www.cya.unam.mx/index.php/cya

Contaduría y Administración

Accounting & Management

Contaduría y Administración 68 (3), 2023, 1-28

Representation and management of data of Mexican patients with COVID-19 using knowledge graphs

Representación y gestión de datos de pacientes mexicanos con COVID-19 mediante gráficos de conocimiento

Maricela Bravo*, Darinel González, José A. Reyes Ortiz, Luis Fernando Hoyos Reyes

Universidad Autónoma Metropolitana, México

Received October 9, 2022; accepted April 26, 2023
Available online September 30, 2024

Abstract

This paper presents the development and evaluation of a knowledge graph for the representation and data management from Mexican patients who were registered as positive COVID-19 cases. The use of knowledge graphs offers advantages over other representation models, knowledge graphs facilitate quick access to data through various query and access technologies, which is relevant for the management of health systems. On the other hand, the use of knowledge graphs favors the integration and expansion of knowledge by linking it with other graphs. This paper describes a construction approach that ranges from the design of the general graph, the search and reuse of existing graphs, as well as the registration of COVID-19 patient information. As a result, an integral, efficient and expandable graph model was obtained; capable of representing and extending by incorporating information from other knowledge graphs in the medical domain related to the concepts of the COVID-19 disease, such as vaccines, laboratory tests, symptoms, medications, as well as clinical diagnosis, which can impact on the improvement of the health systems management. The data set used during the tests shows the usefulness of the knowledge graph to record real data from Mexican patients diagnosed with COVID-19.

*Corresponding author.
E-mail address: mcbc@azc.uam.mx (M. Bravo).
Peer Review under the responsibility of Universidad Nacional Autónoma de México.

http://dx.doi.org/10.22201/fca.24488410e.2023.4839

JEL Code: C63, D83, I00, L86
Keywords: integration of knowledge graphs; reuse of graphs; biomedical knowledge graphs

Resumen

En este artículo se presenta el desarrollo y evaluación de un grafo de conocimiento para la representación y gestión de datos de pacientes mexicanos que fueron registrados como casos positivos COVID-19. El empleo de grafos de conocimiento ofrece ventajas sobre otros modelos de representación, los grafos de conocimiento facilitan el acceso ágil a los datos mediante diversas tecnologías de consulta y acceso, lo que resulta relevante para la administración de sistemas de salud. Por otra parte, el empleo de grafos de conocimiento favorece la integración y expansión del conocimiento mediante la vinculación con otros grafos. En este trabajo se describe un enfoque de construcción que abarca desde el diseño del grafo general, la búsqueda y reutilización de grafos existentes, así como el registro de la información de pacientes COVID-19. Como resultado se obtuvo un modelo de grafo integral, eficiente y expandible; capaz de representar y extenderse incorporando información de otros grafos de conocimiento del dominio médico relacionados con los conceptos de la enfermedad del COVID-19, como son las vacunas, pruebas de laboratorio, los síntomas, medicamentos, así como como el diagnóstico clínico, lo que puede impactar en la mejora de la administración de sistemas de salud. El conjunto de datos utilizados durante las pruebas muestra la utilidad que tiene el grafo de conocimiento para registrar datos reales de pacientes mexicanos diagnosticados con COVID-19.

Código JEL: C63, D83, I00, L86
Palabras clave: integración de grafos de conocimiento; reutilización de grafos; grafos de conocimiento biomédico

Introduction

An enormous volume of data has been generated over the last few years on patients who have suffered COVID-19, some resulting in serious complications such as intubation or even death. The Government of Mexico, through the General Directorate of Epidemiology (DGE),[1] registers the data of patients diagnosed with COVID-19. This registry consists of a downloadable CSV file with more than 7 million records. To manage large amounts of data and improve the administration of health systems, Khan and Yairi (2018) reviewed novel technologies that adequately manage scalability and the exploitation of data to generate derived knowledge.

According to Kalra (2006), the automated recording, updating, maintenance, and exploitation of patient data is indispensable when working in health care for better control and management of information. Representation models based on knowledge graphs allow the execution of reasoners that produce new relations between data by defining rules and logical axioms. This article describes the methodology for designing, constructing, and integrating the knowledge graph to represent Mexican patient data with COVID-19. The registry of cases associated with COVID-19 provided by the DGE was

[1] https://datos.gob.mx/busca/dataset/informacion-referente-a-casos-covid-19-en-mexico

used as a data source. Using query tools and logical inference on the knowledge graphs, the data of patients with COVID-19 represented in the knowledge graph are exploited.

The methodology for designing and constructing the knowledge graph was oriented to extracting and reusing pre-existing knowledge graph modules related to the medical area. The graph-based representation has advantages since the handling of triplets <subject, predicate, object> offers great flexibility for data scaling by referencing IRIs[2], allowing definitions from other knowledge graphs to be linked in a simple way. For the evaluation of the knowledge graph described in this article, three approaches were used and evaluated: the competence of the graph, the quality through design principles, and the usability of the graph through the instantiation of a dataset of Mexican patients with positive COVID-19 cases.

The main contributions presented in this article are:

a) A methodology for knowledge graph construction considering the extraction and reuse of knowledge graph modules in the medical area

b) An integrated knowledge graph for Mexican patient data management, which includes concepts related to COVID-19, namely terms such as vaccines, laboratory tests, medications, diseases, symptoms, and diagnoses

c) An integrated knowledge graph evaluation method based on three approaches: evaluation of usability, evaluation based on design principles, and evaluation of graph competence

The rest of the paper is organized as follows. The Related Works Section reviews research articles about the development of methods for creating and evaluating knowledge graphs for the management of clinical patient data or for health data. The Methodology of Knowledge Graph Development Section describes the methodological stages defined for constructing and integrating the knowledge graph. The Evaluation of the General Graph Section describes three evaluation approaches that were applied to the resulting graph. The Results Section describes the integrated knowledge graph populated with the instances. Finally, the conclusions are presented.

Related works

The representation and management of patients' clinical data is a research topic that has gained relevance in recent years. This is mainly due to the search for solutions to problems related to the interoperability and exchange of specialized medical information between different health users. This section presents a review of research works related to the topics of this article.

---

[2] Internationalized Resource Identifier

http://dx.doi.org/10.22201/fca.24488410e.2023.4839

Regarding the electronic recording of clinical patient data, Farion et al. (2009) present the development and evaluation of a representation model to support clinicians' decision making; this model is used in an acute care and emergency setting. The authors describe the implementation of a new mobile emergency protocol design 2, which provides a unifying environment that can handle multiple clinical applications running on multiple platforms, using a knowledge base and derived models to represent key components of a Clinical Decision Support System (CDSS).

