

# Natural Language Semantic Answering Applied to Medicinal Plant and Coronavirus

Alma Delia Cuevas-Rasgado<sup>1</sup>, Maricela Claudia Bravo-Contreras<sup>2</sup>,  
Franz Ludwig Lake-Moctezuma<sup>3</sup>, Adolfo Guzmán-Arenas<sup>3</sup>

<sup>1</sup> Universidad Autónoma del Estado de México,  
Computer Engineering, Estado de México,  
Mexico

<sup>2</sup> Universidad Autónoma Metropolitana,  
Systems Department, Ciudad de México,  
Mexico

<sup>3</sup> Instituto Politécnico Nacional,  
Centro de Investigación en Computación, Ciudad de México,  
Mexico

adcuevasr@uaemex.mx, mcbbc@azc.uam.mx,  
lakemoctezuma@gmail.com, aguzman@ieee.org

**Abstract.** A question answering system that receives as input a question in Spanish and returns the answer is presented. Preguntas y Respuestas {questions and answers} (PryRe) has two main components: 1) An information retrieval component that identifies the meaning of the question using its semantic properties. This component transforms the question into a triplet:  $R(C, V)$ , where  $R$  is the relation or link,  $C$  is the concept or main idea, and  $V$  is the value of the concept. Example: ¿Cuál es la hierba que mejora la digestión? {What is the herb that improves digestion?} becomes  $R(C, V) = \text{mejora (hierba, digestión)}$  {improves(herb, digestion)}. This component uses natural language processing modules; 2) a component that uses the triplet to carry out a query analysis on PryRe's ontology, to identify the answer, which in the example is Manzanilla {Chamomile}. This component performs the semantic identification of the question while traveling on parts of the ontology. Details of the PryRe system are given, as well as tests on herbalism and Coronavirus. It shows an acceptable accuracy (82%). Resources used in this work are (A) a notation used to describe ontologies, and (B) the deductive capability of PryRe.

**Keywords.** Semantic analysis, ontology, question-answering, knowledge retrieval, natural language processing.

## 1 Introduction

PryRe is a question-answering system that integrates Natural Language Processing (NLP), ontologies, and Information Retrieval (IR) methods. It interprets a given question in Spanish and finds an answer using as a source an ontology of the interest domain.

People who at some time need to ask specific questions about a knowledge domain for support inspired it.

The coronavirus pandemic was its target. Its deductive functionality PryRe (Sections 5.1 to 5.5) usually provides correct, nontrivial answers (82% accuracy). PryRe can be used with other ontologies.

Information is important in the Technology Age, and one way to obtain it is by questions in natural language.

Nevertheless, people usually do not express themselves rigorously when communicating, because the receiver is another person who provides the lacking information by using context and common sense. However, this does not apply with programs and computers. If the receiver is a

computer, then all the details should be explicitly described compare "John is a blonde engineer" with "The color of the hair of John is blond and his profession is engineer."

For this reason, ontologies, which are interpretations for computers, must have explicit and well-defined conceptualizations.

NLP research, an area of Artificial Intelligence, handles the automation of communication between human beings and "artificial agents." Parts of the NLP repertoire used in this paper are:

- Tagging (nouns, verbs, articles, adverbs, and named entities) to identify the elements of a sentence.
- Tools, such as WordNet [1], set of tags for probabilistic tagging based on the Expert Advisory Group on Language Engineering Standards (EAGLES)<sup>1</sup> group. Since WordNet does not provide disambiguation resources for prepositions in Spanish, a previously developed solution [2] for disambiguation of prepositions in Spanish is used.
- Ontologies [3], a type of knowledge base mainly developed in Knowledge Representation in Artificial -Intelligence, which represents a concept (usually described by a word or a set of words) according to its meaning. PryRe uses the Ontology Merging notation [4], a formal notation based on a local logic foundation that facilitates the representation of ontology concepts. Since it uses an XML-based syntax, ontology engineers find it easy to understand and use.
- Freeling [5], an open-source language analysis tool suite. It provides effective use of the dependency tree, easing the translation from Spanish phrases into triplets.

To evaluate its functionality and efficiency, PryRe uses the following:

- Coronavirus;
- Health information concerning the nervous system was used for experimentation. Specifically, the use of the Galphimia Glauca plant (growing in Mexico) was selected;

- Nutritional and medicinal aspects of fruits, vegetables, herbs, and seeds are incorporated, as well as their compositions: minerals, essential amino acids, vitamins, and other useful elements.

All examples are in Spanish, followed by their English translation in brackets {}, manually introduced.

The main contributions of this work are:

- The way in which PryRe splits the natural language question into portions and transforms these into a triplet  $R(C, V)$ , and its deductive capabilities, embodied in the search that PryRe performs on the ontology to provide the correct answer.

The organization of the paper is as follows. Section 2 presents related work and a comparison analysis of question-answering approaches. The methodology is explained in Section 3, while Section 4 contains a description about OM notation; the PryRe system is detailed in Section 5. Section 6 deals with an evaluation that employs documents about medicinal plants and Coronavirus. Finally, Section 7 contains the results, conclusions, and future works.

## 2 Related Work

Question-Answering systems have become very popular since the evolution of semantic models based on ontologies and Knowledge Graphs (KG). This section presents a review of related work, and a comparative analysis of solution approaches (see Table 1).

**Qanda** [6] is a question-answering system that integrates IR, Knowledge Representation (KR) and NLP methods into a hybrid algorithm whose input is a question in natural language. Qanda converts its input into a logical representation, a proposition enhanced by inferencing hierarchical relations of the variables involved. In these propositions, each variable is replaced by a keyword from the question. A small set of the most relevant documents containing the keywords are retrieved, ranked by its relevance. The closest answer (the

<sup>1</sup> <https://www.cs.upc.edu/~nlp/tools/parole-sp.html>

most common answer, according to the search) is returned. The main characteristic of Qanda is its domain independence, since it does not use knowledge bases. The disadvantage is the time consumed, because the approach requires the preprocessing, indexing, and ranking of the collection of documents to execute queries.

**AQUA** [7] is a question-answering system that integrates NLP, Logic, Ontologies, and IR methods. Its process model has four phases:

- User interaction. The user introduces a question in natural language and obtains a list of ranked answers.
- Question processing, which executes a NLP parser (an interpreter that uses the unification and resolution algorithms to find a logical proof of the query). It uses the AQUA lexical resource and the Ontology. The Question processing phase has a failure analysis system, and a module to classify and reformulate questions.
- Document processing, that extracts a set of paragraphs from a collection of documents by identifying the focus of the question.
- Answer processing, which produces answers by extracting passages from the documents, groups, and scores answers.

AQUA uses the AKT reference ontology to provide more information about each question. AQUA's main features are the translation of English questions into a logical form of Query Logic Language (QLL), and enhanced answers using the AKT ontology. The disadvantage is the translation of English questions into the QLL form.

**A multi-agent question-answering system** [8] uses students to support learners in collaborative environments. A set of agents forward the question to students. Their responses are analyzed using extracted documents from course materials and other resources. This analysis verifies and ranks the responses. All questions and answers are stored, verified, and reused. The main features of this work are:

- It is an agent-based system; it supports collaborative learning, and it uses answers verification. Its disadvantage is the lack of a logic-based formal language for query processing and knowledge representation,

which prevents the use of reasoners to obtain more information.

**A Machine Learning (ML) question-answering system** [9] is trained with a set of question-answers pairs from the WebQuestions dataset. It generates answers using Freebase, a general fact knowledge base represented by N-triples in Resource Description File (RDF). The system learns low-dimensional vector embeddings of words that appear in questions and produces a joint embedding space where questions and answers are close. In doing so, it takes advantage of the graph-based representation of knowledge to obtain answer paths and subgraph representations. Its main features are: a vector space model to represent pairs of questions and answers and the use of graph knowledge representation to obtain more information. Its disadvantages are: the Freebase API has been shut down, and its training depends on the use of a benchmark of paired questions and answers.

**Aqqu** [10] is a question-answering system that uses the Freebase API. It identifies all entities from the knowledge base that matches part of the question. Then, based on tree templates, it generates a set of SPARQL Protocol and RDF Query Language (SPARQL) query candidates, from which the answers are obtained. Its main feature is its translation of a natural language question to a SPARQL query. Its disadvantages are: the Freebase API has been shut down, and Aqqu depends on the use of a benchmark of paired questions and answers for training.

**Convolutional Neural Networks (CNN)** are used in [11], a question-answering system that does inference on the knowledge base (entity linking and relation extraction), and then refines the answer. For relation extraction, it relies on a Multi-Channel Convolutional Neural Network (MCCNN), which uses one channel for syntactic, and the other for sentential information.

For answer refinement, it uses Freebase to retrieve candidate answers, and validates answers using Wikipedia. Its main feature is the use of a MCCNN. Its disadvantages are: the Freebase API has been shut down, and its training depends on the use of a benchmark of paired questions and answers.

**Table 1.** Comparison of related work

Question-Answering Approach	Dataset and additional resources	Knowledge Base Representation Language
Hybrid algorithm based on IR, KR and NLP [6].	Text REtrieval Conference (TREC) Question Answering collection.	Formal logic representation of the query.
NLP, Logic, Ontologies, and Information Retrieval methods [7].	Documents retrieved from the Web by using the search engine Google. AKT reference ontology.	Query Logic Language (QLL).
Multi-agent system that coordinates the reception and forwarding of questions. It uses IR and NLP [8].	Course materials, a collection of Frequently Asked Questions (FAQ) and the learners' answers.	There is not a formal language representation of questions and answers.
Learning vector embeddings of words appearing in questions and answers [9].	WebQuestions based on Freebase, using 5,810 question answer pairs.	N-triples RDF
Template matching and relation matching [10].	Freebase 917 and WebQuestions.	RDF and SPARQL
Learning approach based on MCCNN. Use of dependency tree patterns to decompose questions [11].	Freebase and Wikipedia to validate the results.	RDF and SPARQL
Template-based over a knowledge base using a learning approach [12].	ClueWeb09-FACC1, a corpus of 500M Web pages annotated with Freebase	RDF and SPARQL
Template-based over the SNOMED KG [13].	Uses SNOMED KG. WordNet for word similarity.	SNOMED query templates

**QUINT** [12] is a question-answering approach that implements a learning approach to automatically generate question-query templates and manage complex questions.

