VIỆN HÀN LÂM KHOA HỌC CÔNG NGHỆ VIỆT NAM

VIỆN ĐỊA LÝ TÀI NGUYÊN THÀNH PHỐ HỒ CHÍ MINH

BÁO CÁO

**THUYẾT MINH XÂY DỰNG ONTOLOGY**

**PHỤC VỤ TÌM KIẾM DỮ LIỆU**

ĐỀ TÀI

**“NGHIÊN CỨU XÂY DỰNG HỆ THỐNG THÔNG TIN NÔNG NGHIỆP PHỤC VỤ TÁI CƠ CẤU NÔNG NGHIỆP   
TỈNH AN GIANG”**

Mã số: 372.2016.14

**Cơ quan chủ trì: Viện Địa lý tài nguyên TP. Hồ Chí Minh**

**Chủ nhiệm đề tài: TS. Trần Thái Bình**

Người thực hiện

**TS. Trần Thái Bình**

TP. HỒ CHÍ MINH - 2017

VIỆN HÀN LÂM KHOA HỌC CÔNG NGHỆ VIỆT NAM

VIỆN ĐỊA LÝ TÀI NGUYÊN THÀNH PHỐ HỒ CHÍ MINH

BÁO CÁO

**THUYẾT MINH XÂY DỰNG ONTOLOGY**

**PHỤC VỤ TÌM KIẾM DỮ LIỆU**

ĐỀ TÀI

**NGHIÊN CỨU XÂY DỰNG HỆ THỐNG THÔNG TIN NÔNG NGHIỆP PHỤC VỤ TÁI CƠ CẤU NÔNG NGHIỆP TỈNH AN GIANG**

Mã số: 372.2016.14

**Cơ quan chủ trì: Viện Địa lý tài nguyên TP. Hồ Chí Minh**

**Chủ nhiệm đề tài: TS. Trần Thái Bình**

Người thực hiện

**TS. Trần Thái Bình**

TP. HỒ CHÍ MINH - 2017

DANH MỤC CHỮ VIẾT TẮT

|  |  |
| --- | --- |
| OWL | Ontology Web Language |
| RDF | Resource Description Framework |
| RDF(S) | Resource Desscription Framework (Schema) |
| HTML | HyperTeXt Markup Language (Ngôn ngữ đánh dấu siêu văn bản) |
| W3C | The World Wide Web Consortium (Tiêu chuẩn thiết kế Web) |
| XML | eXtensible Markup Language (Ngôn ngữ đánh dấu mở rộng) |
| URI | Uniform Resource Identifier (Định danh tài nguyên) |
| CSS | Cascading Style Sheets (Tập tin định dạng theo tầng) |

DANH MỤC HÌNH ẢNH

[Hình 1: Mô hình phân cấp trong Ontology [31] 1](#_Toc514203649)

[Hình 2: Mô hình phân cấp và mối quan hệ trong Ontology cây trồng [23] 3](#_Toc514203650)

[Hình 3: Ontology Learning Layer Cake [34] 8](#_Toc514203651)

[Hình 4: Cấu trúc phân tầng trong tiếp cận xây dựng Ontology [50] 8](#_Toc514203652)

[Hình 5: Chi tiết cấu trúc Ontology [50] 9](#_Toc514203653)

[Hình 6: Sơ đồ cách tiếp cận toàn hệ thống [35] 11](#_Toc514203654)

[Hình 7: Sơ đồ phân tích của nhiệm vụ xây dựng Ontology [35] 11](#_Toc514203655)

[Hình 8: Các cấp độ biểu diễn tri thức [33] 13](#_Toc514203656)

[Hình 9: Các quan hệ được sử dụng trong toàn bộ Ontology lúa gạo Thái Lan [23] 15](#_Toc514203657)

[Hình 10: Các yếu tố chính liên quan đến Ontology lúa gạo [23] 18](#_Toc514203658)

[Hình 11: Ví dụ một Ontology bằng ngôn ngữ OWL về Pizza - W3C 25](#_Toc514203659)

[Hình 12: Ảnh chụp màn hình giao diện làm việc Protege: phân lớp trong lĩnh vực nông nghiệp An Giang 27](#_Toc514203660)

[Hình 13: Ảnh chụp màn hình giao diện làm việc Protege: phân lớp trong lĩnh vực nông nghiệp An Giang (tiếp) 28](#_Toc514203661)

[Hình 14: Ảnh chụp màn hình giao diện làm việc Protege: phân lớp trong lĩnh vực nông nghiệp An Giang (tiếp) 29](#_Toc514203662)

[Hình 15: Ảnh chụp màn hình giao diện làm việc Protege: thuộc tính cho các lớp trong Ontology nông nghiệp An Giang 30](#_Toc514203663)

[Hình 16: Ảnh chụp màn hình giao diện làm việc Protege: khía cạnh của thuộc tính trong Ontology nông nghiệp An Giang 30](#_Toc514203664)

[Hình 17: Ảnh chụp màn hình giao diện làm việc Protege: Các loại dữ liệu của các khía cạnh trong Ontology nông nghiệp An Giang 31](#_Toc514203665)

# Tổng quan về Ontology

## Lịch sử phát triển và thuật ngữ Ontology

Thuật ngữ Ontology có nguồn gốc và phát triển từ môn khoa học Metaphysics (Siêu hình học) xuất hiện từ rất sớm và có nguồn gốc từ triết học [18], lần đầu tiên được nhắc đến trong quan điểm triết học của Aristotle (cha đẻ của ngành khoa học Metaphysics). Theo Aritotle, Metaphysics là một môn khoa học nghiên cứu, quan sát thế giới thực, tức là những vật thể đang tồn tại hoặc không tồn tại. Các vật thể tồn tại phải có đặc điểm nhận dạng, tính đa dạng, mối quan hệ [19], [20].

Mặc dù thuật ngữ Ontology có gốc từ xuất phát từ Triết học nhưng có thể phân biệt Ontology thành hai loại khác nhau, một là Ontology thuần túy trong Triết học, hai là Ontology dành cho áp dụng trong khoa học. Chúng có những điểm giao nhau về mặt khái niệm nhưng trong chuyên đề này thuật ngữ Ontology phải được hiểu theo nghĩa thứ hai – Ontology được áp dụng cho khoa học.

Thuật ngữ Ontology trong khoa học được giới thiệu lần đầu tiên bởi Gruber [21] vào năm 1995, theo đó Ontology được miêu tả là sự định nghĩa rõ ràng của một khái niệm đã được chia sẻ. Trong *“Theory and History of Ontology”* [18]*,* Raul Corazzon cũng đã định nghĩa Ontology là lý thuyết về các vật chất và mối quan hệ của chúng, nó cung cấp các tiêu chí để phân biệt các loại đối tượng khác nhau (cụ thể và trừu tượng, tồn tại và không tồn tại, thực tế và lý tưởng, độc lập và phụ thuộc) và các mối quan hệ của chúng (quan hệ, phụ thuộc, xác nhận). Còn đối với ngành khoa học máy tính, Ontology mang ý nghĩa hiện đại hơn, nó là tập hợp các loại thực thể được định nghĩa và các quan hệ giữa chúng được sắp xếp trong một hệ thống có cấu trúc theo từng ngữ cảnh mà người dùng đang quan tâm. Hay nói cách khác, Ontology cung cấp một bộ từ vựng chung bao gồm các khái niệm, các thuộc tính cơ bản và quan trọng để nhận dạng được đối tượng, các định nghĩa của các khái niệm và các thuộc tính, ngoài ra nó còn cung cấp các ràng buộc, quan hệ giữa các thực thể, đối tượng, đôi khi các ràng buộc này là giả định theo một ý nghĩa mong muốn của bộ từ vựng được sử dụng trong một lĩnh vực nào đó, từ đó có thể giao tiếp được giữa người với người, người với máy tính, máy tính với máy tính từ các hệ thống phân tán khác nhau.

Ngày nay, Ontology được ứng dụng phổ biến trong nhiều lĩnh vực khác nhau, nổi bật là trong các nghiên cứu về trí tuệ nhân tạo (AI), hoặc làm nền tảng cho Semantic Web (SW), các hệ thống kỹ thuật, tin học y sinh… Mục đích của Ontology đó là tạo ra được bộ từ vựng có cấu trúc giúp cho các hệ thống phân tán có thể cùng hiểu và khai thác tài nguyên của nhau.

## Thành phần chính của Ontology

Theo quan điểm của tổ chức W3C, một Ontology thường bao gồm các thành phần chính sau (W3C, 2009):

***Cá thể (individual)***: hay gọi là trường hợp (Instances) của các đối tượng (Object) là các thành phần cơ bản, nền tảng của một Ontology, các cá thể có thể là các đối tượng cụ thể như con người, xe máy, cái ly…nhưng cũng có thể mang ý nghĩa trừu tượng. Một cá thể có thể thuộc một lớp hoặc nhiều lớp. Một ví dụ cho cá thể của một Ontology, cụ thể là Ontology về phim ảnh (movie ontology) thì một bộ phim cụ thể (Harry Porter hoặc War of Star) là một cá thể thuộc Ontology phim ảnh, một đạo diễn cụ thể (George Lucas) là một cá thể thuộc Ontology phim ảnh, tuy nhiên Thể loại phim (hành động, tâm lý…) lại không phải là một cá thể.

***Lớp (class):***  hay cũng có tài liệu gọi là Khái niệm (concept) trong một Ontology là cách để giới thiệu về những đặc tính chung của một nhóm các đối tượng. Cũng có thể xem một Class trong Ontology là thể loại (types) hay hạng mục (categories), ví dụ như thể loại phim (hành động, tâm lý, tài liệu…) là một Class trong Movie Ontology. Các Classes được tổ chức theo một cấu trúc có thứ bậc để mô tả các loại đối tượng trong một miền quan tâm. Ví dụ, "sinh vật" là một lớp trong phạm vi sinh học. Một lớp có thể có các lớp con như "động vật" và "thực vật" là các lớp con của “sinh vật”.

***Thuộc tính (attribute)*** là những tính chất mà một đối tượng có thể có. Một khái niệm hay một lớp được mô tả bằng một loạt các thuộc tính thông thường (chính). Một ví dụ, một bộ phim có thể được mô tả bởi các thuộc tính cấu thành lên nó như kịch bản, đạo diễn, diễn viên, âm nhạc…

***Quan hệ (relation)*** là cách mà các khái niệm có thể ràng buộc với các khái niệm khác. Một ví dụ để thấy rõ hơn về một quan hệ trong một Ontology như sau: ( nghĩa là và có mối quan hệ với nhau thông qua , và có quan hệ với nhau thông qua .

Ngoài các thành phần trên, một số Ontology còn có thể chứa các giới hạn (restrictions), đây là phần mô tả những khó khăn khi định nghĩa các cá thể hay đối tượng; những chân lý hay tiên đề (axioms) mà theo đó chính là sự tiên nghiệm về những chân lý luôn luôn đúng; các sự kiện (events) là sự thay đổi của các thuộc tính hay các mối quan hệ của các đối tượng.

Còn với quan điểm của Gruber [5][6] thì một Ontology bao gồm 5 thành tố chính đó là: lớp (classes), quan hệ (relations), chức năng (functions), tiên đề (axioms) và trường hợp (instances).

* Các lớp mô tả cho các khái niệm, có thể được xem như các thực thể chung nhất theo nghĩa rộng.
* Quan hệ mô tả loại liên kết giữa các khái niệm với toàn bộ lĩnh vực.
* Chức năng và một trường hợp cụ thể của quan hệ.
* Tiên đề dùng để chỉ những câu mang nội dung luôn đúng. Nó thường được sử dụng để mô tả các tri thức không thể được định nghĩa chính thức bằng các thành phần trong Ontology.
* Trường hợp được sử dụng để mô tả các yếu tố hoặc cá thể trong Ontology.

Một Ontology có thể được mô tả bằng 4 thành tố chính sau: [7]

O= <C, R, I, A.>

C: là tập hợp các lớp mô tả về các khái niệm mà chúng sẽ lý giải cho một lĩnh vực cụ thể, ví dụ như: hóa đơn, thanh toán, sản phẩm, giá cả…

R: là tập hợp các quan hệ liên kết giữa các lớp với nhau, như là quan hệ *“ProductHasPrice” (sản phẩm có giá – quan hệ giữa sản phẩm và giá)*.

I: là tập hợp các trường hợp, trong đó mỗi trường hợp có thể là một cá thể của một lớp hoặc nhiều lớp và có thể liên kết với các cá thể khác bằng các mối quan hệ, ví dụ như: *“Sản phẩm A hasPrice 170 nghìn đồng.”*

A: là tập hợp các tiên đề như là: *“Nếu một sản phẩm có giá hơn 200 nghìn thì giao hàng miễn phí.”* .

## Phân loại Ontology

Ontology thường được phân loại dựa vào tính tổng quát của các khái niệm bên trong nó, phạm vi của các khái niệm và mục đích sử dụng [7]:

* Ontology cấp cao (Upper-level Ontology): thường mô tả cho một mô hình chung của thế giới, có thể đáp ứng được nhiều nhiệm vụ, lĩnh vực nghiên cứu, lĩnh vực ứng dụng rộng lớn.
* Ontology của một lĩnh vực cụ thể (Domain Ontology) thường mô tả cho một lĩnh vực cụ thể như là đường xá, dược phẩm, lúa gạo…
* Ontology cho một ứng dụng hoặc một nhiệm vụ (Application and Task Ontology) cụ thể phù hợp với phạm vi nghiên cứu của một đề tài hay một nhiệm vụ rõ ràng.

Hay với [8] đã phân loại Ontology theo cấp độ của chúng phụ thuộc vào một nhiệm vụ cụ thể hay quan điểm như sau:

* Các Ontology cấp cao (Top-level Ontology) mô tả những khái niệm chung chung, tổng quát.
* Ontology theo một lĩnh vực cụ thể (Domain Ontology) cung cấp bộ từ vựng về các khái niệm mà ở đó có các quan hệ, hoạt động, lý thuyết và các thành tố chính.
* Ontology thực hiện theo một nhiệm vụ (Task Ontology) mô tả bộ từ vựng và cấu trúc của các tri thức liên quan đến nhiệm vụ hay hoạt động chính.
* Ontology ứng dụng (Application Ontology) chứa tất cả các định nghĩa cần thiết cho một mô hình các tri thức được yêu cầu từ một ứng dụng cụ thể.

Hay trong nghiên cứu của Mizoguchi và các đồng nghiệp [9] đã đề xuất 4 loại Ontology như sau:

* Ontology nội dung (Content Ontology) dành cho việc tái sử dụng lại các tri thức bao gồm: Ontology cho một lĩnh vực, ontology cho một nhiệm vụ và Ontology chung và phổ biến.
  + Ontology cho một lĩnh vực cụ thể (Domain Ontology) được chia thành: Ontology đối tượng (Object Ontology), Ontology các hoạt đông (Activity Ontology) và Ontology lĩnh vực (Field Ontology).
  + Ontology nhiệm vụ (Task Ontology) được chia thành Ontology các danh từ chung (Generic Noun Ontology), Ontology các động từ chung (Generic Verb Ontology) và Ontology các tính từ chung (Generic Adjectives Ontology).
  + Ontology tổng quát/ phổ biến (General/ Common Ontology) chứa các yếu tố chung chung như là các vật, sự kiên, thời gian, không gian, quan hệ nhân quả, hành vi, chức năng…
* Ontology giao tiếp (Nói và hỏi) – Communication Ontology (“Tell” & “Ask”) cho việc chia sẻ các tri thức.
* Ontology chỉ mục (Indexing Ontology) để hỗ trợ việc truy tìm thông tin.
* Ontology các thuật ngữ và Ontology mô tả tri thức (meta-Ontology and Knowledge Representation Ontology).