Köhler S. et al. (2009) describe a system called Phenomizer, which is not an expert system but rather a system for experts who can use the system for support during the process of differential diagnosis in human genetics. The Phenomizer system can indicate whether the characteristics of the clinical data entered by the physician are highly suggestive of a given diagnosis.

Celi G. et al. (2012) present a dynamic tool for representing and managing patient profiles, facilitating evidence-based decision support. In this approach, probabilistic modeling is performed on subsets of patients from the institution itself rather than heterogeneous patient populations from different centers. In this model, the term Collective Experience is adopted for this approach, as the information is extracted from several physicians' experience and is stored in the electronic medical record system.

Elizabeth D. Hermsen et al. (2012) describe the advantages of representing patient data using a CDSS, as it allows for improved antimicrobial treatment decision making through the availability of a combination of patient-specific data and costs. There are several benefits to using this system, including reduced adverse events, reduced length of a patient's stay in the health care facility, reduced costs during the stay, and more appropriate antimicrobial use for the patient.

Sherimon and Krishnan (2016) describe OntoDiabetic, a knowledge-base-based CDSS for assessing risk factors and generating treatment suggestions for patients with diabetes. The authors define a set of inference rules to obtain information about the patient's health status. This ontology model lacks up-to-date information about drugs, diseases, and important relations between them.

Zhang, Gou, et al. (2017) address the complex interactions between risk factors, diseases, patient conditions, and treatment modalities. These actions include epidemiology and surveillance strategies to monitor trends and track progress, policies, and environmental approaches to promote health.

Ajami and Mcheick (2018) present the development of a model that enables the representation of patient information to create safe environments for patients with chronic obstructive pulmonary disease (COPD). This model is based on the formal description of knowledge bases of a health-related domain and uses the Semantic Web Rule Language (SWRL). The knowledge base contains all relevant COPD-related concepts, including the patient's personal information, location, activity, symptoms, risk factors, laboratory test results, and treatment plan.

Oyelade et al. (2020) introduce the construction of a knowledge base for patient profile representation; the proposed framework is achieved by knowledge formalization. The result showed interesting performance compared to similar state-of-the-art case-based reasoning (CBR) studies using fuzzy CBR. The knowledge base is general enough for adoption and generalization to diagnostic problems associated with other medical purposes and diseases in patients.

Govindan, Mina, and Alavi (2020) present a practical system for representing patient profiles as CDSS to classify community members and, consequently, to manage demand and control epidemic outbreaks in the healthcare supply chain. In the proposed approach, users are first grouped according to two criteria: age range and pre-existing diseases (such as diabetes, heart problems, or high blood pressure).

Harry et al. (2020) describe a system for representing patient profiles for Electronic Health Record (EHR)-based decision making that can improve cancer prevention and screening in primary care. The Consolidated Framework for Implementation Research (CFIR) was used for this purpose.

According to Bravo et al. (2020), the representation of patient profiles through a semantic representation model offers benefits for inference and automatic reasoning, facilitating the identification of risk cases.

Reyes-Peña et al. (2021) describe a methodology for constructing an ontology network to represent clinical records of patients with type 2 diabetes mellitus.

A similar approach to the one in this article is presented by Mythili et al. (2022), who construct a knowledge graph from clinical records and identify the relations between patient, disease, and drug entities. The graph is constructed from different data sources, and the information is extracted as queries. The steps implemented to build the knowledge graph are data collection, named entity recognition, entity normalization, entity rationing, and graph-based neural networks.

For a broader review of works related to semantic interoperability between electronic health records, De Mello et al. (2022) describe the main application areas as quality assurance in healthcare, reduction of medical errors, disease control and monitoring, and individualized patient care.

This article reports a methodology that enables the construction of the knowledge graph for the representation of diabetic patient profiles, considering the existence of medical terminologies for its reuse, as well as data sets of patients who suffered from COVID-19. As a result, the construction methodology covers specific needs of the health system in Mexico, as well as cases of real patients.

Methodology of knowledge graph development

The design, development, and integration process of a new Knowledge Graph implies the definition of a methodology that considers everything from the identification of the key concepts to be represented and the incorporation and reuse of existing graphs to the evaluation of the resulting graph. This section describes the methodology implemented (see Figure 1). This methodology includes the following stages:

1. Collection and analysis of information related to the representation domain
2. Search and selection of existing domain-related knowledge graphs
3. Design of the General Knowledge Graph model
4. Construction of the General Knowledge Graph
5. Extraction of modules from existing graphs for reuse
6. Integration of modules in the main Graph
7. Definition of general axioms and rules of inference

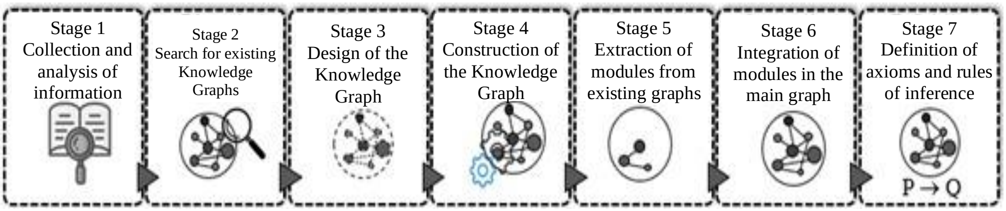

Figure 1. Stages of the Methodology of Knowledge Graph Development
Source: created by the authors.

Stage 1: Collection and analysis of information

The first step is to have valid and reliable sources of information. This is even more relevant when the information in the graph corresponds to the medical and healthcare domain. Therefore, the first step is to search for reliable sources of information about the SARS-CoV-2 coronavirus to identify the key terms or conceptual entities and the relations that may exist between the concepts.

The sources of information considered were the following:

a) Official page with information about COVID-19[3], published by the Government of Mexico's health authorities. In this portal, it is possible to consult official information regarding vaccination programs, access to open data, and recommendations for the population, among others.

b) Official Mexican Standard NOM017 SSA2 2012[4], which establishes the criteria, specifications, and guidelines for the operation of the National Epidemiological Surveillance System.

[3] https://coronavirus.gob.mx/covid-19/
[4]https://epidemiologia.salud.gob.mx/gobmx/salud/documentos/manuales/00_NOM-017-SSA2-2012_para_vig_epidemiologica.pdf

http://dx.doi.org/10.22201/fca.24488410e.2023.4839

c) Standardized Guidelines for Epidemiological and Laboratory Surveillance of Viral Respiratory Disease[5].

d) Clinical guidelines for the treatment of COVID-19 in Mexico[6]. These guidelines were developed with representatives from all public institutions in the health sector. They describe the drugs that can be used to manage COVID-19 and the treatment algorithm for patients with COVID-19, among other relevant aspects and guidelines.

e) International and national guidelines and manuals for the treatment and follow-up of patients. The Government of Mexico has published the national guidelines in conjunction with the open data of the DGE.

f) Knowledge Graphs described in the related works developed to represent patient profiles in the medical domain.

