ISSN 2007-9737

Development of Agricultural Knowledge Graphs through the Reuse of Ontologies

Fernando Pech-May[1], German Rios-Toledo[2]

[1] Instituto Tecnológico Superior de los Ríos, Tabasco, Mexico

[2] Tecnológico Nacional de México, Campus Tuxtla Gutiérrez, Mexico

fernando.pech@cinvestav.mx, german.rt@tuxtla.tecnm.mx

Abstract. Agriculture plays a very important role due to economic and food dependence. This sector generates large volumes of data represented in the form of text, tables, etc., which lacks meaning or value to be used in decision making by agricultural entrepreneurs, government agencies or researchers in the area. Due to the need to generate agricultural knowledge, some organizations, led by the United Nations Food and Agriculture Organization, implement technologies to structure and formalize knowledge through semantic resources. This has led to the development of a large amount of agricultural semantic resources. However, these resources have been developed with characteristics such as climate, soil, relief, vegetation, fauna and hydrology of a particular country or region, which makes it impossible to use them in another country. Since in Mexico, agriculture is an important sector, it is necessary to make semantic resources of the most common crops in Mexico. This article presents the development of 3 semantic resources of the most common crops in Mexico through the reuse of existing ontologies in the literature.

Keywords. Ontology agriculture, agricultural knowledge base, agricultural knowledge graphs.

1 Introduction

Agriculture is the most important sector in the world, providing food sustainability for millions of people. However, the sector faces multiple challenges, such as population growth, lack of investment, natural disasters and, in some areas of the planet, the low availability of natural resources such as water, essential for the development of crops [19].

Currently, a large amount of data related to agriculture is generated, which is obtained from different sources such as soil sensors, drones, weather stations, etc.

But commonly, this data is stored in relational databases or CSV files in in tabular representation. Although these data are useful for decision-making, the link between these data is made when structured queries are made or when generating reports, graphs, or tables.

To achieve a greater impact on decision-making, knowledge of this data must be formalized. In this sense, different organizations, governments, and companies, led by the Food and Agriculture Organization of the United Nations (FAO), have created initiatives to formalize the knowledge of large volumes of data from different branches of agriculture, knowledge that is valuable for future generations.

For data to be useful, it needs to be formalized to provide context or meaning. Formalization requires knowledge engineering: The technologies and tools of the semantic web such as ontologies, vocabularies, taxonomies, etc. They are very useful for the development of applications to achieve sustainable and precise crops [7].

ISSN 2007-9737

Currently, research is being carried out to develop semantic resources that include vocabularies, taxonomies and ontologies in different domains of agriculture such as Agrovoc[1], Tesauro agrícola de la biblioteca nacional de agricultura (NALT)[2], etc. These are generalized vocabularies that contain terms specific to different agricultural subdomains such as fertilizers, climate, pesticides, etc.

On the other hand, knowledge graphs for specific crops have been developed such as Plant Ontology [6], AgriOnt [16], OntoAGroHidro [5], etc. All of these semantic resources are adapted to specific European geographical areas, which makes them unsuitable for use in Mexico because such characteristics are different, such as language, soil, climate, hydrology, and other concepts specific to the region.

Therefore, it is necessary to develop ontologies that provide organized information regarding the main crops in Mexico and that favor decision-making for users interested in agricultural activities. This article presents the development of 3 semantic resources through the reuse of ontologies to use them in support systems for agricultural decision-making.

2 Background and Related Works

The Semantic Web is an extension of the current web and its purpose is to organize and structure knowledge so that it is understood by computers [4]. The Semantic Web has tools to represent knowledge in the form of graphs through a hierarchy of concepts, attributes, and relationships. To design and/or develop semantic websites, there are languages like RDF and OWL [3] stand out for the development of ontologies [18].

An ontology is used to formally represent and share knowledge [18, 15] of different types of data, and in different areas. An ontology is composed of a schema and instances that represent the description of concepts and their relationships (see Figure 1). This knowledge is encoded using ontological languages (OWL).

[1] aims.fao.org/vest-registry/vocabularies/agrovoc
[2] agclass.nal.usda.gov/

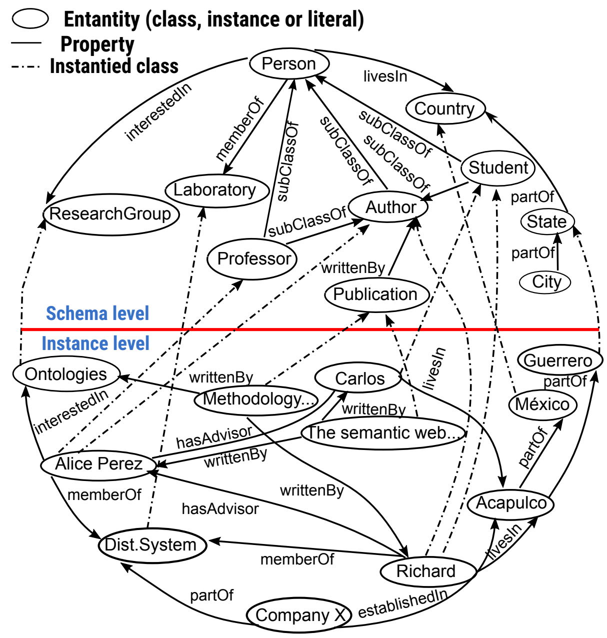

Fig. 1. View of an ontology at two levels: schema and instance

Ontologies are used for the representation of knowledge and the development of applications that are classified as:

– Knowledge-based systems (applications that perform reasoning and suggest solutions through knowledge bases).

– Remote sensors (that integrate data through remote devices).

– Decision support system (which represent knowledge and suggest recommendations).

– Expert systems (applications that make decisions through information reasoning).

This structured knowledge enables new scientific discoveries by linking different data sets. Agronomy, agriculture, food, plant science, and biotechnology benefit from these discoveries.

Nowadays, there are ontologies related to agriculture, both general and specific. In addition, some portals allow you to view and manage agricultural semantic resources.

Table 1. The Metrics of an ontology

Name	Description
Axioms	Properties about the behavior of a class
Classes	Individuals/objects with common characteristics
Object Properties	Description of attributes of an object
Individuals	Domain Particular Objects
Annotation Properties	Detailed description of classes
Subclasses	Child class of another class

Related works have been classified into two categories: 1) agricultural semantic resources and 2) agricultural web portals.

– Agricultural semantic resources. Semantic resources for agriculture use semantic technologies to describe the knowledge collected by an organization or individual.

  Currently, there are a variety of semantic resources in the agricultural domain such as vocabularies, taxonomies, thesauri, and knowledge graphs. The most relevant ones are mentioned below.

  – Agrovoc[3] is one of the largest resources. Contains terms and concepts about agriculture, food, nutrition, fishing, forestry, and the environment. It is available in 27 languages and has 35,000 concepts and 40,000 terms; It supports Linked Open Data (LOD) and is linked to 16 resources.

    Agrovoc contains different irregularities in the relationships between concepts. However, it is used to create new ontological resources and adapt them to a particular region.

  – Chinese Agricultural Thesaurus (CAT) [13] contains 40 main categories, more than 63 thousand concepts, and includes more than 130 thousand semantic relationships. Much of the concepts are in English and it is serialized in RDF/SKOS-XL.

  – Tesauro Agrícola de la Biblioteca Nacional de Agricultura (NALT)[4] contains 128.25 agricultural terms in English and Spanish, 40 categories, such as crop classification, and 63,000 agricultural concepts. legume crops, beans, etc. It also supports linked open data schemas and is integrated with other semantic resources.

  – Plant Ontology (PO) [6]. Related to multispecies anatomy and developed for the annotation of genes and phenotypes. It is developed for different languages.

    Additionally, it contains more than 2.2 million annotations linking PO terms to 110,000 data objects representing genes or gene models, proteins, ARN, germplasm, and quantitative trait loci (QTL) from 22 plant species.

  – OntoAgroHidro [5]. Domain ontology was created by expert researchers interested in water resources, climate change, and land use. It has 6 main classes which are broken down into other subclasses, relationships, properties, and cases.

  – Agricultural Ontology (AgriOnt) [16]. Provides an overview of the domain of agriculture, agricultural concepts, and life cycles between seeds, plants, harvesting, transportation, and consumption. It also provides relationships between agricultural concepts and other concepts related to climate, soil conditions, and fertilizers. In addition, it is made up of 447 classes and more than 700 axioms.

– Agricultural web portals. There are web portals that host a set of semantic resources in the agricultural domain.

  – Agroportal [10]. It contains 98 ontologies and thesauri. In addition, it allows semantic annotation and stores and explores linked data.

  – Crop Ontology [9]. It hosts a set of ontologies focused on the agricultural domain, and allows loading dictionaries of traits for