Lassila và McGuinness [10] đã phân chia Ontology dựa vào tính chất phong phú của cấu trúc bên trong. Họ đã chỉ ra rằng có: Bộ từ vựng có được kiểm soát (Controlled Vocabularies), bảng chú giải (Glossaries), từ điển (Thesauri), Thứ bậc theo is-a không chính thức, thứ bậc is-a chính thức, thứ bậc is-a chính thức mà có chứa các cá thể, khuôn khổ, các giá trị hạn chế và các ràng buộc chung về logic.

Trong khi đó Van Heijst và các đồng nghiệp [11] đã xác định các loại của nền tảng Ontology dựa trên các đối tượng và khái niệm hóa như sau:

* Ontology mô tả tri thức (Knowledge Representation – KB Ontology) đó là nắm bắt các mô tả chính được sử dụng để hình thành tri thức theo một mô hình tri thức nhất định.
* Ontology tổng quát hay thông dụng được sử dụng để mô tả khả năng tái sử dụng các tri thức chung theo một lĩnh vực nào đó.

## Tổng quan nghiên cứu về Ontology trong lĩnh vực nông nghiệp

### Nghiên cứu Ontology trên thế giới

***Nghiên cứu Ontology trong lĩnh vực Nông nghiệp***

* Bộ từ điển AGROVOC

Trong lĩnh vực Nông nghiệp, Tổ chức FAO cũng đã khởi đầu trước tiên với việc xây dựng bộ từ vựng AGROVOC [22]–[24] bao gồm tất cả các khái niệm trong tất cả các lĩnh vực quan tâm của của Tổ chức FAO, bao gồm: dinh dưỡng, nông nghiệp, thủy sản, lâm nghiệp, môi trường…AGROVOC bao gồm hơn 32.000 khái niệm và được diễn đạt bằng 23 ngôn ngữ như tiếng Ả Rập, Trung Quốc, Séc, Anh, Pháp, Đức, Hindi, Hungary, Ý, Nhật, Hàn Quốc, Lào, Malay, Tiếng Thái, tiếng Thổ Nhĩ Kỳ và tiếng Ukraina. Với mục đích hỗ trợ cho người dùng có thể tra cứu bất cứ thông tin mà chưa nắm vững trong các lĩnh vực nói trên, chẳng hạn như tra cứu về tên gọi phổ thông của một loại cây trồng bằng một ngôn ngữ cụ thể nào đó, hoặc có thể tìm mối quan hệ giữa sản phẩm và cây trồng sản xuất ra sản phẩm đó. Bộ từ vựng này có thể được sử dụng như kế thừa cho nhiều hệ thống tìm kiếm thông tin. Ngay sau đó, FAO cũng đã triển khai dự án AOS (Agricultural Ontology Service) [25], [26] kể từ năm 2001. Bằng cách sử dụng các kiến thức có trong hệ thống từ vựng và từ điển như là AGROVOC, AOS cam kết phát triển các thuật ngữ, khái niệm chuyên ngành cụ thể theo từng lĩnh vực, từ đó hỗ trợ tốt hơn cho hệ thống quản lý thông tin trong Nông nghiệp trên môi trường web. AOS với mục đích là cung cấp một dịch vụ kiến thức bởi bộ Ontology trong lĩnh vực nông nghiệp. Nó có thể là thông tin về thời gian gieo hạt cho hệ thống thông tin nông nghiệp hoặc là hệ thống quản lý thông tin về tình hình dịch bệnh trong nông nghiệp.

* CAB Thesaurus

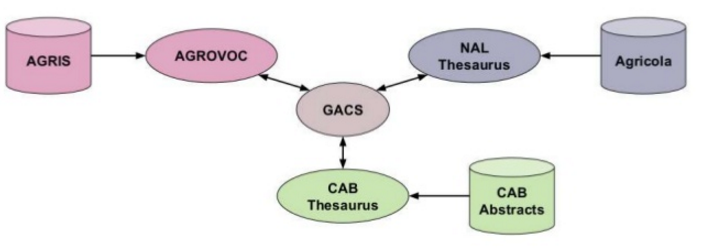
CAB Thesaurus [24], [27] là bộ từ điển các khái niệm, thuật ngữ trong nhiều lĩnh vực khác nhau như: nông nghiệp, lâm nghiệp, trồng trọt, nghiên cứu về đất, côn trùng, sinh lý học, thuốc thú y, dinh dưỡng và nghiên cứu nông thôn. Nó cung cấp công cụ tìm kiếm cho người dùng khi sử dụng dựa trên bộ cơ sở dữ liệu CAB ABSTRACTS và GLOBAL Health, cũng như các sản phẩm có liên quan. Có thể tham khảo bộ từ vựng tại địa chỉ trang web sau: www.cabi.org/cabthesaurus/. CAB có những điểm mạnh sau đây:

* Cung cấp bộ từ vựng bắt đầu được cập nhật từ năm 1983, có nghĩa là trải qua một thời gian rất dài nên bao hàm nhiều thuật ngữ tương đồng, thuận lợi trong nghiên cứu.
* Được cập nhật thường xuyên và bản cập nhật mới nhất là vào tháng 7/2017.
* Phạm vi tri thức của bộ từ vựng bao trùm trên nhiều lĩnh vực khoa học, đời sống, khoa học công nghệ và khoa học xã hội thuần túy cũng như ứng dụng.
* Tổng số các thuật ngữ hiện nay đã đạt gần 2.7 triệu thuật ngữ.
* Các thuật ngữ bao gồm đầy đủ mọi mức độ chi tiết cũng như các thuật ngữ liên quan.
* Bộ từ vựng đa ngôn ngữ với khả năng diễn đạt bằng tiếng Hà Lan, tiếng Bồ Đào Nha và tiếng Tây Ban Nha và hầu hết bằng tiếng Anh, cộng với nội dung nhỏ hơn bằng tiếng Đan Mạch, tiếng Phần Lan, tiếng Pháp, tiếng Đức, tiếng Ý, tiếng Na Uy và tiếng Thụy Điển.
* NALT

Thư viện Thuật ngữ Nông nghiệp Quốc gia (National Agriculture Library Thesaurus – NALT) [28], [29] là một bộ từ vựng trực tuyến về các thuật ngữ nông nghiệp diễn đạt bằng tiếng Anh và Tây Ban Nha, và cũng là một sản phẩm của sự phối hợp giữa Thư viện Nông nghiệp Quốc Gia, tổ chức USDA và Viện Hợp tác Hoa Kỳ về lĩnh vực phối hợp trong nông nghiệp cũng như các Viện Nông nghiệp Mỹ Latin khác. Danh mục thuật ngữ về sinh học chiếm phần lớn thuật ngữ trong bộ từ điển NALT. Trong khi đó, danh mục Địa lý chính trị chủ yếu được mô tả ở cấp quốc gia thuộc Từ điển Nông nghiệp NAL bao gồm các cấu trúc phân cấp, tương đương và mối quan hệ kết hợp giữa các khái niệm. Hiện tại, bộ từ điển này được xem như chỉ mục chính cho cơ sở dữ liệu của Thư viện Nông nghiệp Quốc gia (NAL) về trích dẫn các nguồn tài nguyên nông nghiệp.

* GACS [24]

Một dự án kết hợp tất cả các khái niệm của 3 bộ từ điển trên gồm có AGROVOC, CAB và NAL thành một – Dự án khái niệm Nông nghiệp Toàn cầu (Global Agricultural Concept Scheme – GACS). GACS đóng vai trò như cầu nối các tài liệu nghiên cứu của AGROVOC, CAB và NAL tương ứng với sự phối hợp của FAO, Trung tâm Nông nghiệp và Khoa học Quốc Tế (CABI) và Thư viện Nông nghiệp Quốc Gia Hoa Kỳ (NAL).



Hình 7. Sơ đồ kết hợp các hệ thống Ontology [24]

* FOM và FORS [30], [31]

Trong những năm gần đây, nhiều đề tài nghiên cứu về việc áp dụng kỹ thuật Ontology vào trong quản lý nông nghiệp và đã đạt được nhiều kết quả, điển hình như một số dự án như Mô hình FOM (floricultural ontology model) ứng dụng Ontology trong lĩnh vực trồng hoa tại Trung Quốc, nối tiếp dự án trên là FORS (Floricultural Ontology Retrieve System). Hay như dự án LODE (large scale ontology construction, an ontology development and management system) cũng là một dự án nghiên cứu ứng dụng Ontology trong nông nghiệp như với quy mô lớn.

* RO

Dự án Ontology về Gạo (Rice Ontology – RO) [32] được phát triển ở Nhật Bản là một Ontology chuyên về các thông tin về bộ gen của gạo và đã được phát triển như là một Ontology trong lĩnh vực Sinh học, có thể được áp dụng để trao đổi những thông tin về nguồn gen. Trong Ontology về Sinh học, nguồn Gen và các quá trình sinh học được kết nối để bao phủ các tri thức về di truyền học cũng như thông tin về Gen [33].

* PO

Dự án Ontology về thực vật (Plant Ontology – PO) [34] được phát triển bởi Jaiswal và các đồng nghiệp, nó chứa đựng bộ từ vựng có kiểm soát, đối tượng của dự án là các loại thực vật có hoa để lấy ra những chú giải về bộ gen và kiểu hình của Gen để tập hợp thành bộ dữ liệu. Bộ Ontology PO [34] mô tả về đặc điểm hình thái và giải phẫu, xác định các giai đoạn khác nhau của chu kỳ cuộc sống của cây (giai đoạn sinh trưởng) và các cấu trúc cây (giai đoạn phát triển) như là: giai đoạn sinh dưỡng, giai đoạn sinh sản, nẩy mầm,…Bộ từ vựng được sử dụng để mô tả khái quát mô hình giải phẫu thực vật, hình thái học, giai đoạn lớn lên và phát triển cũng như các chú thích thêm các bộ dự liệu về nguồn gen và di truyền học [34].

* Zea mays Ontology

Dự án Zea mays Ontology [33]–[35] là một Ontology mẫu trong lĩnh vực khoa học thực vật. Nó giới thiệu về cấu trúc của cây, bao gồm giải phẫu học và hình thái học của ngô. Nó bao gồm cả các thuật ngữ quốc tế về thực vật, các tài liệu tham khảo, từ đồng nghĩa và thông tin hệ thống các loài và các nguồn mở rộng. Sự sắp xếp các thuật ngữ từ vựng được kiểm soát phản ánh sự hiểu biết hiện tại về mối quan hệ sinh học giữa các bộ phận thực vật liên kết ở cơ quan, tế bào, mô và cơ quan. Bộ Ontology Zea mays có thể được truy cập thông qua trang web Ontology Thực vật học (Plant Ontology).

* CWR

Dự án Ontology về các cây trồng hoang dại (The Crop Wild Relatives Ontology - CWR) [36] là dự án liên quan giữa FAO, Trung tâm Đa dạng Sinh học Quốc tế (Bioversity International) và các bên liên quan khác. Điểm khởi đầu của CWR là bộ 11407 thuật ngữ được thu thập từ nguồn Online. Các thuật ngữ có liên quan cao được nhóm lại vào một chủ đề, tương ứng với danh mục AGROVOC (Ontology cấp cao – Top-level) hoặc các chỉ mục của các nguồn khác (từ điển về sinh học, địa lý online).

Nghiên cứu phát triển Ontology cho quản lý kiến thức nghiên cứu trong nông nghiệp: Điển cứu cho gạo Thái Lan (Ontology development for agricultural research knowledge management: a case study for Thai rice) của tác giả Aree Thunkijjanukij [33]. Thông tin nghiên cứu là một trong những yếu tố quan trọng cho sự phát triển trong nghiên cứu cả trong việc xây dựng chính sách và nâng cao năng lực của các nhà nghiên cứu. Vì vậy tất cả các nghiên cứu và kết quả nghiên cứu trong quá khứ được coi là một phần có giá trị cơ sở tri thức cho sự phát triển nghiên cứu. Tuy nhiên, một công cụ tìm kiếm thông thường không thể đáp ứng đầy đủ các nhu cầu của các nhà nghiên cứu khi thực hiện tìm kiếm các khái niệm, kết quả truy vấn thường dẫn đến kết quả mơ hồ khi đi kèm theo là rất nhiều thông tin không liên quan. Và điều này cũng tương tự trong trường hợp của các nhà nghiên cứu về gạo ở Thái Lan.

Đối với các nghiên cứu về Ontology tại Việt Nam chủ yếu tập trung đi sâu vào lĩnh vực Công nghệ thông tin, các ứng dụng đặc trưng như xây dựng công cụ tìm kiếm, quản lý thông tin… Một số đề tài nghiên cứu điển hình trong ứng dụng Ontology có thể kể đến như đề tài “Viet Nam - Knowledge and Information Management (VN-KIM)”[40]. Miền dữ liệu mà VN-KIM hướng đến là các thực thể được đề cập đến trong các thông tin quốc tế hàng ngày. Ontology của VN-KIM hiện tại có khoảng 250 lớp và 100 thuộc tính. Cơ sở tri thức hiện có khoảng 80,000 thực thể về các nhân vật, thành phố, công ty, tổ chức quan trọng và phổ biến trên thế giới.

Chức năng chính của Viet Nam - Knowledge and Information Management là rút trích và chú thích tự động lớp và danh hiệu của các thực thể có tên xuất hiện trong các trang báo điện tử tiếng Việt. VN-KIM bao gồm những khối chính sau:

* Cơ sở tri thức về các nhân vật tổ chức núi non sông ngòi và địa điểm phổ biến ở Việt Nam.
* Khối rút trích thông tin tự động từ các trang báo điện tử tiếng Việt.
* Khối truy hồi thông tin và các trang Web về các thực thể có tên ở Việt Nam.

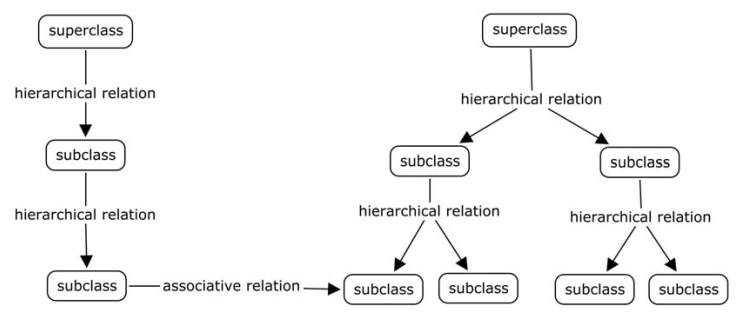
Ngoài ra còn có các đề tài nghiên cứu áp dụng công nghệ Ontology trong khai thác dữ liệu trực tuyến như khai thác văn bản pháp luật, khai thác tài liệu thư viện trực tuyến, xây dựng cổng thông tin.