Based on this information, the following terms relevant to patient data representation and management were identified: Patient, Diagnosis, Disease, Medication, Symptom, Laboratory Test, and Vaccine. Moreover, it was considered appropriate to include the registration of information on the municipality and state where the patients reside and where they were born.

Stage 2: Search and selection of existing knowledge graphs

The objective of this stage is to search for knowledge graphs to determine if there are graphs with the concepts required for the general graph to be integrated. Bioportal[7] is one of the most complete repositories of biomedical knowledge bases. Once the graphs with related representations have been identified, they should be reviewed and studied to know their structure in more detail, so that it can be decided whether it is convenient to reuse them partially or totally. Therefore, it is important to know the medical classes and concepts related to COVID-19 and their properties, as mentioned by Whetzel et al. (2011).

According to the conceptual requirements considered, the comprehensive model should include graphs related to the concepts of Disease, Drug, Sign and Symptom, Laboratory Test, and Vaccine. The first graph consulted was NDFRT (National Drug File - Reference Terminology), as it includes concepts related to drugs such as physiological effects, therapeutic categories, chemical structure, diseases it can treat, and diseases it can prevent, among others. NDFRT has approximately 36 200 classes. On the other hand, LOINC (Logical Observation Identifier Names and Codes), a graph that includes concepts of all

---

[5] https://www.gob.mx/cms/uploads/attachment/file/715464/Lineamiento_VE_y_Lab_Enf_Viral_05042022.pdf
[6] https://www.gob.mx/cms/uploads/attachment/file/659911/GuiaTx_COVID19_Consenso_2021.08.02_compressed.pdf
[7] https://bioportal.bioontology.org/

types of laboratory tests, including those to rule out or confirm positive cases of COVID-19, has approximately 281 878 classes.

Another graph reviewed is SYMP (Symptom). This graph describes a broad set of symptoms in a general way and symptoms caused by the disease, with approximately 1 000 classes. Finally, the NCIT (National Cancer Institute Thesaurus) graph was consulted. This graph includes a broad set of biomedical concepts, from the representation of biological processes and biomedical materials from which the vaccine module for COVID-19 is derived, with more than 174 278 classes.

Table 1 shows the name (or acronym), meaning, and Internationalized Resource Identifier (IRI) of each graph selected for reuse.

Table 1
Biomedical knowledge graphs considered for partial reuse

Graph	Stands for	IRI
NDFRT	National Drug File - Reference Terminology	https://bioportal.bioontology.org/ontologies/NDFRT
LOINC	Logical Observation Identifier Names and Codes	https://bioportal.bioontology.org/ontologies/LOINC
SYMP	Symptom	https://bioportal.bioontology.org/ontologies/SYMP
NCIT	National Cancer Institute Thesaurus	https://bioportal.bioontology.org/ontologies/NCIT

Source: created by the authors.

Stage 3: Design of the general knowledge graph model

The design that supports the General Knowledge Graph was carried out considering the result of the analysis of the information related to national and international standards for managing the information of patients with COVID-19. The main concepts to be represented in the general knowledge graph were identified through a process of elicitation of terms. This section describes the design criteria for each of these concepts.

a) The general knowledge graph was designed as a module-based model, where each module represents a main concept in such a way that it allows both the construction of graphs from scratch and the reuse of other graphs. This general graph is called PatientElectronicRecord and describes the classes and properties for representing patient data and diagnosing COVID-19. It also defines the meta-relations between the different modules or knowledge networks to be reused.

b) For the information on vaccines, it was decided to reuse the information contained in the NCIT graph, which represents the information on vaccines developed and approved for emergency use by national and international organizations. The graph module called Vaccine was designed, which includes the concepts and relations to represent the set of vaccines for COVID-19 prevention.

c) For the representation of drugs, it was decided to reuse the NDFRT graph. This graph includes the attributes of the drugs, their relations with diseases, and chemical components, among others. Therefore, a graph module was designed that includes the concepts of disease, drug, physiological effect, and chemical component, among others. Likewise, the semantic relations between these concepts were defined; this graph module was named Medicament.

d) It was decided to reuse the information from the LOINC graph to represent laboratory tests. Therefore, a graph module called LaboratoryTest was designed to represent the types of laboratory tests that exist to diagnose or rule out any disease, as well as the tests to determine if there is a positive case of COVID-19 infection.

e) Finally, a graph module called Symptom was designed to represent disease symptoms, including the semantic relation between symptoms and diseases.

Figure 2 shows the five subgraphs or modules integrated into the knowledge graph and the meta-relations between these graphs.

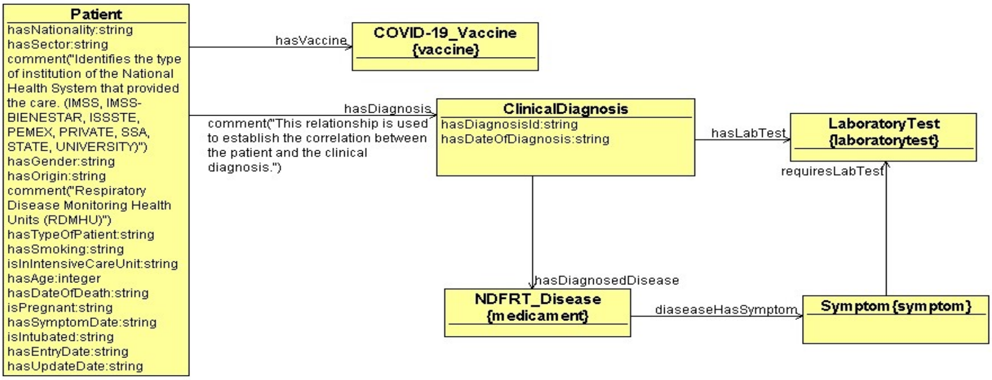

Figure 2. General model of the modules comprising the PatientElectronicRecord knowledge graph
Source: created by the authors.

http://dx.doi.org/10.22201/fca.24488410e.2023.4839

Stage 4: Construction of the general knowledge graph

The PatientElectronicRecord graph includes the concepts shown in Table 2.

Table 2
Concepts defined in the knowledge graph PatientElectronicRecord

Concept	Description
ClinicalDiagnosis	Record of the patient's clinical diagnosis.
Patient	Data from the patient's record.
StateFederative	Federal State of Mexico.
Municipality	State Municipality.

Source: created by the authors.

For each concept in Table 2, the data properties that characterize the individuals belonging to each concept were defined (see Table 3).

