống	Hiển thị kết quả truy vấn	
Luồng sự kiện thay thế	STT	Thực hiện bởi	Hành động
Hậu điều kiện	Không		

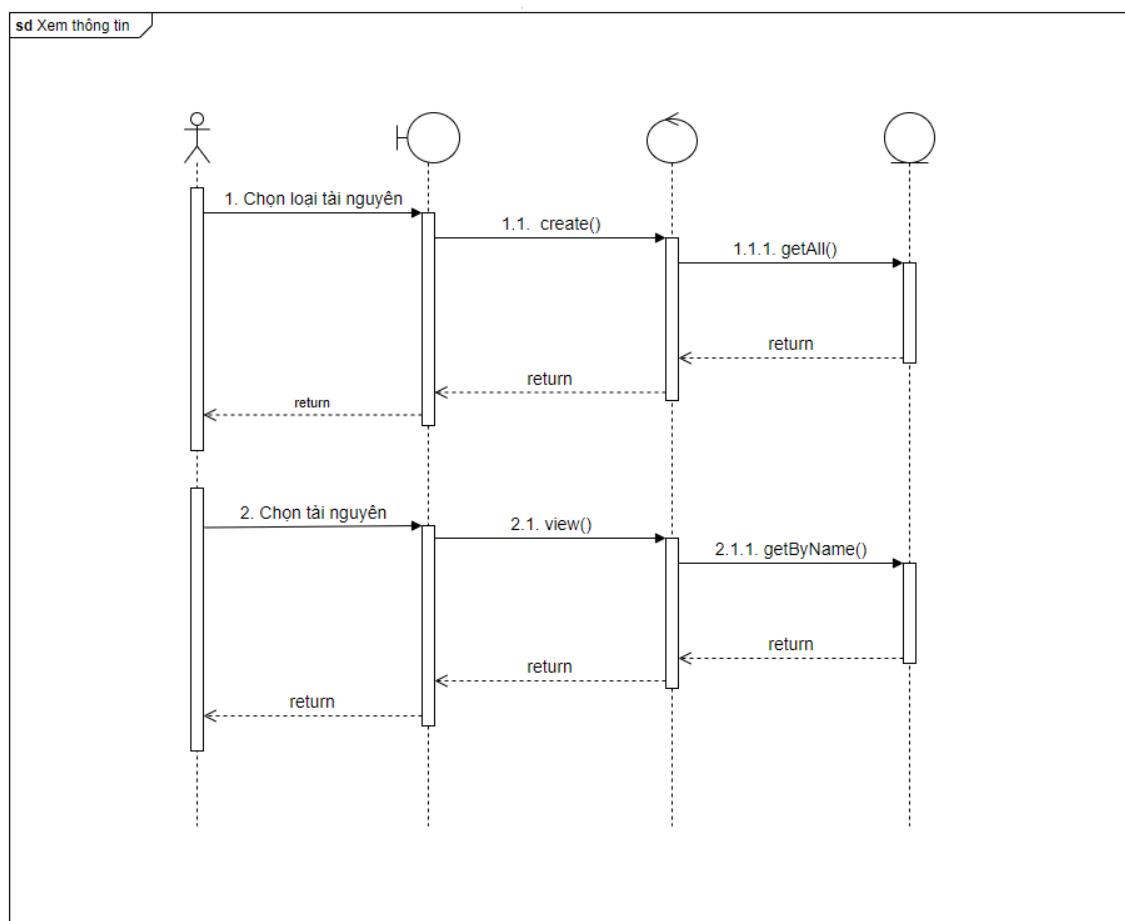
Dữ liệu đầu vào gồm các trường dữ liệu sau:

**Bảng 3.5:** Danh sách dữ liệu đầu vào

STT	Trường dữ liệu	Mô tả	Bắt buộc?	Ví dụ
1	Câu truy vấn	Truy vấn Ontology bằng ngôn ngữ SPARQL	Có	SELECT DISTINCT ?name WHERE { ?x ex:tenCho ?name. }

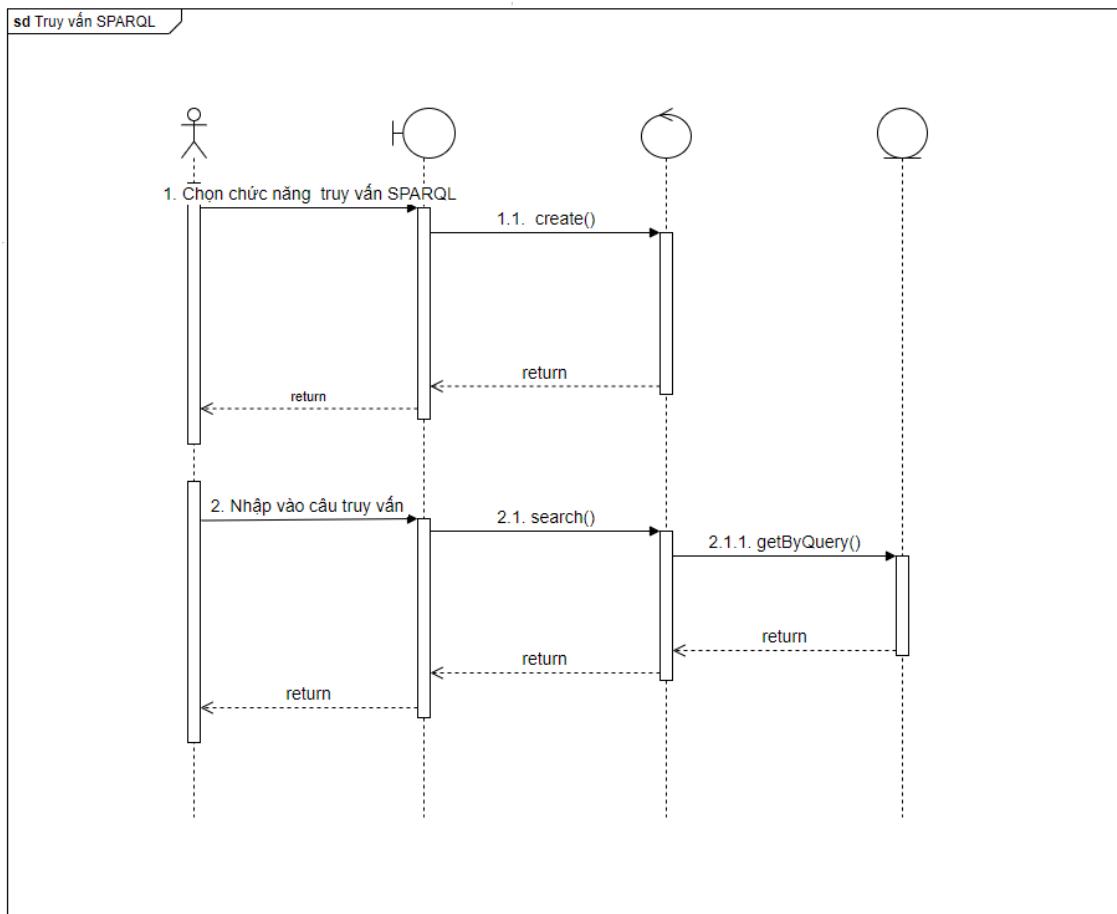
### 3.5.2 Mô hình hóa hành vi

#### a, Chức năng "Xem thông tin"



**Hình 3.17:** Biểu đồ tuần tự Xem thông tin

### b, Chức năng "Truy vấn SPARQL"



**Hình 3.18:** Biểu đồ tuần tự Truy vấn SPARQL

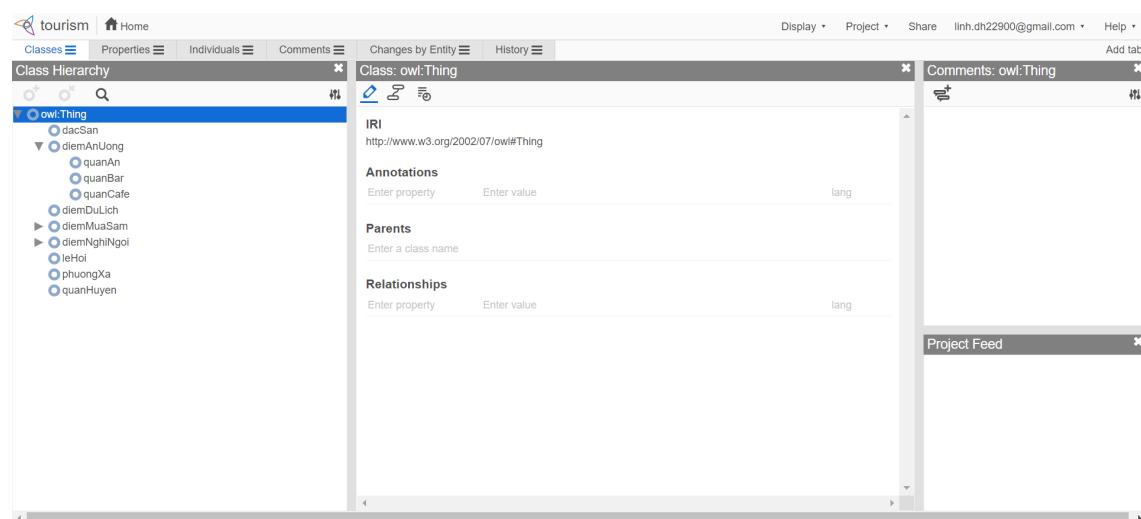
## CHƯƠNG 4. TRIỂN KHAI HỆ THỐNG

### 4.1 Công nghệ sử dụng

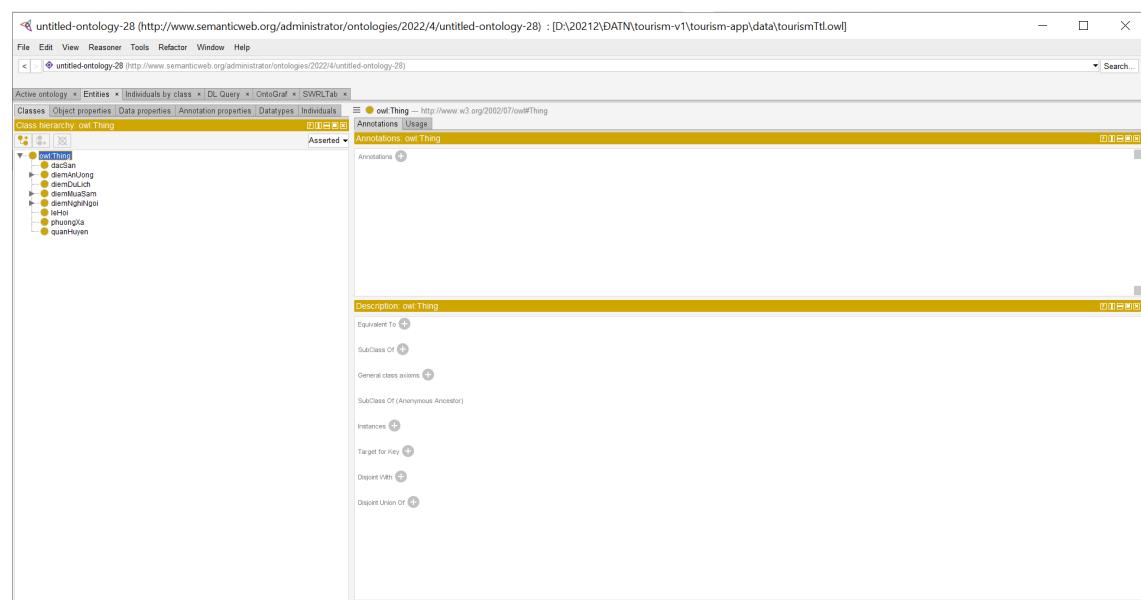
#### a, Protégé

Protégé là một nền tảng mã nguồn mở miễn phí cung cấp cho người dùng một bộ công cụ để xây dựng các mô hình và các ứng dụng dựa trên cơ sở tri thức với Ontology.

Protégé cung cấp hai phiên bản là WebProtégé dành cho website và Protégé Desktop dành cho ứng dụng trên máy tính.