### Quy trình thiết kế và xây dựng Ontology

* Xác định phạm vi và lĩnh vực của Ontology
* Xem xét kế thừa các ontology
* Thiết kế và lựa chọn các khái niệm trong phạm vi và lĩnh vực
* Lựa chọn ngôn ngữ và công cụ xây dựng
* Xác định và xây dựng mô hình phân cấp các khái niệm
* Xác định các khía cạnh của thuộc tính
* Xác định và xây dựng quan hệ giữa các lớp và thuộc tính

### Cách tiếp cận trong xây dựng cấu trúc các khái niệm và quan hệ trong Ontology

Một Ontology được cấu thành như một mô hình dữ liệu mà ở đó mô tả các bộ khái niệm trong một miền (lĩnh vực) xác định và các quan hệ giữa các khái niệm đó. Chính vì vậy, thành tố chính cấu thành nên một Ontology và khái niệm và quan hệ. Các khái niệm được tổ chức theo mô hình phân cấp, phân lớp, tức là các lớp khái niệm trong lĩnh vực được sắp xếp theo thứ tự cấp bậc. Từ đó hình thành các quan hệ giữa các lớp khái niệm theo 2 dạng chính là: các mối quan hệ phân cấp thứ bậc và mối quan hệ liên kết, tác động. Trong đó mối quan hệ phân cấp thứ bậc là mối quan hệ giữa các lớp chính (superclasses) và các lớp phụ (subclasses), trong khi đó mối quan hệ liên kết, tác động không theo chiều của cấp bậc mà theo chiều ngang giữa các lớp có sự liên kết hoặc tác động lẫn nhau. Mô hình cấu trúc được sử dụng trong thiết kế và xây dựng bộ Ontology nông nghiệp cho An Giang được tham khảo và kế thừa dựa vào mô hình dưới đây:



Hình 1.1: Mô hình phân cấp trong Ontology [31]

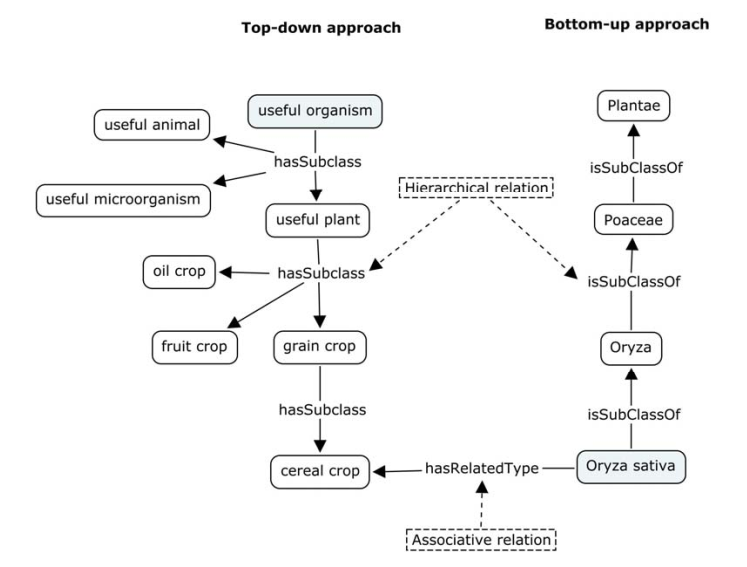
Có một vài cách tiếp cận khả thi trong xây dựng hệ thống phân cấp lớp [31], Có thể kết hợp cả hai cách: Top-down (từ trên xuống) hoặc Bottom-up (từ dưới lên) như trong một số nghiên cứu và quan điểm trước đây:

1. Liệt kê ra tất cả các khái niệm của lĩnh vực quan tâm từ các nghiên cứu trước đây. Sau đó, chia các khái niệm thành 2 nhóm: Nhóm các khái niệm bên trên (Top level) và nhóm các khái niệm bên dưới (Bottom level).
2. Sử dụng cách tiếp cận từ trên xuống để xác định cấu trúc theo cấp bậc. Các khái niệm ở lóp trên được xem như là lớp chính (Superclasses). Ví dụ như trong sản phẩm lúa gạo sẽ sử dụng các lớp chính là: Đặc điểm lúa gạo, các yếu tố sinh học, Yếu tố môi trường, tác nhân quản lý, quy trình quản lý. Và sau đó xác định tiếp các lớp phụ (subclasses).
3. Sử dụng cách tiếp cận từ dưới lên (bottom – up). Xác định các khái niệm ở tầng dưới, hầu hết là các khái niệm về các đối tượng, thực thể, ví dụ như giống gạo cụ thể, loại cỏ cụ thể, côn trùng gây hại cụ thể…Sau khi đã có được các khái niệm ở cấp thấp nhất, tiếp tục xác định các khái niệm cao hơn cho đến khi tiếp cận được các khái niệm ở tầng cao nhất (Top level).
4. Cách tiếp cận bắt đầu từ các khái niệm tầng giữa (middle – level) [31] khi cùng sử dụng cả hai cách trên.

Cách tiếp cận trong xác định các mối quan hệ:

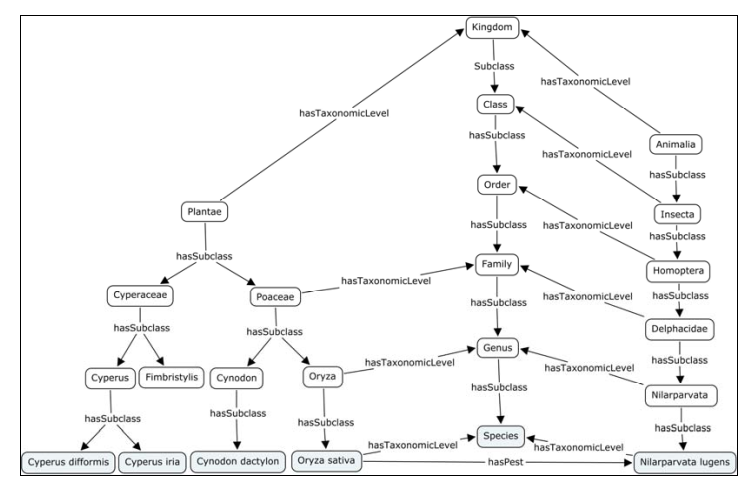
1. Đối với quan hệ thứ bậc. Chỉ có một quan hệ tên là “HasSubclass”. Xác định các quan hệ của tất cả các khái niệm thứ bậc.
2. Đối với quan hệ liên kết, tác động được xác định bởi các động từ liên quan giữa các khái niệm với nhau. Việc gán tên các quan hệ được tuân theo một quy tắc, quy chuẩn. Các quan hệ liên kết, tác động trong đề tài này sẽ được dựa trên các quan hệ của các Ontology đã tồn tại trước đây. Đối với các mối quan hệ kiểu mới sẽ được tham khảo ý kiến của chuyên gia nhằm xác định tên mối quan hệ một cách đúng đắn và khoa học nhất.
3. Tạo các quan hệ liên kết giữa các khái niệm trong sự khác nhau về phân cấp nhưng có liên quan với nhau.

Tạo các quan hệ đảo ngược nếu cần thiết để mô tả những thông tin theo cả hai chiều.



Hình 1.2: Mô hình phân cấp và mối quan hệ trong Ontology cây trồng [23]

Để xây dựng được một hệ thống các khái niệm có các mối quan hệ như đã nêu trên cần xây dựng mô hình “bản nháp” trước khi đi vào xây dựng chính thức nhằm kiểm tra tính đầy đủ và toàn vẹn của hệ thống. Có thể sử dụng các công cụ như là MindManage và công cụ Cmaptools để phác thảo mô hình trong suốt quá trình nghiên cứu và triển khai. Một ví dụ cho kết quả xây dựng bản phác thảo mô hình:



Hình 1.3: Phân cấp các lớp và các quan hệ trong Ontology thuộc lĩnh vực nông nghiệp

Soergel và các đồng nghiệp [32] đã phát triển một bản kê các mối quan hệ cụ thể với ngữ nghĩa rõ ràng cho lĩnh vực Nông nghiệp, ví dụ như các mối quan hệ giữa các khái niệm được chỉ ra sau đây:

X, Y là hai khái niệm

Quan hệ ***“Isa”***

X <includesSpecific> Y / Y <isa> X

X <inheritsTo> Y / Y <inheritsFrom> X

Quan hệ ***tổng thể và thành phần***

X <containsSubstance> Y / Y <substanceContainedIn> X

X <hasIngredient> Y / Y <ingredientOf> X

X <madeFrom> Y / Y <usedToMake> X

X <yieldsPortion> Y / Y <portionOf> X

X <spatiallyIncludes> Y / Y <spatiallyIncludedIn> X

X <hasComponent> Y / Y <componentOf> X

X <includesSubprocess> Y / Y <subprocessOf> X

X <hasMember> Y / Y <memberOf> X

Quan hệ ***tương lai*** (lấy một số từ nghiên cứu của Schmitz-Esser 1999)

X <causes> Y / Y <causedBy> X

X <instrumentFor> Y / Y <performedByInstrument> X

X <processFor> Y / Y <usesProcess> X

X <beneficialFor> Y / Y <benefitsFrom> X

X <treatmentFor> Y / Y <treatedWith> X

X <harmfulFor> Y / Y <harmedBy> X

X <hasPest> Y / Y <afflicts> X

X <growsIn> Y / Y <growthEnvironmentFor> X

X <hasProperty> Y / Y <propertyOf> X

X <hasSymptom> Y / Y <indicates> X

X <similarTo> Y / Y <similarTo> X

X <oppositeTo> Y / Y <oppositeTo> X

X <hasPhase> Y / Y <phaseOf> X

X <growsIn> Y / Y <EnvironmentForGrowing> X

X <ingests> Y / Y <ingestedBy> Y

Còn trong nghiên cứu bộ Ontology trong nông nghiệp AGROVOC của FAO đã tạo ra các quan hệ cơ bản giữa các khái niệm trong hệ thống CS (Concept Server) (FAO, 2008).

### Cách tiếp cận trong việc thiết kế Ontology

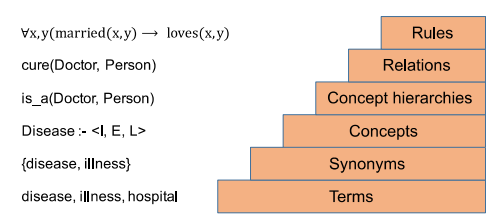
Về bản chất, đây chính là bước xây dựng Ontology chính thức bằng một phần mềm biên tập (trong dự án này là Protégé). Để thực hiện được cần phải chuyển đổi mô hình khái niệm thành mô hình chính thức. Mô hình khái niệm đã được tạo ra từ bước trước sẽ được chuyển thành mô hình chính thức. Các khái niệm được mô tả bởi các thuật ngữ và quan hệ có thứ bậc thông qua mô hình quan hệ cấu trúc. Các bước thực hiện trong giai đoạn này như sau:

* Tất cả các khái niệm và quan hệ được xác định trong mô hình khái niệm hóa ở bước trước sẽ được liệt kê thành bảng dữ liệu.
* Xác định các thuật ngữ trong từng khái niệm. Ở bước này cần phân biệt rằng “Khái niệm” (Concept) là một ý tưởng có thể được mô tả bằng nhiều từ và bằng nhiều ngôn ngữ khác nhau. Còn “Thuật ngữ” (Term) là một từ đơn hoặc một từ phức, ký tự hay công thức chỉ về một khái niệm cụ thể trong một chủ đề nhất định. Còn thuật ngữ ưu tiên là thuật ngữ được lựa chọn để mô tả một khái niệm. Từ đồng nghĩa (thuật ngữ không ưu tiên) là một thuật ngữ trong cùng một khái niệm (giống nhau về ý nghĩa). Một khái niệm có thể có nhiều thuật ngữ, để tránh sự mơ hồ và nhầm lẫn, cần sắp xếp và xây dựng như sau:
  + Liệt kê tất cả các thuật ngữ của mỗi khái niệm từ mô hình khái niệm hóa.
  + Lựa chọn các thuật ngữ ưu tiên để đại diện cho khái niệm. Các thuật ngữ không liên quan sẽ được xác định là từ đồng nghĩa.
  + Từ đồng nghĩa có thể được xác định bằng cấp độ thuật ngữ hoặc cấp độ chuỗi, ví dụ như: từ viết tắt, thuật ngữ viết tắt, thuật ngữ chính tả, số nhiều hoặc số ít, tên phổ thông, tên địa phương, tên khoa học, biểu tượng hóa học, công thức hóa học, tên thương mại…
* Xác định các thuật ngữ về mối quan hệ. Tất cả tên của các quan hệ nên được viết bắt đầu bằng chữ hoa và viết hoa chữ đầu của các từ khác và ở giữa chúng không có khoảng trống [23]. Có ba loại thuật ngữ về mối quan hệ đó là:
  + Mối quan hệ từ khái niệm đến thuật ngữ có tên là “hasLexicallization”. Đây là một mối quan hệ giữa khái niệm và thuật ngữ ưu tiên. Ví dụ như khái niệm về *“Lúa”* hasLexicallization (có chính xác về mặt từ vựng) thuật ngữ *“Oryza sativa”* (Đây là từ vựng về lúa trong tiếng anh – tên khoa học).
  + Quan hệ từ thuật ngữ đến thuật ngữ. Tất cả các quan hệ này được sử dụng cho các thuật ngữ ưu tiên và từ đồng nghĩa của chúng như là: hasAcronym, hasAbbreviation, hasSpellingVariant, hasPural hoặc hasSingular, hasCommonName, hasLocalName, hasScientificName, hasTradeName, hasChemicalSymbol, hasChemicalFormula, hasTranslation, hasSynonym. Một ví dụ thuật ngữ lúa [rice] hasPlural [rices]; Thuật ngữ [Sulphur] hasSpellingVariant (biến thể chính tả) thuật ngữ là [sulfer] hay thuật ngữ [Oryza sativa] hasCommonName thuật ngữ [rice].
  + Quan hệ từ khái niệm đến khái niệm. Quan hệ này kết nối giữa các khái niệm (được mô tả bởi các thuật ngữ ưu tiên) ở các vị trí thứ bậc khác nhau, như là hasPest, hasDisease, hasPathogen, hasRelatedType.
* Xác định thuộc tính các khái niệm, ví dụ như: trạng thái, định nghĩa, ghi chú về phạm vi. Một thuộc tính của loại dữ liệu có thể được sử dụng để gắn với một khái niệm hoặc một cá thể đến một giá trị cụ thể. Dữ liệu được kết nối với khái niệm thông qua quan hệ này có thể là một loại dữ liệu cụ thể.