Table 3
Data properties of the knowledge graph PatientElectronicRecord

Data property	Domain	Rank	Description
hasOrigin	Patient	integer	USMER[8] of origin of patient
hasSector	Patient	string	Health System Institution that provided care
hasGender	Patient	string	Patient's gender
hasAge	Patient	integer	Patient's age
hasTypeOfPatient	Patient	string	Outpatient or inpatient
hasNationality	Patient	string	Patient's nationality
hasEntryDate	Patient	string	Date of patient admission to USMER
hasSymptomDate	Patient	string	Date of patient's symptoms
hasSymptomUpdate	Patient	string	Date of patient's symptom update
hasDateOfDeath	Patient	string	Date of patient's death
hasSmoking	Patient	string	The patient smokes
isInIntensiveCareUnit	Patient	string	The patient is in the intensive care unit
isPregnant	Patient	string	The patient is pregnant
isIntubated	Patient	string	The patient is intubated
hasAbbreviationState	StateFederative	string	Abbreviation of the state name
hasKeyState	StateFederative	string	State key
hasNameState	StateFederative	string	State name
hasState	Municipality	string	Municipality belongs to state
hasNameMunicipality	Municipality	string	Municipality name
hasKeyMunicipality	Municipality	string	Municipality key

Source: created by the authors.

[8]Viral Respiratory Disease Monitoring Health Unit (Spanish: Unidad de Salud Monitora de Enfermedad Respiratoria viral)

Table 4 shows the characteristics between concepts (meta-relations) of the general knowledge graph PatientElectronicRecord

Table 4
Characteristics between concepts of the knowledge graph PatientElectronicRecord

Characteristic	Domain	Rank	Description
belongsTo	Municipality	StateFederative	Municipality belongs to state
hasDiagnosis	Patient	ClinicalDiagnosis	Patient has clinical diagnosis
hasMunicipality	StateFederative	Municipality	Federal State has municipality
hasResidenceMunicipality	Patient	Municipality	Municipality of residence
hasResidenceState	Patient	StateFederative	Patient’s state of residence
hasStateFederativeBirth	Patient	StateFederative	Patient’s state of birth
hasStateMedicalUnit	Patient	StateFederative	Patient has medical unit

Source: created by the authors.

Stage 5: Extraction of modules from existing graphs for reuse

The Semantic Web community has accepted module-based reuse since generating a more simplified model than the original one is very helpful. As pointed out in Stage 3, a set of graphs to be reused was identified; therefore, a reuse method based on extracting modules from the graphs listed in Table 1 was defined. This method consists of extracting one or several modules from a knowledge graph, designing and automatically generating a simplified model, and reusing it by importing it into the general graph, as shown in Figure 3.

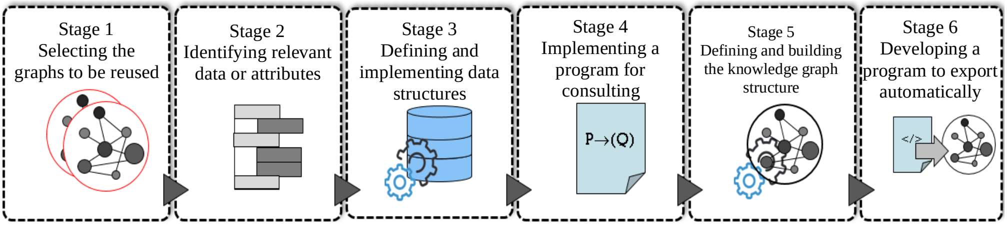

Figure 3. Stages of the method for reusing graphs
Source: created by the authors.

1) Selecting the graph to be reused, which must be downloaded and analyzed locally with the help of the ontology editor.

2) Identifying the relevant attributes to be reused to complete the integrated graph conceptualization.

3) Defining and implementing data structures using the Object-Oriented Programming paradigm for clearer and more efficient handling of the concepts of interest.

4) Implementing a program for consulting the ontologies using SPARQL to automatically obtain the list of concepts of interest and store it in the data structure.

5) Defining and building the T-Box of the knowledge graph module that will be used to represent the concepts and characteristics of interest.

6) Developing a program that automatically generates the knowledge graph with the list of concepts obtained.

The construction of the Symptom knowledge graph module is described to illustrate the execution of this reuse method. One of the important medical terminologies required to construct the general knowledge graph PatientElectronicRecord is disease symptoms. Symptoms are crucial for accurate disease diagnosis for COVID-19.

Selecting the graph to be reused

The SYMPT[9] graph was selected and downloaded, which defines a symptom as: "a perceived change in function, sensation, or appearance, which is reported by the patient, and which may be indicative of a disease." This graph considers the relation between the concept of sign and symptom so that signs are also included, and any term is considered to be either a sign or a symptom. The left side of Figure 4 shows the metrics of the SYMPT ontology, which has a total of 6 271 axioms and 993 classes of symptom types, and a total of 889 subclasses. Similarly, on the right-hand side, the SYMP:0000613 class can be seen, which refers to the "fever" symptom, which in turn is a subclass of the SYMP_0000410 "neurological and psychological symptom" class. All classes in this graph have annotations and characteristics that are useful to know the specific details when selecting the part of the model to export.

One of the important transformations implemented during this modularization process was to use the subclasses of symptoms as instances (or individuals) in the new model. The purpose was to have a model with a hierarchy of the main symptom classes, in which all specific symptoms are included as individuals and not as subclasses, thus facilitating the establishment of relations between diseases and their signs or symptoms.

[9] https://bioportal.bioontology.org/ontologies/SYMP

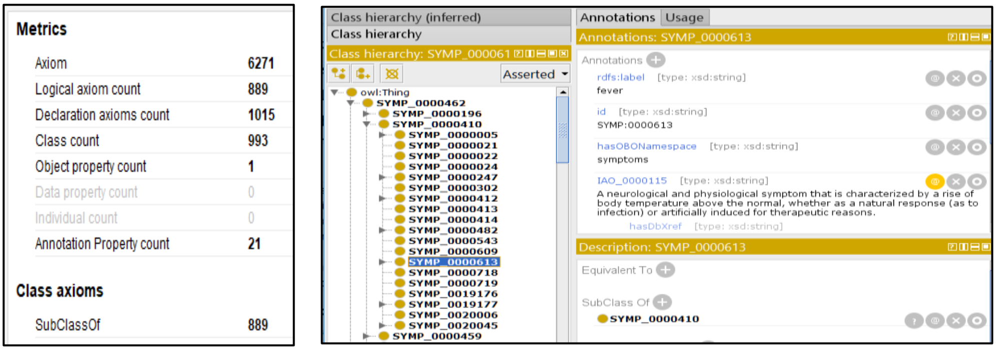

Figure 4. SYMPT ontology visualization (with Protégé)
Source: created by the authors.