Hình 4.1: WebProtégé

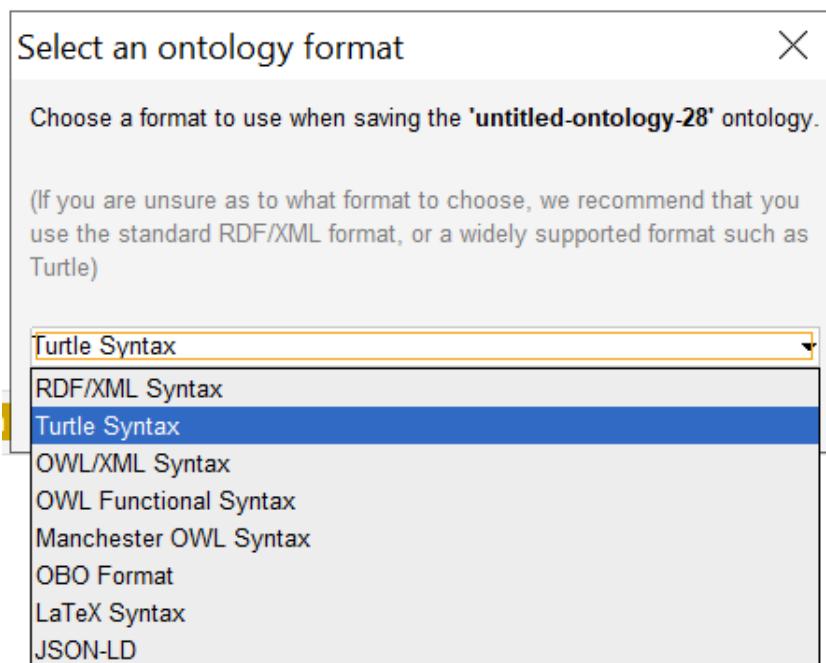


Hình 4.2: Protégé Desktop

Trong đồ án tốt nghiệp này, tôi lựa chọn sử dụng Protégé Desktop - một môi trường phát triển Ontology phong phú về tính năng với sự hỗ trợ đầy đủ cho ngôn ngữ OWL 2 mới nhất và kết nối trực tiếp tới các trình lập luận logic mô tả như HermiT và Pellet.

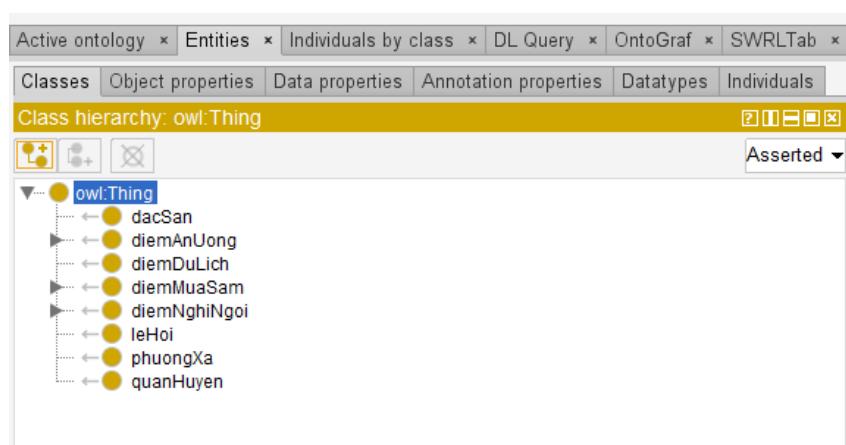
Một số chức năng chính được sử dụng trong phạm vi đồ án:

- Tạo mới một project trong Protégé và lưu trữ dưới nhiều định dạng như RDF/XML, Turtle, OWL/XML,...



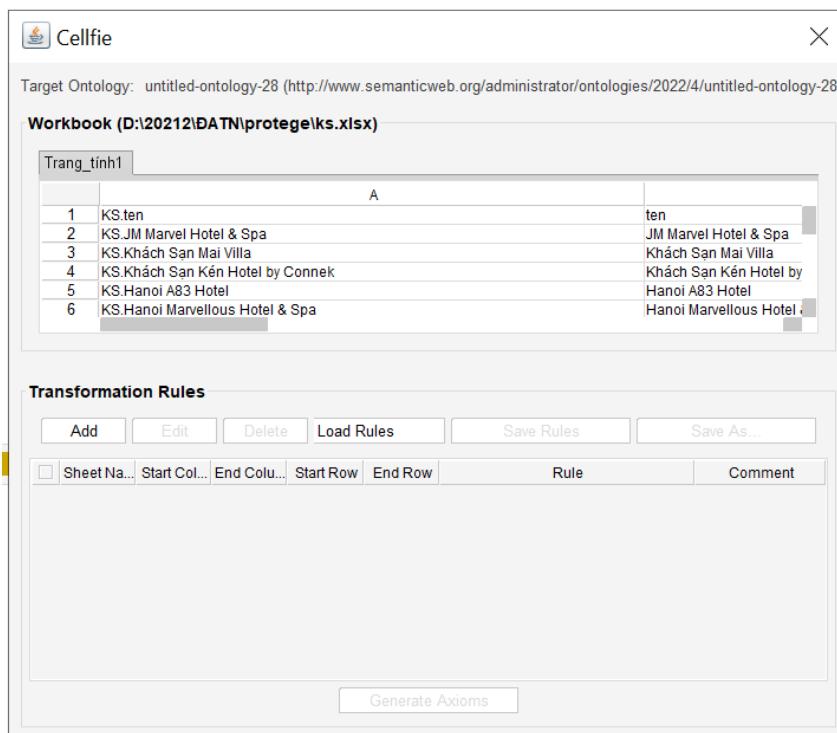
**Hình 4.3:** Lưu trữ Ontology dưới nhiều định dạng

- Xây dựng các lớp, thuộc tính và thể hiện cho Ontology đã tạo thông qua các tab Classes, Object properties, Data properties, Individuals mà ứng dụng đã định nghĩa sẵn.



**Hình 4.4:** Minh họa tab chức năng trong Protégé

- Nhập dữ liệu tự động thông qua Cellfie plugin.



**Hình 4.5:** Minh họa giao diện Cellfie plugin

- Thêm luật suy diễn thông qua ngôn ngữ luật cho Web ngữ nghĩa - SWRL.

Name	
<input checked="" type="checkbox"/> dac_san	coDacSan(?x, ?y) -> taiDiemDuLich(?y, ?x)
<input checked="" type="checkbox"/> le_hoi	coLeHoi(?x, ?y) -> taiDiemDuLich(?y, ?x)

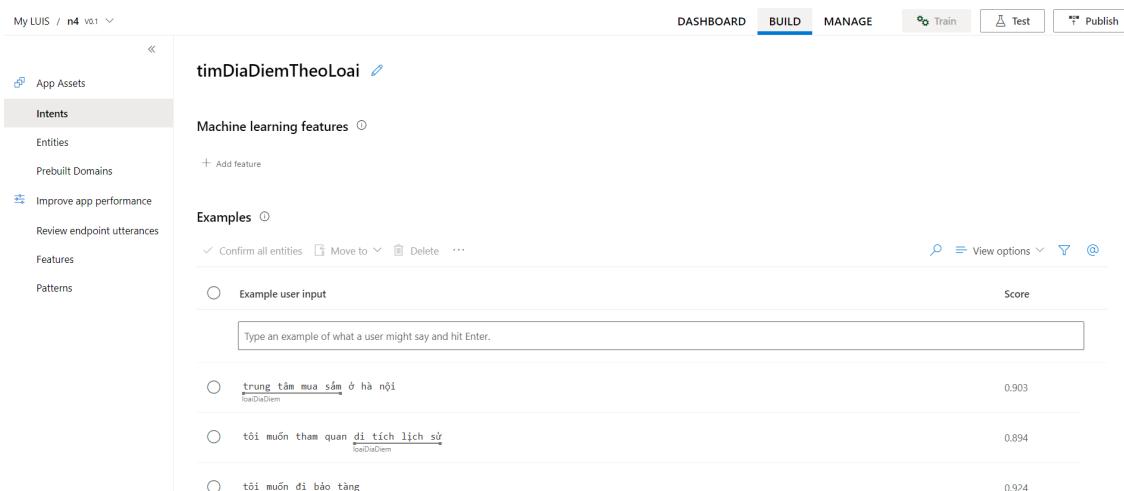
**Hình 4.6:** Minh họa tab SWRL

### b, LUIS

Language Understanding Intelligent Service (LUIS) là một dịch vụ AI, lưu trữ các đoạn hội thoại trên nền tảng đám mây và áp dụng học máy để tùy chỉnh vào văn bản, ngôn ngữ tự nhiên, cách trò chuyện của người dùng để dự đoán ý nghĩa tổng thể, và đưa ra các thông tin chi tiết, liên quan, giúp quá trình trò chuyện trở nên dễ dàng, tự nhiên, linh hoạt hơn.[14]

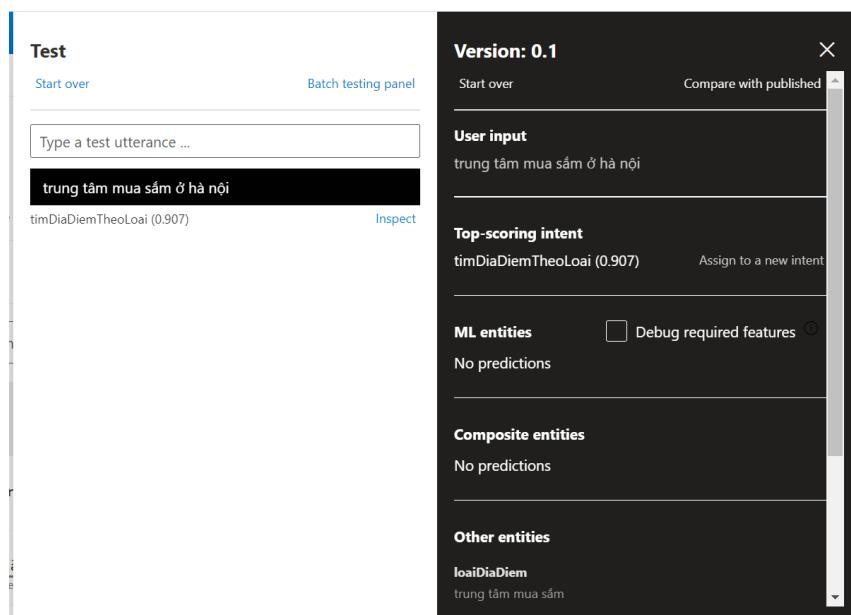
Vòng đời phát triển ứng dụng:

- Xác định các tình huống mà người dùng có thể sử dụng từ đó xây dựng hệ thống intents (ý định của người dùng) và entities (thực thể cần trích xuất).
- Tạo các câu ví dụ mô phỏng các câu đầu vào từ người dùng.



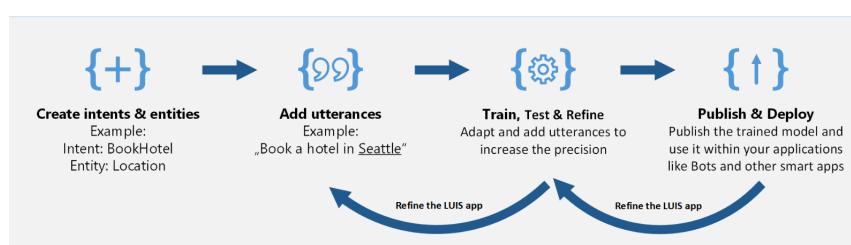
**Hình 4.7:** Minh họa chức năng tạo intents, entities và utterances

- Chọn chế độ huấn luyện và kiểm tra kết quả.



**Hình 4.8:** Minh họa chức năng Test của LUIS

- Triển khai mô hình và sử dụng nó trong các ứng dụng Bots hay các ứng dụng thông minh khác.



**Hình 4.9:** Minh họa vòng đời phát triển ứng dụng

## 4.2 Kết quả thực nghiệm

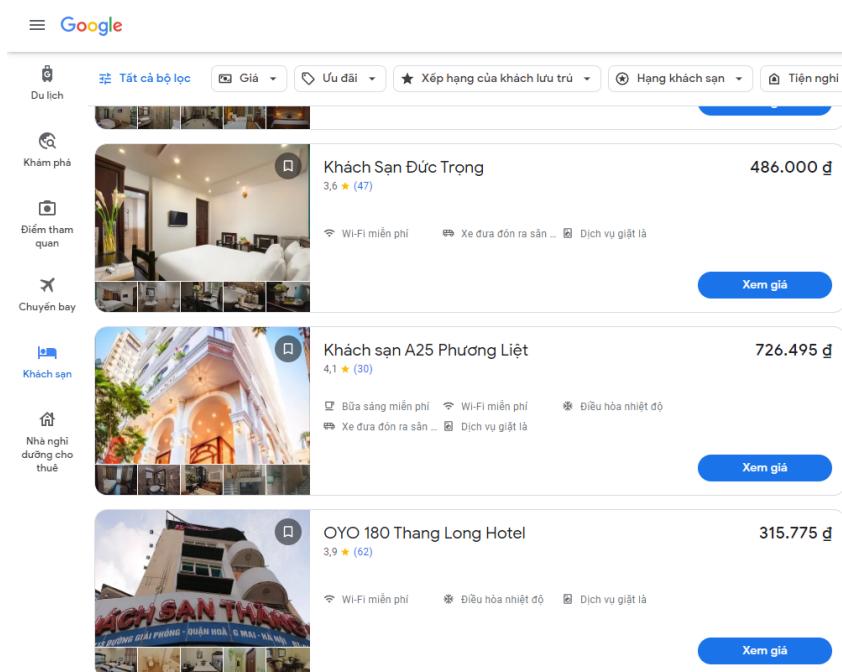
### 4.2.1 Kết quả thu thập dữ liệu

Để xây dựng Ontology về du lịch Hà Nội, tôi đã tiến hành crawl dữ liệu từ các bài viết trên trang Wikipedia, Google Travel và một số trang web khác. Thông kê về lượng dữ liệu đã thu thập được trình bày trong bảng dưới đây.