[3] aims.fao.org/vest-registry/vocabularies/agrovoc

[4] https://agclass.nal.usda.gov/

doi: 10.13053/CyS-27-4-4789

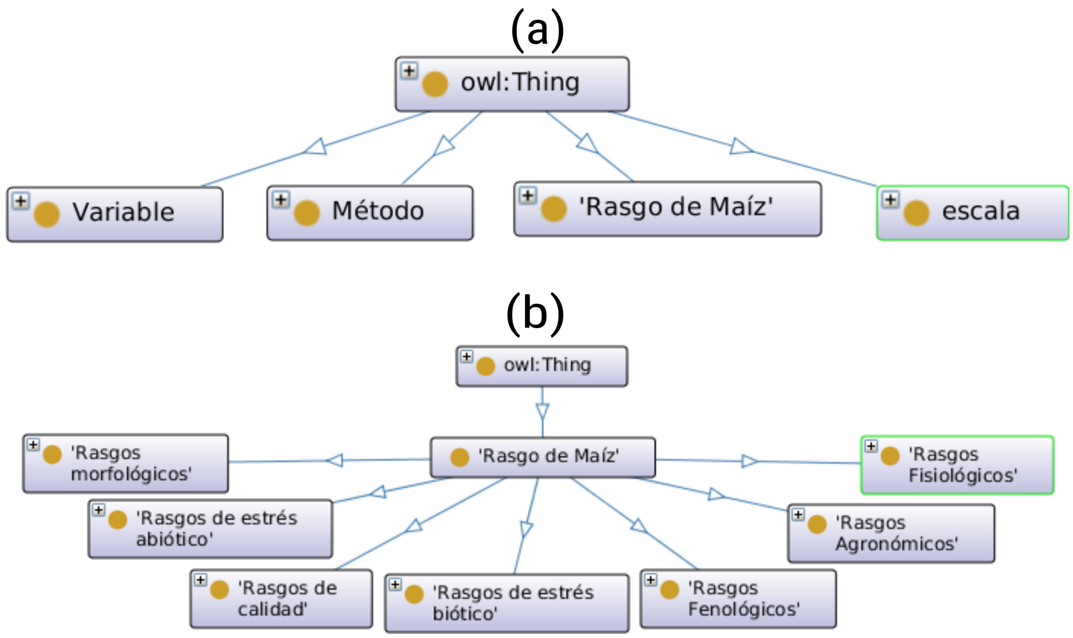

Fig. 2. Maize ontology classes. (a) Main Classes and (b) subclasses of Rasgos del Maíz

crop improvement and the direct creation of ontologies; Likewise, it allows you to consult, create, update, or delete ontologies.

– CIARD Ring [17]. It is a portal with semantic web services that contains 3201 data sets and 5327 data services. In addition, it allows indexing a series of software tools to analyze such data.

– Vest[5]. It is a repository that belongs to FAO and Godan. It has 398 resources and a description of graphs with LOD. In addition, it allows queries to be made in the SPARQL language [8].

In addition, proposals have been made through the use of semantic resources from the agricultural domain. Jonquet et al. [10] created a repository of semantic resources called Agroportal focused on agriculture; they reuse the semantic knowledge of the biomedical domain in the agronomic, food, flora and biodiversity sectors.

This portal provides ontology hosting, search, version management, visualizations, comments, and recommendations. Likewise, it allows semantic annotation, stores and exploits ontological links. Ngo et al. [16] developed AgriOnt, an agricultural ontology for intelligent agricultural systems.

This ontology is made up of 4 related thematic subdomains: agricultural, geographical, IoT (internet of things), and the business subdomain.

Chatterjee and N. Kaushik [11] developed techniques for automatic extraction of vocabulary and relationships between terms. A fundamental activity in the creation of ontologies. Furthermore, they proposed a scheme for an ontology of the agricultural domain.

Such a scheme uses domain-dependent regular expressions and natural language processing techniques. They also identify the semantic relationships between the extracted terms and phrases.

Likewise, they propose a rule-based reasoning algorithm RelExOnt (Relation Extraction for Ontology) using term similarity. B. Drury et al. [7] developed a web application to address agricultural problems.

[5]aims.fao.org/vest-registry

The application allows a comprehensive review of existing semantic resources and their construction methods, and data exchange standards, as well as a survey of current applications of semantic web technologies.

B. Drury et al. stated that there is a limited adaptation of these technologies in academic literature, unlike complementary domains such as biomedicine. Kawtrakul [12] present the design of a system implemented an agricultural domain knowledge service system called Cyber-Brain.

This system combines approaches based on knowledge engineering and language engineering to collect knowledge from various sources to provide an effective knowledge service and enable advanced search agriculture.

R. Almadhoun and Abu-Naser [2] proposed an expert system that allows farmers and specialists to diagnose and provide appropriate advice on banana diseases.

The expert system was developed with the CLIPS language. Ahsan et al. [1] developed an approach to the agriculture domain. The knowledge management approach helps stakeholders share and understand domain and knowledge levels.

Furthermore, it does not require integration with other sources of knowledge compared to traditional approaches. Malik and Sharan [14] proposed the design of a generic approach for agricultural domain entities that includes five important classes: plants, diseases, pests, pesticides, and fertilizers. They described the use of an ontology to extract text and recover the source of information.

3 Development of Agricultural Ontologies

The development of ontologies requires experts in the knowledge domain and considerable time. Generally, ontologies are made considering the characteristics of an area.

Therefore, they contain elements and values typical of a region or country. Currently, there are semantic resources that can be reused to create new agricultural resources adapted to a particular region.