Identifying relevant data and attributes

In this stage, an analysis is performed on the SYMPT ontology to identify the concepts, attributes (characteristics), and data types that must be imported and reused in the PatientElectronicRecord general knowledge graph. For ease of reading by the analyst, it is recommended that the SYMPT ontology be saved using the Turtle syntax[10]. For a sample of the data and attributes of this ontology, below is an excerpt from the neurological and psychological symptom category, specifically the concept of SYMP:0000613, which corresponds to the symptom of fever.

http://purl.obolibrary.org/obo/SYMP_0000613

obo:SYMP_0000613 rdf:type owl:Class ;

rdfs:subClassOf obo:SYMP_0000410 ;

obo:IAO_0000115 “Fever is a neurological and physiological symptom characterized by a rise of body temperature above the normal whether a natural response (as to infection) or artificially induced for therapeutic reasons.”^^xsd:string ;

oboInOwl:hasDbXref

“ICD9CM_2005:780.6”^^xsd:string ,

“UMLS_CUI:C0015967”^^xsd:string ,

“UMLS_ICD9CM_2005_AUI:A0058972”^^xsd:string ;

oboInOwl:hasExactSynonym “pyrexia”^^xsd:string ;

oboInOwl:hasOBONamespace “symptoms”^^xsd:string ;

oboInOwl:id “SYMP:0000613”^^xsd:string ;

[10] https://dbpedia.org/page/Turtle_(syntax)

rdfs:label "fever"^^xsd:string .

[rdf:type owl:Axiom ;

owl:annotatedSource obo:SYMP_0000613 ;

owl:annotatedProperty obo:IAO_0000115 ;

owl:annotatedTarget "Fever is a neurological and physiological symptom characterized by a rise of body temperature above the normal whether a natural response (as to infection) or artificially induced for therapeutic reasons."^^xsd:string ;

oboInOwl:hasDbXref "url: http://www2.merriam-webster.com/cgi-bin/mwmednlm?book=Medical&va=fever"^^xsd:string]

The properties that exist in this graph are: auto-generated-by, created_by, creation_date, database_cross_reference, date, default, namespace, deprecated, description, has_alternative_id, has_obo_format_version, has_obo_namespace, id, label, license, notation, part_of, saved-by, term replaced by title. Of which the following were included and adapted as data properties: hasSymptomDefinition, hasSymptomID, hasSymptomName, in addition to the IRI of the class.

Defining and implementing data structures

A Java program was developed based on the OWL API library for symptom extraction and representation. This program consists of the classes Symptom, SymptomOntologyCreation, and OntologyUtils, as shown in the class diagram in Figure 5.

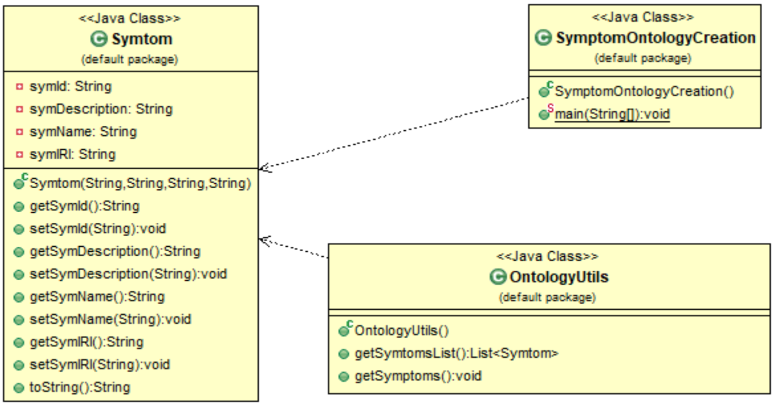

Figure 5. Class diagram of the program that extracts the symptoms and registers them in the new module
Source: created by the authors.

http://dx.doi.org/10.22201/fca.24488410e.2023.4839

Implementing a program for consulting SYMPT ontology

During this stage, a set of methods was implemented in the OntologyUtils class using the RDF4J API to execute queries in the SPARQL language. The purpose was to automate the extraction of symptom information, generate the list of symptom objects, and save this information in a new graph called Symptom.owl. The SPARQL query used is shown below:

```
PREFIX obo: <http://purl.obolibrary.org/obo/>
PREFIX oboInOwl: <http://www.geneontology.org/formats/oboInOwl#>
SELECT ?s ?label ?id ?description
WHERE { ?s rdf:type owl:Class .
s rdfs:subClassOf obo:SYMP_0000462 .
s obo:IAO_0000115 ?description .
s rdfs:label ?label .
s oboInOwl:id ?id .}
```

Defining and building the T-Box of the knowledge graph module

The new model in which the symptom instances are registered was designed. The definition of the Symptom class adheres to the type of data handled by the program that extracts the list of symptoms. In this new graph model, the data characteristics are simplified. Symptom concepts registered as classes and subclasses in the original ontology are registered as instances or individuals in the new graph to make this module lighter and easier to integrate into the general knowledge graph PatientElectronicRecord.

Developing a program that automatically generates the knowledge graph module

Finally, in the OntologyUtils class, methods for extracting symptom information from the graph (getSymptomsList) and for the automatic generation of the graph module to be reused (populateSymptomOntology) were integrated. These methods are executed from a main application (SymptomOntologyCreation) as shown in the following Java code:

```
public class SymptomOntologyCreation {
        public static void main(String[] args) {
        OntologyUtils util = new OntologyUtils();
        //Extraction of the symptom list through SPARQL
```

```
        List<Symtom> symptomList = util.getSymptomsList();
//Symptom list entry in the new graph module
        util.populateSymptomsOntology("Ontologies/Symptom.owl",
symptomList);       }}
```

As a result of the implementation of the method for reuse, a new knowledge graph module called Symptom.owl is obtained, which consists of 1 class, 3 data properties, 259 individuals, and a total of 1292 axioms, as shown in Figure 6.

Metrics

Axiom	1292
Logical axiom count	1028
Declaration axioms count	264
Class count	1
Object property count	0
Data property count	3
Individual count	259
Annotation Property count	1

Figure 6. Symptom graph module metrics
Source: created by the authors.

The reuse method described above (as shown in Figure 3) was implemented for the extraction of modules from the following graphs: NDFRT, National Drug File - Reference Terminology; LOINC, Logical Observation Identifier Names and Codes; NCIT, National Cancer Institute Thesaurus. As a result, the following Knowledge Graphs (GC; Spanish: Grafos de Conocimiento) were obtained:

a) GC Medicament. This graph module includes the definitions of the data properties and object properties identified as relevant from the analysis of the NDF-RT knowledge graph. As a result, the following classes were defined in this Medicament graph: Medicament, ChemicalStructure, MechanismOfAction, PharmacologicalTreatment, NDFRT_Disease, TerapeuticCategory, and PhysiologicalEffect. In these classes, individuals related to Medicament and diseases are recorded.

b) GC LaboratoryTest. The properties identified as relevant in the original LOINC knowledge graph are included in this graph. In particular, laboratory tests for detecting COVID-19 are included. According to the analysis of the information consulted in the original graph, the LaboratoryTest class is defined, which consists of 73 subclasses of the types of laboratory tests in general.