**Bảng 4.1:** Bảng thống kê dữ liệu thu thập được

STT	Tên tài nguyên	Số lượng
1	Địa Điểm Du Lịch	75
2	Đặc Sản	20
3	Lễ Hội	54
4	Khách Sạn	243
5	Chợ	454
6	Siêu Thị	197
7	Quán Ăn	329
8	Quán Bar	217
9	Quán Cafe	285

Ví dụ khi muốn lấy thông tin về khách sạn trên Google Travel.



**Hình 4.10:** Ví dụ một số khách sạn trên Google Travel

Sau thu thập bằng công cụ selenium bài viết được lưu trữ dưới dạng trang tính như hình dưới:

	A	B	C	D	E
1	khachSan	ten	diaChi	sdt	gia
2	H01	JM Marvel Hotel & Spa	16 Phố P. Hàng Da, Cửa Đông 024 3823 8855	1.318.665 đ	
3	H02	Khách Sạn Mai Villa	39 P. Đặng Văn Ngữ, Trung Tự, Đống Đa, Hà Nội	510.400 đ	
4	H03	Khách Sạn Kén Hotel by Connekt	98 Hàng Gai, Hoàn Kiếm, Hà Nội 096 995 34 43	495.001 đ	
5	H04	Hanoi A&S Hotel	Ngõ 83 P. Xã Đàn, Phương Liệt 024 3742 5555	240.000 đ	
6	H05	Giường ngủ và bữa sáng 5 phòng		657.330 đ	
7	H06	Hanoi Marvellous Hotel & Spa	55 Đường Thành, Cửa Đông, 024 3266 9240	1.091.469 đ	
8	H07	Khách sạn Hilton Hanoi Opera	1 P. Lê Thánh Tông, Phan Chu Trinh 024 3933 0500	2.114.844 đ	
9	H08	Le Grand Hotel Hanoi	1b9 Đầm Trầu, Bạch Đằng, H 024 3984 5350	289.710 đ	
10	H09	Sofitel Legend Metropole Hotel	15 P. Ngõ Quyền, Street, Hoàn Kiếm, Hà Nội 024 3826 6919	5.231.819 đ	
11	H010	Khách Sạn Đức Trọng	Số 38, Phố Vọng, Phường Phúc Đồng 098 276 68 68	232.875 đ	
12	H011	Bluebell Hotel	41 Ng. Huyện, Hàng Trống, H093 634 51 23	150.800 đ	
13	H012	Khách sạn Hanoi La Siesta Hotel	94 P. Mã Mây, Hàng Buồm, H 024 3926 3641	886.500 đ	
14	H013	Căn hộ dịch vụ, dành cho 2 người		595.902 đ	
15	H014	Khách Sạn Blue 29	325 P. Giảng Võ, Chợ Dừa, Đ 024 3512 4013	246.628 đ	
16	H015	Khách Sạn Royal Palace 2	95 P. Hàng Bông, Hàng Bông, 024 3824 7443	164.850 đ	

Hình 4.11: Một số dữ liệu khách sạn được crawl

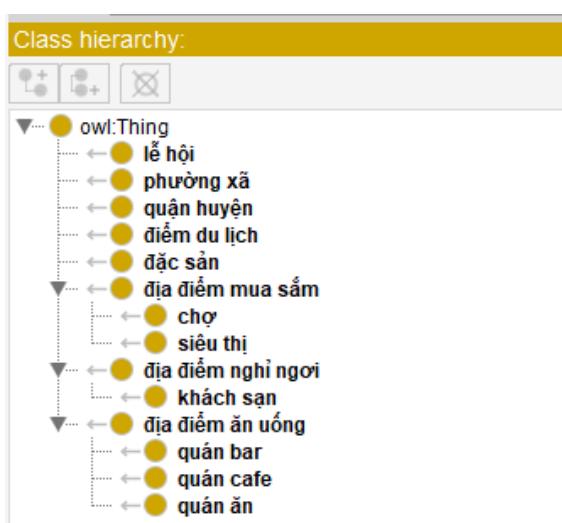
Cuối cùng, tiến hành xử lý dữ liệu, trích xuất ra các dữ liệu cần thiết, chuyển địa chỉ thành tọa độ để hiển thị trên bản đồ nhờ Geocoding by SmartMonkey.

A	B	C	D	E	F	G	H	I	J
1	KS.ten	ten	diaChi	sdt	gia	lat	lng	PX.phuongXa	QH.quanHuyen
2	KS.JM JM Marvel Hotel & Spa	16 Phố P. Hàng Da, Củ 024 3823 8855	1.318.665 đ		21.0313856	105.8467474	PX.Hàng Da	QH.Hoàn Kiếm	
3	KS.Khách Sạn Mai Khách Sạn Mai Villa	39 P. Đặng Văn Ngữ, Trung Tự, Đống Đa, Hà Nội 024 3826 6919	510.400 đ		21.0082251	105.8320227	PX.Trung Tự	QH.Đống Đa	
4	KS.Khách Sạn Kén Khách Sạn Kén Hotel by 98 Hàng Gai	Hoàn Kiếm 096 995 34 43	495.001 đ		21.0316273	105.8494042	PX.Hàng Gai	QH.Hoàn Kiếm	
5	KS.Hanoi A&S Hotel	Ngõ 83 P. Xã Đàn, Phúc 024 3742 5555	240.000 đ		21.0126934	105.8352371	PX.Phường Liền	QH.Đống Đa	
6	KS.Hanoi Marvellous Hotel	55 Đường Thành, Cửa Đông, 024 3266 9240	1.091.469 đ		21.0312778	105.8471416	PX.Hàng Bông	QH.Hoàn Kiếm	
7	KS.Khách sạn HiltcKhách sạn Hilton Hanoi 1 P. Lê Thánh Tông	024 3933 0500	2.114.844 đ		21.0231955	105.8577909	PX.Phản Chu Trinh	QH.Hoàn Kiếm	
8	KS.Le Grand Hotel Le Grand Hotel Hanoi	1b9 Đầm Trầu, Bạch Đằng 024 3984 5350	289.710 đ		21.0119554	105.8652486	PX.Bạch Đằng	QH.Hai Bà Trưng	
9	KS.Sofitel Legend Sofitel Legend Metropole	15 P. Ngõ Quyền, Street 024 3826 6919	5.231.819 đ		21.02559	105.8564147	PX.Tràng Tiền	QH.Hoàn Kiếm	
10	KS.Khách Sạn Đức Trọng Khách Sạn Đức Trọng	Số 38, Phố Vọng, Phúc 098 276 68 68	232.875 đ		20.999071	105.841833	PX.Phường Định	QH.Hai Bà Trưng	
11	KS.Bluebell Hotel Bluebell Hotel	41 Ng. Huyện, Hàng T 093 634 51 23	150.800 đ		21.0291582	105.8484237	PX.Hàng Trống	QH.Hoàn Kiếm	
12	KS.Khách sạn Hanoi La Siesta Hotel	94 P. Mã Mây, Hàng E 024 3926 3641	886.500 đ		21.0344048	105.8529512	PX.Hàng Buồm	QH.Hoàn Kiếm	
13	KS.Khách Sạn Blue Khách Sạn Blue 29	325 P. Giảng Võ, Ô Chợ Dừa 024 3512 4013	246.628 đ		21.0238363	105.8206156	PX.Ô Chợ Dừa	QH.Đống Đa	
14	KS.Khách Sạn Royal Palace Khách Sạn Royal Palace	95 P. Hàng Bông, Hàng 024 3824 7443	164.850 đ		21.0299217	105.8467342	PX.Hàng Bông	QH.Hoàn Kiếm	
15	KS.A25 Hotel & Sp A25 Hotel & Spa - 684 N 684 P. Minh Khai, Vinh 024 3987 4720	096 157 55 99	1.079.208 đ		21.000662	105.8710033	PX.Vĩnh Phú	QH.Hai Bà Trưng	
16	KS.ANATOLE HOTEANATOLE HOTEL HANOI 26-28-30 P. Nhà Chung 096 157 55 99				21.0282151	105.8497072	PX.Hàng Trống	QH.Hoàn Kiếm	

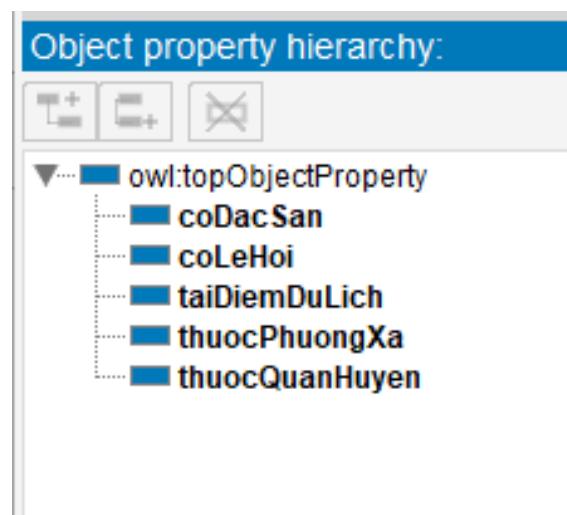
Hình 4.12: Một số dữ liệu khách sạn sau khi được xử lý và lưu trữ

#### 4.2.2 Kết quả xây dựng cơ sở tri thức

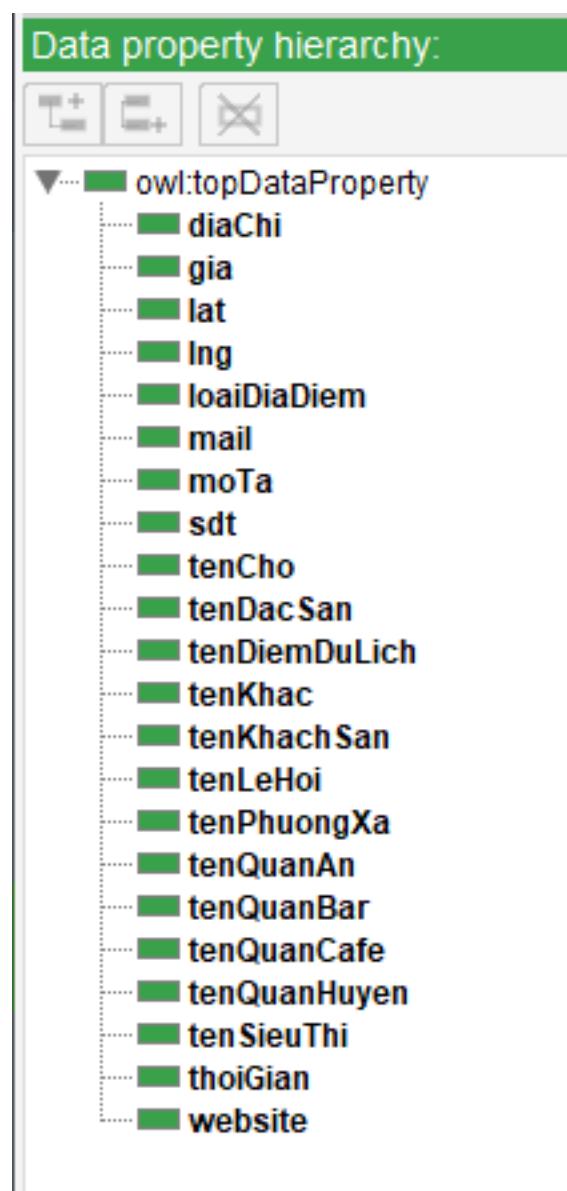
##### a, Kết quả xây dựng lớp và thuộc tính của Ontology



Hình 4.13: Các lớp của Ontology



**Hình 4.14:** Object property của Ontology



**Hình 4.15:** Data property của Ontology

### b, Kết quả xây dựng cơ sở tri thức

Dữ liệu sau khi thu thập được thêm vào Ontology sử dụng Cellfie của Protégé. Dưới đây là minh họa các thể hiện của lớp địa điểm du lịch và các thuộc tính, quan hệ của một thể hiện cụ thể trong lớp địa điểm du lịch.