For experimentation, the authors used Freebase and WebQuestions benchmarks. The general process is divided into two: template generation and question-answering.

The main difference with previous efforts (and its main feature) is the fully automatic generation of question templates that link a question to a triplet pattern query over a KG.

**A question-answering system for SNOMED** medical ontology [13] uses a question-template library based on ontology definitions. Its main feature is the use of a template matching inference method combined with semantic similarity.

Considering the different solution approaches analyzed, it can be observed that some of them strongly depend on the use of pre-configured pairs of question-answering benchmarks to train a learning model. Additionally, the question-answer benchmarks use a knowledge base that may no longer be publicly available. Two questions arise: is it feasible to offer an open knowledge base about any topic? Is it possible to build question-answer approaches that do not require a training stage using previously validated benchmarks?

Template-based approaches have shown good results, some based on previously defined template libraries and others, more advanced, generate question templates using ML methods.

PryRe represents knowledge using KG, identifies and generates templates using graphs, generates answers by identifying and mapping question types, provides a friendly NL interface, translates the NL query into a logic query (a triplet), answers that query using a simple inference engine that exploits the laws of equivalence of inference logic.

### 3 Methodology

In this section, the research and development methodology consisting of five stages described in Figure 1.

**Descriptions about medicinal plants and Coronavirus:** First, data from plants related to the central nervous system was obtained [14].

**Manual analysis:** From them, the concepts, implicit and explicit relationship, classifications, and partitions are identified, to obtain an idea about how to organize the ontology.

**Building the ontology:** To have an ontology that contains information about a specific domain, relevant aspects must be decided. First, it is necessary to determine the formal language in which the ontology will be implemented.

In this area, there are several methods to represent ontologies such as Protegè [15]. A complete study of this area is presented in [16], where an analysis of different recent tools to visualize ontologies is given, as well as its main features. For the representation of the ontology, the OM notation is used. For the manual design, a

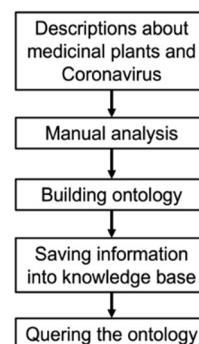


Fig. 1. PryRe method

graphic ontology builder, OM Edit [17] is presented, a graphical editing tool through which an ontology can be implemented without having to deal with the strict syntax of the language, it automatically generates code in OM notation [17]. This language is supported with tools to edit and verify the conceptualizations that will be included. It was decided to use the OM notation, which has a formal basis based on local logic.

**Saving the information in the knowledge base:** There are two ways to save the ontology: 1) graphically, creating a JPG file, or 2) saving the OM notation in an ONT file (a text file).

**Querying the ontology:** Using the ontology in OM notation, allows different types of searches to be performed, considering several scenarios and domains. A tag-based language (XML) was used to implement OM notation. Thus, the query algorithms using tree representations and making use of the semantic relationships described in the language formalization were built.

In addition, a method of analysis and translation of the input questions made in natural language into a representation in triplets  $R(C, V)$  was developed.

This process is simple because the question is typed in natural language. For example, *¿Cuál es la relación entre Organismo y Planta Vascolar?* {*What is the relationship between Organism and Vascolar plant?*}. **PryRe** converts the main objective of the search into a triplet  $R(C, V)$ .  $R$  represents the name of the relationship or link,  $C$  is the concept where the relationship starts, and  $V$  is

the concept where the relationship finishes. In this example, PryRe returns the triplet: subset (Planta vascular, organismo), R=subset, C=Planta vascular and V=organismo. The **PryRe** answer is: *Vascular plant es un subconjunto de Organismo {Vascular plant is a subset of Organism}*.

At the end, a whole set of test cases and scenarios was designed to evaluate both the efficiency of the algorithm and the responses obtained.

## 4 Ontology Construction

This section describes OM notation, the language selected for the implementation of the ontologies. First, the formal foundations of OM notation are described, and then its implementation is given.

### 4.1 Description of the OM Notation

The Ontology Merge (OM) notation was introduced in 2010 by [4] because of [18]. The purpose of the OM notation is to design ontologies with concepts and relationships described in a high degree of detail. These descriptions help the question-answering process since the OM notation can represent nested concepts, synonyms, implicit and explicit relationships (a relationship is a link that connects the concepts among themselves).

Some of types of relationships provided by OM notation are explained below (these appear between <> symbols, for example: <subset>, <word>) and shown in Figures 11 to 14:

- 1 **Implicit relations.** They represent the hierarchical structure of the ontology. The concepts in an ontology are classified into sets and subsets, parts of, members of and type of. For instance, the concept; *Girasol {Sunflower}* is a subset (<subset>) of *Flor {Flower}* but it is possible that *Girasol {Sunflower}* is a type (<type>) of *Flor {Flower}*. Referring to linguistics, the *subset* is known as Hyponym. In the same order *Pistilo {Pistil}* is a part of the *Girasol {Sunflower}*. The *Pistilo {Pistil}* is known as Meronym and *Girasol {Sunflower}* is known as its Holonym.

<sup>2</sup> <http://w3.org/OWL/>

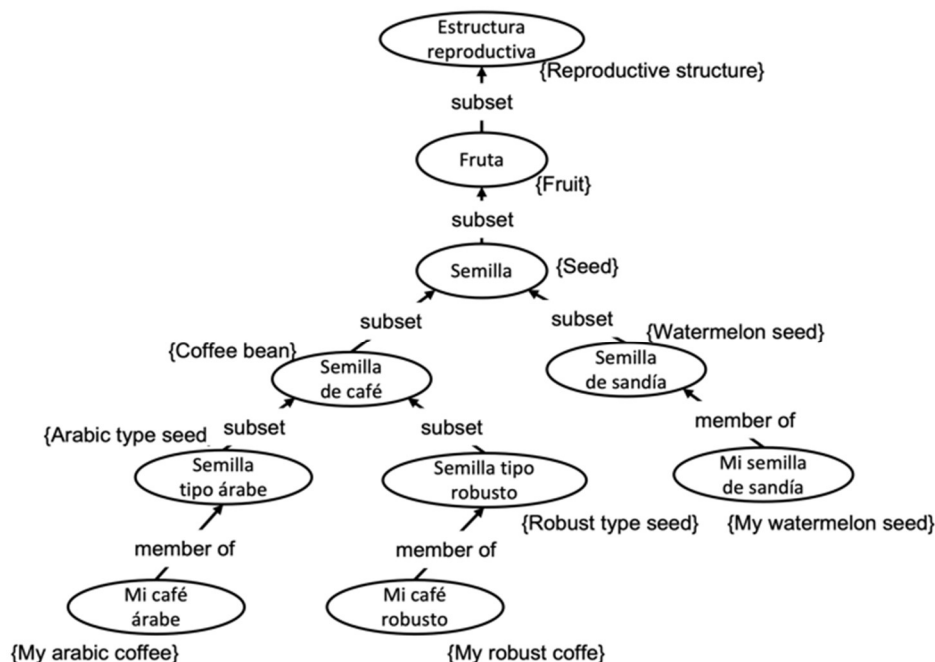
- 2 **Synonyms Concepts.** Many concepts have synonyms. In reference to [19], the OM notation represents synonyms with the tag <word>.
- 3 **Explicit Relations.** These types of relations increase the meaning of the concepts. For example, it is known that *Pistil* is a part of *Girasol {Sunflower}* but to know what tamaño {size}, grosor {thickness}, and textura {texture} the *Pistil* has, the <relation> tag is used.
- 4 **Definition.** It is the set of words that describes the concept, taken from dictionaries and expressed in natural language, its tag is <gloss>.

### 4.2 OM Notation and the Ontology Languages

OM notation, like Ontology Web Language (OWL)<sup>2</sup> are structured languages, using tags as in XML. These structures allow their interpretation by computers, as well as by people.

OM notation supports:

- 1 Conjunctions of classes and cardinality (arity) relations.
- 2 Simplicity in categorizing (object taxonomies) and relationships.
- 3 Greater expressiveness using <relation>, <word> and <gloss> that define the concept in detail, where the relationship can be both a concept and a relation. They express its meaning, including logical properties such as transitivity, symmetry, and inheritance, as OWL and Description Logic (DL).
- 4 Linking several ontologies at the same time, merging them (in a binary way). This is the degree of freedom that OM notation provides but guaranteeing the consistency of ontologies.



**Fig. 2.** In the ontology some instances are Mi Café Árabe {My Arabic coffee}, Mi café Robusto {My Robust coffee} and Mi semilla de Sandía {My Watermelon seed} and some types are Semilla Tipo Árabe {Arabic type seed}, Semilla tipo Robusto {Robust type seed}. The link member of Mi café Robusto {My Robust coffee} member of Semilla tipo Robusto {Robust type seed} means that the first is an instance, object or example of the second

In OWL 2 the OWL ontology languages have been improved with more expressiveness, extensions on data types and annotation capacity, simple metamodeling, among others. The OM notation has those capabilities, too, except metamodeling, and is simple to use.

The names of the concepts in this article are their usual names. However, they can be changed by URL, or memory addresses. All concepts have one address, as in OWL. For example, [20] shows the large number of nodes that the Facebook Ontology can have. Google, Instagram, and Facebook find relationships between users and events. Their challenge is to take advantage of the data found in its pages and combine them with other more structured databases to strengthen its knowledge. For this they use URL addresses. Obviously, the idea is to have control of links and a large amount of structured data. Facebook has the resources to store 50 million entities (primary

nodes), Microsoft 2 billion entities, Google 1 billion, eBay around 100 million, IBM more than 100 million entities.

Our idea is to build small ontology fragments of descriptions of an event and a user. It is small because it is formed by a set of concepts that are born from a theme, broken down into triplets (relationship, object, value) and merge these small ontologies until obtaining a large common ontology as in [18].