GC Vaccine. This graph includes the definitions of classes, data characteristics, and objects consulted in the original NCIT knowledge graph. Specifically, the conceptual definitions were revised and included to incorporate the various types of vaccines aimed at preventing COVID-19[11]. As a result, the Vaccine graph includes the following vaccine types: COVID-19_Vaccine, mRNA_COVID-19_Vaccine, and SARS- CoV-2_mRNA_Vaccine.

Stage 6: Integration of modules in the main knowledge graph

Once the extraction stage is completed, the integration process is performed by importing the modules into the PatientElectronicRecord general knowledge graph and defining meta-relations between the concepts of the imported graphs according to the requirements of the general model. As a result of this integration, a general model will later allow the representation of Mexican patient cases.

The definition of meta-relations is an important task for integrating external concepts. These meta-relations can be defined using properties between objects or instances of classes. It is important to remember that the properties between objects establish binary relations, i.e., for each property, the domain and rank must be defined. During the integration of the general graph PatientElectronicRecord, the properties' names, domains, and ranks were defined and shown in Table 5.

Table 5
Object characteristics resulting from the integrated model

Object property	Domain class	Rank class	Description
diseaseHasSymptom	NDFRT_Disease	Symptom	Disease has symptoms
hasDiagnosedDisease	ClinicalDiagnosis	NDFRT_Disease	Clinical diagnosis has disease
hasLabTest	ClinicalDiagnosis	LaboratoryTest	Clinical diagnosis has laboratory test
hasPharmalogicalTreatment	ClinicalDiagnosis	PharmalogicalTreatment	Clinical diagnosis has clinical treatment
hasVaccine	Patient	COVID-19_Vaccine	Patient has been vaccinated
requiresLabTest	NDF-RT_Disease	LaboratoryTest	Symptom requires laboratory test

Source: created by the authors.

To view the general integrated model, see Figure 7. Similarly, Figure 8 shows a fragment of the general knowledge graph PatientElectronicRecord Manchester[12] notation code. The header shows the prefixes of the imported graphs followed by the annotation properties and data properties, among others.

[11] https://www.gob.mx/cofepris/acciones-y-programas/vacunas-covid-19-autorizadas
[12] https://www.w3.org/2007/OWL/wiki/ManchesterSyntax

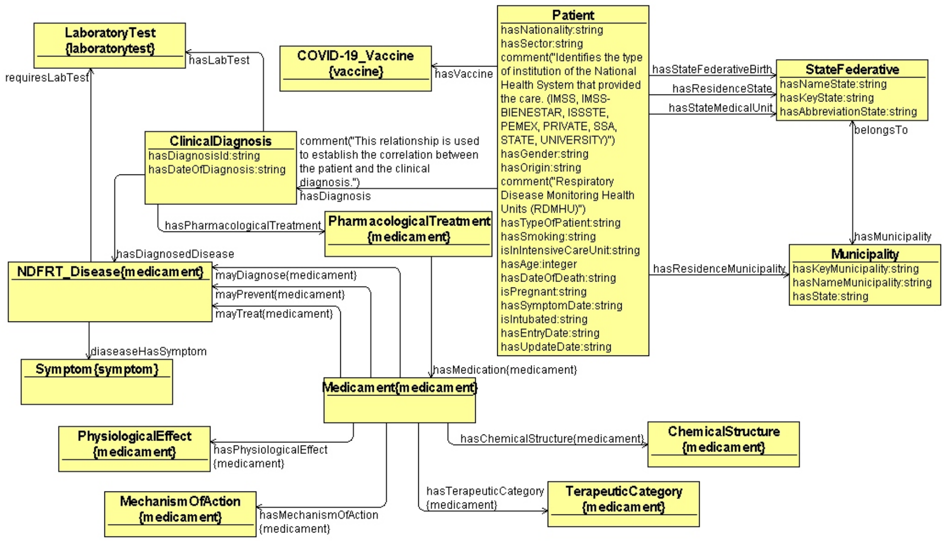

Figure 7. Integrated PatientElectronicRecord knowledge graph
Source: created by the authors.

```
Prefix: : <http://www.semanticweb.org/darinelstein/ontologies/2021/8/PatientElectronicRecord>
Prefix: laboratorytest: <http://www.semanticweb.org/darinelstein/ontologies/2021/LaboratoryTest#>
Prefix: medicament: <http://www.semanticweb.org/darinelstein/ontologies/2021/8/Medicament#>
Prefix: owl: <http://www.w3.org/2002/07/owl#>
Prefix: owlapi: <http://www.semanticweb.org/owlapi#>
Prefix: patientelectronicrecord: <http://www.semanticweb.org/darinelstein/ontologies/2021/8/PatientElectronicRecord#>
Prefix: rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
Prefix: rdfs: <http://www.w3.org/2000/01/rdf-schema#>
Prefix: swrl: <http://www.w3.org/2003/11/swrl#>
Prefix: swrla: <http://swrl.stanford.edu/ontologies/3.3/swrla.owl#>
Prefix: swrlb: <http://www.w3.org/2003/11/swrlb#>
Prefix: symptom: <http://www.semanticweb.org/darinelstein/ontologies/2021/8/Symptom#>
Prefix: vaccine: <http://www.semanticweb.org/darinelstein/ontologies/2021/8/Vaccine#>
Prefix: xml: <http://www.w3.org/XML/1998/namespace>
Prefix: xsd: <http://www.w3.org/2001/XMLSchema#>
Ontology: <http://www.semanticweb.org/darinelstein/ontologies/2021/8/PatientElectronicRecord>
<http://www.semanticweb.org/darinelstein/ontologies/2021/8/PatientElectronicRecord>
Import: <http://www.semanticweb.org/darinelstein/ontologies/2021/LaboratoryTest>
Import: <http://www.semanticweb.org/darinelstein/ontologies/2021/8/Medicament>
Import: <http://www.semanticweb.org/darinelstein/ontologies/2021/8/Symptom>
Import: <http://www.semanticweb.org/darinelstein/ontologies/2021/8/Vaccine>

AnnotationProperty: owlapi:nodeID
AnnotationProperty: rdfs:comment
AnnotationProperty: rdfs:label
AnnotationProperty: swrla:isRuleEnabled
Datatype: rdf:PlainLiteral
Datatype: xsd:boolean
Datatype: xsd:integer
Datatype: xsd:string

ObjectProperty: patientelectronicrecord:belongsTo
    Domain:
        patientelectronicrecord:Municipality
    Range:
        patientelectronicrecord:StateFederative
```

Figure 8. Code fragment of the PatientElectronicRecord general knowledge graph
Source: created by the authors.