The screenshot shows the Protégé Cellfie interface with the following details:

- Description:** diemDuLich
- Instances:**
  - DD.AoVua
  - DD.BaoTangCachMangViệtNam
  - DD.BaoTangDânTộcHọcViệtNam
  - DD.BaoTangHàNội
  - DD.BaoTangHồChiMinh
  - DD.BaoTangLịchSửQuânSựViệtNam
  - DD.BaoTangLịchSửQuốcGia
  - DD.BaoTangMỹThuậtViệtNam
  - DD.BaoTangPhụNữViệtNam
  - DD.ChùaHòeNhai
  - DD.ChùaHương
  - DD.ChùaLiênPhái
  - DD.ChùaLáng
  - DD.ChùaMộtCột
  - DD.ChùaNonNước
  - DD.ChùaSài
  - DD.CôngViênHồTây
  - DD.CôngViênThiênĐườngBảoSơn
  - DD.CôngViênThốngNhất
  - DD.CôngViênThủLệ

**Hình 4.16:** Các thể hiện của lớp điểm du lịch

The screenshot shows the Protégé Cellfie interface for the individual 'DD.BaoTangCachMangViệtNam' with the following details:

- Description:** DD.BaoTangCachMangViệtNam
- Type:** diemDuLich
- Object property assertions:**
  - thuocQuanHuyen QH.HoànKiem
  - thuocPhuongXa PX.TràngTiền
- Data property assertions:**
  - loaiDiaDiem "Bảo tàng"^^xsd:string
  - tenDiaDiem "Bảo tàng Cách mạng Việt Nam"^^xsd:string
  - long "105.8561402"^^xsd:string
  - lat "21.0259446"^^xsd:string
  - thoiGian "8h30-11h45 và 13h30-16h15 các ngày trong tuần trừ thứ 2 hàng tuần và Tết Nguyên đán"^^xsd:string
  - diaChi "25 phố Tông Đản, phường Tràng Tiền, quận Hoàn Kiếm, thành phố Hà Nội"^^xsd:string
  - moTa "Viện Bảo tàng Cách mạng Việt Nam là nơi tái hiện lịch sử đấu tranh hàng trăm năm qua của nhân dân Việt Nam chống Pháp, chống Nhật, chống Mỹ, chống phong kiến, giành lại nền độc lập tự do cho Việt Nam"^^xsd:string

**Hình 4.17:** Các thuộc tính và quan hệ của một thể hiện trong lớp điểm du lịch

### c, Tổng kết về cơ sở tri thức

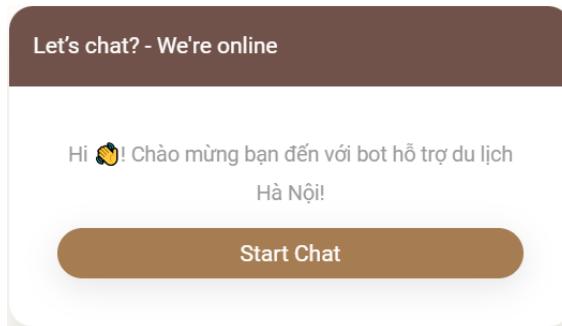
Kết quả xây dựng cơ sở tri thức về du lịch tại Hà Nội:

- Tổng số bộ ba RDF: 20450
- Tổng số lớp: 14

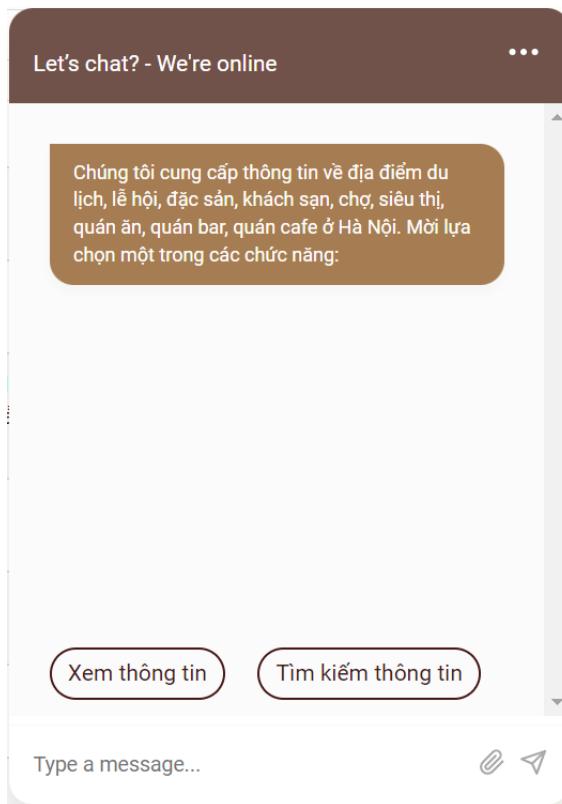
- Tổng số thuộc tính đối tượng: 5
- Tổng số thuộc tính giá trị: 22
- Tổng số thể hiện: 2500

### 4.2.3 Kết quả cài đặt Chatbot hỗ trợ du lịch

#### a, Lời chào hệ thống



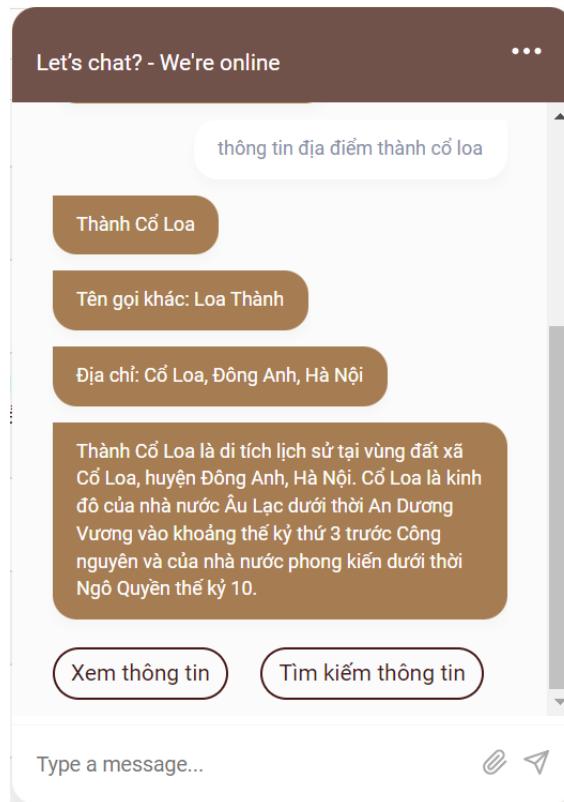
**Hình 4.18:** Lời chào hệ thống



**Hình 4.19:** Hướng dẫn của Chatbot

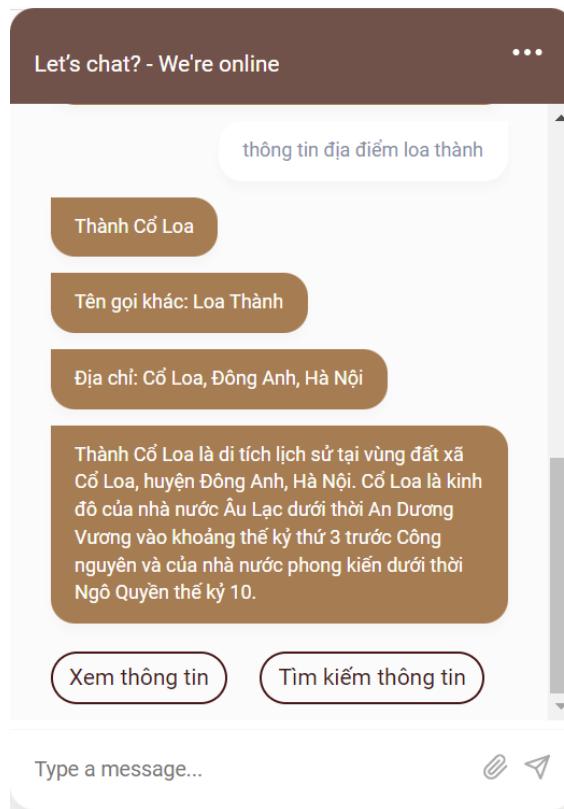
#### b, Chức năng "Xem thông tin"

Sau khi người dùng chọn chức năng "Xem thông tin", Chatbot sẽ yêu cầu người dùng nhập thông tin tài nguyên và Chatbot sẽ cung cấp các thông tin về tài nguyên đó.



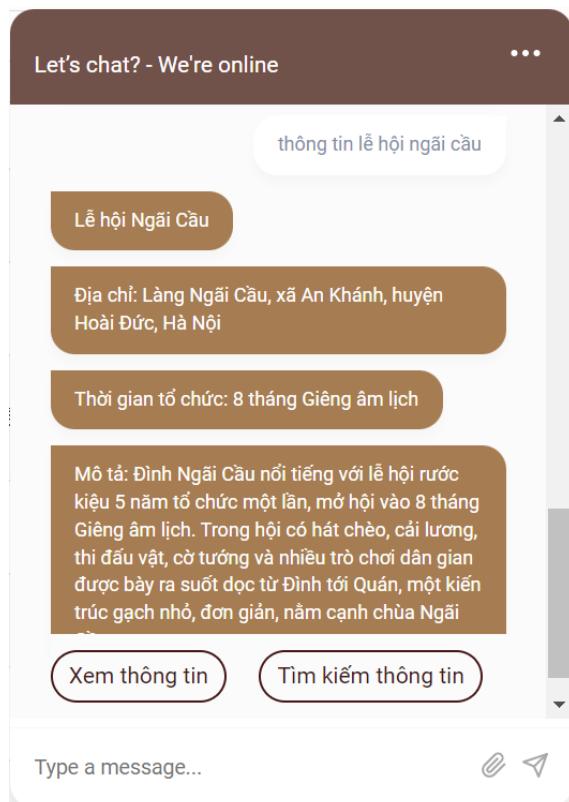
**Hình 4.20:** Minh họa chức năng xem thông tin địa điểm du lịch (1)

Chatbot còn hỗ trợ tìm kiếm với tên khác nếu địa điểm du lịch có nhiều tên khác nhau.

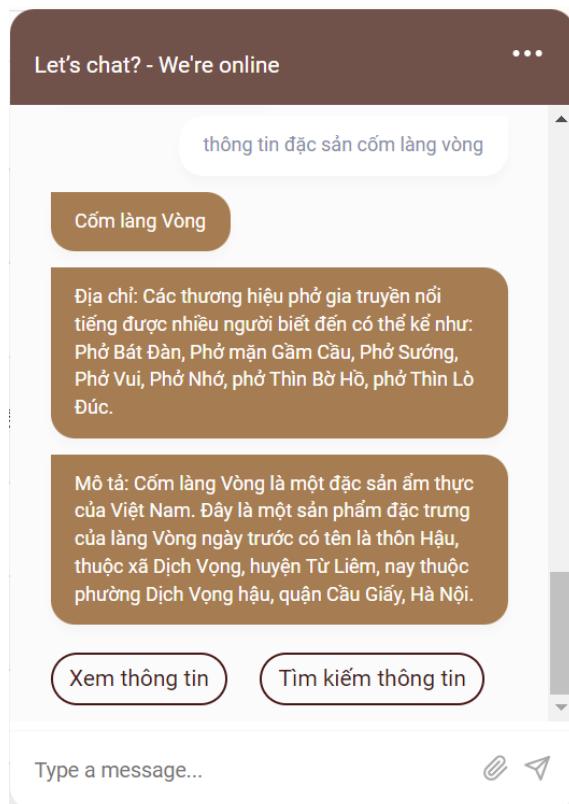


**Hình 4.21:** Minh họa chức năng xem thông tin địa điểm du lịch (2)

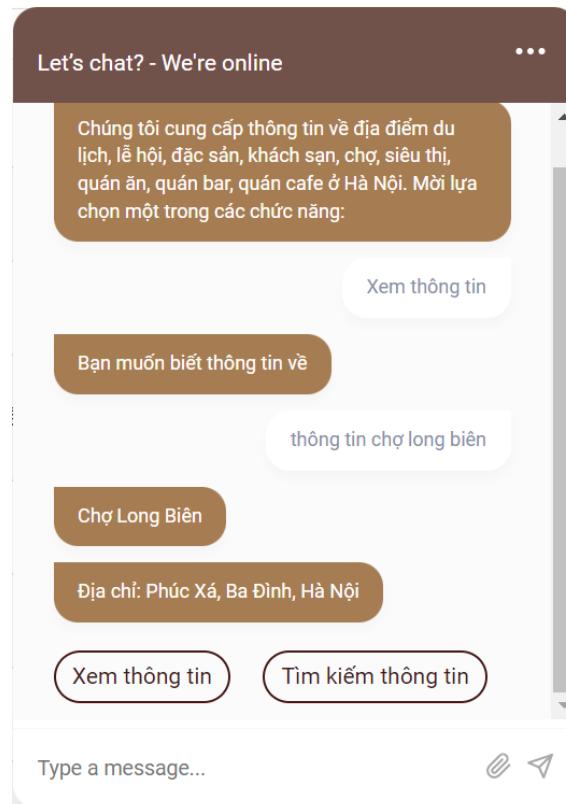
Minh họa chức năng xem thông tin của một số tài nguyên khác.