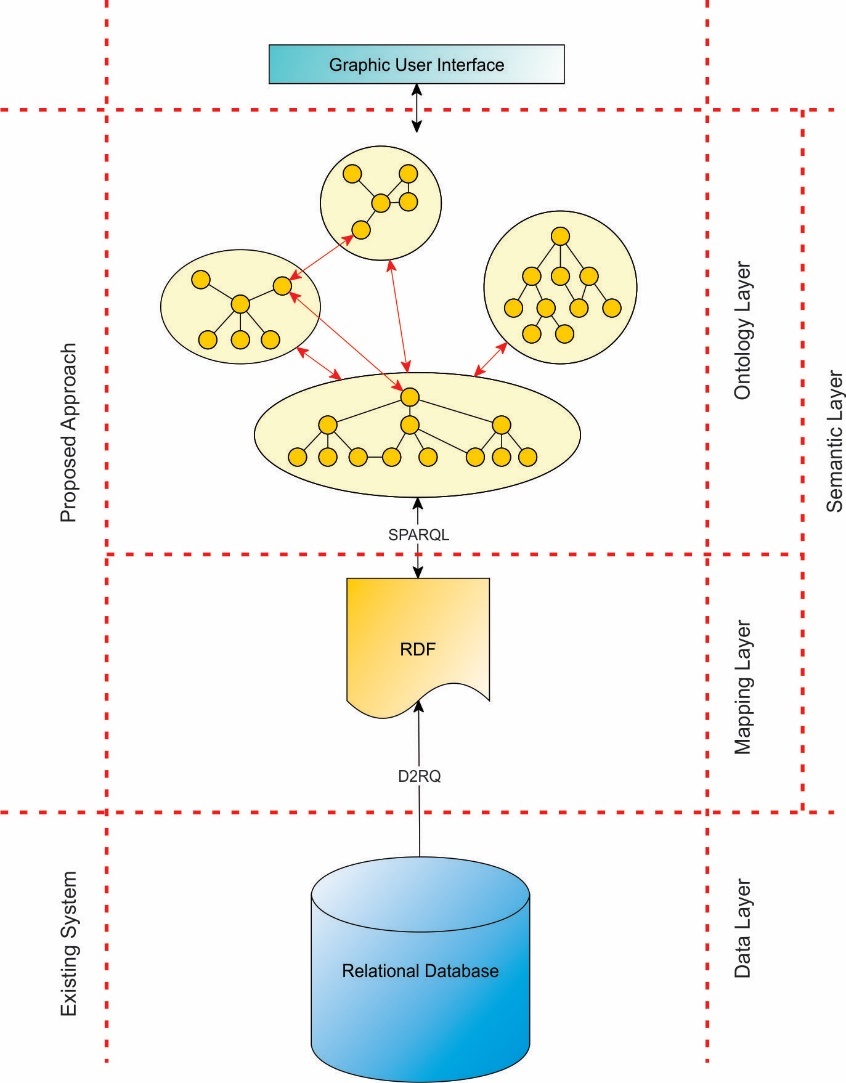
Điền các thông tin vào bảng dữ liệu để hình thành các khái niệm. Các bảng dữ liệu được thiết kế để tạo thuận lợi cho các nhà xây dựng Ontology, đặc biệt là các chuyên gia trong việc chuyển đổi mô hình khái niệm sang mô hình chính thức. Có ba loại bảng dữ liệu được thiết kế đó là: bảng dữ liệu cho các khái niệm và từ vựng hóa; bảng dữ liệu cho hình thành các khái niệm và các mối quan hệ liên quan; bảng dữ liệu cho các khái niệm và mối quan hệ cấp bậc.

### Cách tiếp cận trong xây dựng Ontology

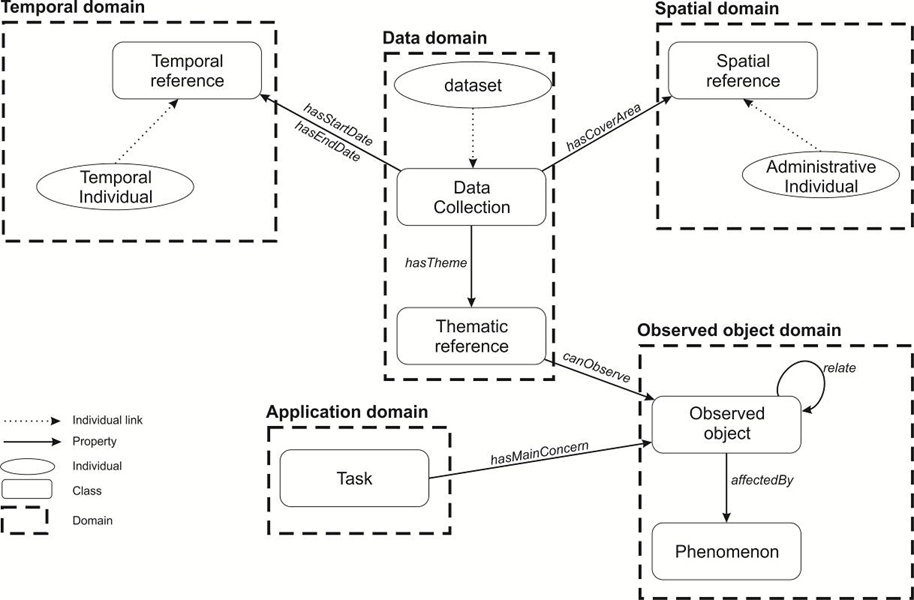
Có thể phân thành hai phương pháp xây dựng ontology là xây dựng thủ công và “học” (tự động, bán tự động). Xây dựng ontology thủ công thường mất nhiều thời gian và đòi hỏi phải có các chuyên gia trong lĩnh vực đó. Do đó, phương pháp này chỉ phù hợp để xây dựng các ontology cho các miền tri thức nhỏ, giới hạn và ít thay đổi. Đối với các ontology lớn hay thường thay đổi thì cần có phương pháp khả thi hơn, tiết kiệm thời gian, nhân lực và đáp ứng với các thay đổi tốt hơn, đó chính là phương pháp xây dựng một cách tự động hay bán tự động hay còn gọi là “học” ontology (ontology learning). Học ontologgy là quá trình xác định các thuật ngữ (term), các khái niệm (concept), các quan hệ phân loại hay cấp bậc (taxonomy relation) và quan hệ không cấp bậc (non-taxonomy relation), các tiên đề (axiom). Đây chính là các thành phần của một ontology. Tuy nhiên, tùy vào từng cấp độ chi tiết, vào qui mô của ontology mà quá trình học có thể chỉ xác định một số trong các thành phần trên. Vị trí của các thành phần trên trong một ontology được gọi là “Ontology Learning Layer Cake” và được mô tả như hình sau đây[33]:



Hình 1.4: Ontology Learning Layer Cake [34]



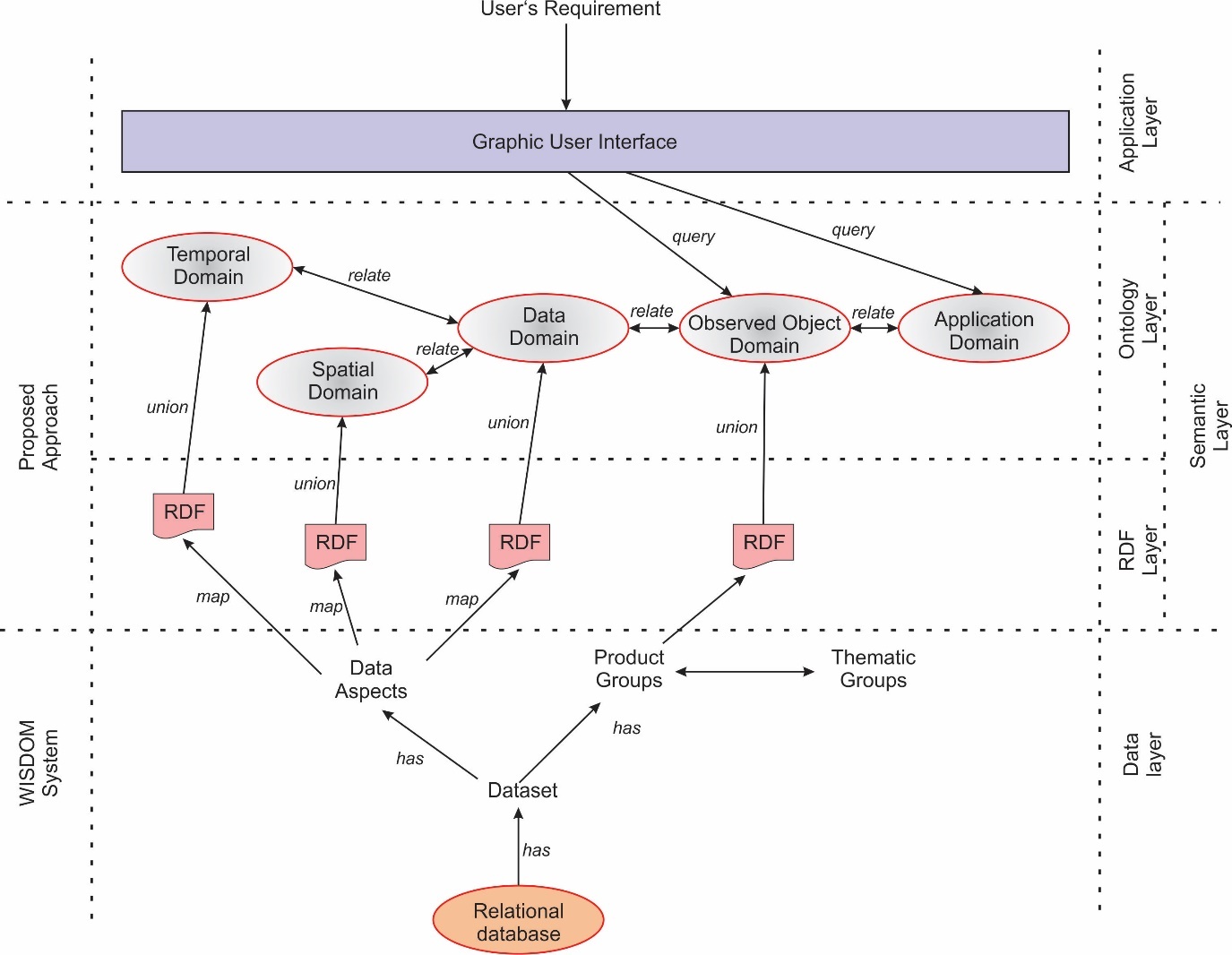
Hình 1.5: Cấu trúc phân tầng trong tiếp cận xây dựng Ontology [50]



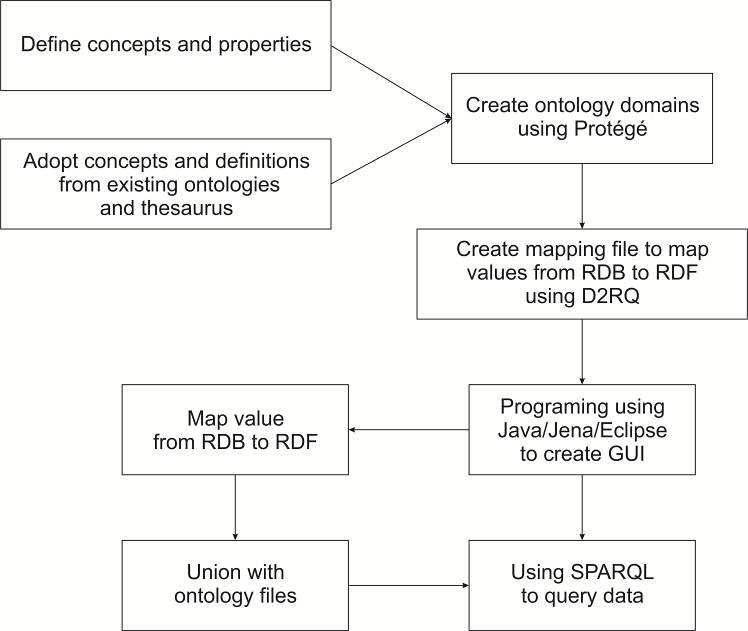
Hình 1.6: Chi tiết cấu trúc Ontology [50]

Trong đó:

* **Data Domain** chứa các lớp mô tả thuộc tính của bộ dữ liệu, ví dụ như về loại format: độ phân giải không gian (kích thước pixcel), biểu diễn không gian (đường, điểm hay vùng) và quan hệ không gian. Các dữ liệu liên quan đến không gian và thời gian được lưu trữ trong Cơ sở dữ liệu quan hệ (RD) với thư mục datacollection.
* **Observed Object Domain** bao gồm các lớp mô tả các đối tượng vật lý và phi vật lý liên quan đến nông nghiệp, ví dụ các yếu tố nhân tạo, tự nhiên, xã hội cũng được gọi là các đối tượng quan sát. Các quan hệ trong domain này được mô tả độc lập trong mục Tasks. Vì vậy, việc xác định các khái niệm trong Domain này rất dễ để gộp với bất cứ nhiệm vụ nào trong việc ứng dụng.
* **Application Domain** mô tả các tác vụ của người dùng, được chia thành các loại: nhiệm vụ hồi đáp, nhiệm vụ giám sát…Các nhiệm vụ của người sử dụng được mô tả liên quan đến các đối tượng quan sát được mối quan tâm chính của các nhiệm vụ này. Tác vụ này hoạt động như những khó khăn để hạn chế kết quả trả về một hiện tượng nhất định.
* **Temporal Domain** có lớp chính là Temporal Entity chứa các giá trị về thời gian của bộ dữ liệu. Các phần trong này là các ngày bắt đầu và ngày kết thúc, hiển thị khoảng thời gian mà bộ dữ liệu bắt đầu có hiệu lực và kết thúc.
* **Spatial Domain** lưu giữ các dữ liệu không gian như hành chính, lớp nền bao phủ lên toàn vùng nghiên cứu. Trong lớp ngữ nghĩa, các miền quan hệ với nhau bằng cách kết nối các trường thuộc tính. Sơ đồ ở trên chính là mô tả các miền chính của hệ thống và các quan hệ giữa các miền với nhau. Miền dữ liệu liên kết với miền đối tượng quan sát bằng thuộc tính “canObserve” bằng cách liên kết các lớp chủ đề với các đối tượng đã quan sát được. Tác vụ người dùng trong miền ứng dụng liên kết các đối tượng quan sát bằng thuộc tính "isMainConcern". Các vùng thời gian và không gian chứa các giá trị thể hiện các thuộc tính của bộ dữ liệu kết nối hai miền vào miền dữ liệu với "hasStartDate", "hadEndDate" và "hasCoverArea".



Hình 1.7: Sơ đồ cách tiếp cận toàn hệ thống [35]

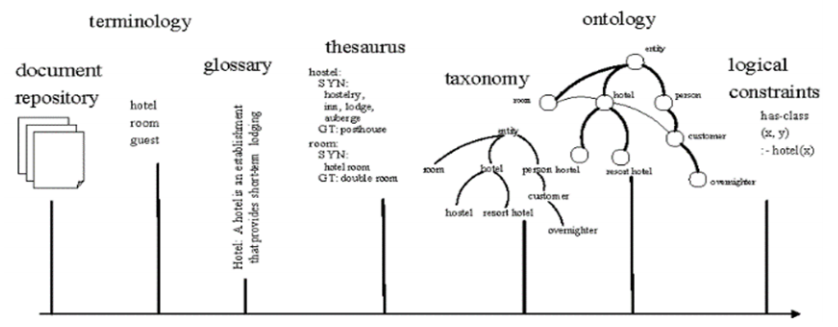


Hình 1.8: Sơ đồ phân tích của nhiệm vụ xây dựng Ontology [35]

# Thiết kế Ontology cho Hệ thống thông tin Nông nghiệp An Giang

### Xác định phạm vi và lĩnh vực của Ontology

Lĩnh vực cần xây dựng Ontology ở đây là thông tin nông nghiệp ở tỉnh An Giang với 5 mặt hàng chủ lực là: Lúa, Cá tra, Nấm, Rau màu, Bò thịt.