The work of OM notation is to represent useful information, describing the concepts found. For instance, the information that a biography, or descriptive or technical documents contain. It permits merging such information locally and uses it. It is not intended to represent figurative language: poetry, stories, or political speeches. OM notation uses a set of labels that have a specific function. There is no graphical user interface like Protégé [15], that allows us to

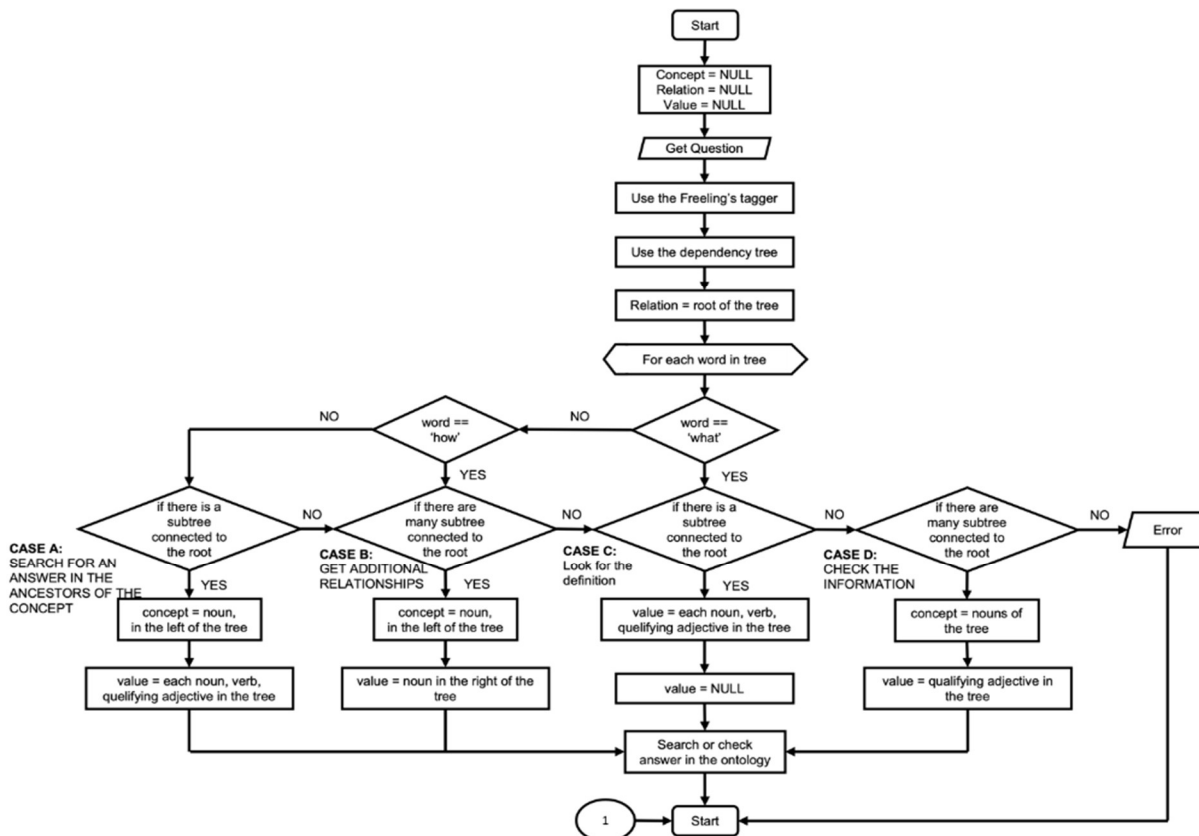


Fig. 3. The *PryRe* system (first part), showing how it creates the triplet in cases A to D

visualize the structure of the ontology, restrictions, and classifications of each class or instance.

All answers of *PryRe* are concrete responses, except when the query includes a concept definition taken from dictionaries.

## 5 Design of *PryRe*

The general purpose of *PryRe* is to interpret a question in natural language and generate a response. Figure 3 shows in detail the control flow of the *PryRe* question interpreter. *PryRe* first receives a question in natural language, then performs a syntactic and semantic analysis of the question, labels the grammatical elements of the question using a PoS tagger, and integrates a dependency tree with the elements of the question.

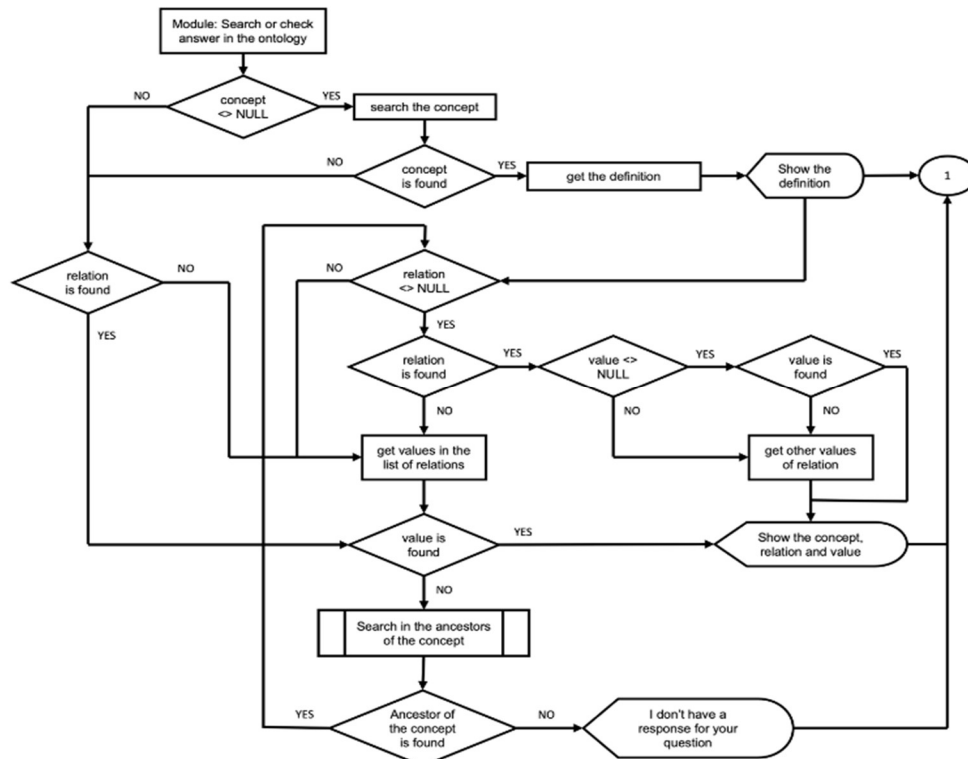
Those elements are identified as Concept, Relationship, and Value to obtain a triplet

representation of the question. Once the triplet is obtained, *PryRe* identifies the type of question and obtains a response according to the following cases:

- Search for an answer in the ancestors of the concept.
- Obtain additional relationships.
- Look for the definition.
- Check the information.
- Search relation and value in all the ontology.

The second part of *PryRe* is shown in Figure 4: finding the possible answer (case E) and looking for the triplet in the ontology. The answer is produced with two precisions: low or high, as explained below.





**Fig. 4.** The PryRe system, second part. Case E: Search for relation and value in all the ontology. This part of PryRe the case E is applied if the answer is not found inside of ancestors of the Concept. The recursion of the algorithm happens when PryRe needs to look for the same triplet in the whole ontology and applies the whole process to whatever it found related to this triplet. The feedback going from the “YES” exit of the bottom diamond “relation <=> null” in this figure, and the box “Search in the ancestors of the concept”, show the recursion of PryRe

### 5.1 Case A. Search for an Answer in the Ancestors of the Concept

This case occurs when searching for the answer using the inheritance relationships of the concept. For this, three subcases are considered:

- a. Subcase A.1: when the concept is found in the ontology, but the relationship or the value is not found. For instance, the triplet *forma (Galphimia Glauca, arbusto verde) {form (Galphimia Glauca, green bush)}*, has low possibilities of precision when searching among its ancestors. In this example, the relationship *form* and *the green bush* is searched. If *Plant* is an ancestor of *Galphimia Glauca*, it may have a relationship: *form*, since *Galphimia* is a *bush* and a *bush* is a *plant*, therefore the triplet *form (Plant, green bush)*

represents a closer response, semantically. The accuracy of this example is:  $C = 1, R = 0, V = 0$ , it is equal to:  $(1 + 0 + 0) / 3 = 0.3$ , it would change to  $C = 0.5$  because it was found in an ancestor. The ancestors are semantically more general than the concept,  $R = 1, V = 1$  it is equal to:  $(0.5 + 1 + 1) / 3 = 0.8$ . The result ranges from 0 to 1, the worst is 0 and the best is 1.

- b. Subcase A.2: when the concept is found in the ontology, but the relationships are not found in the concept, and it is not among the synonyms of the relationship (subcase E.1). In this case, *PryRe* searches for the most common relationship of the concept. For example, consider the triplet: *forma (Galphimia Glauca, arbusto verde) {form (Galphimia Glauca, green bush)}* whose most common relationship is:

orden(*Galphimia Glauca*, *Polygalales*) {order (*Galphimia Glauca*, *Polygalales*)}. The result of this example is low: C = 1, R = 0, V = 0 therefore:  $(1.0 + 0 + 0) / 3 = 0.3$ .

- c. Subcase A.3: when the relationship was not found in any of the ancestors of C. Therefore, the answer is: *No tengo respuesta para su pregunta* {I have no answer for your question}. Result: C = 0, R = 0, V = 0 has:  $(0 + 0 + 0) / 3 = 0.0$ .