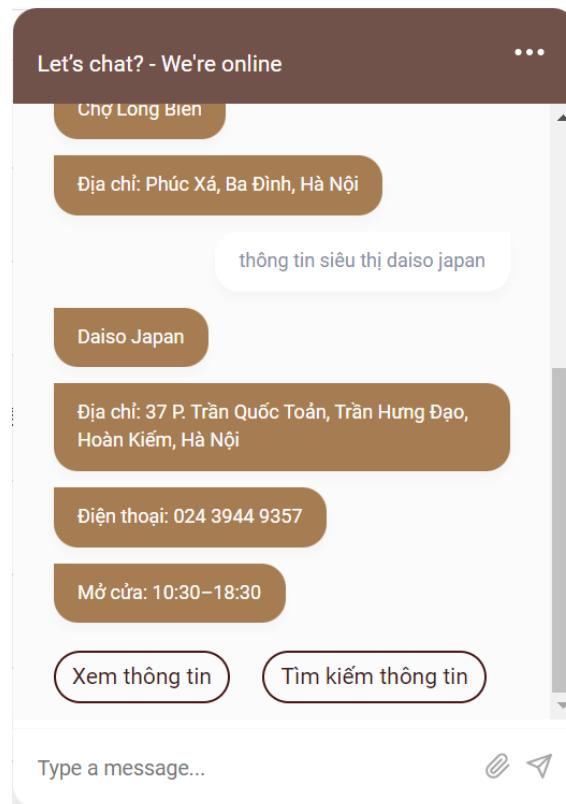
**Hình 4.22:** Minh họa chức năng xem thông tin lễ hội



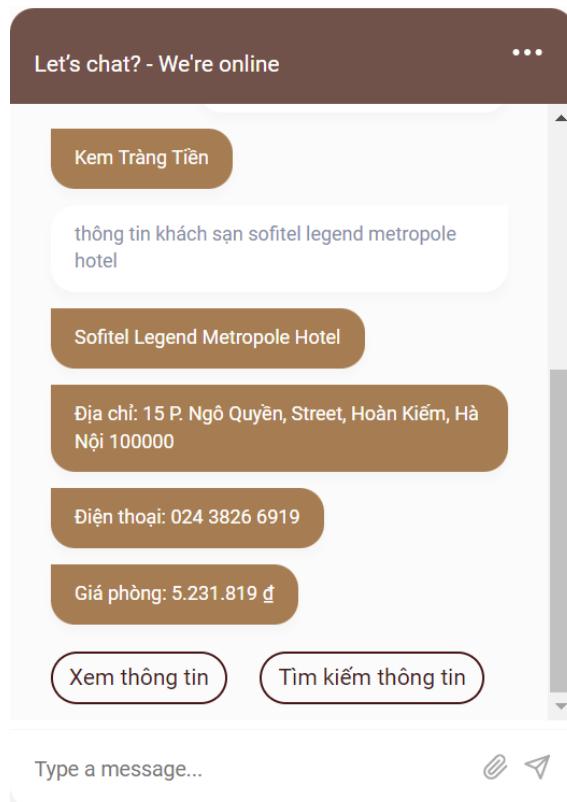
**Hình 4.23:** Minh họa chức năng xem thông tin đặc sản



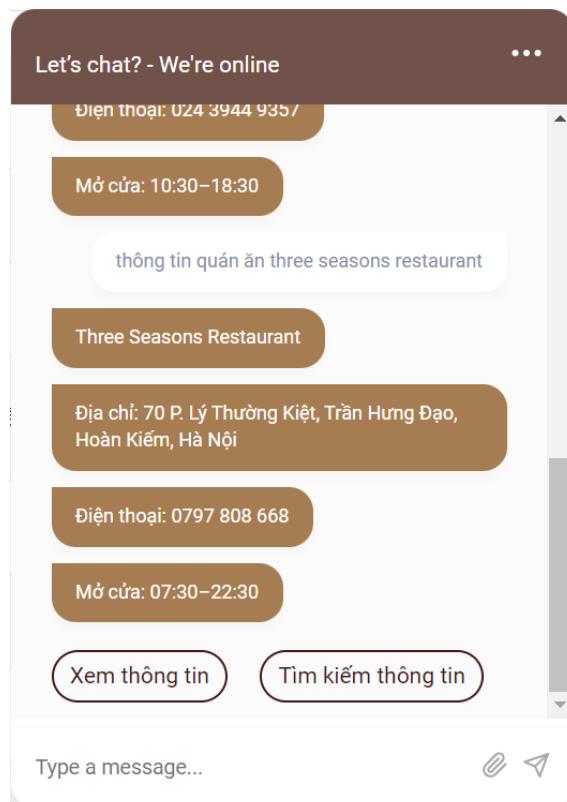
**Hình 4.24:** Minh họa chức năng xem thông tin chợ



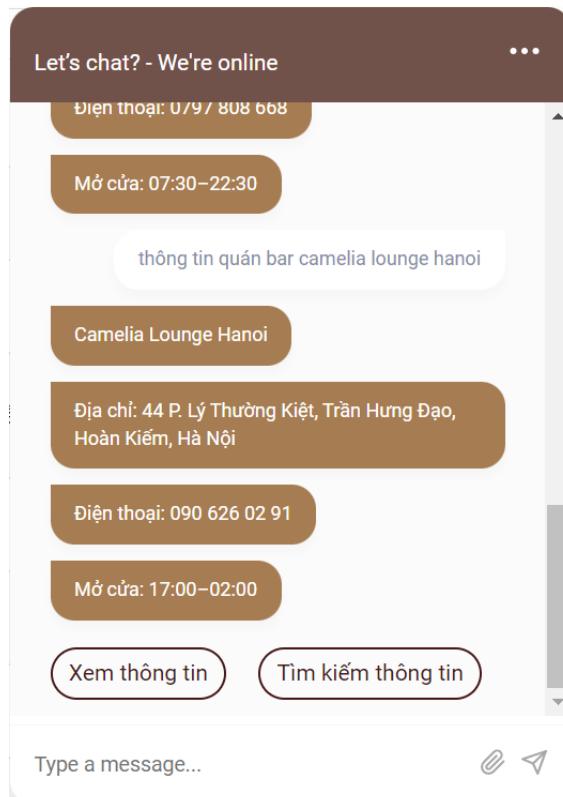
**Hình 4.25:** Minh họa chức năng xem thông tin siêu thị



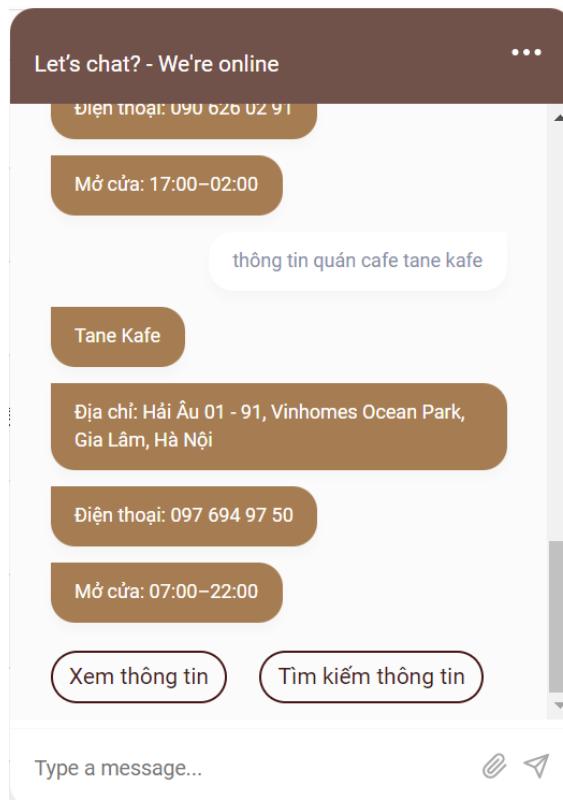
**Hình 4.26:** Minh họa chức năng xem thông tin khách sạn



**Hình 4.27:** Minh họa chức năng xem thông tin quán ăn



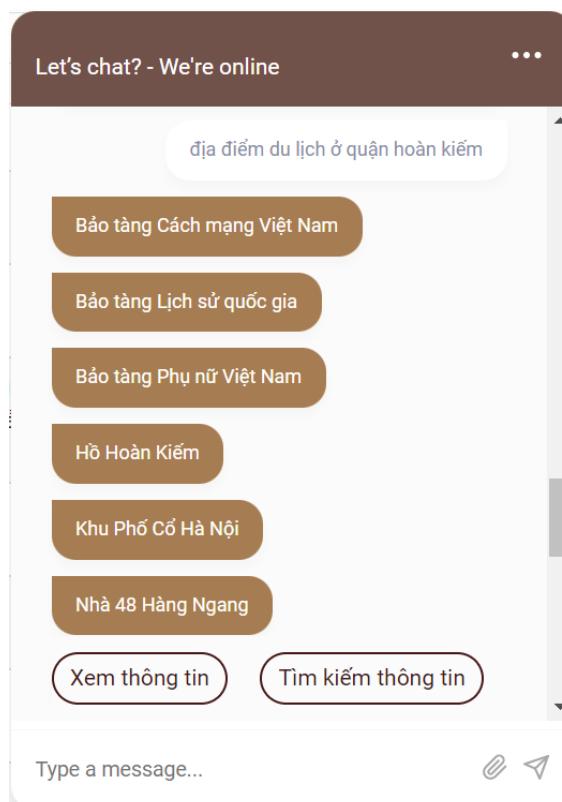
**Hình 4.28:** Minh họa chức năng xem thông tin quán bar



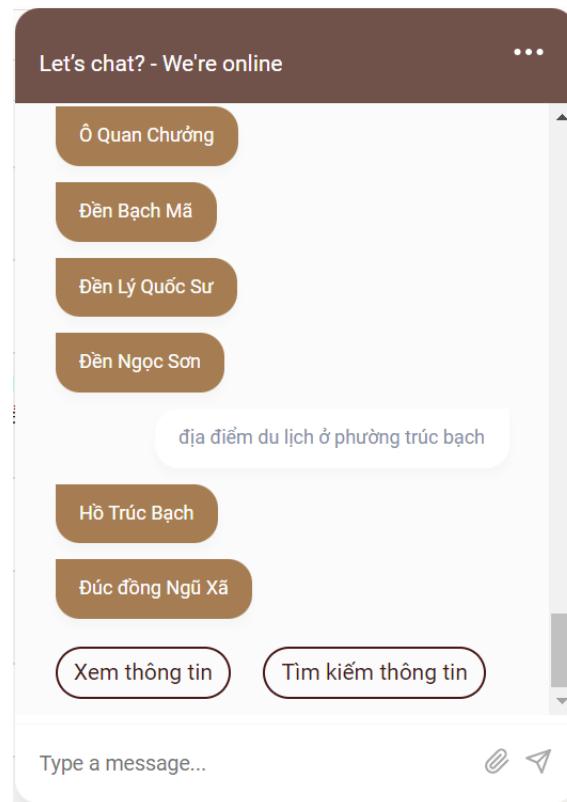
**Hình 4.29:** Minh họa chức năng xem thông tin quán cafe

### c, Chức năng "Tìm kiếm thông tin"

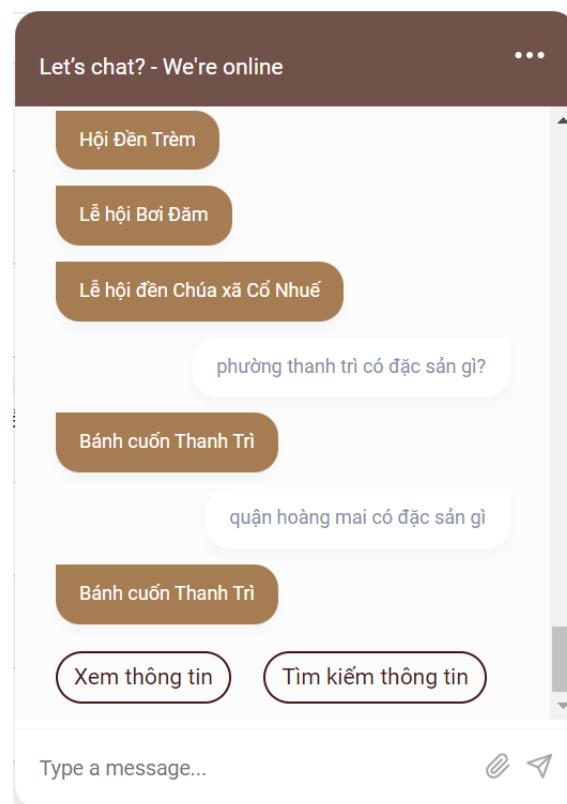
Chức năng hỗ trợ tìm kiếm thông tin các tài nguyên thông qua các thông tin liên quan như địa chỉ, loại hình, lễ hội, đặc sản và địa điểm du lịch hoạt động trong lễ hội.