Hình 2.1: Các cấp độ biểu diễn tri thức [33]

|  |  |
| --- | --- |
| Code | Đối tượng |
| B | **Bò thịt** |
| C | **Cá tra** |
| L | **Lúa** |
| N | **Nấm** |
| R | **Rau màu** |

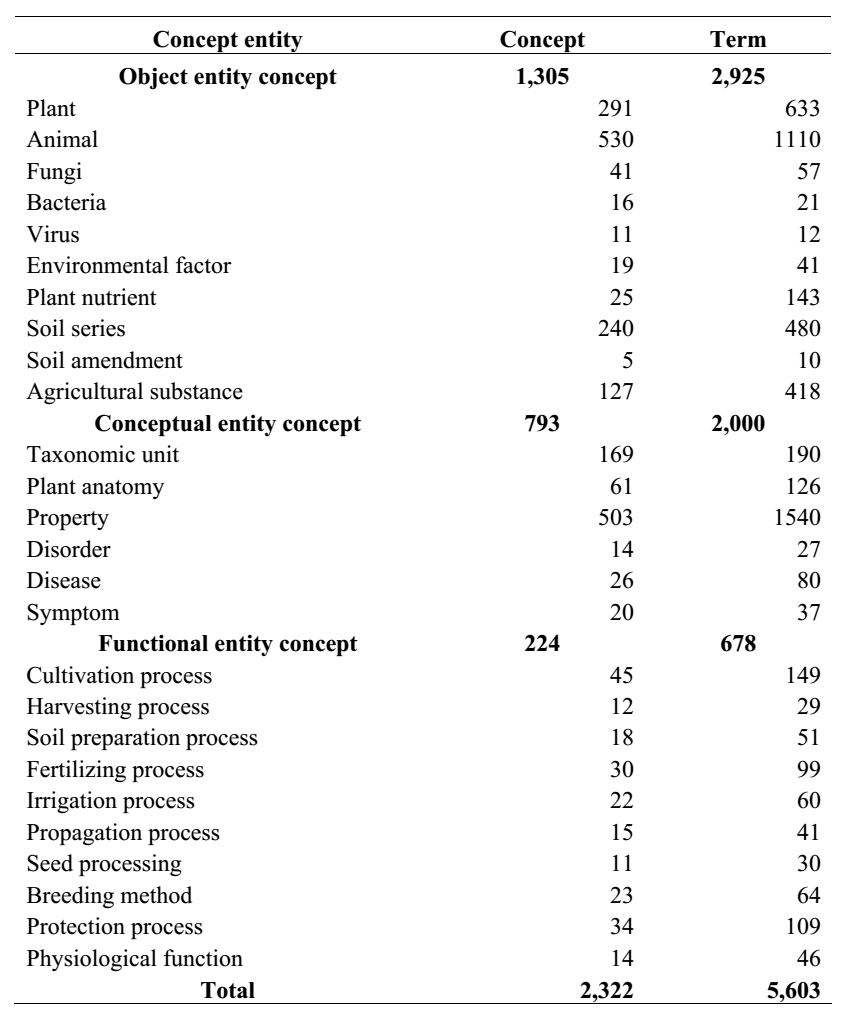
### Xem xét kế thừa các Ontology

Việc tái sử dụng các ontology có sẵn là rất quan trọng vì giảm bớt nhiều thời gian và công sức cho quá trình xây dựng. Trong lĩnh vực nông nghiệp, đã có nhiều nghiên cứu xây dựng Ontology trên thế giới, đặc biệt là tổ chức FAO, chính vì vậy có thể tận dụng một số ít tài liệu cũng như kiến thức để phục vụ cho việc xây dựng Ontology nông nghiệp cho tỉnh An Giang.

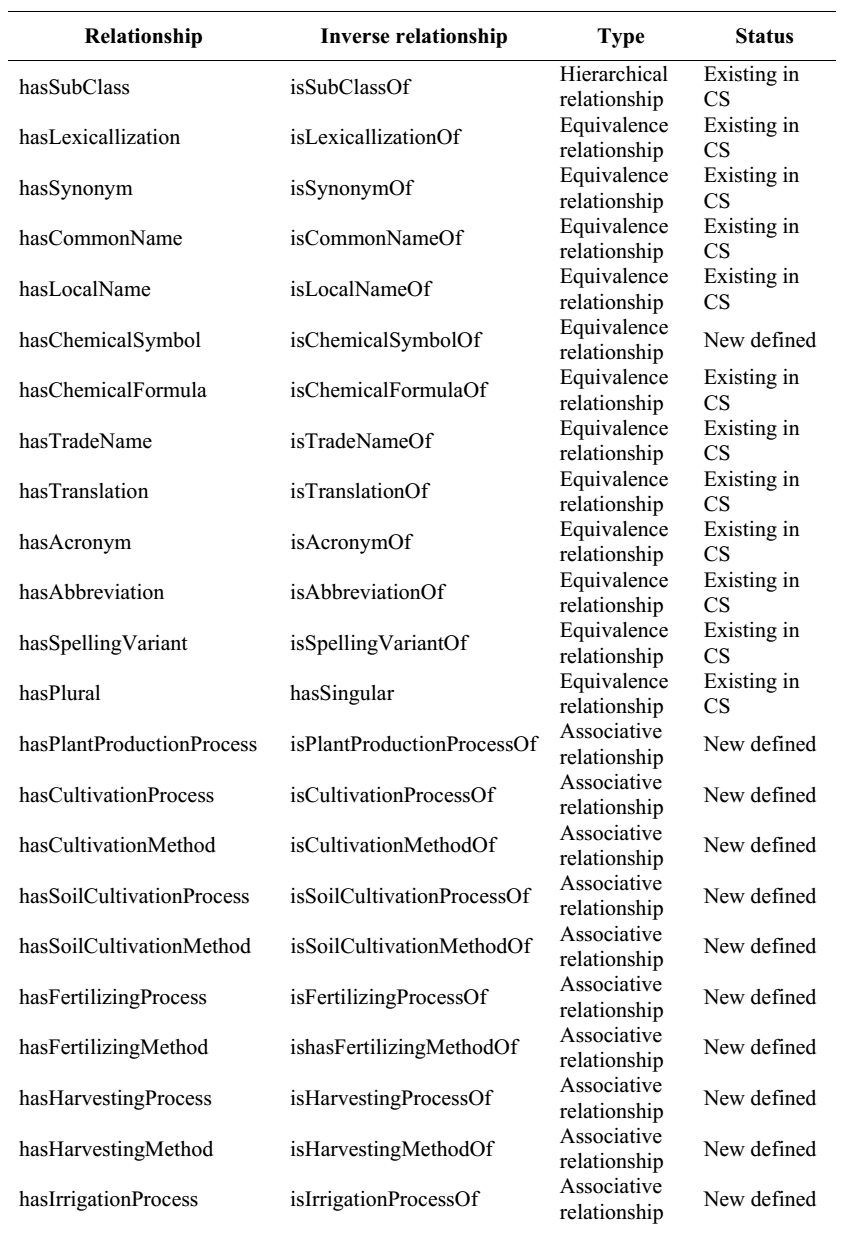
Đối với đối tượng lúa gạo, nhóm nghiên cứu sẽ dựa vào kết quả nghiên cứu của đề tài xây dựng Ontology về Gạo Thái Lan đã đưa ra được bộ các khái niệm chính trong lĩnh vực nghiên cứu, cụ thể như sau:

|  |  |
| --- | --- |
| **Khái niệm của đối tượng** | **Khái niệm của quá trình** |
| Thực vật (lúa, cỏ dại và sâu bệnh gây hại)  Động vật (sâu bệnh và thiên địch)  Nấm (sâu bệnh và thiên địch)  Vi khuẩn (sâu bệnh và thiên địch)  Virus (sâu bệnh)  Chất hóa học và các nguyên tố (phân bón, thuốc trừ sâu, chất kích thích tăng trưởng thực vật, nguyên tố dinh dưỡng)  Đất  Giải phẫu học thực vật và hình thái cây trồng  Dịch bệnh  Rối loạn sinh lý  Yếu tố môi trường (nước, ánh sáng, nhiệt độ,…) | Quá trình trồng trọt  Quá trình thu hoạch  Quá trình bảo vệ và kiểm soát  Quá trình bón phân  Quá trình tưới tiêu  Quá trình chuẩn bị đất  Giống và nhân giống cây trồng  Chức năng sinh lý |

Như đã đề cập trong những phần trước, cấu trúc của một Ontology bao gồm khái niệm, thuật ngữ và quan hệ. Các khái niệm trong toàn bộ Ontology của dự án nói chung và đối với lúa gạo nói riêng được xác định gồm 3 nhóm chính như sau: Các khái niệm thực thể đối tượng (Object entity concepts), các khái niệm thực thể khái niệm (Conceptual entity concepts), các khái niệm thực thể chức năng (Funtional entity concepts). Theo như kết quả của nghiên cứu xây dựng bộ Ontology lúa gạo Thái Lan thì toàn bộ Ontology có 2,322 khái niệm, 5,603 thuật ngữ và 70 quan hệ. Trong đó có khoảng một nửa là khái niệm thực thể đối tượng, phần còn lại là khái niệm thực thể khái niệm và khái niệm thực thể chức năng.



Hình 2.2: Các quan hệ được sử dụng trong toàn bộ Ontology lúa gạo Thái Lan [23]







Hình 2.3: Các yếu tố chính liên quan đến Ontology lúa gạo [23]

Bảng 1: Các yếu tố liên quan đến Lúa gạo

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Đặc điểm của gạo** | **Các yếu tố sinh học** | **Yếu tố môi trường** | **Tác nhân quản lý** | **Quy trình quản lý** |
| - Hình thái của lúa gạo (Rice morphology)  - Sinh trưởng của cây lúa (Rice Physiology) | - Cỏ dại (weed)  - Côn trùng và động vật gây hại (Pest insect and animal)  - Dịch bệnh (Disease)  - Thiên địch (Natural Emeny) | - Đất  - Chất dinh dưỡng  - Ánh sáng  - Độ ẩm không khí  - Độ ẩm của đất  - Nước  - Nhiệt độ  - Không khí | - Phân bón  - Thuốc trừ sâu  - Yếu tố khác | - Trồng lúa  - Giống lúa gạo  - Kiểm soát và bảo vệ cây lúa  - Phân bón  - Thủy lợi  - Hệ thống mùa vụ  - Chuẩn bị đất |

### Thiết kế và lựa chọn các khái niệm trong lĩnh vực nông nghiệp tỉnh An Giang

Căn cứ vào các báo cáo đánh giá “Tình hình hoạt động ngành nông nghiệp và PTNT từ năm 2011 – 2016 của tỉnh An Giang” có thể nhận thấy rằng nông nghiệp An Giang tập trung vào 5 mặt hang chủ lực là Lúa, rau màu, cá tra, thịt bò và nấm (nấm ăn, nấm dược liệu). Như vậy, bộ từ vựng cơ sở cho Ontology nông nghiệp An Giang sẽ tập trung vào 6 yếu tố đó là Lúa, rau màu, cá tra, thịt bò, nấm (nấm ăn và dược liệu) và dịch bệnh trong nông nghiệp.

Trong bước 1 ta đã xác định được miền và phạm vi của ontology, trong bước này là liệt kê các thuật ngữ, khái niệm quan trọng nằm trong miền và phạm vi đó.

|  |  |
| --- | --- |
| **Code** | **Khái niệm** |
| B | Bò thịt |
| C | Cá tra |
| L | Lúa |
| N | Nấm |
| R | Rau màu |

Bảng 2: Các thuật ngữ quan trọng trong khái niệm cây lúa

|  |  |  |
| --- | --- | --- |
| LÚA |  |  |
| Code | **Sub-code** | **Đối tượng** |
| LG |  | **Giống lúa** |
|  | LG01 | *Phân loại theo đặc tính thực vật học* |
|  | LG02 | *Phân loại theo sinh thái địa lý* |
|  | LG03 | *Phân loại theo đặc tính sinh lý* |
|  | LG04 | *Phân loại theo điều kiện môi trường canh tác* |
| LĐ |  | **Đặc điểm sinh thái cây lúa** |
|  | LĐ01 | *Điều kiện khí hậu* |
|  | LĐ02 | *Điều kiện đất đai* |
| LH |  | **Hình thể học cây lúa** |
|  | LH01 | *Các giai đoạn sinh trưởng* |
|  | LH02 | *Hạt lúa* |
|  | LH03 | *Sự nảy mầm* |
|  | LH04 | *Mầm lúa và mạ non* |
|  | LH05 | *Rễ lúa* |
|  | LH06 | *Thân lúa* |
|  | LH07 | *Lá lúa* |
|  | LH08 | *Bông lúa* |
|  | LH09 | *Hoa lúa* |
| LD |  | **Dinh dưỡng cho cây lúa** |
|  | LD01 | *Đất ngập nước và dinh dưỡng khoáng* |
|  | LD02 | *Chất đạm* |
|  | LD03 | *Chất lân* |
|  | LD04 | *Chất Kali* |
|  | LD05 | *Chất Silic* |
|  | LD06 | *Chất sắt* |
| LC |  | **Côn trùng** |
|  | LC01 | *Rầy nâu* |
|  | LC02 | *Rầy lưng trắng* |
|  | LC03 | *Rầy xanh* |
|  | LC04 | *Rầy bông* |
|  | LC05 | *Bọ xít hôi* |
|  | LC06 | *Bọ gai* |
|  | LC07 | *Bọ xít đen* |
|  | LC08 | *Bù lạch* |
|  | LC09 | *Dế nhũi* |
|  | LC10 | *Sâu đục thân* |
|  | LC11 | *Sâu cuốn lá – sếp lá* |
|  | LC12 | *Sâu sừng xanh* |
|  | LC13 | *Sâu phao* |
|  | LC14 | *Sâu keo* |
|  | LC15 | *Dòi đục lá* |
|  | LC16 | *Muỗi gây lá hành* |
| LB |  | **Bệnh dịch** |
| LBN |  | ***Bệnh do nấm*** |
|  | LBN01 | *Bệnh cháy lá* |
|  | LBN02 | *Bệnh đốm nâu* |
|  | LBN03 | *Bệnh gạch nâu* |
|  | LBN04 | *Bệnh than vàng* |
|  | LBN05 | *Bệnh đốm vằn* |
|  | LBN06 | *Bệnh thối bẹ* |
|  | LBN07 | *Bệnh thối thân* |
|  | LBN08 | *Bệnh lúa von* |
| LBV |  | ***Bệnh do vi khuẩn*** |
|  | LBV01 | *Bệnh cháy bìa lá* |
|  | LBV02 | *Bệnh sọc trong* |
| LBS |  | ***Bệnh do siêu vi khuẩn*** |
|  | LBS01 | *Bệnh do rầy nâu truyền* |
|  | LBS02 | *Bệnh do rầy xanh truyền* |
|  | LBS03 | *Bệnh do rầy bông truyền* |
|  | LBS04 | *Bệnh do tuyến trung* |
|  | LBS05 | *Bệnh tiêm đọt sần* |
|  | LBS06 | *Bệnh bứu rễ* |
| LBK |  | ***Những bệnh và vật gây hại khác*** |
|  | LBK01 | *Bệnh vàng chín lá sớm* |
|  | LBK02 | *Bệnh lem lép hạt* |
|  | LBK03 | *Nhện gié* |
|  | LBK04 | *Ốc bưu vàng* |
|  | LBK05 | *Chim và chuột* |
| LT |  | **Triệu chứng dinh dưỡng bất thường** |
|  | LT01 | *Độc do mặn* |
|  | LT02 | *Độc do phèn* |
|  | LT03 | *Độc do chất hữu cơ* |
| LN |  | **Nhu cầu nước của cây lúa** |
| LS |  | **Sản phẩm từ lúa gạo** |
| LM |  | **Thu hoạch và bảo quản sau thu hoạch** |
|  | LM01 | *Thu hoạch* |
|  | LM02 | *Bảo quản sau thu hoạch* |