## 5.2 Case B. Get Additional Relationship

- a. Case B.1: when *PryRe* searches for the answer using the implicit relationships *subset* and *type* of the concept. Recalling that the *subset* is known as Hyponym, while *type* is the Hyperonym. For example, consider the question *¿Qué animales salvajes portan coronavirus?* {What wild animals carry coronaviruses?}. The resulting dependency tree of this question is shown in Figure 5, which is represented with the triplet: *portan (animales salvajes, coronavirus)* {carry (wild animals, coronaviruses)}. For this question *PryRe* will search for the concept: *animales salvajes* {wild animals}, for the relation: *portan* {carry} and for the value: *coronavirus*. Then *PryRe* will search using the implicit relationships *subset* and *type* of the concept: *wild animals*. The resulting relationships are *subset (animales salvajes, murciélago)* {subset (wild animals, bat)}, and *type (animales salvajes, rata)* {type (wild animals, rat)}. Therefore, the answer for this question will be: *murciélago, rata* {bat, rat}. That is, all *subset* and *type* relationships that are contained in the concept: *animales salvajes* {wild animals} will be displayed. The result of this example is: C = 1, R = 1, V = 1 therefore:  $(1 + 1 + 1) / 3 = 1$ .
- b. Case B.2: This case happens when *PryRe* searches for the answer using the explicit relation and *partition* of the concept. For example, consider the question: *¿Cuál es la forma y la estructura del Coronavirus?* {What is the shape and structure of Coronavirus?}. The question word is: *Cuál* {What}, the root is: *ser* {is}, the nouns connected (relation label) with the conjunction are: *forma* {form} and *estructura* {structure}. *PryRe* will search in the

ontology for the *Coronavirus* concept and the relation between *forma* {form} and *estructura* {structure} (they are explicit relations in the ontology). Then, it will obtain the values of these two relationships. Figure 6 shows the code (A) and dependency tree (B) of the question. The triplet is: *forma(Coronavirus, null)* {form(Coronavirus, null)}. The precision is: C=1, R=1, V=1 therefore:  $(1 + 1 + 1) / 3 = 1$ . For form, its precision is: C=1, R=1, V=1 although *part* was found instead of *structure* (as synonym), rendering a precision of 1, and the overall result will be  $(1 + 1) / 2 = 1$ .

## 5.3 Case C. Look for the Definition

This case occurs when *PryRe* searches for the definition of a given concept. Given a triplet, if the concept is found in the ontology, *PryRe* searches in all the relationships of the concept using the following keywords: *significa* {it means} (and its different lemmas) *es* {it is}, *tener* {to have}, *entender* {to understand}, using the question word: *Cuál* {What} and the value: null.

For instance, consider the following questions:

- a. *¿Qué significa Galphimia Glauca?* {What does Galphimia Glauca mean?} The triplet representation that corresponds to this question is: *significa (Galphimia Glauca, null)* {means (Galphimia Glauca, null)}.
- b. *¿Cuál es el significado de Galphimia Glauca?* {What is the meaning of Galphimia Glauca?} The triplet representation that corresponds to this question is: *significa (Galphimia Glauca, null)* {meaning has (Galphimia Glauca, null)}.
- c. *¿Qué se entiende por Gauphimia Glauca?* {What is meant by Gauphimia Glauca?}. The triplet representation that corresponds to this question is: *entendido por (Gauphimia Glauca, null)* {understand by (Gauphimia Glauca, null)}.

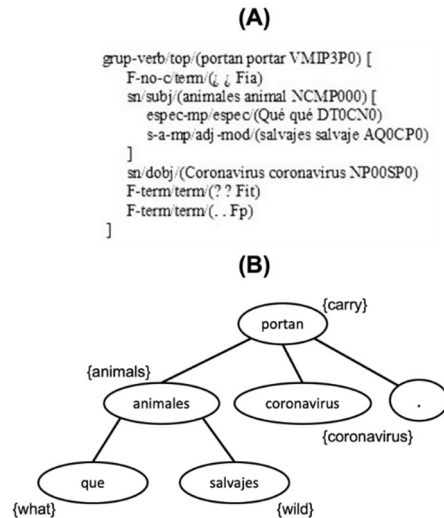
In the three cases above, the relationship indicates the *type* for search. Another way to answer the same questions is by obtaining the definition in the <gloss> tag. The answer for the questions is: *Arbusto o planta medicinal mexicana para el sistema nervioso central. También llamado en el Bajío mexicano como: "ojo de gallina"* {Shrub

or Mexican medicinal plant for the central nervous system. Also known in the Mexican Central Area as: “eye of the hen”. The accuracy of this example is:  $C = 1, R = 1, V = 0, (1 + 1 + 0) / 3 = 0.6$ . However, the question word is a key to obtain the precise answer. In this case, it is *Qué* {*What*} and complements the word in the relationship *significa* {*means*}. The correct answer is found in the definition. Therefore, the value is  $V = 1$ . The result of this example is:  $C = 1, R = 1, V = 1, (1 + 1 + 1) / 3 = 1$ .

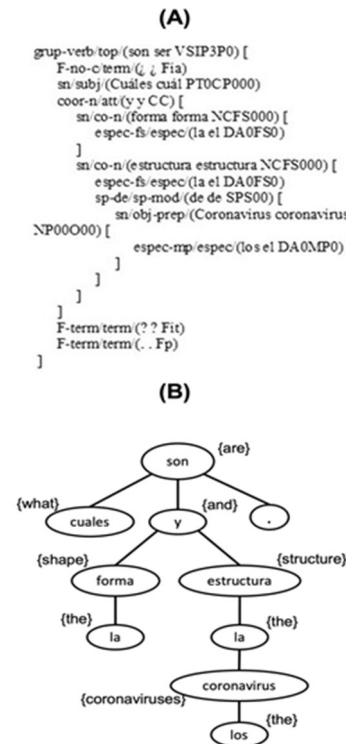
**5.4 Case D. Check the Information**

This case occurs when *PryRe* searches for the information verifying that the triplet is found. There are two possibilities:

- a. Subcase D.1. If the concept, relationship, and value are found (which represents the best case), only the question needs to be confirmed. For instance, consider the question *¿Es Galphimia Glauca un arbusto verde?* {*Is Galphimia Glauca a green shrub?*} The triplet corresponding to this question is *forma (Galphimia Glauca, arbusto verde)* {*shape (Galphimia Glauca, green bush)*}. The result of this example is  $C = 1, R = 1, V = 1$ . Then accuracy =  $(1 + 1 + 1) / 3 = 1$ .
- b. Subcase D.2. If the concept and its relationship are found, but the value is not. For instance, consider the question *¿Cuál es la forma de la Galphimia Glauca?* {*What is the shape of Galphimia Glauca?*}, Triplet corresponding to this question: *forma (Galphimia Glauca, null)* {*shape (Galphimia Glauca, null)*}. In this example, unlike subcase A.1, the *forma* {*shape*} relation is not obtained from the definition, or it would be difficult to find it since not all the definitions cite the shapes of the defined objects. Therefore, the value of *V* is null, and the value of *V* will be found in the set of values of the shape relation. The answer is *arbusto verde* {*green bush*}. The triplet representation of this question is: *forma(Galphimia Glauca, arbusto verde)* {*shape(Galphimia Glauca, green bush)*}. The result of this example is:  $C = 1, R = 1, V = 0, (1 + 1 + 0) / 3 = 0.6$ .



**Fig. 5.** Dependency tree of the question *¿Qué animales salvajes portan coronavirus?* {*What wild animals carry coronavirus?*}. The dependency tree is represented in a code (A). This code is explained in [5], but for simplicity, the graphical mode is used (B)



**Fig. 6.** Dependency tree of the question: *What is the shape and structure of Coronavirus?*

### 5.5 Case E. Search Relation and Value in All the Ontology

This case occurs when *PryRe* searches for the relation and value of a concept. The following subcases can occur:

- a. Subcase E.1. If the concept is found in the ontology, the relation is not found, and the value is found, *PryRe* searches for a relationship that is synonymous with the missing relationship. For example, consider the triplet: *forma (hoja, redonda) {shape (leaf, round)}* a synonym is: *forma (hoja, circular) {form (leaf, circular)}*. Synonyms can be found in the *<word>* tag only if the relationship is also a concept (it is one of the advantages of OM notation). For this case, the result is:  $C = 1, R = 1, V = 1, (1 + 1 + 1) / 3 = 1$ .
- b. Subcase E.2. Based on the previous case, if a synonym is not found, the value is searched in all the relations of the concept, but it can be erroneous. For example, *forma (hoja, redonda) {form (leaf, round)}* and *forma (tallo, redondo) {stem (leaf, round)}*. This solution is ambiguous since the shape of the bush is round, but the stem of the bush is also round. Therefore, the result is:  $C = 1, R = 0, V = 1, (1 + 0 + 1) / 3 = 0.6$ .
- c. Subcase E.3. If the concept is not found in the ontology, but the relation and value are found in the ontology, *PryRe* will find a concept in the ontology that satisfies the relation and the value of the same triplet. However, there is a risk to find a more general or semantically different concept. For example, consider the triplet: *forma (Galphimia Glauca, arbusto verde) {shape (Galphimia Glauca, green bush)}*. If *Galphimia Glauca* is not found in the ontology, the triplet may be misinterpreted as: *la planta es verde {plant is green}* (too general) *forma (planta, arbusto verde) {shape (plant, green bush)}*. The precision is:  $C = 0, R = 1, V = 1, (0 + 1 + 1) / 3 = 0.6$ . In this case, although the answer has been found, there is not a guarantee that the result will be semantically close to the correct answer. The result would be: 0.6.

## 6 Evaluation

To evaluate *PryRe*, this section describes experiments from diverse domains: medicinal plants, and Coronavirus. For each of the tests, the question, the tagging of the question, the dependency tree, the integration of the triplet, the search for the triplet in the ontology, and the correct answer are shown.

### 6.1 Test 1 Meaning

*¿Que significa Citrus Aurantium? {What does Citrus Aurantium mean?}*, after Freeling's labeling, Table 2 shows the tags assigned.

Figure 7 shows the dependency tree generated.

The next step is to choose nouns, verbs, and adjectives. Only one sub-tree remains (Figure 8).

#### Triplet Integration

For each sub-tree of the dependency tree, the root is taken as the relation, that is, the link to be searched in the ontology. The first noun is taken as the main concept: *Citrus Aurantium*, since there are no more nouns, the value is null. Therefore, the triplet is as follows:

*Significa (Citrus Aurantium, null)*  
*{means (Citrus Aurantium, null)}*

Now *PryRe* will look for the answer in the ontology.

There are two ways to find it:

- a) In the ontology, the relationship *significa {means}* of the concept *Citrus Aurantium* points to the value *Naranja amarga {Bitter orange}* because the relationship *significa {means or meaning}*, would be found (the lemma is taken from the word), then this value is connected to the relation, see Figure 9 (A). *PryRe* would apply case B.
- b) Another option is: that the value of V is found in its gloss or description. In this case, only the value of the *<gloss>* tag is obtained in the *Citrus Aurantium* concept, see Figure 9 (B). In this case, *PryRe* would apply case D.