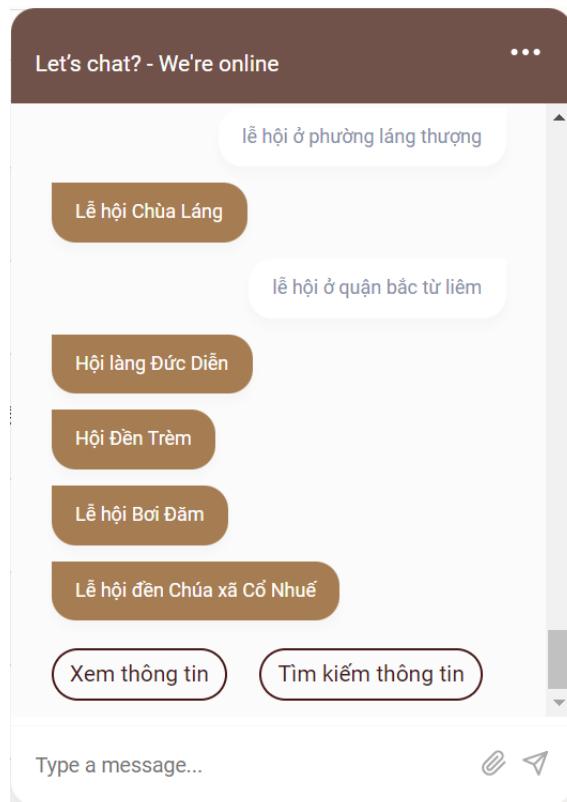
**Hình 4.30:** Minh họa chức năng tìm kiếm thông tin địa điểm du lịch thông qua quận huyện



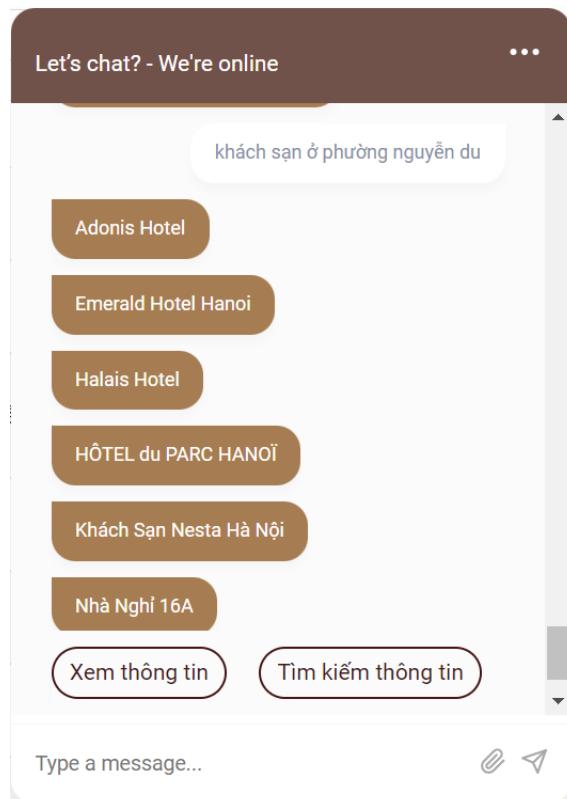
**Hình 4.31:** Minh họa chức năng tìm kiếm thông tin địa điểm du lịch thông qua phường xã



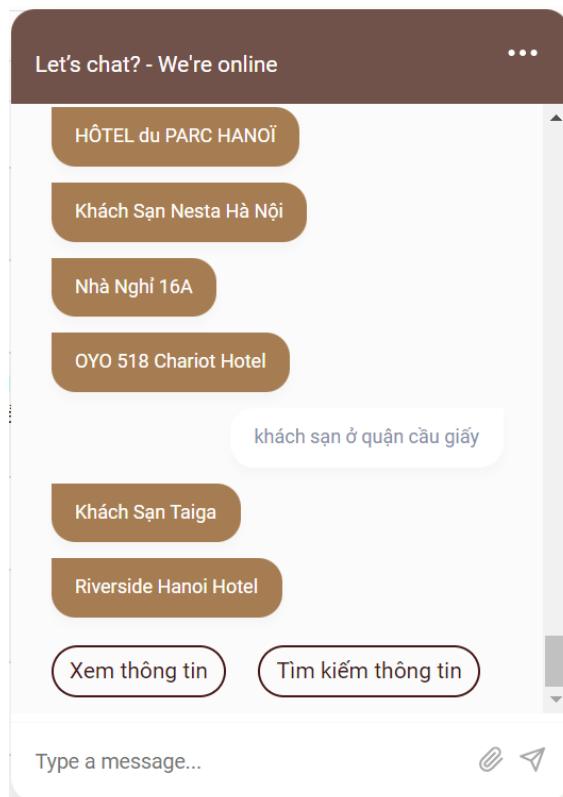
**Hình 4.32:** Minh họa chức năng tìm kiếm thông tin đặc sản thông qua phường xã hoặc quận huyện



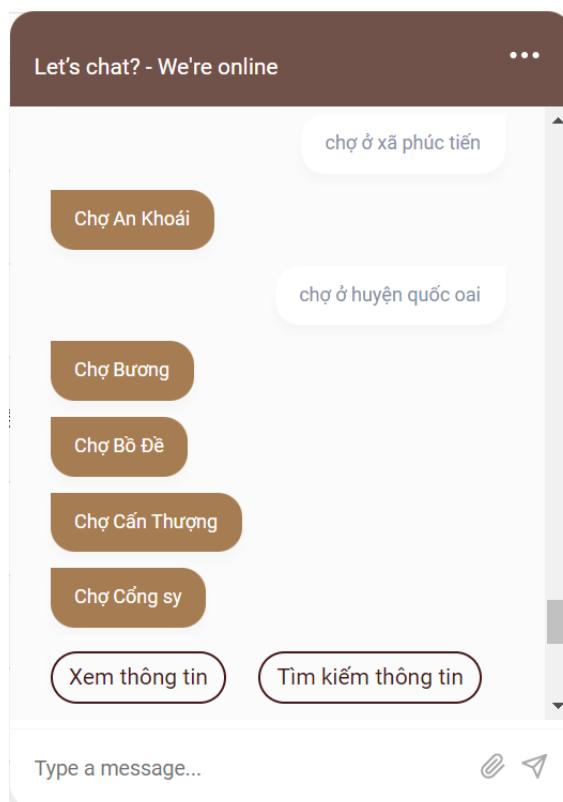
**Hình 4.33:** Minh họa chức năng tìm kiếm thông tin lễ hội thông qua phường xã hoặc quận huyện



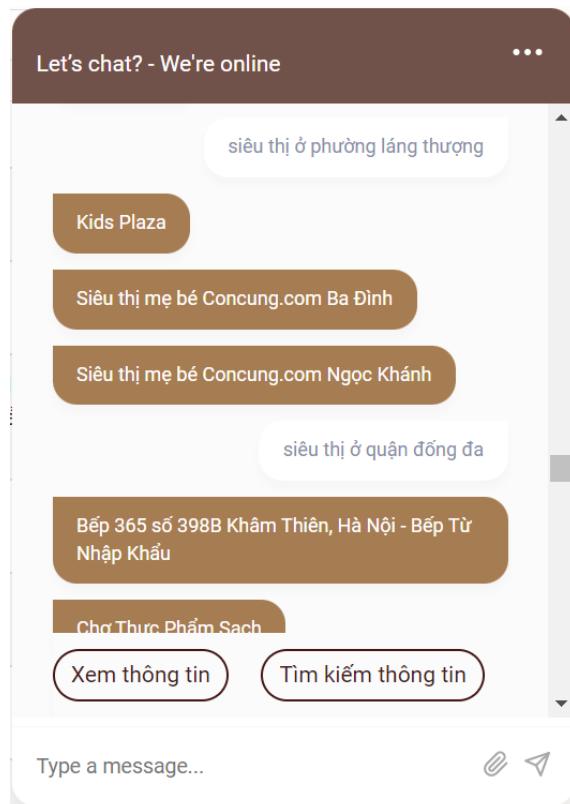
**Hình 4.34:** Minh họa chức năng tìm kiếm thông tin khách sạn thông qua phường xã



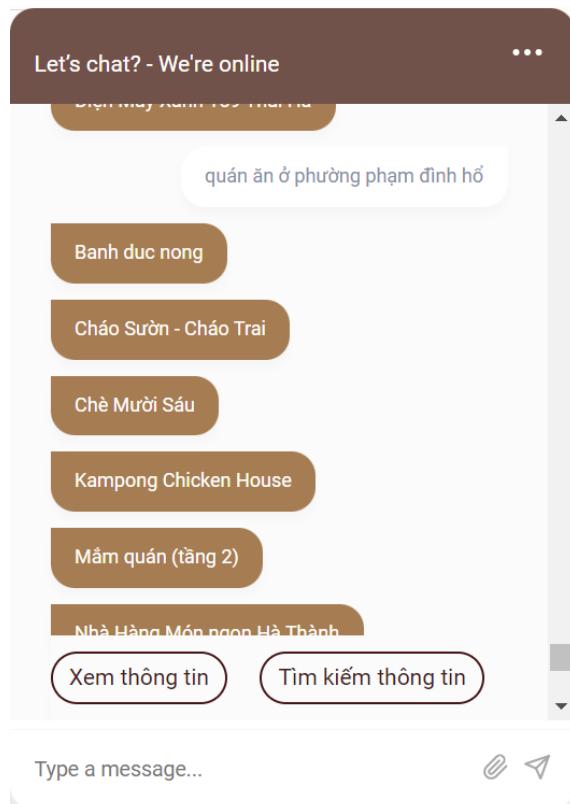
**Hình 4.35:** Minh họa chức năng tìm kiếm thông tin khách sạn thông qua quận huyện



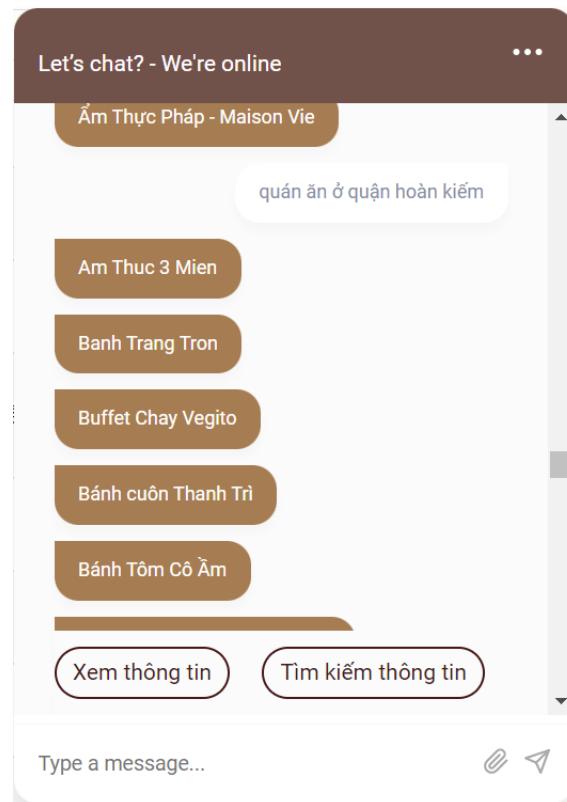
**Hình 4.36:** Minh họa chức năng tìm kiếm thông tin chợ thông qua phường xã hoặc quận huyện



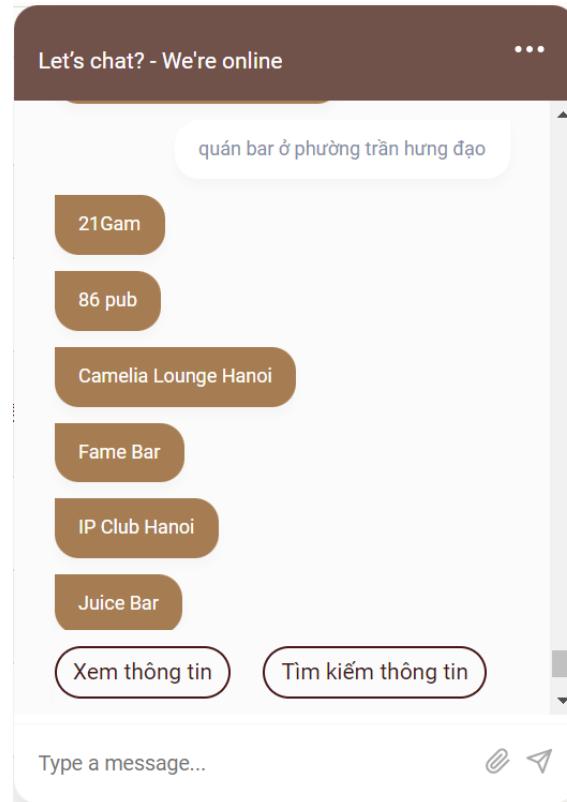
**Hình 4.37:** Minh họa chức năng tìm kiếm thông tin siêu thị thông qua phường xã hoặc quận huyện