|  |  |  |
| --- | --- | --- |
| BÒ THỊT |  |  |
| Code | **Sub-code** | **Đối tượng** |
| BG |  | **Bò giống** |
|  | BG01 | *Bò vàng* |
|  | BG02 | *Bò lai sind* |
| BB |  | **Các bệnh thường gặp trên bò** |
| BBS |  | ***Bệnh sản khoa*** |
|  | BBS01 | *Viêm tử cung* |
|  | BBS02 | *Viêm âm đạo* |
| BBR |  | ***Bệnh rối loạn tiêu hóa và ngộ độc*** |
|  | BBR01 | *Bệnh chướng bụng, đầy hơi* |
|  | BBR02 | *Ngộ độc ở bò* |
| BBT |  | ***Bệnh truyền nhiễm*** |
|  | BBT01 | *Bệnh lở mồm long móng* |
|  | BBT02 | *Bệnh dịch tả* |
|  | BBT03 | *Bệnh tụ huyết trùng* |
| BBK |  | ***Bệnh ký sinh trùng*** |
|  | BBK01 | *Bệnh tiên mao trùng* |
|  | BBK02 | *Bệnh sán lá gan* |
|  | BBK03 | *Bệnh giun đũa ở bê* |
|  | BBK04 | *Bệnh giun phổi ở bê* |
|  | BBK05 | *Bệnh cầu trùng* |
| BT |  | **Thức ăn cho bò** |
| BTH |  | ***Thức ăn thô*** |
|  | BTH01 | *Thức ăn xanh* |
|  | BTH02 | *Thức ăn ủ ướp* |
|  | BTH03 | *Cỏ khô và rơm lúa* |
|  | BTH04 | *Thức ăn củ quả* |
|  | BTH05 | *Phế phụ phẩm công nghiệp chế biến* |
| BBT |  | ***Thức ăn tinh*** |
|  | BBT01 | *Cám gạo* |
|  | BBT02 | *Bột ngô* |
|  | BBT03 | *Bột sắn* |
|  | BBT04 | *Khô dầu* |
| BBS |  | ***Thức ăn bổ sung*** |
|  | BBS01 | *Urê* |
|  | BBS02 | *Thức ăn bổ sung khoáng* |
| BBC |  | ***Cỏ trồng cho bò*** |

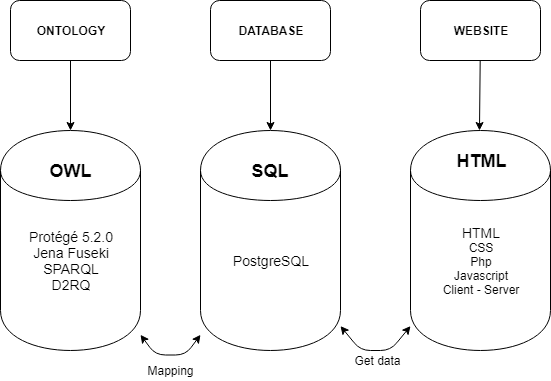
|  |  |  |
| --- | --- | --- |
| RAU MÀU |  |  |
| Code | **Sub-code** | **Đối tượng** |
| RG |  | **Giống rau** |
| RGQ |  | ***Rau ăn quả*** |
|  | RGQ01 | *Cà chua* |
|  | RGQ02 | Dưa leo |
| RGL |  | ***Rau ăn lá*** |
|  | RGL01 | *Cải bẹ xanh* |
| RB |  | **Một số bệnh trên rau** |
|  | RB01 | *Bệnh sương mai* |
|  | RB02 | *Bệnh héo rũ do nấm* |
|  | RB03 | *Bệnh héo xanh* |
|  | RB04 | *Bệnh đốm nâu* |
|  | RB05 | *Bệnh lở cổ rễ* |
|  | RB06 | *Bệnh bã trầu* |
| RC |  | **Một số loại côn trùng hại rau màu** |
|  | RC01 | *Sâu xanh đục quả* |
|  | RC02 | *Sâu khoang đục quả* |
|  | RC03 | *Bọ phấn* |
|  | RC04 | *Sâu tơ* |
|  | RC05 | *Bọ nhảy sọc cong* |
|  | RC06 | *Bọ rùa vàng* |
|  | RC07 | *Rầy lửa* |
| RK |  | **Kỹ thuật trồng** |
| RKC |  | ***Kỹ thuật trồng cà chua*** |
|  | RKC01 | *Luân canh* |
|  | RKC02 | *Thời vụ* |
|  | RKC03 | *Ươm cây non* |
|  | RKC04 | *Mật độ trồng* |
|  | RKC05 | *Bón phân* |
|  | RKC06 | *Thu hoạch* |
| RKD |  | ***Kỹ thuật trồng dưa leo*** |
|  | RKD01 | *Làm đất, gieo hạt* |
|  | RKD02 | *Vun xới, làm giàn, sửa dây* |
|  | RKD03 | *Bón phân* |
|  | RKD04 | *Tưới nước* |
|  | RKD05 | *Thu hoạch* |

|  |  |  |
| --- | --- | --- |
| CÁ TRA |  |  |
| Code | **Sub-code** | **Đối tượng** |
| CG |  | **Giống cá tra** |
| CB |  | **Bệnh trên cá tra** |
| CBV |  | ***Bệnh do vi khuẩn*** |
|  | CBV01 | *Bệnh xuất huyết* |
|  | CBV02 | *Bệnh gan thân mủ* |
|  | CBV03 | *Bệnh trắng đuôi* |
| CBN |  | ***Bệnh do vi nấm*** |
|  | CBN01 | *Bệnh trương bụng* |
| CBK |  | ***Bệnh do ký sinh trùng*** |
|  | CBK01 | *Bệnh do lớp sán lá đơn chủ Monogenea* |
|  | CBK02 | *Bệnh do trùng bánh xe Trichodina* |
|  | CBK03 | *Bệnh do trùng loa kèn Apinosoma* |
|  | CBK04 | *Bệnh đốm trắng* |
|  | CBK05 | *Bệnh do vi bào tử và bào tử trùng* |
|  | CBK06 | *Bênh do trùng lông Balantidium và Ichthyonyctus* |
|  |  | *Bệnh do giun sán kí sinh* |
| CBC |  | ***Bệnh chưa rõ nguyên nhân*** |
|  | CBC01 | *Hội chứng đỏ cơ* |
|  | CBC02 | *Hội chứng vàng da* |
|  | CBC03 | *Bệnh trắng gan trắng mang* |
| CBP |  | ***Phòng bệnh ở cá*** |

|  |  |  |
| --- | --- | --- |
| NẤM |  |  |
| Code | **Sub-code** | **Đối tượng** |
| NP |  | **Phân loại nấm** |
|  | NA | **Nấm ăn** |
|  | ND | **Nấm dược liệu** |
| NAG |  | **Các giống nấm ăn** |
|  | NAGR | *Nấm rơm* |
|  | NAGS | *Nấm sò* |
| NDG |  | **Các giống nấm dược liệu** |
|  | NDGL | *Nấm linh chi* |
| NAGR |  | ***Nấm rơm*** |
|  | NAGR01 | *Phân loại thực vật học* |
|  | NAGR02 | *Chu kỳ sống* |
|  | NAGR03 | *Sinh trưởng* |
|  | NAGR04 | *Kỹ thuật trồng* |
| NAGS |  | ***Nấm sò*** |
|  | NAGS01 | *Phân loại thực vật học* |
|  | NAGS02 | *Đặc tính sinh học* |
|  | NAGS03 | *Quy trình trồng* |
| NDGL |  | ***Nấm linh chi*** |
|  | NDGL01 | *Phân loại thực vật học* |
|  | NDGL02 | *Đặc tính sinh học* |
|  | NDGL03 | *Quy trình trồng* |

# Xây dựng Ontology phục vụ tìm kiếm trong Hệ thống thông tin Nông nghiệp An Giang

### Công cụ hỗ trợ xây dựng Ontology

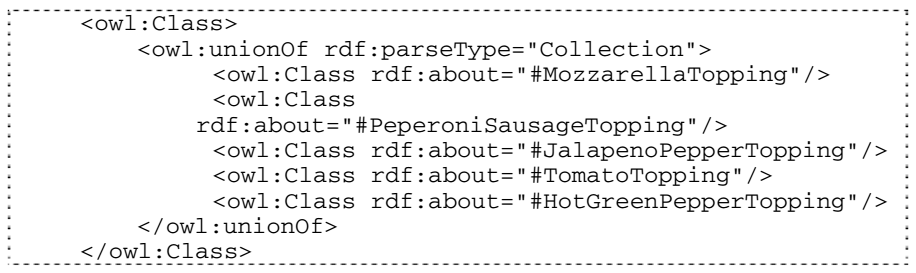


Hình 23. Các tầng của hệ thống và công nghệ sử dụng

* ***Phần mềm Protégé:*** Protégé là bộ phần mềm mã nguồn mở Java nổi tiếng. Protégé được nghiên cứu và phát triển từ năm 1998 bởi nhóm nghiên cứu của Mark Musen, ĐH. Stanford nhằm quản lý các thông tin trong lĩnh vực sinh y học. Phần mềm Protégé: là phần mềm miễn phí Java nổi tiếng. Hiện tại Protégé đã có phiên bản 5.2 được phát hành vào 17/5/2015.
* ***Jena Fuseki:*** Jena là một Java framework dùng để phát triển một Web ngữ nghĩa hoặc công cụ tìm kiếm dựa trên Ontology. Jena có vai trò giao tiếp với định dạng Ontology là OWL/RDF thông qua các câu lệnh từ SPARQL, hỗ trợ cho việc thực hiện các câu lệnh truy vấn SPARQL.
* ***SPARQL:*** được phát triển bởi nhóm RDF Data Access Working Group - một phần trong hoạt động của Semantic Web và đã được W3C - tổ chức chịu trách nhiệm xây dựng, quản lý đưa ra các chuẩn liên quan đến World Wide Web - khuyến nghị vào năm 2008. SPARQL là một ngôn ngữ để truy cập thông tin từ các đồ thị RDF. SPARQL cung cấp các tính năng sau: trích thông tin trong các dạng của URI, các nút trống hay giá trị nguyên thủy hoặc các kiểu được định nghĩa từ các giá trị nguyên thủy, trích thông tin từ các đồ thị con và xây dựng một đồ thị RDF mới dựa trên thông tin trong đồ thị truy vấn.
* ***D2RQ:*** mapping giữa Ontology và dữ liệu quan hệ trên nền tảng web.

### Ngôn ngữ Ontology OWL

OWL là ngôn ngữ Ontology tiêu chuẩn cho một Web ngữ nghĩa. Nó là ngôn ngữ mở rộng từ nền tảng của RDF và cũng là ngôn ngữ có khả năng tương thích với ngôn ngữ Ontology sớm, bao gồm SHOE, DAML + OIL và cung cấp những khả năng để biểu diễn ngữ nghĩa. OWL bao gồm sự liên kết, phân tách, tồn tại và sự phổ biến về số lượng, chính điều đó có thể được sử dụng để thực hiện các nhận định, kết luận Logic và những hiểu biết về nguồn gốc của từ.



Hình 3.1: Ví dụ một Ontology bằng ngôn ngữ OWL về Pizza - W3C

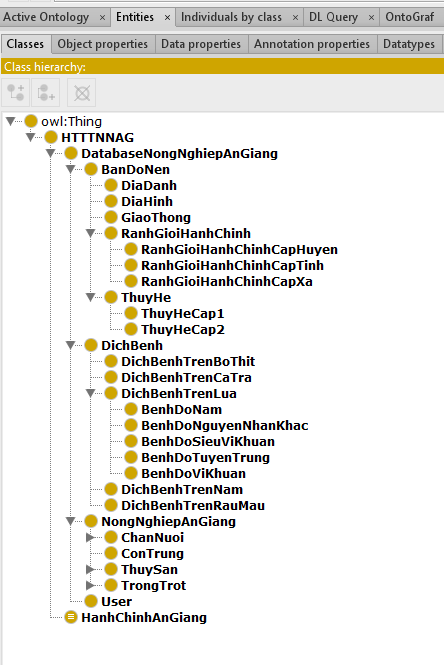
Tuy nhiên, OWL vẫn có một số hạn chế như sau[[1]](#footnote-1):

1. Cấu trúc của OWL phức tạp, chính vì vậy đã hình thành ra 3 ngôn ngữ phụ cho OWL:

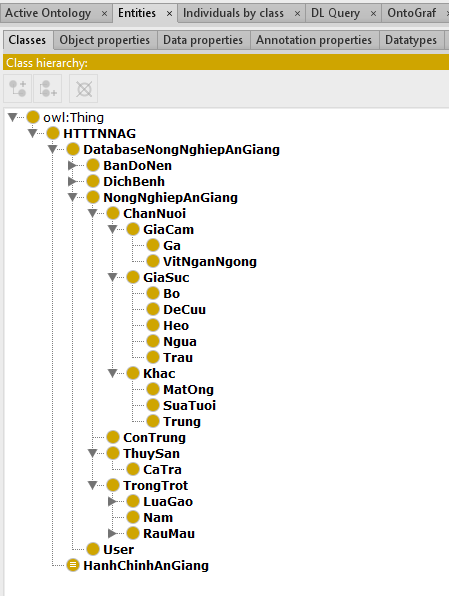
* OWL Lite – hỗ trợ sự phân loại có cấp bậc và các đặc trưng bắt buộc đơn giản như một tập hợp các yếu tố bắt buộc. Lợi thế của loại này đó là dễ để thấu hiểu (cho người dùng) và dễ để thực hiện (cho người xây dựng)[[2]](#footnote-2), bên cạnh đó cũng có hạn chế về mặt biểu hiện.
* OWL DL – bao gồm toàn bộ cấu trúc ngôn ngữ OWL với một số giới hạn như là sự chia thành các loại, nó phù hợp với Logic mô tả. Ưu thế của loại này là cho phép hỗ trợ cho lập luận có hiệu quả. Tuy nhiên sự tương thích với RDF cũng bị mất đi.
* OWL Full – hỗ trợ biểu cảm tối đa và tự do cú pháp của RDF mà không đảm bảo tính toán. OWL Full tương thích hoàn toàn với RDF, cả về cú pháp và ngữ nghĩa. Tuy nhiên, nó không thể giải quyết được và không thích hợp cho lý luận đầy đủ trong một số trường hợp cụ thể.