### 6.2 Test 2 Explanation

*¿Cómo se conocen los Espasmolíticos?*

(A)  
 grup-verb/top/(significa significar VMIP3S0) [  
 F-no-c/term/(¿ ¿ Fia)  
 sn/subj/(Qué qué PTOCN000)  
 sn/dobj/(Citrus\_Aurantium citrus\_aurantium NP00SP0)  
 F-term/term/(? ? Fit)  
 F-term/term/(. . Fp)  
 ]

sadv/cc/(Cómo cómo PT000000)  
 morfema-verbal/es/(se se P00CN000)  
 sn/dobj/(Espasmolíticos espasmolíticos NP00O00) [  
 espec-mp/espec/(los el DA0MP0)  
 ]  
 F-term/term/(? ? Fit)  
 F-term/term/(. . Fp)  
 ]

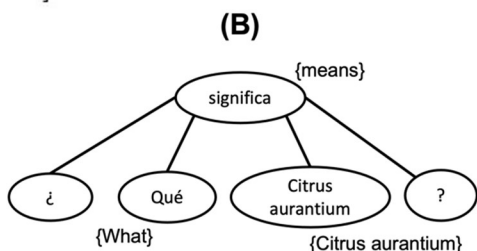


Fig. 7. Dependency tree of the question: ¿Qué significa Citrus Aurantium? {What does Citrus Aurantium mean?}

Table 2. Words and its tags accord to EAGLES

Word	Tag
¿	Fia
Qué {What}	PTOCN000
Significa {mean}	VMIP3S0
Citrus_Aurantium {Citrus Aurantium}	NP00SP0
?	Fit

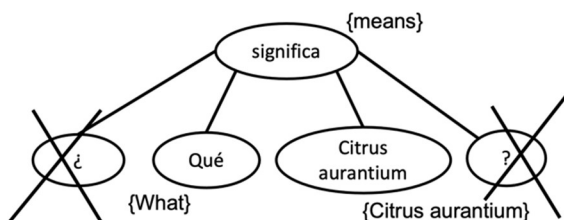


Fig. 8. Question marks are removed from the tree

{How are Spasmolytics known?}

The labels for each word are shown in table 3. The dependency tree of the question is as follows:

grup-verb/top/(conocen conocer VMIP3P0) [  
 F-no-c/term/(¿ ¿ Fia)

For now, no more details of the graphical representation of the dependency tree will be given. The triplet that corresponds to the question is *conocido como* (*Espasmolíticos*, null) {*known as* (*Spasmolytic*, null)}, as seen in Figure 10.

In ontology there are two ways to find the answer: (A) with the link: *conocido como* {*known as*} and (B) with the link: *mejor conocido como* {*best known as*}. Both answers are valid. PryRe has applied the Case B (section 5.2).

### 6.3 Other Ways to Find Answer in the Ontology

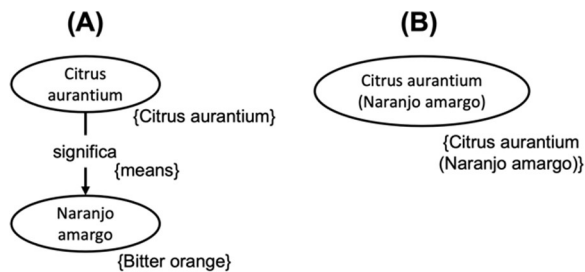
Several examples of inferences and queries using Medicinal Plants are explained below. Details about the inference process can be found in [21].

**Test 3: implicit relationship:** *hoja suave* {*soft leaf*} is a part of *Galphimia Glauca*, the set of *Galphimia Glauca* has a part that is *hoja suave* {*soft leaf*} that is also type of *hoja* {*leaf*}. See Figure 11.

**Test 4: Redundancy correction in triplet:** If *Galphimia Glauca* is a subset of *planta medicinal* {*medicinal plant*} and this is a subset of *planta vascular* {*vascular plant*} then *Galphimia Glauca* is a subset of *planta vascular* {*vascular plant*}. It is not necessary to set a relation between *Galphimia Glauca* and *Planta vascular* {*vascular plant*} because shifting through the trajectories: by the route of *planta medicinal* {*medicinal plant*}, *Planta vascular* {*vascular plant*} ancestor is arrived at (see Figure 12).

**Test 5: Explicit relationship in triplet:** *Galphimia Glauca* is shaped like *Arbusto verde* {*Green Bush*} and blooms in *tiempos lluviosos* {*rainy times*}. The explicit relationship gives more semantics to the concepts. (See Figure 13).

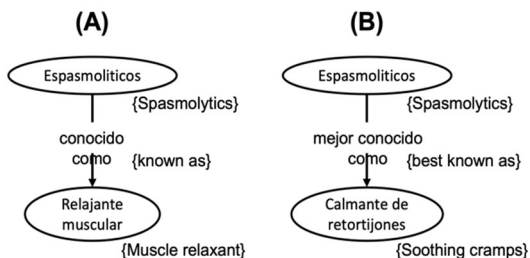
**Test 6: Synonyms:** *Galphimia Glauca* has the following synonyms enclosed in parentheses



**Fig. 9.** The Concept Citrus Aurantium and its definition in (A) are found (means points to bitter orange), and in (B), this is in the gloss (Naranja amargo {bitter orange}) inside of Citrus aurantium

**Table 3.** Words and its tags, accord to EAGLES [5]

Word	Tag
Cómo {How}	PT000000
Se {are}	P00CN000
Conocen {known}	VMIP3P0
Los {the}	DA0MP0
Espasmolíticos {Spasmolytics}	NP00000
?	Fit
.	Fp



**Fig. 10.** Question marks are removed from the tree

(*Thyallis glauca*, *Galphimia gracilis*, *Galphimia humboldtiana*, *Galphimia multicaulis*) (see Figure 14).

**Test 7: Requested Information is not part of the response in the ontology.** In the question: *¿Las ratas de bambú portan coronavirus? {Do bamboo rats carry coronavirus?}* PryRe says No, because Bamboo rat is not part of *animal salvaje {wild animal}* whose carries Coronavirus. *Rata de bambú {bamboo rat}* is subset of *rata {rat}* that is a subset of Animal (above of *animal salvaje {wild animal}*) in Figure 15. This *rata de bambú {bamboo rat}* does not appear in the *animal salvaje {wild animal}* concept, for this reason, the answer is: No, although by common sense it is true.

**6.4 Examples Related to Coronavirus**

In the early days of the outbreak, China quickly shared its understanding of the Coronavirus with the world through the World Health Organization. (WHO)<sup>3</sup>. Tested and tempered by the viral epidemic such as the SARS epidemic, professionals, and experts in the first line focused in the "epicenter" Wuhan, China. They decided to share their invaluable experiences and lessons from the current outbreak as well as during their internships and experiences in China and various countries making it possible to edit The Coronavirus Prevention Handbook [22].

Pryre has been tested with 100 questions about coronavirus obtained from this book [22]. Only 16 questions are shown in this section, all of them appear here<sup>4</sup>.

**1 Question:** *¿Qué son los virus asociados con las infecciones respiratorias? {What are viruses associated with respiratory infections?}* Case applied: C.

**Book answer:** "Viruses associated with respiratory infections" refer to viruses that invade and proliferate in the epithelial cells of the respiratory tract that can cause respiratory and systemic symptoms.

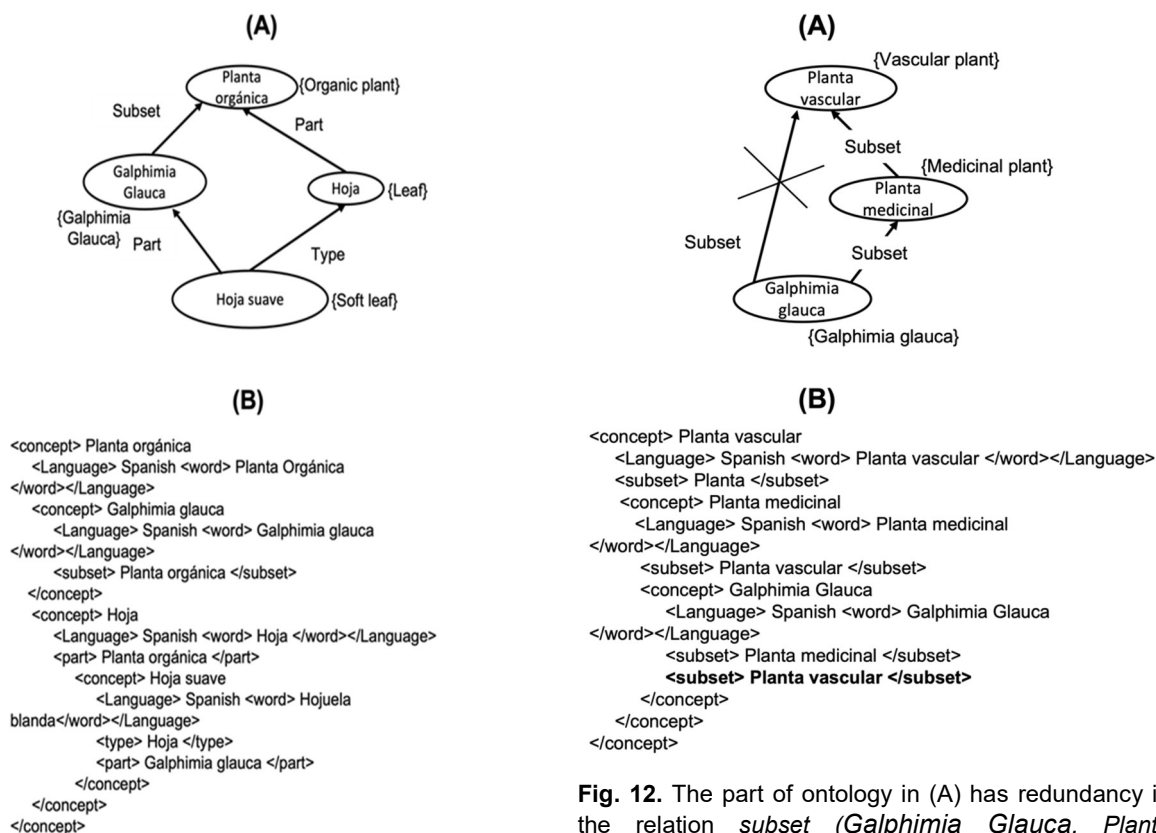
**PryRe triplet:** *significa (Virus asociados con infecciones respiratorias, null) {means (viruses associated with respiratory infections, null)}*.