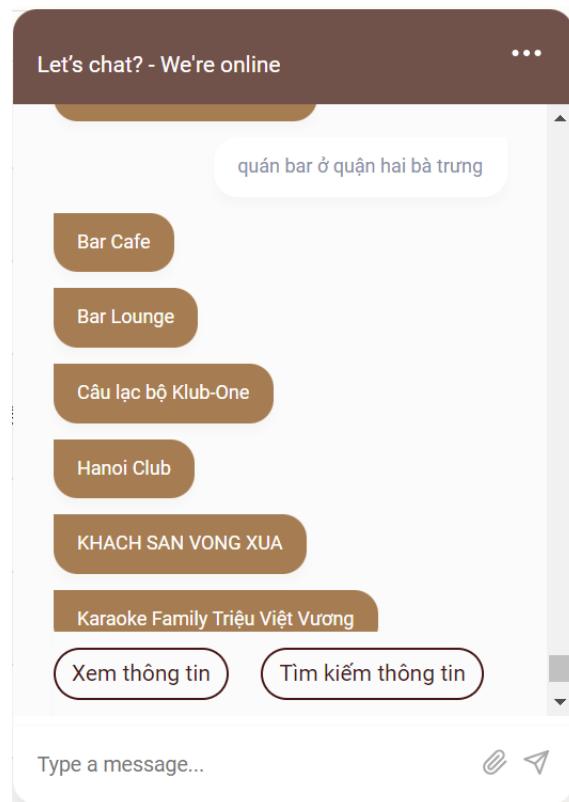
**Hình 4.38:** Minh họa chức năng tìm kiếm thông tin quán ăn thông qua phường xã



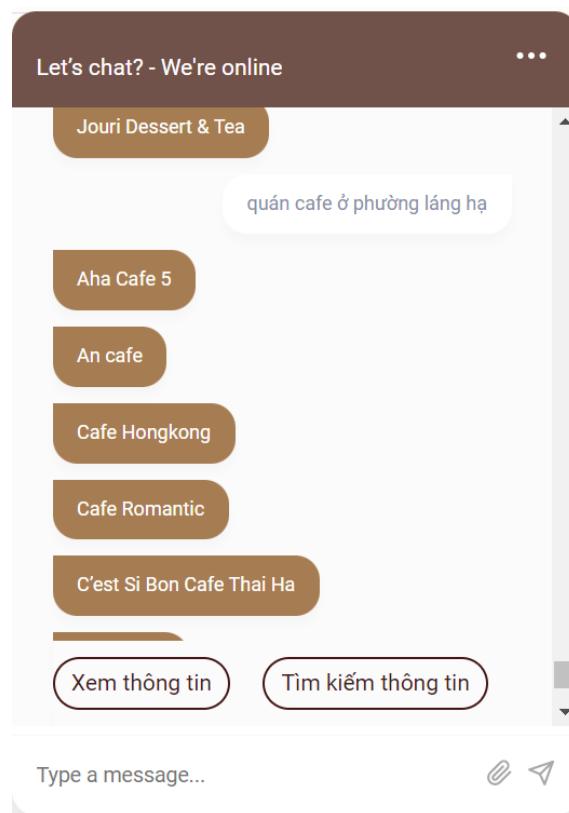
**Hình 4.39:** Minh họa chức năng tìm kiếm thông tin quán ăn thông qua quận huyện



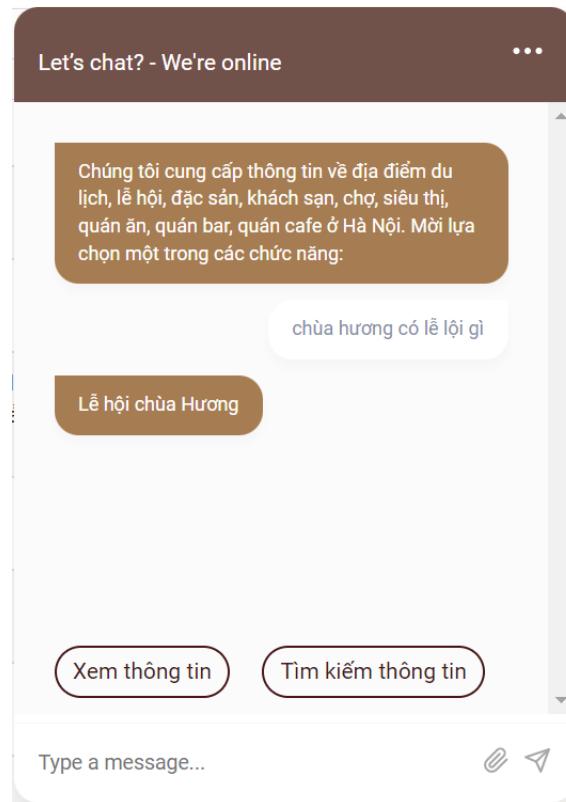
**Hình 4.40:** Minh họa chức năng tìm kiếm thông tin quán bar thông qua phường xã



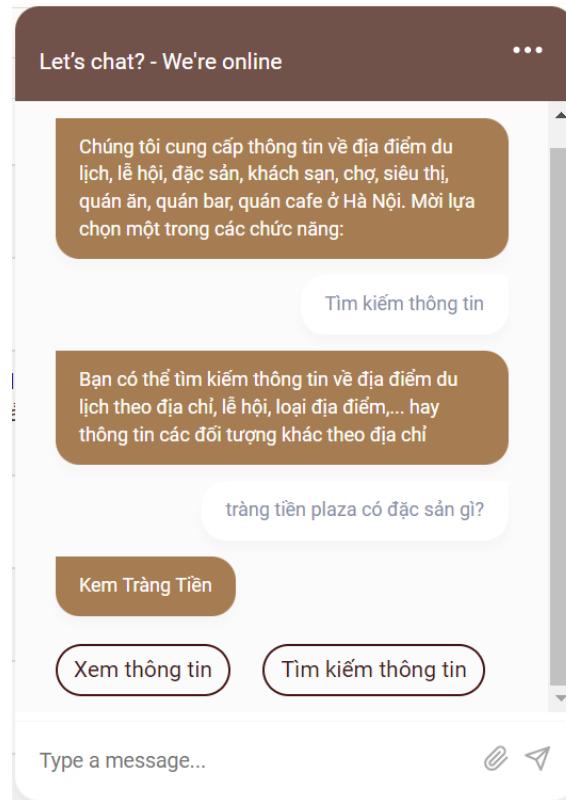
**Hình 4.41:** Minh họa chức năng tìm kiếm thông tin quán bar thông qua quận huyện



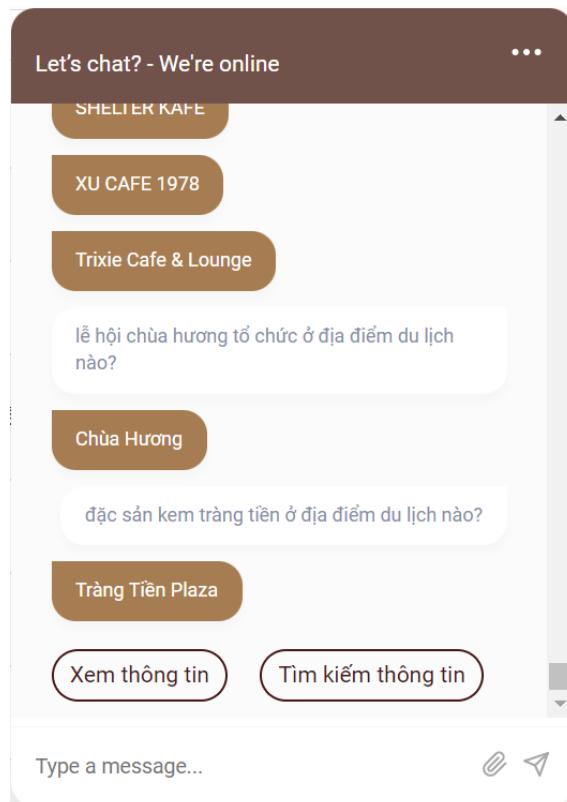
**Hình 4.42:** Minh họa chức năng tìm kiếm thông tin quán cafe thông qua phường xã



**Hình 4.43:** Minh họa chức năng tìm kiếm thông tin lễ hội thông qua địa điểm du lịch



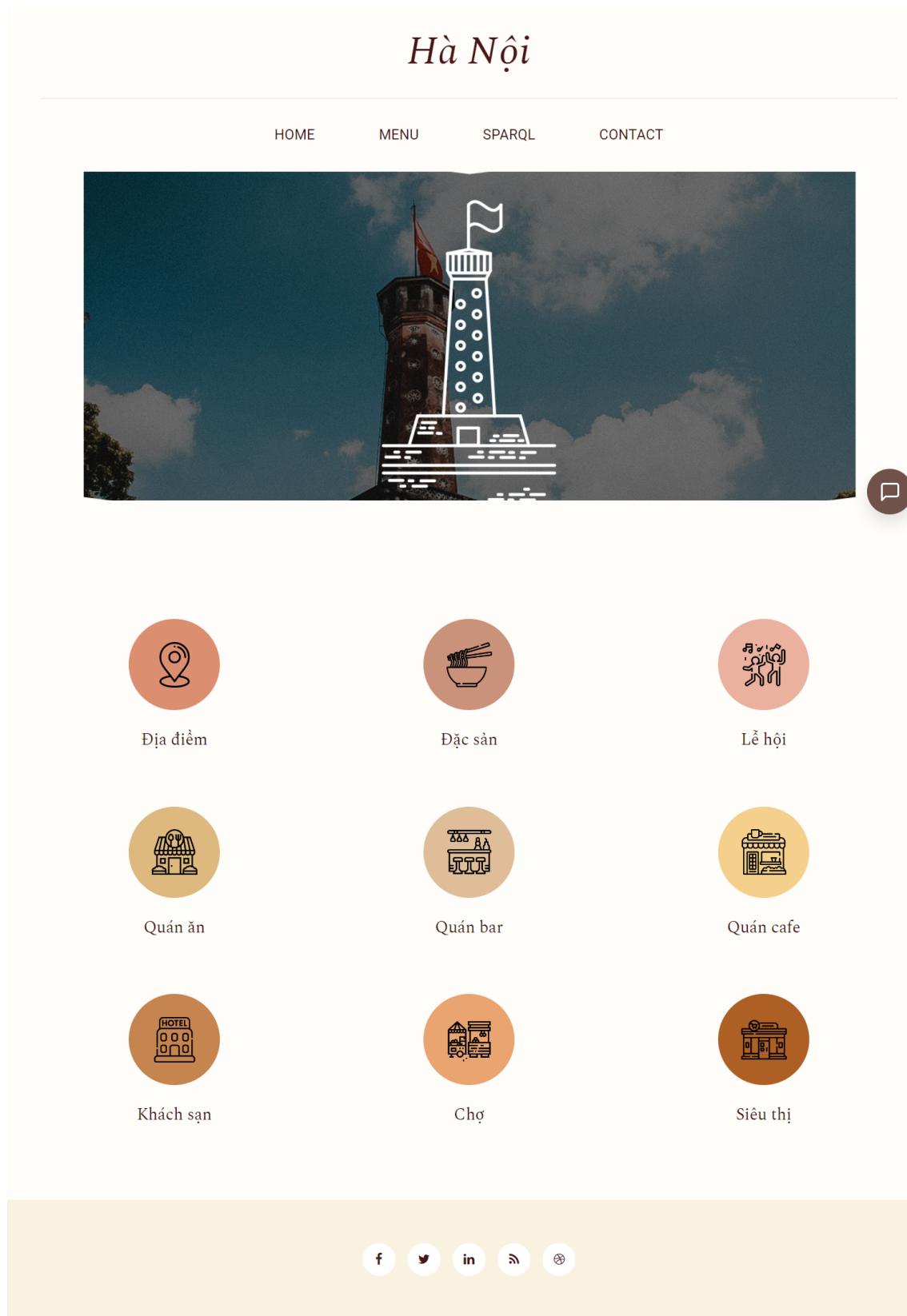
**Hình 4.44:** Minh họa chức năng tìm kiếm thông tin đặc sản thông qua địa điểm du lịch



**Hình 4.45:** Minh họa chức năng tìm kiếm thông tin địa điểm du lịch thông qua lễ hội hoặc đặc sản

#### 4.2.4 Kết quả cài đặt Website tra cứu thông tin

Trang web cho phép người dùng chọn biểu tượng để xem thông tin tài nguyên hoặc chọn SPARQL để truy vấn dữ liệu.



**Hình 4.46:** Minh họa trang chủ của website

#### a, Chức năng "Xem thông tin"

Sau khi danh sách tài nguyên được hiển thị, người dùng có thể chọn tài nguyên muốn xem hoặc tìm kiếm tài nguyên trên thanh Search.

*Hà Nội*

---

HOME MENU SPARQL CONTACT



Show 10 entries Search:

Tên chợ	Địa chỉ
Chợ 19/12	Số 41, Phố Hai Bà Trưng, Phường Trần Hưng Đạo, Hoàn Kiếm, Hà Nội
Chợ 230	Phường Sài Đồng, Long Biên, Hà Nội
Chợ 337 Dịch Vọng	Phường Dịch Vọng, Cầu Giấy, Hà Nội
Chợ A12 Khương Thượng	Tôn Thất Tùng, Đống Đa, Hà Nội
Chợ An Khoái	Phúc Tiến, Phú Xuyên, Hà Nội
Chợ Ao Đông	Phường Trung Hưng, Sơn Tây, Hà Nội
Chợ Ba Lăng	Dũng Tiến, Thường Tín, Hà Nội
Chợ Ba Thá	Viên An, Ứng Hòa, Hà Nội
Chợ Ba Trại	Ba Trại, Ba Vì, Hà Nội
Chợ Bá	Hồng Hà, Đan Phượng, Hà Nội
Tên chợ	Địa chỉ

Showing 1 to 10 of 475 entries

Previous 1 2 3 4 5 ... 48 Next

✉

f t in r e

**Hình 4.47:** Minh họa trang danh sách chợ

Trang thông tin chi tiết của tài nguyên được minh họa trong hình dưới đây.