Table 2. Ontological Maize Metrics

Metric	Quantity
Axioms	11542
logical axioms	3810
statement axioms	2196
Classes	1103
Object properties	3
Individuals	1088
Annotation properties	9
Subclasses	2722

In this work, corn, bean and general agricultural ontologies were developed through the reuse of semantic resources from Crop Ontology.

To do this, the ontologies were translated, their structures (classes, axioms, properties, etc.) were analyzed and different subdomains were created to adapt them to the particular characteristics of Mexico.

3.1 Maize Ontology

The ontology has the purpose of measuring corn traits through different variables according to the method or scale used. The ontology provides harmonized breeder trait names, measurement methods, scales and standard variables.

The SKOS[6] resource is used for transitivity documentation, definition, labeling, acronym assignment, and prefix assignment.

Likewise, it adopts terms from open biomedical ontologies. Figure 2 shows the main classes of the ontology and Table 2 shows the total metrics of the ontology. Each class is described below:

– Rasgo de Maíz. This Class contains physical and chemical characteristics of the plant (see Figure 2b) such as:

  1. Rasgos Agronómicos, are the physical characteristics of the plant (height of the leaf ear, aspects, number of ears, number of grains, among others).

[6] skos.um.es/

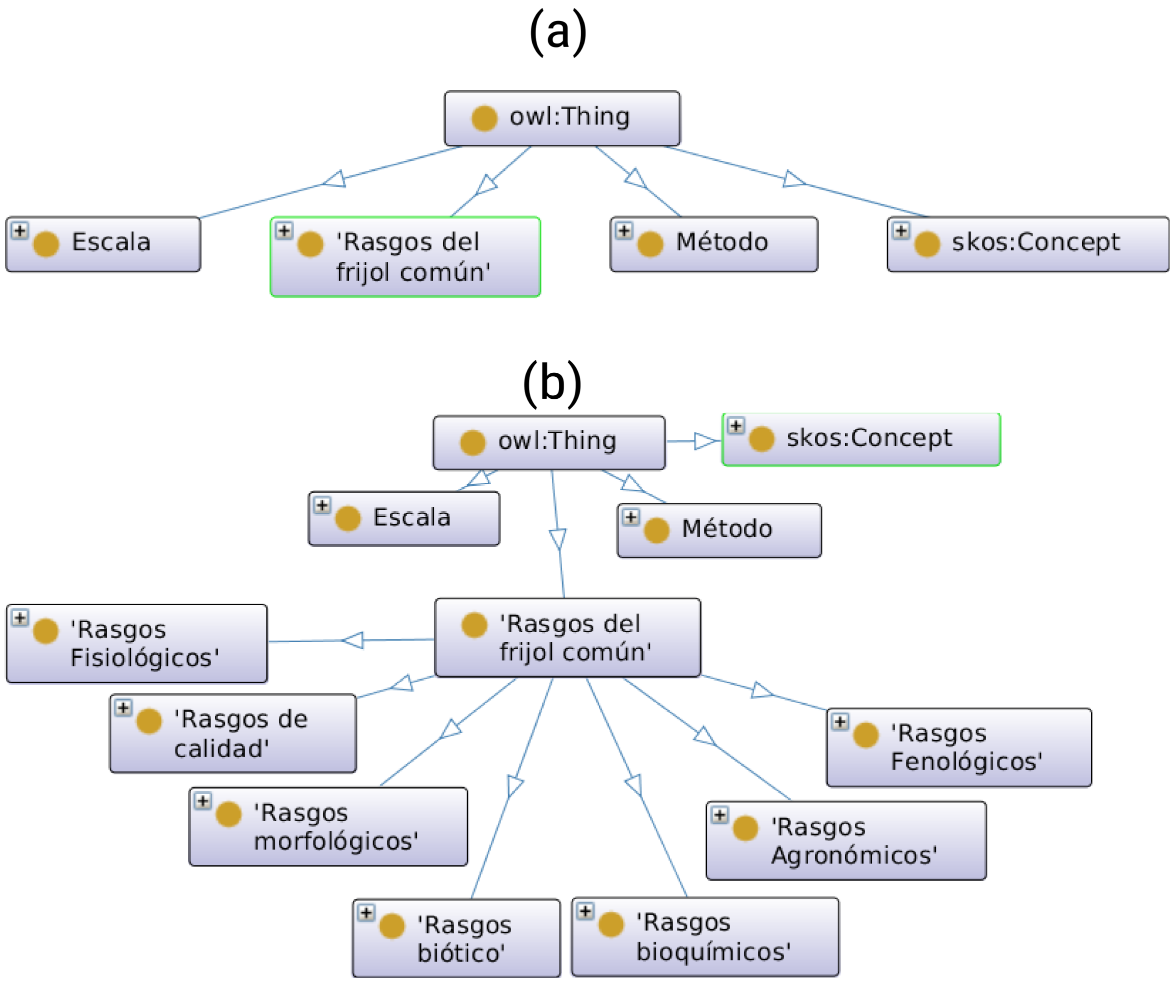

Fig. 3. Main classes of bean ontology. (a) Main Classes and (b) Rasgos del frijol común class structure