1. Cần phải có sự hiểu biết về OWL mới có thể xây dựng được kiến trúc Ontology hiệu quả.

### Xác định và xây dựng mô hình phân cấp các khái niệm trong Ontology nông nghiệp



Hình 3.2: Ảnh chụp màn hình giao diện làm việc Protege: phân lớp trong lĩnh vực nông nghiệp An Giang

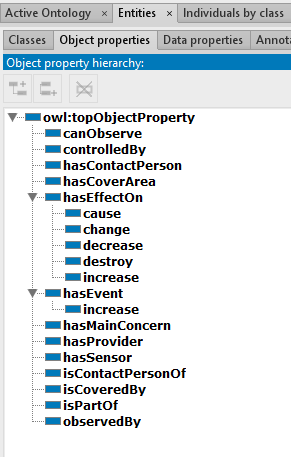


Hình 3.3: Ảnh chụp màn hình giao diện làm việc Protege: phân lớp trong lĩnh vực nông nghiệp An Giang (tiếp)



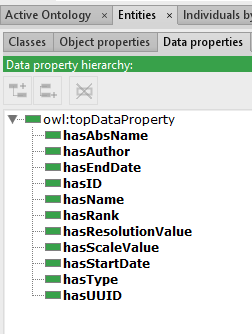
Hình 3.4: Ảnh chụp màn hình giao diện làm việc Protege: phân lớp trong lĩnh vực nông nghiệp An Giang (tiếp)

### Xác định và xây dựng thuộc tính cho các lớp trong Ontology nông nghiệp

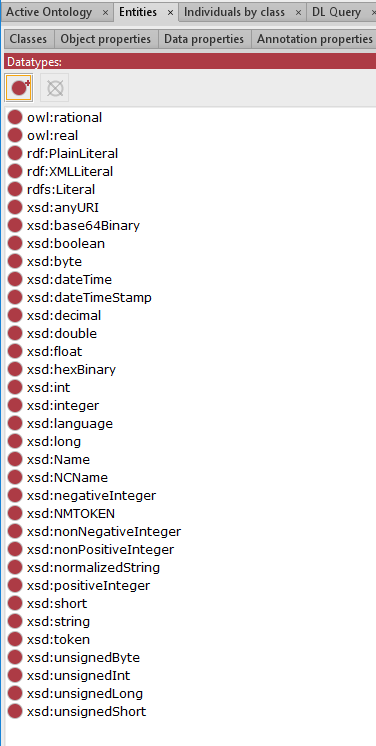


Hình 3.5: Ảnh chụp màn hình giao diện làm việc Protege: thuộc tính cho các lớp trong Ontology nông nghiệp An Giang

### Xác định các khía cạnh của thuộc tính



Hình 3.6: Ảnh chụp màn hình giao diện làm việc Protege: khía cạnh của thuộc tính trong Ontology nông nghiệp An Giang



Hình 3.7: Ảnh chụp màn hình giao diện làm việc Protege: Các loại dữ liệu của các khía cạnh trong Ontology nông nghiệp An Giang

# Thiết kế và xây dựng công cụ tìm kiếm dựa trên Ontology

## Thiết kế hệ thống tìm kiếm Ontology



Hình 4.1. Mô hình thiết kế hệ thống phục vụ chức năng tìm kiếm Ontology

Hệ thống phục vụ tìm kiếm Ontology được thiết kế gồm 2 tầng chính là tầng máy chủ (Server) và tầng máy khách (Client) theo mô hình Clients -Server – Database.

**Tầng Server**: tầng này được thiết kế để đáp ứng các yêu cầu về tìm kiếm từ phía Client gửi đến và trả các kết quả tìm kiếm về cho người dùng thông qua giao thức HTTP;

Về cơ bản thì tầng này là một máy chủ Web với phần mềm Apache Tomcat được cài đặt như là một Web Server. Apache Tomcat có chức năng thực thi các phần mềm hỗ trợ cung cấp các dịch vụ hỗ trợ truy vấn SparQL cho Ontology như Jena fuseki hay D2RQ.

Đồng thời, tâng này cũng là tầng lưu trữ và thực thi các kết nối đến cơ sở dữ liệu (database), giúp truy vấn và trả về các kết quả phù hợp với yêu cầu của người dùng.

**Tầng Client**: Tầng này là nơi giao tiếp với người dùng thông qua giao diện chức năng tìm kiếm trên các trình duyệt Web khác nhau như Chrome, Firefox, Opera, MS Edge.. Tại đây, nội dung truy vấn của người dùng sẽ được gửi đến Server để xử lý, sau đó nhận kết quả từ Server và hiển thị ra cho người dùng thông qua kỹ thuật Ajax.

## Xây dựng máy chủ Ontology

Máy chủ Ontology là một máy chủ đã được thiết lập môi trường và cài đặt các phần mềm hỗ trợ cung cấp các dịch vụ truy vấn Ontology như mô hình 5.1

Do hệ thống sẽ sử dụng hai phần mềm Jena Fuseki và D2RQ là hai phần mềm Java-based Web để kết nối với Database, truy vấn SparQL và cung cấp các dịch vụ khác cho công cụ tìm kiếm, do vậy cần cài đặt môi trường Web Server Java-based để thực thi chúng. Apache Tomcat là một phần mềm máy chủ Web mã nguồn mở phổ biến, hỗ trợ thực thi các ứng dụng Web được viết trên nền Java như jsp, servlet,.. được khuyến khích sử dụng để thực thi hai ứng dụng trên.

### Phần mềm sử dụng

**Apache Tomcat:** Đây là một phần mềm mã nguồn mở có chức năng thực thi các ứng dụng sử dụng công nghệ Java Servlet, JavaServer Pages, Java Expression Languate và Java WebSocket. Tomcat là một ứng dụng máy chủ gọn nhẹ, thường dùng để deploy các ứng dụng Java Web. Nó được phát triển bởi Apache và hoàn toàn miễn phí.

**Apache Jena**: Đây là một phần mềm cung cấp các thư viện, chức năng hỗ trợ cho Web ngữ nghĩa (Semantic Web) được xây dựng trên nền Java. Nó cung cấp các API cho phép trích xuất và ghi dữ liệu vào biểu đồ RDF thông qua cú pháp SparQL.

**Jena Fuseki**: Có thể xem đây là giao diện HTTP của Jena cho dữ liệu RDF. Nó cũng hỗ trợ ngôn ngữ SparQL cho việc cập nhật và truy vấn dữ liệu. Người dùng có thể import các file RDF hay OWL vào để sử dụng. Đồng thời Jena Fuseki cũng cung cấp dịch vụ cho phép người dùng truy vấn dữ liệu thông qua Web Services của nó. Web Services này hỗ trợ chuẩn dữ liệu XML và JSON.

**D2RQ**: Đây là một bộ các công cụ cho phép người dùng “mapping” giữa dữ liệu quan hệ và dữ liệu Ontology, đồng thời cũng có thể xuất dữ liệu ra các kiểu dữ liệu Ontology như TURTLE, RDF/XML, RDF/XML-ABBREV, N3, N-TRIPLE,.. Có rất nhiều hệ quản trị cơ sở dữ liệu quan hệ khác nhau mà bộ công cụ này hỗ trợ như PostgreSQl, MySQL, SQL Server, Oracle,..

**PostgreSQL**: Đây là một phần mềm quản trị hệ cơ sở dữ liệu quan hệ mã nguồn mở phổ biến, đặc biệt là trong lĩnh vực dữ liệu không gian. Phần mềm này được sử dụng để lưu trữ và quản lý dữ liệu cho toàn hệ thống thông tin Nông nghiệp An Giang nói chung và chức năng tìm kiếm dựa trên Ontology nói riêng. Tuy nhiên, điểm cần lưu ý là dữ liệu được lưu trữ ở PostgreSQL là dữ liệu quan hệ, còn dữ liệu để xây dựng hệ thống Ontology phải được chuyển đổi sang dữ liệu ngữ nghĩa.

Trong hệ thống thông tin nông nghiệp An Giang thì D2RQ có chức năng kết nối giữa Database của hệ thống được lưu trữ bằng phần mềm quản trị dữ liệu quan hệ là PostgreSQL và máy chủ cung cấp dịch vụ Ontology do Jena đảm nhiệm.

### Ngôn ngữ sử dụng

**PHP**: viết tắt của "Hypertext Preprocessor", là một ngôn ngữ lập trình kịch bản được chạy ở phía server nhằm sinh ra mã html trên client. PHP đã trải qua rất nhiều phiên bản và được tối ưu hóa cho các ứng dụng web, với cách viết mã rõ rãng, tốc độ nhanh, dễ học nên PHP đã trở thành một ngôn ngữ lập trình web rất phổ biến và được ưa chuộng.

Phía Back-end của hệ thống thông tin nông nghiệp An Giang được xây dựng chủ yếu bằng ngôn ngữ PHP, PHP cũng được dùng để xây dựng các Services riêng kết nối đến máy chủ Ontology cho công cụ tìm kiếm bằng Ontoloy.

**SparQL**: là ngôn ngữ và giao thức truy vấn tiêu chuẩn cho Dữ liệu Mở Liên kết (Linked Open Data) trên web hoặc trong cơ sở dữ liệu đồ họa ngữ nghĩa (còn được gọi là RDF triplestore - bộ 3 RDF).

SPARQL, là viết tắt của Giao thức SPARQL và Ngôn ngữ Truy vấn RDF bằng tiếng Anh, "SPARQL Protocol And RDF Query Language", cho phép những người sử dụng truy vấn thông tin từ các cơ sở dữ liệu hoặc bất kỳ nguồn dữ liệu nào có thể được ánh xạ tới RDF.

Tiêu chuẩn SPARQL được W3C thiết kế và phê chuẩn và giúp những người sử dụng và những người phát triển tập trung vào những gì họ muốn biết thay vì cách mà cơ sở dữ liệu được tổ chức.

Trong việc xây dựng chức năng tìm kiếm bằng Ontology, ngôn ngữ SparQL giúp truy vấn được dữ liệu từ các cơ sở dữ liệu ngữ nghĩa RDF. Kết hợp với PHP, SparQl giúp xây dựng các Services truy vấn dữ liệu từ máy chủ Ontology.

## Xây dựng công cụ tìm kiếm phía người dùng

### Ngôn ngữ

Công cụ tìm kiếm phía người dùng được xây dựng chủ yếu bằng ngôn ngữ HTML, CSS và Javascript.

**HTML**: viết tắt cho HyperText Markup Language, hay là "Ngôn ngữ Đánh dấu Siêu văn bản" là một ngôn ngữ đánh dấu được thiết kế ra để tạo nên các trang web với các mẩu thông tin được trình bày trên World Wide Web. Cùng với CSS và JavaScript, HTML tạo ra bộ ba nền tảng kỹ thuật cho World Wide Web. HTML được định nghĩa như là một ứng dụng đơn giản của SGML và được sử dụng trong các tổ chức cần đến các yêu cầu xuất bản phức tạp. HTML đã trở thành một chuẩn Internet do tổ chức World Wide Web Consortium (W3C) duy trì.

Khung giao diện của chức năng tìm kiếm dữ liệu dựa trên Ontology như các nút chức năng, các khu vực lựa chọn, bảng dữ liệu trả về,.. được xây dựng bằng ngôn ngữ HTML.

**CSS**: viết tắt của Cascading Style Sheets – được dùng để miêu tả cách trình bày các tài liệu viết bằng ngôn ngữ HTML và XHTML. Ngoài ra ngôn ngữ định kiểu theo tầng cũng có thể dùng cho XML, SVG, XUL. Các đặc điểm kỹ thuật của CSS được duy trì bởi World Wide Web Consortium (W3C). CSS giúp quy định kiểu dáng cho văn bản HTML (hoặc XHTML) như màu sắc, định dạng, bố cục, kích thước, font chữ,...

Ở công cụ tìm kiếm, CSS giúp quy định màu sắc và bố cục cho chức năng. Giúp định dạng bố cục và bảng dữ liệu trả về.

**Javascript**: Javascript là một ngôn ngữ lập trình kịch bản dựa vào đối tượng phát triển có sẵn hoặc tự định nghĩa ra, javascript được sử dụng rộng rãi trong các ứng dụng Website. Javascript được hỗ trợ hầu như trên tất cả các trình duyệt như Firefox, Chrome, ... thậm chí các trình duyệt trên thiết bị di động cũng có hỗ trợ.

Đây là ngôn ngữ lập trình chính được sử dụng ở phía người dùng. Ngôn ngữ này giúp xây dựng các chức năng có tính tương tác trực tiếp với người dùng như tìm kiếm, chọn dữ liệu, hiển thị dữ lieuj. Đây cũng là nền tảng để thực hiện kỹ thuật Ajax. Một kỹ thuật chính yếu được sử dụng để xây dựng công cụ tìm kiếm dữ liệu dựa trên Ontology.

### Kỹ thuật sử dụng

Chức năng tìm kiếm dữ liệu dựa trên Ontology chủ yếu sử dụng kỹ thuật Ajax để gửi yêu cầu, nhận kết quả trả về từ Server và hiển thị ra cho người dùng.

Ajax là viết tắt của Asynchronous JavaScript and XML – đây là kỹ thuật kết hợp hai tính năng mạnh của JavaScript được các nhà phát triển đánh giá rất cao:

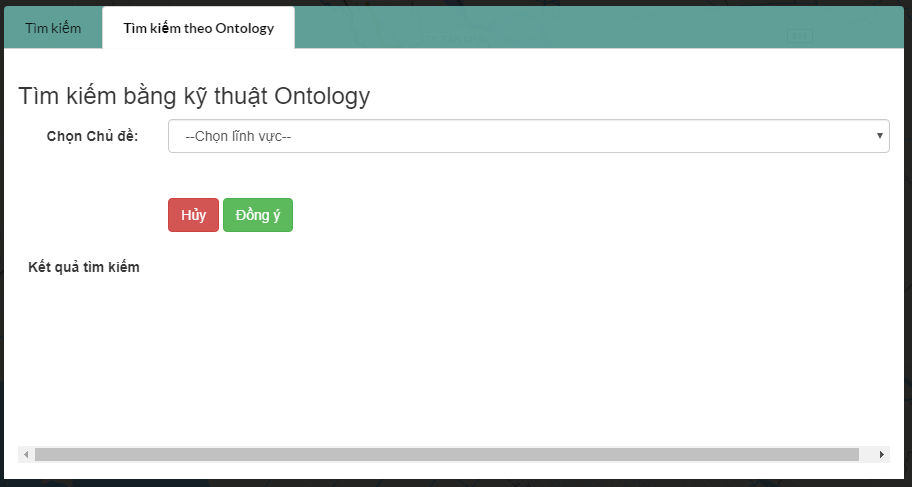
• Gửi yêu cầu (request) đến máy chủ mà không cần nạp lại trang

• Phân tách và làm việc với XML

Các ứng dụng Ajax xoay quanh một tính năng có tên là XMLHttpRequest. Khi đó, các yêu cầu gửi resquest và nhận response do Ajax Engine thực hiện. Thay vì trả dữ liệu dưới dạng HTML và CSS trực tiếp cho trình duyệt, web server có thể gửi trả dữ liệu dạng XML và Ajax Engine sẽ tiếp nhận, phân tách và chuyển hóa thành XHTML+CSS cho trình duyệt hiển thị. Việc này được thực hiện trên client nên giảm tải rất nhiều cho server, đồng thời người sử dụng cảm thấy kết quả xử lý được hiển thị tức thì mà không cần nạp lại trang. Mặt khác, sự kết hợp của các công nghệ web như CSS và XHTML làm cho việc trình bày giao diện trang web tốt hơn nhiều và giảm đáng kể dung lượng trang phải nạp

### Giao diện chức năng

Chức năng tìm kiếm dữ liệu dựa trên Ontology được xây dựng thành một tab riêng biệt với chức năng tìm kiếm thông thường như hình dưới.