**PryRe answer:** *Los virus que invaden y proliferan en las células epiteliales de las vías respiratorias que pueden causar síntomas*

<sup>3</sup> World Health Organization, www.who.int/

<sup>4</sup> https://tinyurl.com/4pk7r2sj





**Fig. 11.** (A) shows a part of the ontology (predecessor of Galphimia Glauca) and (B) shows the code of this part in OM notation

**Fig. 12.** The part of ontology in (A) has redundancy in the relation *subset* (*Galphimia Glauca*, *Planta vascular*) (*subset* (*Galphimia Glauca*, *Vascular plant*)). Although the relation appears in the code (B), it is eliminated in the ontology in memory. PryRe eliminates redundant relations.

*respiratorios y sistémicos* {Viruses that invade and proliferate in the epithelial cells of the respiratory tract that can cause respiratory and systemic symptoms}.

- 2 **Question:** *¿Cuáles son los virus comunes asociados con las infecciones respiratorias?* {What are the common viruses associated with respiratory infections?}

Case applied: B.

**Book answer:** Influenza virus, syncytial virus respiratory, measles virus, mumps virus, virus Hendra, Nipah virus, rubella virus, rhinovirus, SARS coronavirus.

**PryRe triplet:** *type* (*virus comunes asociados con las infecciones respiratorias*, null) {*type*

(*common viruses associated with respiratory infections*, null)}.

**PryRe answer:** *Virus de la influenza, virus sincicial respiratorio, virus del sarampión, virus de la parotiditis, virus Hendra, virus Nipah y metapneumovirus humano, virus de la rubéola, la familia Picornaviridae (rinovirus), y la familia Coronaviridae (coronavirus del SARS).*

- 3 **Question:** *¿Qué es el síndrome respiratorio de Oriente Medio (MERS)?* {What is Middle East Respiratory Syndrome (MERS)?}

Case applied: C.

**Book answer:** it is an illness caused by MERS-CoV.

**PryRe triplet:** *significa* (*Síndrome Respiratorio de Medio Oriente (MERS)*, null) {*means*

(Middle East Respiratory Syndrome (MERS)), null).

**PryRe answer:** Es una enfermedad causada por MERS-CoV {Is an illness caused by MERSCoV}.

**4 Question:** ¿Qué es el nuevo coronavirus? {What is the new coronavirus?}

Case applied: C.

**Book answer:** It is a mutated novel coronavirus (genus B), which is named 2019-nCoV by WHO and SARS-CoV-2 by ICTV.

**PryRe triplet:** significa (nuevo coronavirus, null) {means (new coronavirus, null)}.

**PryRe answer:** Es un nuevo coronavirus mutado (género B), que la OMS denomina 2019-nCoV y la ICTV el SARS-CoV-2 {It is a novel mutated coronavirus (genus B), which is named 2019-nCoV by WHO and SARS-CoV-2 by ICTV}.

**5 Question:** ¿Qué es la neumonía adquirida en la comunidad? {What is community-acquired pneumonia?}

Case applied: C.

**Book answer:** Refers to infectious pneumonia of the lung parenchyma (included in the alveolar wall, which belongs to the pulmonary interstitium in a broad sense) contracted outside the hospital, including pneumonia from known pathogens that occur after admission within their period of average incubation.

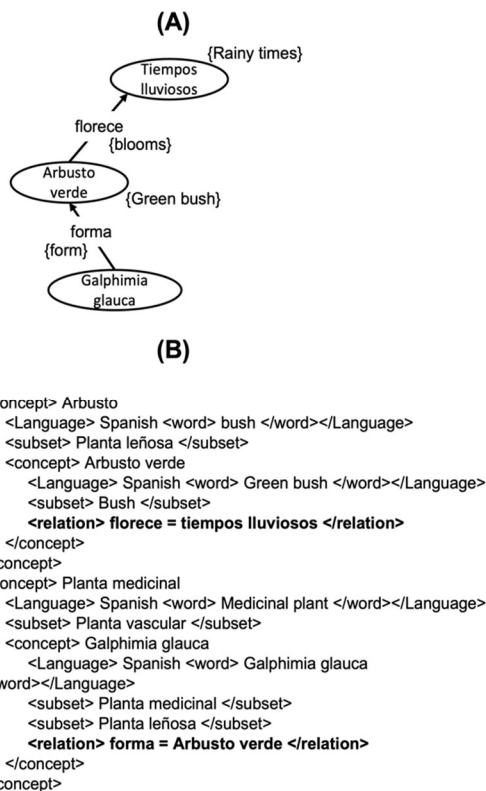
**PryRe triplet:** significa (neumonía adquirida en la comunidad, null) {means (community-acquired pneumonia, null)}.

**PryRe answer:** Se refiere a la neumonía infecciosa del parénquima pulmonar, contraída fuera del hospital {Refers to infectious pneumonia of the lung parenchyma, contracted outside the hospital}.

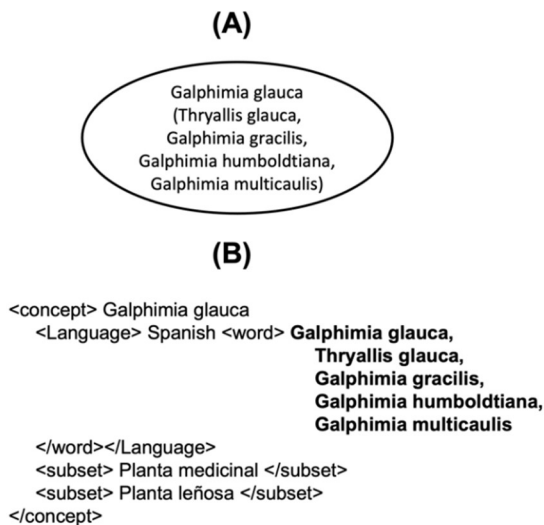
**6 Question:** ¿Qué patógenos causan neumonía adquirida en la comunidad? {What pathogens cause community-acquired pneumonia?}

Case applied: B.

**Book answer:** The most common pathogens causing acute respiratory diseases include bacteria, viruses, or a combination of bacteria and viruses. New pathogens, such as the new



**Fig. 13.** Explicit relations (bold letter) give more information about concepts



**Fig. 14.** In (A) synonyms inside parentheses are shown. In (B) these synonyms are represented with <word> label (bold letter)



coronavirus, can cause an epidemic or pandemic of acute respiratory disease. Bacteria is the leading cause of community-acquired pneumonia. Strep pneumonia is one of the most common types of bacterial pneumonia. Other bacterial pathogens include Mycoplasma, Chlamydia, Klebsiella pneumonia, Escherichia coli, Staphylococcus aureus, and Acinetobacter baumannii; Pseudonia caused by Pseudomonas aeruginosa has also been reported.

**PryRe triplet:** causa (neumonía adquirida en la comunidad, null) {*cause (community-acquired pneumonia, null)*},

**PryRe answer:** Bacteria, virus, nuevo coronavirus, Mycoplasma, Chlamydia, Klebsiella pneumonia, Escherichia coli, Staphylococcus aureus, Pseudomonas aeruginosa, Acinetobacter baumannii {*Bacteria, virus, new coronavirus, Mycoplasma, Chlamydia, Klebsiella pneumonia, Escherichia coli, Staphylococcus aureus, Pseudomonas aeruginosa, Acinetobacter baumannii*}.

- 7 **Question:** ¿Cómo es transmitida la neumonía adquirida en la comunidad? {*How is community-acquired pneumonia transmitted?*}  
Case applied: B.

**Book answer:** Theoretically, all pathogens that cause community-acquired pneumonia have the potential for transmission from person to person. Transmission routes from the source of infection to susceptible populations are transmission by droplets, contact transmission, and air transmission.

**PryRe triplet:** transmisión (neumonía adquirida en la comunidad, null) {*transmission (community-acquired pneumonia, null)*}.

**PryRe answer:** Persona a persona, contacto, gotitas, por aire {*Person to person, contact, droplets, by air*}.

- 8 **Question:** ¿Cuáles son los factores de riesgo de transmisión de la neumonía adquirida en la comunidad? {*What are the risk factors for the transmission of community-acquired pneumonia?*}  
Case applied: B.

**Book answer:** The Fall and Winter are seasons prone to seeing the prevalence of respiratory viruses such as influenza, and when other respiratory infections occur. These factors are: Environmental conditions, accessibility, and effectiveness of medical care services and Infection prevention measures, host factors, pathogen characteristics.

**PryRe triplet:** factores de riesgo de transmisión (neumonía adquirida en la comunidad, null) {*transmission risk factors (community acquired pneumonia, null)*}.

**PryRe answer:** Condiciones ambientales, Accesibilidad y efectividad de los servicios de atención médica, medidas de prevención e infección, factores de huésped, características del patógeno {*Environmental conditions, Accessibility, and effectiveness of medical care services and infection prevention measures, host factors, pathogen characteristics*} (See Figure 16).