Stage 7: Definition of general axioms and rules of inference

Knowledge graphs are implemented by defining logical axioms; strictly, an axiom is an unobjectionable truth about a concept or a fact. From a practical approach, axioms are formal definitions of concepts; that is, they are the set of formal definitions that restrict the belonging of individuals to classes. A logical axiom seen as a restriction establishes the necessary or sufficient conditions for an individual to be a member of or belong to a class. Therefore, the definition of axioms allows the fine-grained and detailed characterization of concepts to be implemented. During this stage of the knowledge graph development, the axioms included in the model were defined, as well as a set of inference rules that allow the automatic generation of new semantic relations and new definitions.

Evaluation of the PatientElectronicRecord general graph

This section describes the implementation of three evaluation approaches: the evaluation of the usability of the graph, the evaluation of the design principles, and the evaluation of the competence of the graph.

Evaluation of the usability of the knowledge graph

The usability of a knowledge graph is intended to measure how easy it is for the user to take advantage of the model by instantiating real-world data. Employing usability, it is possible to know how "adequate" the design of a model is to represent real-world cases. This section presents the method implemented to represent a "real" COVID-19 patient data set, as shown in Figure 9. The DGE of the Mexican government makes available to the general population the data concerning the cases associated with COVID-19 to facilitate access, use, reuse, and redistribution of these data to all users who require it.

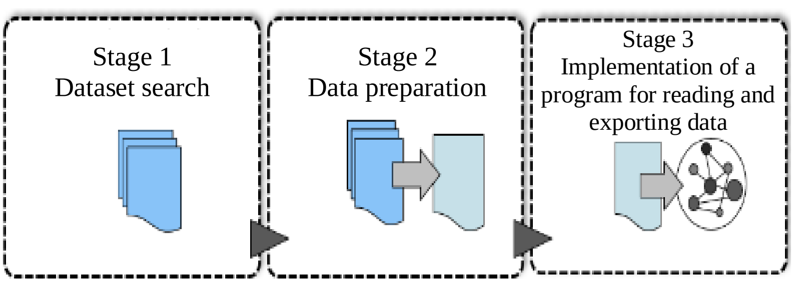

Figure 9. Method for the representation of COVID-19 patient data using the PatientElectronicRecord graph.
Source: created by the authors.

Dataset search

In the usability evaluation stage of the model, public COVID-19 patient datasets were sought to instantiate the general graph and evaluate its usability according to the needs initially proposed. The data used were obtained from the General Directorate of Epidemiology (DGE).

Data preparation

The CSV file with patient data was downloaded. This file contains more than 7 000 000 records of patients treated in the Health Units Monitoring Viral Respiratory Disease (USMER). 500 records were used to test the usability of the model, selecting those of Mexican nationality. The patient records' attributes that coincide with those defined in the general graph were selected to carry out the registration. The source file is shown in Figure 10.

```
"FECHA_ACTUALIZACION","ID_REGISTRO","ORIGEN","SECTOR","ENTIDAD_UM","SEXO","ENTIDAD_NAC","ENTIDAD_RES","MUNICIPIO_RES","TIPO_PACIENTE","FECHA_INGRESO","FECHA_SINTOMAS","FECHA_DEF"
"2022-08-04","z21c19",1,12,"19",2,"19","19","026",1,"2020-11-01","2020-11-01","9999-99-99",97,2,54,1,97,2,2,2,2,2,2,2,2,2,2,2,2,2,2,97,2,97,6,99,"México","97",97
"2022-08-04","z1e370",1,12,"14",1,"14","14","085",1,"2020-04-22","2020-04-18","9999-99-99",97,2,42,1,2,2,2,2,2,1,2,2,2,2,2,2,2,2,1,2,2,97,7,99,"México","97",97
"2022-08-04","z4bf07",2,12,"31",1,"31","31","041",1,"2020-06-13","2020-06-10","9999-99-99",97,2,23,1,2,2,2,2,2,1,2,2,2,2,2,2,2,1,1,2,2,97,7,99,"México","97",97
"2022-08-04","z45dcb",2,12,"09",2,"09","09","003",1,"2020-08-18","2020-08-16","9999-99-99",97,2,42,1,97,2,2,2,2,2,2,2,2,2,2,2,2,1,1,2,2,97,7,99,"México","97",97
"2022-08-04","z5c20f",1,12,"31",2,"31","31","102",1,"2020-05-25","2020-05-22","9999-99-99",97,2,40,1,97,1,1,2,2,2,2,2,2,2,2,2,2,2,1,1,2,97,3,99,"México","97",97
"2022-08-04","z1da2d",1,12,"02",1,"02","02","004",2,"2020-06-18","2020-06-12","2020-07-19",2,1,33,1,2,2,2,98,98,98,98,98,1,98,98,98,98,99,2,97,2,97,6,99,"México","97",2
"2022-08-04","z4dc5f",1,12,"19",1,"19","19","026",1,"2020-07-13","2020-07-13","9999-99-99",97,2,55,1,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,97,2,97,6,99,"México","97",97
"2022-08-04","z23c2e",2,12,"09",1,"09","09","005",1,"2020-02-26","2020-02-20","9999-99-99",97,2,47,1,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,97,2,97,6,99,"México","97",97
"2022-08-04","z2215d",1,12,"32",1,"32","32","056",1,"2020-11-28","2020-11-22","9999-99-99",97,2,40,1,2,2,2,2,2,1,2,1,2,2,2,2,2,2,1,2,2,97,7,99,"México","97",97
"2022-08-04","zzz6fd",1,12,"01",2,"14","32","019",1,"2020-12-24","2020-12-24","9999-99-99",97,2,62,1,97,2,2,2,2,2,2,2,2,2,2,2,2,2,2,97,1,2,7,99,"México","97",97
"2022-08-04","z405fd",1,12,"09",2,"09","09","011",1,"2021-01-25","2021-01-17","9999-99-99",97,2,25,1,97,2,2,2,2,2,2,2,2,2,2,2,2,1,2,97,1,2,7,99,"México","97",97
"2022-08-04","z1b053",2,12,"09",2,"21","15","033",1,"2020-10-10","2020-10-01","9999-99-99",97,2,42,1,97,2,2,2,2,2,2,2,2,2,2,2,2,2,2,97,2,97,1,99,"México","97",97
"2022-08-04","z2a1ad",2,12,"04",2,"04","04","002",1,"2020-06-09","2020-06-05","9999-99-99",97,2,47,1,97,2,2,2,2,1,2,2,2,2,2,2,1,1,1,1,2,97,3,99,"México","97",97
"2022-08-04","z1caed",2,12,"04",2,"04","04","009",1,"2020-08-17","2020-08-14","9999-99-99",97,2,53,1,97,2,2,2,2,2,1,2,2,1,2,2,1,1,2,2,97,7,99,"México","97",97
"2022-08-04","z1b4b0",2,12,"09",2,"09","09","015",1,"2020-12-09","2020-12-04","9999-99-99",97,2,56,1,97,2,2,2,2,2,2,2,2,2,2,2,2,1,2,97,1,2,7,99,"México","97",97
"2022-08-04","z21332",1,12,"31",1,"31","31","050",1,"2020-05-15","2020-05-13","9999-99-99",97,2,19,1,2,2,2,2,2,2,2,2,2,2,2,2,1,1,2,2,97,7,99,"México","97",97
"2022-08-04","z4533d",1,12,"29",1,"21","21","117",2,"2020-07-02","2020-06-28","9999-99-99",2,2,11,1,2,2,2,1,2,2,2,2,2,2,1,2,2,2,2,97,2,97,6,99,"México","97",2
```

Figure 10. Fragment of raw data set
Source: created by the authors.