2. Rasgos de Calidad, contains characteristics of the quality of the corn grain and is established through the physical constitution to determine the texture, hardness and chemical composition (iron content, zinc, grain texture, etc.).
3. Rasgos de Estres Abiotico, made up of two subclasses that determine the negative traits of plant adaptation to different stress conditions (leaf severity and blasting).
4. Rasgos de Estres Biótico, subclasses that determine the negative impact of factors such as fungi, viruses and herbivores that damage the plant (severity of common rust, incidence of black streak, incidence of stem rot, incidence of virus of corn streak, presence of fungal diseases, severity of tropical streak, etc.).
5. Rasgos Fenológicos, provides a list of characteristics that determine the life cycle of the plant (anthesis sedation interval, anthesis time, maturity, silk and ear leaf senescence).
6. Rasgos Fisiológicos, presents a list of characteristics regarding chemical and physical processes associated with the life of the plant (chlorophyll, glucose and proline content in the grain, osmotic adjustment, water index, etc.).

Table 3. Bean Ontology Metrics

Metric	Quantity
Axioms	8315
logical axioms	2490
Statement axioms	2131
Clases	1070
Object properties	2
Individuals	1056
Annotation properties	8
Subclasses	1434

7. Rasgos Morfológicos, is made up of a series of characteristics regarding the organs that make up the body of the plant (stem, leaf, roots, etc.).

– The Método class has measurement values that allow the crop to be evaluated and identify problems such as:

1. Count, a subclass that is made up of methods that allow the counting of some plant traits (number of plants damaged by busseola and plot, number of plants affected by the corn stunt complex per plot, corn weevils in a sample, etc.).
2. Calculation, is made up of a series of methods that allow the calculation of some of the plant traits (days to anthesis, senescence, ear, days to maturity, dry weights of the base, ear leaf senescence date, etc.).
3. Estimate, subclasses that determine an estimate of some corn traits (sedation date, maturity, etc.).
4. Measurement, is made up of a series of methods that allow measuring chemical and physical traits of the plant (sugar, sucrose, glucose, zinc, dry weight of the cob, etc.).

– Escala. Composed of 7 subclasses:

1. Time, allows setting the day, month, and years of some of the pre-established methods (estimation or calculation of certain plant traits).
2. Nominal, allows the measurement of specific physical traits of the plant (ear shape scale, leaf grain row arrangement, stem color, grain color, etc.).
3. Numeric, contains a series of numerical scales (percentages, ear position ratio, proportion scale, rank number, etc.).
4. Ordinal, is made up of a series of intervals that allow measuring the physical and chemical traits of the plant (senescence scale, colors of the cob, position of the ear, grain texture, standard germination, opening of the tassel, etc.).
5. Text, allows to relate the estimate of the presence of fungal diseases of corn.
6. Code.
7. Duration, serve as support for extra methods.

– Variable. It has more than 100 subclasses with modifiable data depending on the type of relationship between concepts.

3.2 Bean Ontology

It was developed to measure bean traits, measurement methods, and scales. SKOS is used for transitivity, labeling, acronym assignment, etc. It consists of 3 main classes (see Figure 3) called: 1) Escala, 2) Rasgos del Frijol común, and 3) Método. Table 3 shows the total metrics of the ontology. Each one is described below.

– Rasgos del frijol común. This class is made up of subclasses (see Figure 3b) that define physical and chemical characteristics of the bean plant:

1. Rasgos Agronómicos, is a list of physical characteristics of the plant (height , aspects, pods per plant, seed size, etc.).
2. Rasgos de Calidad, a list of characteristics related to the quality of the bean seed (iron, phosphorus, protein, Zinc, etc.).
3. Rasgos Bioquímicos; determine the chemical composition of the plant (presence of the DNA marker linked to the gene bc-3, bgm-1, BGYMV, etc.).