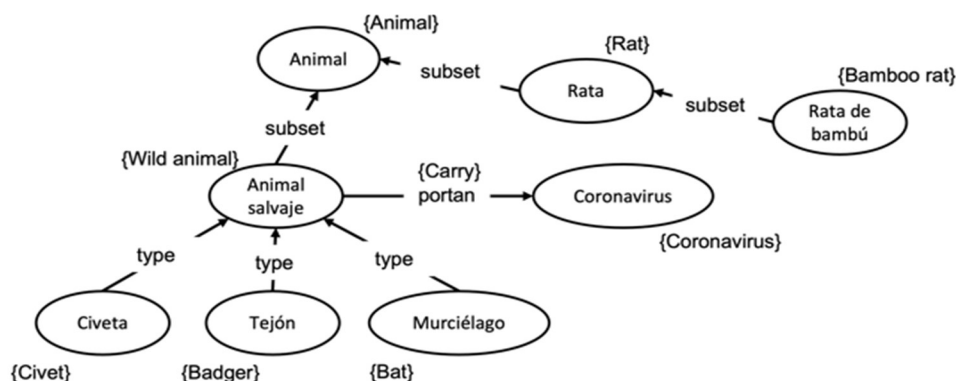
- 9 **Question:** ¿Cómo prevenir la neumonía adquirida en la comunidad? {*How to prevent community-acquired pneumonia?*}

Case applied: B.

**Book answer:** (1) Maintaining a balanced diet, ensuring proper nutrition, and maintaining oral health can help prevent infection. (2) Exercise regularly to increase immunity. (3) Quit smoking, limit alcohol consumption, and try to be positive. (4) Ensure interior ventilation, natural ventilation and/or use of exhaust fans for better airflow. (5) Get vaccinated if available.

**PryRe triplet:** prevención (neumonía adquirida en la comunidad, null) {*prevention (community-acquired pneumonia, null)*}.

**PryRe answer:** Mantener una dieta equilibrada, hacer ejercicio con regularidad, DEJAR de fumar, dejar de consumir alcohol, mantenerse de buen humor, asegurar una ventilación interior natural, utilizar extractores de aire, vacunarse {*Maintain a balanced diet, Exercise regularly, Stop smoking, stop consuming alcohol, stay in a good mood, ensure natural indoor ventilation, use air extractors, get vaccinated*}.



**Fig. 15.** The problem here is the lack of the *rata de bambú* {bamboo rat} concept in *animal salvaje* {wild animal} that are related to the *coronavirus*. If *rata* {rat} was connected to *animal salvaje* {wild animal}, then Case A could look up the answer and the answer would be: *Sí* {Yes}

- 10 Question:** *¿Quién es susceptible al 2019-nCoV?* {Who is susceptible to 2019-nCoV?} Case applied: B.

**Book answer:** The coronavirus is newly emerging in humans. Therefore, the general population is susceptible because they lack immunity against it. 2019-nCoV can infect people with normal or compromised immunity. The amount of exposure to the virus also determines whether it becomes infectious. If one is exposed to many viruses, one can get sick even if their immune function is normal. For people with poor immune function, such as the elderly, pregnant women, or people with liver or kidney dysfunction, the disease progresses relatively quickly, and symptoms are more serious.

**PryRe triplet:** susceptible, vulnerable (2019-nCoV, null).

**PryRe answer:** Población general, población expuesta, personas con función inmunológica deficiente, personas con disfunción hepática o renal. {General population, exposed population, people with poor immune function, people with liver or kidney dysfunction}.

- 11 Question:** *¿Cuáles son las características epidemiológicas de COVID-19?* {What are the epidemiological characteristics of COVID-19?} Case applied: B.

**Book answer:** The emerging epidemic of COVID-19 has undergone three stages: local outbreak, community communication, and

general stage (epidemic). Communication stages: the COVID-19 epidemic went through three stages: 1) the local outbreak stage (the cases of this stage are mainly related to exposure to seafood); 2) the stage of community communication (interpersonal communication and transmission of grouping in communities and families); 3) generalized stage (rapid diffusion, with large population flow, to the whole country of China and even to the world).

**PryRe triplet:** características epidemiológicas (COVID-19, null) {epidemiological characteristics (COVID-19, null)}.

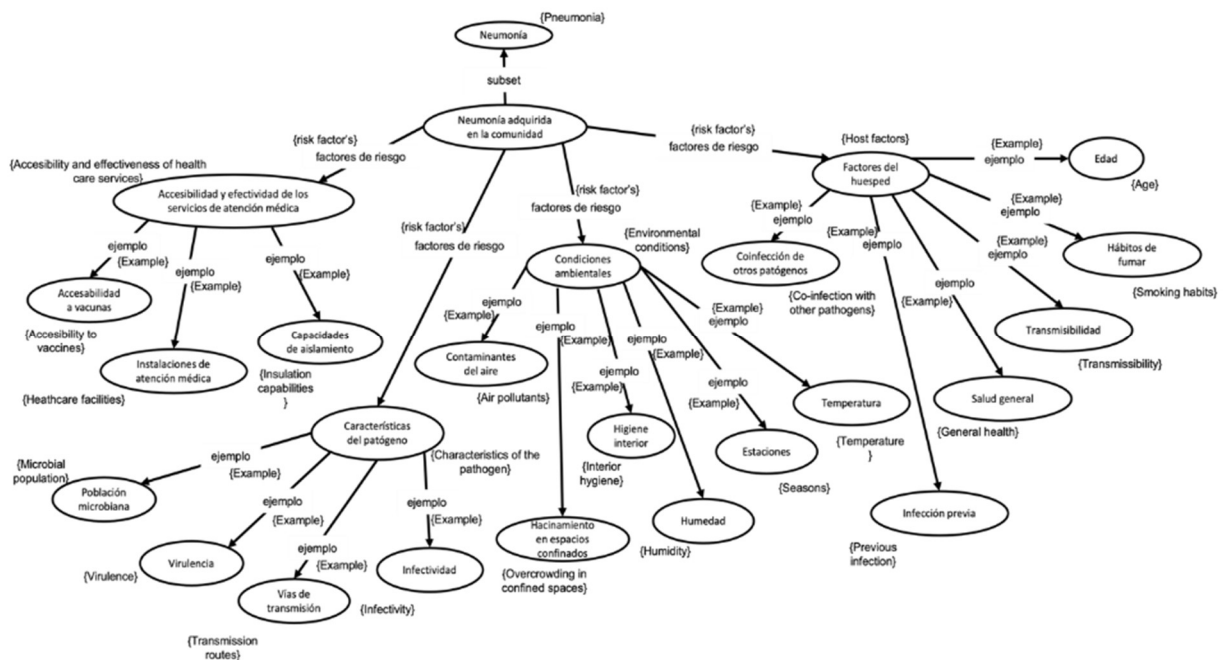
**PryRe answer:** Etapa de brote local, Etapa de comunicación comunitaria, Etapa generalizada {Stage of local outbreak, Stage of community communication, Generalized stage}.

- 12 Question:** *¿Cuáles son las rutas de transmisión de COVID-19?* {What are the 2019-nCoV transmission routes?}

Case applied: B.

**Book answer:** Currently, transmission via droplets and respiratory contacts is believed to be the primary route, but there is a risk of fecal-oral transmission. Aerosol transmission, mother-to-child transmission, and other routes are not yet confirmed.

**PryRe triplet:** Rutas de transmisión (COVID-19, null) {transmission routes (COVID-19, null)}.



**Fig. 16.** Part of the ontology that represents the answer to question number 8, in which only the examples of each factor are presented. However, more information can be shown in each example. For example, air pollutants, overcrowding in confined spaces

**PryRe answer:** Contactos respiratorios, gotitas {*Respiratory contacts, droplets*}.

**13 Question:** ¿Qué es la transmisión de gotas? {*What is the transmission of drops?*}

Case applied: C.

**Book answer:** Drops can enter mucous surfaces within a certain distance (usually 1m). Due to the relatively large size and weight of the droplets, they cannot remain airborne for long.

**PryRe triplet:** significa (transmisión de gotas, null) {*means (droplet transmission, null)*}.

**PryRe answer:** Gota es una partícula que contiene agua con un diámetro mayor a 5 mm, las gotas pueden ingresar a las superficies mucosas dentro de una cierta distancia (generalmente 1 m). Debido al tamaño y peso relativamente grandes de las gotas, no pueden permanecer en el aire por mucho tiempo. {*Droplet is a particle that contains water with a diameter greater than 5 mm, the*

*droplets can enter the mucous surfaces within a certain distance (generally 1 m). Due to the relatively large size and weight of the droplets, they cannot remain airborne for long.*}

**14 Question:** ¿Qué es la transmisión aérea? {*What is air transmission?*}

Case applied: C.

**Book answer:** Also known as Aerosol Transmission. Aerosols are suspensions of small particles or droplets that can be transmitted through the air.

**PryRe triplet:** significa (transmisión aérea, null) {*means (airborne transmission, null)*}.

**PryRe answer:** También conocido como transmisión por aerosol. Los aerosoles son suspensiones de pequeñas partículas o gotitas que se pueden transmitir a través del aire. {*Also known as aerosol transmission. Aerosols are suspensions of small particles or droplets that can be transmitted through the air.*}

**Table 4.** The fifty questions answered by PryRe

Number	Case Applied	Precision
1	C	1.0
2	B	1.0
3	C	1.0
4	C	1.0
5	C	1.0
6	B	1.0
7	B	1.0
8	B	1.0
9	B	1.0
10	B	1.0
11	B	1.0
12	B	1.0
13	C	1.0
14	C	1.0
15	C	1.0
16	C	1.0
17	B	1.0
18	B	1.0
19	B	1.0
20	B	1.0
21	B	1.0
22	Unsolved	0.0
23	Unsolved	0.0
24	C	1.0
25	B	1.0
26	B	1.0

27	B	1.0
28	Unsolved	0.0
29	B	1.0
30	Unsolved	0.0
31	B	1.0
32	B	1.0
33	Unsolved	0.0
34	B	1.0
35	Unsolved	0.0
36	B	1.0
37	B	1.0
38	B	1.0
39	B	1.0
40	C	1.0
41	B	1.0
42	B	1.0
43	B	1.0
44	Unsolved	0.0
45	B	1.0
46	D	1.0
47	B	1.0
48	Unsolved	0.0
49	Unsolved	0.0
50	B	1.0

**15 Question:** *¿Qué es la transmisión por contacto? {What is contact transmission?}*  
Case applied: D.

**Book answer:** Contact transmission refers to the transmission of pathogens through direct

or indirect contact through fomites (objects carrying pathogens): (1) Direct contact. Pathogens are transmitted by direct contact of the mucous or skin with an infected host. Blood or bloody fluids enter the body through mucous membranes or non-intact skin (mainly viruses).