# Hà Nội

[HOME](#)

[MENU](#)

[SPARQL](#)

[CONTACT](#)

## CHỢ BA LĂNG

Show 10 entries

Search:

Property	Value
Tên chợ	Chợ Ba Lăng
Địa chỉ	Dũng Tiến, Thường Tín, Hà Nội
Vĩ độ	20.8186724
Kinh độ	105.8554513
Quận/ Huyện/ Thị xã	Huyện Thường Tín
Phường/ Xã	Xã Dũng Tiến
Property	Value

Showing 1 to 6 of 6 entries

Previous 1 Next

A map showing the location of Ba Lang Market (Chợ Ba Lăng) in Dong Dien, Thuy Tinh District, Hanoi. The market is located at approximately 20°49'07.2"N 105°51'19.6"E, near the intersection of Phan Van Trach and Pham Van Dong roads. The map also shows other landmarks such as Ngastone (Quần áo trẻ em), Siêu Thị Mini Ánh Kim, Siêu thị sữa bim Oanh Baby, UBND xã Dũng Tiến, Lương Ât, Cửa hàng Gỗ Dán, Nghia trang thôn Ba Lăng, Chùa Phúc Tự, and Chùa Đẳng Giác. A red marker indicates the exact location of the market. The map includes a zoom control and a Google logo.

[f](#) [t](#) [in](#) [s](#) [e](#)

**Hình 4.48:** Minh họa trang thông tin chi tiết chợ

### b, Chức năng "Truy vấn SPARQL"

Chức năng truy vấn cho phép người dùng nhập vào truy vấn bằng ngôn ngữ SPARQL và trả về thông tin dưới dạng bảng.

Ví dụ về một truy vấn SPARQL lấy ra tên, địa chỉ, vĩ độ, kinh độ của các chợ trong cơ sở tri thức.

*Hà Nội*

HOME MENU SPARQL CONTACT

SPARQL QUERY

```
SELECT DISTINCT ?name ?location ?lat ?lng
WHERE
{
?x ex:tenCho ?name;
ex:diaChi ?location;
ex:lat ?lat;
ex:lng ?lng.
}
```

EXECUTE QUERY

f t in r e

**Hình 4.49:** Minh họa trang nhập truy vấn SPARQL

Kết quả trang web trả về được minh họa trong hình dưới đây.

The screenshot shows a web application interface for a dataset about markets in Hanoi. At the top, there is a navigation bar with links to HOME, MENU, SPARQL, and CONTACT. Below the navigation bar, a large dark brown header area contains the text "SPARQL". Underneath this, there is a search bar labeled "Search:" followed by a magnifying glass icon. On the left side of the main content area, there is a button labeled "Show 10 entries". To the right of the search bar, there is a "Search:" input field and a small circular icon with a magnifying glass symbol. The main content area displays a table of market data. The table has four columns: "name", "location", "lat", and "lng". The "name" column lists various market names such as Chợ 19/12, Chợ 230, Chợ 337 Dịch Vọng, Chợ A12 Khương Thượng, Chợ An Khoái, Chợ Ao Đông, Chợ Ba Lăng, Chợ Ba Thá, Chợ Ba Trại, and Chợ Bá. The "location" column provides the address for each market. The "lat" and "lng" columns show the geographical coordinates for each market. Below the table, there is a message indicating "Showing 1 to 10 of 475 entries" and a set of page navigation buttons numbered 1 through 48. At the bottom of the page, there is a yellow footer bar with social media sharing icons for Facebook, Twitter, LinkedIn, and RSS, along with a small circular icon.

name	location	lat	lng
Chợ 19/12	Số 41, Phố Hai Bà Trưng, Phường Trần Hưng Đạo, Hoàn Kiếm, Hà Nội	21.0254476	105.8482526
Chợ 230	Phường Sài Đồng, Long Biên, Hà Nội	21.0348453	105.9127108
Chợ 337 Dịch Vọng	Phường Dịch Vọng, Cầu Giấy, Hà Nội	21.0347067	105.7923394
Chợ A12 Khương Thượng	Tôn Thất Tùng, Đống Đa, Hà Nội	21.0025625	105.8298223
Chợ An Khoái	Phúc Tiến, Phú Xuyên, Hà Nội	20.7251289	105.9259271
Chợ Ao Đông	Phường Trung Hưng, Sơn Tây, Hà Nội	21.1313193	105.5049436
Chợ Ba Lăng	Dũng Tiến, Thường Tín, Hà Nội	20.8186724	105.8554513
Chợ Ba Thá	Viên An, Ứng Hòa, Hà Nội	20.8018254	105.7263152
Chợ Ba Trại	Ba Trại, Ba Vì, Hà Nội	21.1256787	105.3467822
Chợ Bá	Hồng Hà, Đan Phượng, Hà Nội	21.1335253	105.6823124
name	location	lat	lng

Showing 1 to 10 of 475 entries

Previous 1 2 3 4 5 ... 48 Next

**Hình 4.50:** Minh họa trang kết quả truy vấn

## CHƯƠNG 5. KẾT LUẬN

### 5.1 Kết luận

Đồ án "Xây dựng cơ sở tri thức ngữ nghĩa và chatbot về du lịch Hà Nội" đã hoàn thành được các mục tiêu được đặt ra ban đầu.

Về mặt lý thuyết, đồ án đã nêu ra được những nét đặc trưng và ưu thế của công nghệ Web ngữ nghĩa; cách lưu trữ dữ liệu có ngữ nghĩa RDF; tìm hiểu về khái niệm Ontology, các thành phần và công cụ để xây dựng Ontology; tìm hiểu cách sử dụng ngôn ngữ SPARQL để truy vấn RDF; nắm được các kiến thức cơ bản về chatbot như khái niệm, thành phần kiến trúc hay cách xây dựng một chatbot.

Đối với kết quả thực nghiệm, đồ án đã thành công xây dựng cơ sở tri thức ngữ nghĩa về du lịch ở Hà Nội và hệ thống tra cứu, hỏi đáp sử dụng kho tri thức đã xây dựng. Cơ sở tri thức lưu trữ một lượng thông tin tương đối đầy đủ về địa điểm (địa điểm du lịch, khách sạn, chợ, siêu thị, quán ăn, quán bar, quán cafe), lễ hội và đặc sản ở Hà Nội. Các dữ liệu này được thu thập chủ yếu từ Wikipedia, Google Travel và được chuyển đổi sang dạng ngữ nghĩa bằng các mô tả về lớp, quan hệ và thuộc tính trong Ontology. Sau khi hoàn thành việc xây dựng cơ sở tri thức, tôi tiến hành tìm hiểu cách xây dựng chatbot và công nghệ xử lý chatbot sử dụng dịch vụ xử lý ngôn ngữ tự nhiên LUIS. Chatbot gồm hai chức năng chính là xem thông tin và tìm kiếm thông tin liên quan đến du lịch cho phép người dùng tìm kiếm địa điểm du lịch cùng với các tiện ích xung quanh như quán ăn, khách sạn, siêu thị,... Trang web tra cứu thông tin được hoàn thành sau cùng với hai chức năng là xem thông tin và truy vấn SPARQL, biểu diễn cơ sở tri thức một cách rõ ràng và chi tiết giúp người dùng truy vấn thông tin dễ dàng hơn. Hệ thống sau khi hoàn tất hoạt động tương đối ổn định, đáp ứng được những yêu cầu phân tích thiết kế đặt ra. Công nghệ web ngữ nghĩa đã chứng minh được ưu điểm của mình là khả năng suy diễn và tính linh hoạt trong sử dụng dữ liệu giúp hệ thống tìm kiếm và thao tác thông tin với độ chính xác cao hơn.

Ngoài ra, quá trình nghiên cứu lý thuyết và xây dựng hệ thống còn giúp tôi học hỏi được các kiến thức, công nghệ mới như Web ngữ nghĩa, dịch vụ xử lý ngôn ngữ tự nhiên LUIS, Python Selenium,... Tôi cũng tích lũy được thêm kinh nghiệm trong việc nghiên cứu khoa học và nâng cao một số kỹ năng như đọc tài liệu, phân tích và giải quyết vấn đề, cách lập luận và trình bày một báo cáo khoa học.

Bên cạnh những thành công đạt được, đồ án vẫn còn một số hạn chế. Lượng dữ liệu chưa đủ lớn, chưa tận dụng được tối đa ưu thế của Web ngữ nghĩa. Khả năng

tương tác của chatbot còn khá giới hạn, cần xây dựng thông minh hơn để tăng trải nghiệm người dùng. Trang web chỉ ở mức độ cơ bản, còn cần cải thiện tính năng.

## 5.2 Hướng phát triển trong tương lai

Trong khuôn khổ đồ án này, đề tài chỉ dừng lại ở tìm hiểu, nghiên cứu và bước đầu xây dựng ứng dụng với quy mô nhỏ. Trong tương lai, đồ án có thể mở rộng để xây dựng hệ thống tra cứu và hỏi đáp thông tin du lịch ở Việt Nam hay thậm chí là toàn thế giới với đầy đủ tính năng hơn, đáng tin cậy và thân thiện với người dùng.

Cơ sở tri thức về du lịch có khả năng mở rộng rất lớn. Thay vì chỉ thu thập dữ liệu du lịch ở Hà Nội, hệ thống có thể mở rộng phạm vi địa lý thành du lịch Việt Nam hay toàn thế giới. Tăng số lượng lớp, thuộc tính trong cơ sở tri thức để cung cấp nhiều thông tin liên quan hơn và nâng cao khả năng suy diễn. Ví dụ, lớp "khách sạn" có thể bổ sung thêm thuộc tính "hạng của khách sạn", "loại phòng",... hay lớp "quán ăn" có thể bổ sung thêm thuộc tính "thực đơn", "đánh giá",...

Cùng với sự mở rộng của cơ sở tri thức, hệ thống tra cứu và hỏi đáp cũng có khả năng bổ sung thêm nhiều chức năng mới để người dùng có thể trải nghiệm. Chatbot có thể thêm ý định và cải thiện độ chính xác nhờ sự gia tăng của dữ liệu. Thu thập các câu đầu vào của người dùng để huấn luyện chatbot, tạo ra nhiều kịch bản mới. Trang web có thể phát triển một số tính năng mới như cho phép các cá nhân hay tổ chức đăng ký thêm thông tin vào cơ sở tri thức như khách sạn, nhà hàng,...; liên kết với các trang web khác để làm giàu nguồn dữ liệu,...

## TÀI LIỆU THAM KHẢO

- [1] T. Berners-Lee, J. Hendler **and** O. Lassila, “The semantic web,” *Scientific american*, **jourvol** 284, **number 5, pages** 34–43, 2001.
- [2] R. W. Group, *W3c semantic web activity*. **url:** <https://www.w3.org/2001/sw/> (**urlseen** 30/06/2022).
- [3] N. C. Nhật, *Semantic web và thư viện số*. **url:** <https://nltv.gov.vn/nghiep-vu-thu-vien/semantic-web-va-thu-vien-so.html> (**urlseen** 30/06/2022).
- [4] W3C, *Architecture*. **url:** <https://www.w3.org/2000/Talks/1206-xml2k-tbl/slides10-0.html> (**urlseen** 30/06/2022).
- [5] N. T. K. Thương, “Đồ án tốt nghiệp "xây dựng chatbot hỗ trợ du lịch lễ hội tại việt nam ứng dụng web ngữ nghĩa",” phdthesis, Đại học Bách Khoa Hà Nội, VN, 2020.
- [6] R. W. Group, *Rdf*. **url:** [w3.org/RDF/](https://w3.org/RDF/) (**urlseen** 30/06/2022).
- [7] T. Gruber, *Ontology*. 2018.
- [8] W3C, *Web ontology language (owl)*. **url:** <https://www.w3.org/OWL/> (**urlseen** 30/06/2022).
- [9] ——, *Sparql 1.1 query language*. **url:** <https://www.w3.org/TR/sparql11-query/> (**urlseen** 30/06/2022).
- [10] Maruti Techlabs, *How do chatbots work? a guide to chatbot architecture*. **url:** <https://marutitech.com/chatbots-work-guide-chatbot-architecture/> (**urlseen** 07/07/2022).
- [11] J. O. C. Trygve Reenskaug, “The dci architecture: A new vision of object-oriented programming,” 2009.
- [12] FirebirD.
- [13] D. M. T. Nguyễn, “Ứng dụng dữ liệu mở liên kết trong việc nâng cao chất lượng các nguồn tài nguyên giáo dục mở,” 2020.
- [14] Microsoft, *What is language understanding (luis)?* **url:** <https://docs.microsoft.com/vi-vn/azure/cognitive-services/luis/what-is-luis> (**urlseen** 07/07/2022).