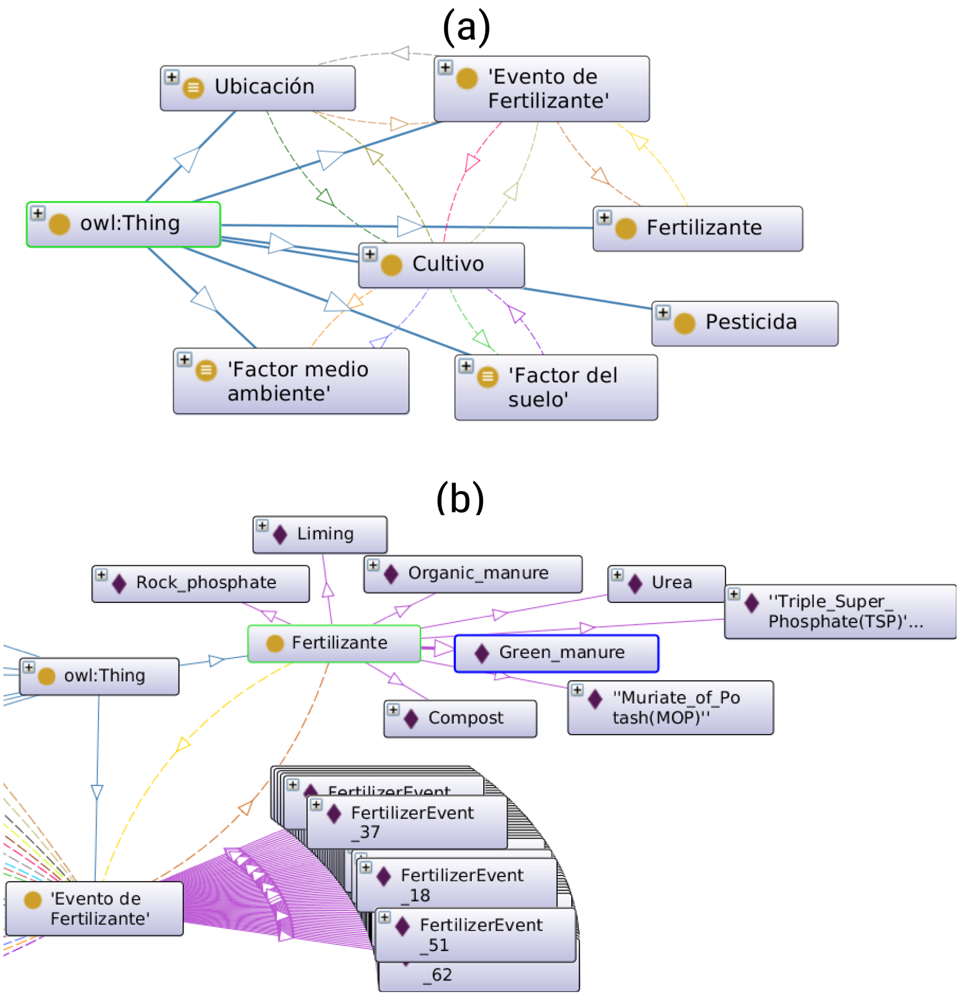

Fig. 4. General agricultural ontology. (a) Main Classes and (b) fertilizante class structure

4. Rasgos Bióticos, determine the negative impact of factors such as fungi, viruses and herbivores that damage the plant (stemworm, common mosaic necrosis virus, mealy leaf spot, etc.).

5. Rasgos Fenológicos, provides characteristics that determine the life cycle of the plant (days to flower, days to physiological maturity and growth stage).

6. Rasgos Fisiológicos, subclass that presents a list of characteristics of chemical and physical processes associated with the life of the plant (discrimination of carbon isotopes of the grain, loss of leaves, number of effective nodules in bush beans, content of seed calcium in the field, etc.).

7. Rasgos Morfológicos, is a series of characteristics that make up the body of the plant (sheath, leaf, roots, etc.).

– **Método**. This class is made up of more than 100 measurement methods that allow you to evaluate the crop and at the same time identify problems (agronomic method of water efficiency, basal length of the root, orientation of the pod peak, etc.).