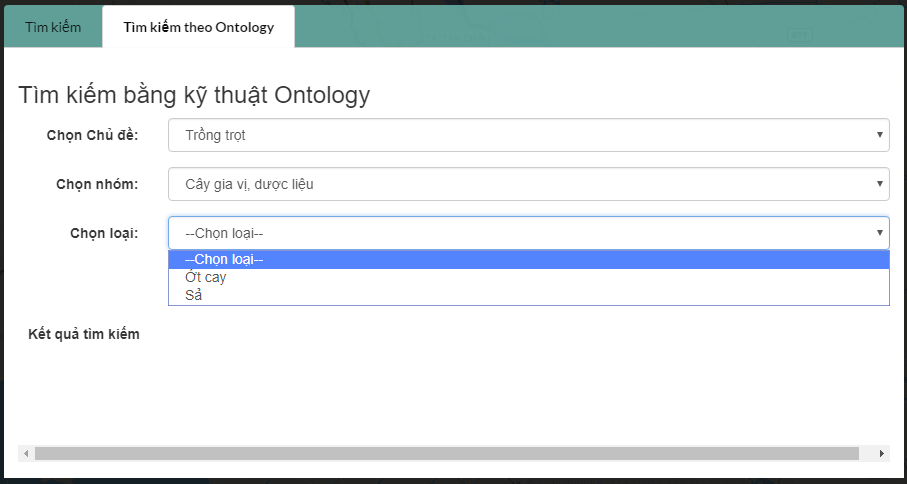


Hình 4.2. Giao diện chức năng tìm kiếm dựa trên Ontology

Giao diện chức năng được phân làm 2 khu vực chính là:

* Khu vực chọn lựa các yêu cầu tìm kiếm và xác nhận tìm kiếm
* Khu vực hiển thị kết quả tìm kiếm.

**Khu vực lựa chọn các nội dung tìm kiếm**

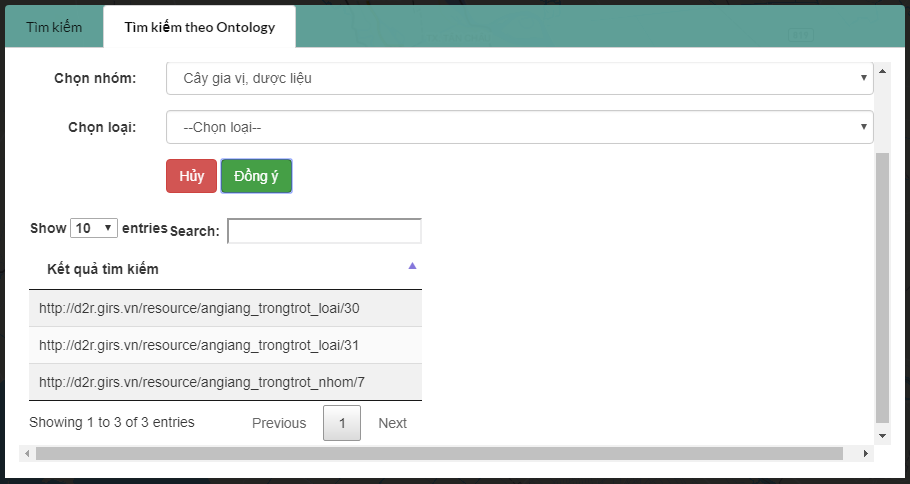


Hình 4.3. khu vực lựa chịn các nội dung tìm kiếm

Khu vực này cho phép người dùng lựa chọn các thông tin mà họ quan tâm như nông nghiệp, chăn nuôi hay thủy sản. Sau khi lựa chọn một chủ đề thì bằng kỹ thuật Ajax, hệ thống sẽ trả về danh sách các nhóm và loại đối tượng tương ứng với mỗi chủ đề cho người dùng lựa chọn.

**Khu vực hiển thị kết quả tìm kiếm**

Đây là khu vực hiển thị danh sách các kết quả tìm kiếm từ những thông tin mà người dùng lựa chọn từ khu vực tìm kiếm.



Hình 4.4. Khu vực hiển thị dữ liệu

Theo mô hình dữ liệu Ontology thì mỗi một dữ liệu sẽ lưu trữ các thông tin sau:

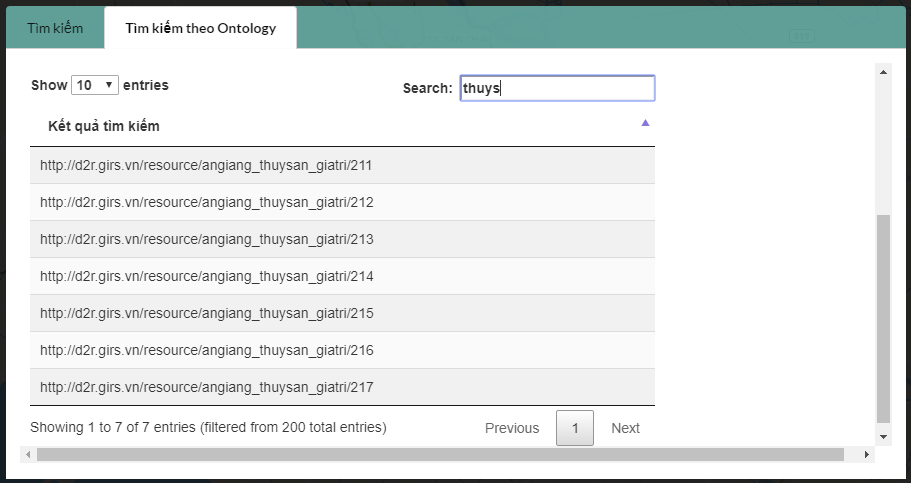
* Subject (S): Đối tượng – thường là một đường dẫn liên kết (Linked open data) đến toàn bộ các thông tin khác của đối tượng đó.
* Property hay Predicate (P): Là các quan hệ, thuộc tính của đối tượng
* Object (O): Tương ứng với mỗi một Property hay Predicate là một giá trị.

Thỉnh thoảng, một O có thể thay thế cho một S và ngược lại.

Do vậy, sau khi người dùng tìm kiếm hệ thống sẽ trả về cho người dùng các Subject (dưới dạng là các link liên kết) có quan hệ với các thông tin mà người dùng đã lựa chọn dưới dạng một bảng dữ liệu.

Số lượng kết quả dữ liệu trả về được giới hạn ở mức 200 records giúp hệ thống hoạt động ổn định.

**Danh sách kết quả trả về**



Hình 4.5. Kết quả truy vấn

Kết quả trả về được hiển thị dưới dạng mọt bảng dữ liệu. Bảng dữ liệu này có các chức năng sau:

* Lọc dữ liệu: Cho phép lọc dữ liệu trả về theo keyword;
* Phân trang: Cho phép tùy chọn số lượng record hiển thị và chọn trang muốn xem;
* Sắp xếp: Cho phép sắp xếp dữ liệu trả về theo ABC.

**Xem chi tiết thông tin**



Hình 4.6. Chi tiết thông tin

Chức năng này cho phép người dùng xem chi tiết các quan hệ cũng như các giá trị của quan hệ đó trong một đối tượng.

TÀI LIỆU THAM KHẢO

[1] R. Corazzon, *Theory and History of Ontology*. 2016.

[2] N. H. Nhật, “Tổng quan về Ontology.”

[3] F. Freitas and N. F. Noy, “Ontology Issues and Applications Guest Editors ’ Introduction.”

[4] T. Gruber, “Toward principles for the desing of ontologies used for knowledge sharing,” *Acad. Press Ltd.*, no. 43, pp. 907–928, 1995.

[5] F. T. Fonseca, “Ontologies and Knowledge Sharing in Urban GIS,” pp. 1–27.

[6] T. R. Gruber, “A translation approach to portable ontology specifications. Knowledge Acquisition,” *Knowl. Acquis.*, vol. 5, no. 2, pp. 199–220, 1993.

[7] J. Davies, R. Studer, and P. Warren, *Semantic Web Technologies: Trends and Research in Ontology-based Systems*. 2006.

[8] N. Guarino and C. Welty, “A Formal Ontology of Properties,” in *Proceedings of the EKAW-2000: The 12th International Conference on Knowledge Engineering and Knowledge Management*, 2000, pp. 97–112.

[9] R. Mizoguchi, M. Ikeda, K. Seta, and J. Vanwelkenhuysen, “Ontology for Modeling the World from Problem Solving Perspectives,” in *Proc. of IJCAI-95 Workshop on Basic OntologicalIssues in Knowledge Sharing*, 1995, pp. 1–12.

[10] O. Lassila and D. McGuinness, “The Role of Frame-Based Representation on the Semantic Web Knowledge Systems Laboratory Report KSL-01-02,” 2001.

[11] V. Heijst, A. T. S. G., and B. J. Wielinga, “Using explicit ontologies in KBS development,” *Int. J. Hum. Comput. Stud.*, vol. 46, no. 2/3, pp. 183–292, 1997.

[12] B. Lauser, M. Sini, A. Liang, J. Keizer, and S. Katz, “From AGROVOC to the Agricultural Ontology Service / Concept Server An OWL model for creating ontologies in the agricultural domain.”

[13] D. Soergel, B. Lauser, A. Liang, F. Fisseha, J. Keizer, and S. Katz, “Reengineering Thesauri for New Applications: the AGROVOC Example,” pp. 1–23.

[14] FAO, “AGROVOC GUIDELINES FOR ENGLISH LANGUAGE EDITORS,” no. May, 2015.

[15] C. Roussey, F. Pinet, M. A. Kang, and O. Corcho, *An Introduction to Ontologies and Ontology Engineering*. 2011.

[16] C. Roussey, V. Soulignac, J. Champomier, V. Abt, and J. Chanet, “Ontologies in Agriculture 2 Ontology Classification,” *AgEng 2010, Int. Conf. Agric. Eng.*, pp. 1–10, 2010.

[17] C. Chang, G. Xian, and G. Li, “Thesaurus and Ontology Technology for the Improvement of Agricultural Information Retrieval,” *World Conf. Agric. Inf. IT*, pp. 531–536, 2008.

[18] V. Ivanova, P. Lambrix, S. Lohmann, C. Pesquita, and S. Lohmann, “VOILA 2015 Visualizations and User Interfaces for Ontologies and Linked Data : Proceedings of the International Workshop on Visualizations and User Interfaces for Ontologies and Linked Data,” 2015.

[19] S. Pokharel, M. A. Sherif, and J. Lehmann, “Ontology Based Data Access and Integration for Improving the Effectiveness of Farming in Nepal.”

[20] X. L. Su, J. Li, Y. P. Cui, X. X. Meng, and Y. Q. Wang, “Review on the Work of Agriculture Ontology Research Group,” *J. Integr. Agric.*, vol. 11, no. 5, pp. 720–730, 2012.

[21] J. Jebaraj, “An Exploratory Study on Agriculture Ontology : A Global Perspective,” no. 6, pp. 202–206, 2017.

[22] T. Submitted, P. Fulfillment, and T. Agriculture, “THESIS ONTOLOGY DEVELOPMENT FOR AGRICULTURAL RESEARCH KNOWLEDGE MANAGEMENT : A CASE STUDY FOR THAI RICE,” 2009.

[23] A. Thunkijjanukij and A. Kawtrakul, “Ontology Development : A Case Study for Thai Rice,” vol. 604, pp. 594–604, 2009.

[24] P. Jaiswal, S. Avraham, K. Ilic, E. A. Kellogg, S. Mccouch, P. Stevens, L. Vincent, D. Ware, and F. Zapata, “Plant Ontology ( PO ): a controlled vocabulary of plant,” pp. 388–397, 2006.

[25] T. Plant and O. Consortium, “The Plant Ontology 2 Consortium and Plant Ontologies,” no. March, pp. 137–142, 2002.

[26] M. Hulden and P. Report, “CWR Ontology,” 2007.

[27] X. Liu, Z. Li, and S. Jiang, “Ontology-based representation and reasoning in building construction cost estimation in China,” *Futur. Internet*, vol. 8, no. 3, 2016.

[28] C. H. Trụ, “VN-KIM cho Web Việt có ngữ nghĩa,” vol. 10, pp. 17–19, 2014.

[29] M. Takeya, H. Numa, and K. Doi, “Ontology Using Role Concept Recognized on Biological Relationships and Its Application,” *Genome Informatics*, vol. 14, pp. 685–686, 2003.

[30] P. L. D. Vincent, E. H. Coe, and M. L. Polacco, “Zea mays ontology–a database of international terms,” *Trends Plant Sci.*, vol. 8, no. 11, pp. 517–520, 2003.

[31] M. Uschold and G. M., “Ontologies: Principles, Methods and Applications,” *Knowl. Eng. Rev.*, 1996.

[32] D. Soergel, B. Lauser, A. Liang, F. Fisseha, J. Keizer, and S. Katz, “Reengineering thesauri for new applications: the AGROVOC example,” *J. Digit. Inf.*, vol. 4, no. 4, 2004.

[33] T. C. Án, T. T. N. Mai, and L. T. T. Lan, “Xây dựng ontology tự động từ bảng chú giải,” *Can Tho Univ. J. Sci.*, vol. Công nghệ, p. 133, 2017.

[34] S. Bird, E. Klein, E. Loper, and J. Baldridge, “Multidisciplinary instruction with the natural language toolkit,” *TechCL*, no. June, pp. 62–70, 2008.

[35] T. T. Binh, “The Knowledge-based search for water-related information system for the Mekong Delta, Vietnam,” Universität Bonn, 2013.

1. Berendt, B., Hotho, A., Mladenič, D., van Someren, M., Spiliopoulou, M., Stumme, G.: A Roadmap for Web Mining: From Web to Semantic Web. In: Berendt, B., Hotho, A., Mladenič, D., van Someren, M., Spiliopoulou, M., Stumme, G. (eds.) EWMF 2003. LNCS (LNAI), vol. 3209, pp. 1–22. Springer, Heidelberg (2004) [↑](#footnote-ref-1)
2. Antoniou, G., Harmelen, F.V.: Web Ontology Language: OWL. Handbook on Ontologies, 67–92 (2004) [↑](#footnote-ref-2)