Transmission is caused by contact with secretions that contain certain pathogens, commonly due to infections by bacteria, viruses, parasites, etc. (2) Indirect contact. Pathogens are transmitted through contaminated objects or people. The pathogens of intestinal infectious diseases are transmitted mainly through indirect contact. (3) Other important pathogens transmitted by indirect contact: MRSA (*Staphylococcus aureus* resistant to benzoxazole/methicillin), VRE (vancomycin-resistant enterococcus), *Clostridium difficile*.

**PryRe triplet:** significa (transmisión por contacto, null) {*means (contact transmission, null)*}.

**PryRe answer:** Se refiere a la transmisión de patógenos por contacto directo o indirecto a través de fomites (objetos portadores de patógenos): (1) Contacto directo. (2) Contacto indirecto. (3) Otros patógenos importantes transmitidos por contacto indirecto {*Refers to the transmission of pathogens through direct or indirect contact through fomites (objects carrying pathogens): (1) Direct contact. (2) Indirect contact. (3) Other important pathogens transmitted by indirect contact*}.

**16 Question:** *¿Qué es un contacto cercano?* {*What is close contact?*}

Case applied: D.

**Book answer:** Close contacts refer to people who have contact with a confirmed or suspected 2019-nCoV infection patient), including the following situations. (1) Those who live, study, work, or have other forms of close contact with a patient, (2) Medical personnel, family members, or others who have had close contact with a patient without taking effective protective measures during diagnosis, treatment, nursing and visits, (3) Other patients and their companions who share the same room with an infected patient, (4) Those who shared the same transport or lift with the patient, (5) Those who are considered as such through field investigations.

**PryRe triplet:** significa (contacto cercano, null) {*means (close contact, null)*}.

**PryRe answer:** Referirse a personas que tienen contacto con un paciente con infección

confirmada o sospechada de 2019-nCoV, incluidas las siguientes situaciones: (1) Quienes viven, estudian, trabajan o tienen otras formas de contacto cercano con un paciente, (2) Personal médico, familiares u otras personas que han tenido contacto cercano con un paciente sin tomar medidas de protección efectivas durante el diagnóstico, tratamiento, enfermería y visitas, (3) Otros pacientes y sus acompañantes que comparten la misma habitación con un paciente infectado, (4) Los que compartieron el mismo transporte o ascensor con el paciente, (5) Los que son considerados como tales a través de investigaciones de campo {*It refers to people who have contact with a confirmed or suspected 2019-nCoV infection patient, including the following situations: (1) Those who live, study, work, or have other forms of close contact with a patient, (2) Medical personnel, family members, or others who have had close contact with a patient without taking effective protective measures during diagnosis, treatment, nursing and visits, (3) Other patients and their companions who share the same room with an infected patient, (4) Those who shared the same transport or lift with the patient, (5) Those who are considered as such through field investigations.*}

In Question 16, PryRe only presents the types of transmission by contact but does not define each of them: only *Contacto directo* {Direct contact} appears, and not its description.

In Question 16 PryRe presents the types of people who can transmit the virus since this answer is an important part of the question. That is, (2) *Personal médico* {*medical personnel*}, *familiares* {*relatives*}, etc.

The two books that were used had different structures. The book about herbs and medical plants [14] is based on tables and lists. The coronavirus book is based on questions/answers [22].

The output of PryRe is text. Thus, the results presented in this section are copied and pasted by hand. 50 questions answered by PryRe, are presented here.

The ontology about both topics (medical plants and coronavirus book) was based in the WordNet

hierarchy; this ontology is downloaded in Spanish from [4].

## 7 Conclusions and Future Work

Based on an ontology, PryRe interprets questions in Spanish, and provides the answers in Spanish. It does this with acceptable precision and can be used, as it is, with other ontologies to answer related questions in natural language.

The knowledge base was built manually obtaining the information from reliable sources [14, 22]: one hundred questions about coronavirus and 85 descriptions of medicinal plants with its features: scientific name, place of collection, location, coordinates, name of the collector, year of collection, altitude, and observations are represented. This yields an average of eight relations or links per description. Building the knowledge base in this manner was a tedious process. Then, PryRe tested (using 185 questions) the complete knowledge. The accuracy (percentage of correct answers) obtained by PryRe has been 82%. Table 4 shows only fifty of these questions.

In Table 4, 41 answers have precision 1 (perfect answer), and 9 have 0 (no match), because the book's answer is ambiguous, not direct or concrete. These 9 responses were considered unsolved as PryRe. The total precision of PryRe for this test, is therefore  $(41)/50 = 0.82$  (82%).

Availability of the complete ontology is here<sup>5</sup>, and the code of PryRe is publicly available here<sup>6</sup>.

The next work is to use crawlers and scrappers to retrieve from the web other useful information, and to review manually its suitability. Then, transform it into OM notation with the help of the Ontology Merger [18]. This path increases the body of knowledge for PryRe, in a semi-automated, supervised way.

## References

- Landes, S., Leacock, C., Teng, R. I. (1998). Building semantic concordances. In: Fellbaum, C., ed., WordNet: an electronic lexical database, chapter 8, MIT Press, pp. 199–216.
- Villanueva, D., Cuevas-Rasgado, A. D., Juárez, O., Guzmán-Arenas, A. (2013). Using frames to disambiguate propositions. Expert Systems with Applications, Vol. 40, No. 2, pp. 598–610. DOI: 10.1016/j.eswa.2012.07.061.
- Gruber, T. R. (1995). Toward Principles for the Design of Ontologies Used for Knowledge Sharing? International Journal of Human-computer Studies, Vol. 43, No. 5–6, pp. 907–928. DOI: 10.1006/ijhc.1995.1081.
- Guzmán-Arenas, A., Cuevas, A. D. (2010). Knowledge accumulation through automatic merging of ontologies. Expert Systems with Applications, Vol. 37, No. 3, pp. 1991–2005. DOI: 10.1016/j.eswa.2009.06.078.
- Padró, L., Stanilovsky, E. (2012). FreeLing 3.0: Towards Wider multilinguality. 8th International Conference on Language Resources and Evaluation (LREC), pp. 2473–2479.
- Breck, E., Burger, J., House, D., Light, M., Mani, I. (1999). Question Answering from Large Document Collections. AAAI Fall Symposium on Question Answering Systems.
- Vargas-Vera, M., Motta, E. (2004). AQUA-ontology-based question answering system. Lecture Notes in Computer Science, Vol. 2972, pp. 468–477. DOI: 10.1007/978-3-540-24694-7\_48.
- Alinaghi, T., Bahreininejad, A. (2011). A multi-agent question-answering system for e-learning and collaborative learning environment. International Journal of Distance Education Technologies (IJDET), Vol. 9, No. 2, pp. 23–39. DOI: 10.4018/jdet.2011040103.
- Bordes, A., Weston, J., Chopra, S. (2014). Question answering with subgraph embeddings. EMNLP: Conference on Empirical Methods in Natural Language Processing, pp. 615–620. arXiv:1406.3676. DOI: 10.48550/arXiv.1406.3676.
- Bast, H., Haussmann, E. (2015). More accurate question answering on freebase.

<sup>5</sup> <https://tinyurl.com/524m8fj4>

<sup>6</sup> <https://tinyurl.com/y37p9tph>

- 24th ACM International Conference on Information and Knowledge Management, pp. 1431–1440. DOI: 10.1145/2806416.2806472.
11. **Xu, K., Reddy, S., Feng, Y., Huang, S., Zhao, D. (2016).** Question answering on freebase via relation extraction and textual evidence. 54th Annual Meeting of the Association for Computational Linguistics, pp. 2326–2336. DOI: 10.18653/v1/P16-1220.
  12. **Abujabal, A., Yahya, M., Riedewald, M., Weikum, G. (2017).** Automated template generation for question answering over knowledge graphs. 26th international conference on world wide web, pp. 1191–1200. DOI: 10.1145/3038912.3052583.
  13. **Zhu, X., Yang, X., Chen, H. (2018).** A biomedical question answering system based on SNOMED-CT. Lecture Notes in Computer Science, Vol. 11061, pp. 16–28. DOI: 10.1007/978-3-319-99365-2\_2.
  14. **Estrada, E., Lara, A. (2008).** Sistema nervioso y herbolaria. Universidad Autónoma Chapingo.
  15. **Noy, N. F., McGuinness, D. L. (2001).** Ontology Development 101: A Guide to Creating Your First Ontology. Stanford Knowledge Systems Laboratory.
  16. **Swaminathan, V., Sivakumar, R. (2012).** A Comparative Study of Recent Ontology Visualization Tools with a Case of Diabetes Data. International Journal of Research in Computer Science (IJORCS), Vol. 2, No. 3, pp. 31–36. DOI: 10.7815/ijorcs.23.2012.026.
  17. **Castañeda, E., Cortés, O. (2011).** Construcción de una base de conocimiento sobre el uso de herramientas de carpintería. Instituto Politécnico Nacional.
  18. **Cuevas, A. D. (2008).** Unión de Ontologías usando propiedades semánticas. Doctoral Dissertation Thesis, Centro de Investigación en Computación, IPN.
  19. **Pavon, L. (1999).** Clases de partículas: preposición, conjunción y adverbio. In: **Bosque, I., DeMonte V., eds.**, Gramática descriptiva de la lengua española, Capítulo 9. Espasa-Calpe. Vol. 1, pp. 565–656.
  20. **Noy, N., Gao, Y., Jain, A., Narayanan, A., Patterson, A., Taylor, J. (2019).** Industry-Scale Knowledge Graphs: Lessons and Challenges: Five diverse technology companies show how it's done. Communications of the ACM, Vol. 62, No. 8, 36-43. DOI: 10.1145/3331166.
  21. **Cuevas, R. A. D., Niño, M. Y. E., Lamont, F. G. (2017).** Semantic analyzer for Spanish, using ontologies. Komputer Sapiens, Vol. 3, pp. 13–36.
  22. **Zhou, W. (2020).** The Coronavirus Prevention Handbook 101 Science-based Tips that Could Save your Life. Skyhorse Publishing.

*Article received on 09/06/2021; accepted on 24/01/2022.  
Corresponding author is Alma Delia Cuevas-Rasgado.*