ISSN 2007-9737

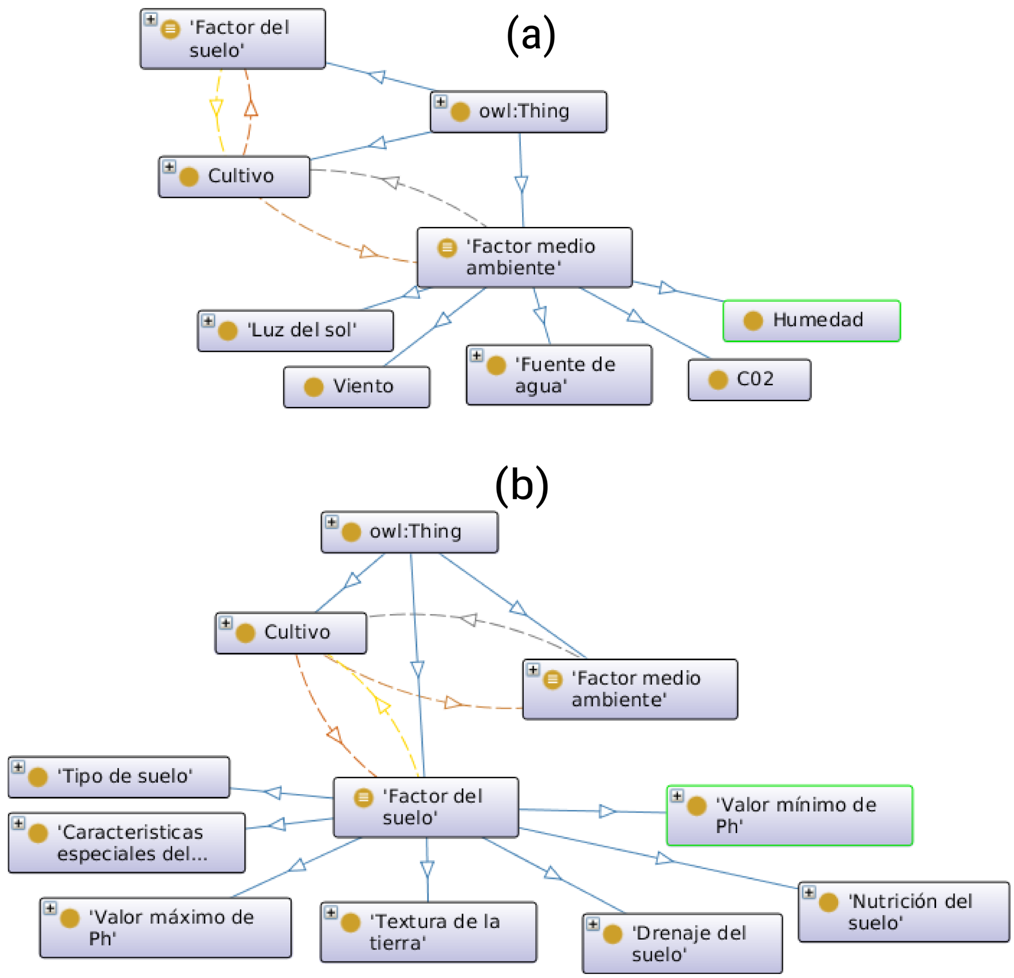

Fig. 5. General agricultural ontology. (a) subclass factor medio ambiente and (b) subclass Factor del suelo

– **Escala**. This class inherits from the Método class OJOJOJO. It is made up of more than 100 types of scales such as percentages, intervals from 1 to 9, evaluation of scales in categories, grams, among others.

3.3 General Agricultural Ontology

This ontology has local characteristics related to climate, culture, languages, and local plant varieties. In addition, it contains subdomains for different crops (see Figure 4), the most relevant ones are Fertilizante, Pesticida and Factor Medio Ambiente.

Table 4 shows the total metrics of the ontology. Some main classes are described as follows:

– Fertilizante. Class that defines organic or inorganic materials that supply essential nutrients for plant growth.

In addition, it contains instances with relevant information about fertilizers such as compost, limed fertilizer, organic fertilizer, green manure, etc.

Likewise, it defines a subclass called Evento de Fertilizante, which contains concepts such as symptom, quantity, time, growing problem, harvesting method, etc. (see Figure 4b).

Table 4. General Ontology metrics

Metric	Quantity
Axioms	5591
logical axioms	977
Statement axioms	2131
Classes	83
Object properties	191
Data properties	45
Individuals	657
Annotation properties	4
Subclasses	33

– Factor Medio Ambiente. This ontology addresses concepts organized in two classes:

1. Factor de medio ambiente (see Figure 5a), which provides information on external factors based on crops or farms, such as humidity, sunlight, wind, CO2 and water source.
2. Factor del Suelo, includes all necessary soil conditions related to crops (soil type, pH value, soil texture and characteristics, etc.).

– Pesticida. This ontology contains a list of organic and inorganic pesticides to be applied to crops.

– Factor del suelo. Class with all the required soil conditions for crop development (see Figure 5b), such as: soil type, maximum pH value, minimum pH value , etc.

For the development of the semantic resources mentioned above, an existing variety was reused. In the ontology of maize, more than 11,500 axioms, 1,103 classes and more than 2,700 subclasses were obtained.

Regarding the ontology of the bean, 1070 classes, 1434 subclasses and 1056 individuals. It should be noted that the subdomain ontology is still in development, so far 1070 classes, 1435 subclasses and 1056 individuals have been obtained.

4 Conclusions

The semantic web plays an important role in structuring knowledge; provides technologies and tools for the formalization of knowledge.

The semantic web has contributed to formalizing agricultural data obtained from soil sensors, drones, weather stations, etc.

On the other hand, despite the large amount of semantic resources available, little work has been done in Mexico to formalize agricultural knowledge.

The implementation of semantic resources would boost the development of applications for sustainable and precise cultivation. The Semantic Web in the agricultural sector contributes to this new stage of the countryside in Mexico and is manifested in aspects such as:

1. The openness of businessmen and technicians in the acceptance of these technologies.
2. The generational change.
3. The professionalization of the sector.
4. The social penetration of computer sciences.
5. The maturity of advanced business management applications (customer management, input management, production control, etc.) that have gone from being products only available to large corporations to be accessible by cooperatives and small and medium-sized agri-food companies.

References

1. **Ahsan, M., Motla, Y. H., Asim, M. (2014).** Knowledge modeling fore-agriculture using ontology. International Conference on Open Source Systems and Technologies, pp. 112–122. DOI: 10.1109/icosst.2014.7029330.
2. **Almadhoun, H. R., Abu-Naser, S. S. (2018).** Banana knowledge based system diagnosis and treatment. International Journal of Academic Pedagogical Research, Vol. 2, No. 7, pp. 1–11.
3. **Arenas, M., Ugarte, M. (2017).** Designing a query language for RDF: Marrying open and closed worlds. ACM Transactions on Database Systems, Vol. 42, No. 4, pp. 1–46. DOI: 10.1145/3129247.

4. **Berners-Lee, T., Hendler, J., Lassila, O. (2001).** The semantic web: A new form of web content that is meaningful to computers will unleash a revolution of new possibilities. ScientificAmerican, Vol. 284, pp. 34–43.

5. **Bonacin, R., Nabuco, O. F., Pierozzi-Junior, I. (2016).** Ontology models of the impacts of agriculture and climate changes on water resources: Scenarios on interoperability and information recovery. Future Generation Computer Systems, Vol. 54, pp. 423–434. DOI: 10.1016/j.future.2015.04.010.

6. **Cooper, L., Jaiswal, P. (2016).** The plant ontology: A tool for plant genomics, Vol. 1374. DOI: 10.1007/978-1-4939-3167-5_5.

7. **Drury, B., Fernandes, R., Moura, M. F., de-Andrade-Lopes, A. (2019).** A survey of semantic web technology for agriculture. Information Processing in Agriculture, Vol. 6, No. 4, pp. 487–501. DOI: 10.1016/j.inpa.2019.02.001.

8. **DuCharme, B. (2013).** Learning SPARQL, Second Edition. O'Reilly Media, Inc.

9. **Duruflé, H., Laporte, M. A., Matteis, L., Valette, L., Agbona, A., Agrama, H., Kumar, S., Beebe, S., Van-den-Bergh, I., Borja, F. N., Boukar, O., Chirwa, R., Crichton, R., Diers, B., Agarwal, D., Drabo, I., Eusebio, W., Nelson, R., Nelson, R., Arnaud, E. (2014).** The crop ontology: Improving the quality of 18 crop trait dictionaries for the breeding management system and adding new crops. Conference: General Research Meeting of the Generation Challenge Programme. DOI: 10.13140/2.1.2398.3363.

10. **Jonquet, C., Toulet, A., Arnaud, E., Aubin, S., Dzalé-Yeumo, E., Emonet, V., Graybeal, J., Laporte, M. A., Musen, M. A., Pesce, V., Larmande, P. (2018).** AgroPortal: A vocabulary and ontology repository for agronomy. Computers and Electronics in Agriculture, Vol. 144, pp. 126–143. DOI: 10.1016/j.compag.2017.10.012.

11. **Kaushik, N., Chatterjee, N. (2018).** Automatic relationship extraction from agricultural text for ontology construction. Information Processing in Agriculture, Vol. 5, No. 1, pp. 60–73. DOI: 10.1016/j.inpa.2017.11.003.

12. **Kawtrakul, A. (2012).** Ontology engineering and knowledge services for agriculture domain. Journal of Integrative Agriculture, Vol. 11, No. 5, pp. 741–751. DOI: 10.1016/s2095-3119(12)60063-x.

13. **Liang, A., Sini, M. (2007).** Mapping agrovoc and the chinese agricultural thesaurus: definitions, tolos, procederus. New Review of Hypermedia and Multimedia, pp. 51–62.

14. **Malik, N., Sharan, A. (2016).** Semantic web oriented framework for knowledge management in agriculture domain. International Journal of Web Applications, Vol. 8, No. 3, pp. 71–79.

15. **McGuinness, D. L., van-Harmelen, F. (2004).** OWL web ontology language overview. Technical report, W3C.

16. **Ngo, Q. H., Le-Khac, N. A., Kechadi, T. (2018).** Ontology based approach for precision agriculture, Vol. 11248. DOI: 10.1007/978-3-030-03014-8_15.

17. **Pesce, V., Maru, A., Archer, P., Malapela, T., Keizer, J. (2015).** Setting up a global linked data catalog of datasets for agriculture. Metadata and Semantics Research, Springer International Publishing, pp. 357–368.

18. **Saha, G. K. (2007).** Web ontology language (OWL) and semantic web. Ubiquity, Vol. 2007, No. September, pp. 1–1. DOI: 10.1145/1295289.1295290.

19. **Trendov, N. M., Varas, S., Zeng, M. (2019).** Tecnologías digitales en la agricultura y las zonas rurales.

Article received on 04/05/2023; accepted on 23/09/2023.
Corresponding author is German Rios-Toledo.