Finally, the patient data were saved in a patients.xlsx file to be processed through a program for automatic data reading and recording.

Program for the automation of data reading and data recording

To automatically populate the patient records in the general graph, a program was developed to read them sequentially and transform them into a data structure, which is subsequently stored in the knowledge graph. Figure 11 shows the class diagram of this program. Methods were developed to read the data from the Excel file using the Apache POI API[13]. A main application was also developed from which the operations are executed.

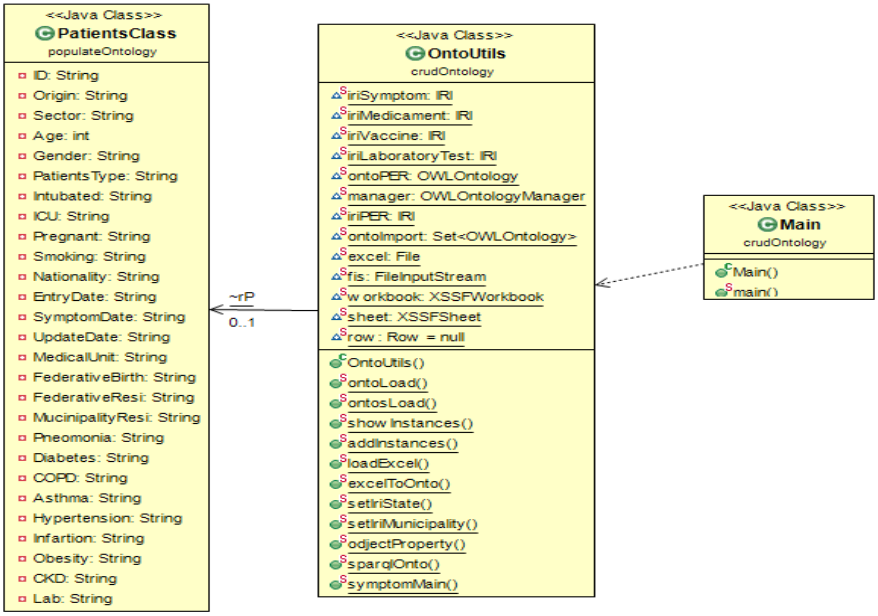

Figure 11. Class diagram for reading and recording patient data in the general graph
Source: created by the authors.

Figure 12 shows an example of a patient registered in the graph PatientElectronicRecord. This is a Mexican patient with the identifier pat01e35a. She is 56 years old, female, from Mexico City in the Coyoacan district, treated in Mexico City at the Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado (ISSSTE). The onset of symptoms is January 08, 2020, and the date of registration of the case is January 09, 2020. Her record states that intubation was not necessary, nor hospitalization in the USMER, although the medical history establishes that the patient has comorbidities of diabetes mellitus and obesity, which represent a high risk of complications. Her record established that the patient should be treated with the medication indicated by the physician.

[13] https://poi.apache.org/

http://dx.doi.org/10.22201/fca.24488410e.2023.4839

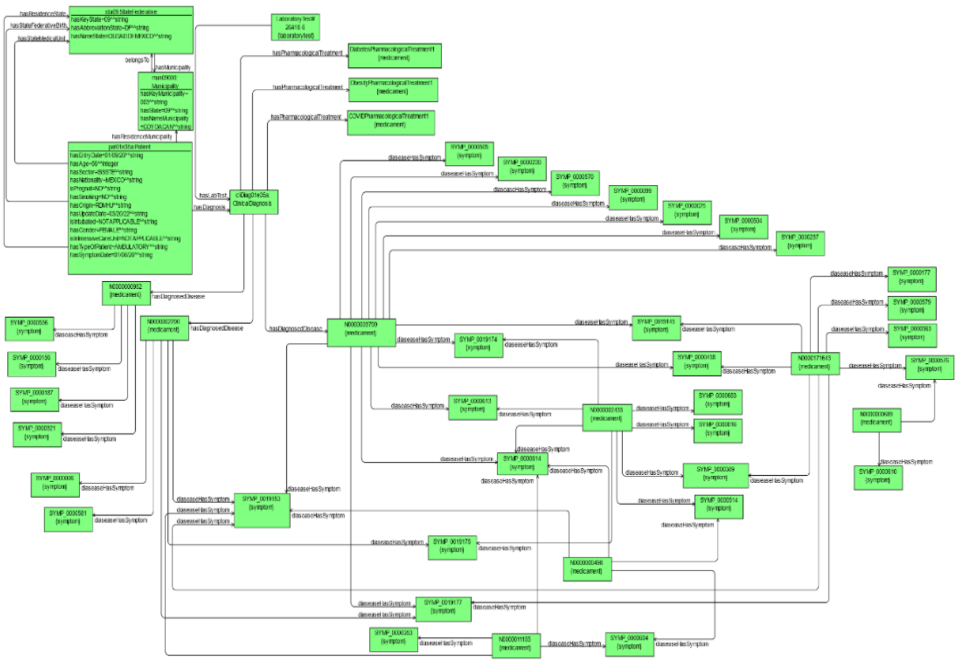

Figure 12. Example of the result of recording data from a Mexican patient with COVID-19
Source: created by the authors.

Evaluation based on the design principles

According to Tom Gruber (1995), design principles are guidelines that determine the design and construction of ontological models. This section reviews the design principles applied in the evaluation of the PatientElectronicRecord graph.

The design principles proposed by Gómez-Pérez (1996) and Duque-Ramos (2014) were selected. As Bravo et al. (2019) mentioned, the construction of ontologies encompasses methods, techniques, and design principles that support the design and construction of efficient ontologies. Therefore, the design principles and methods were considered in the development of this project of the integrated ontology model.

Design principles applied to the evaluation of the knowledge graph

a) Clarity. A knowledge graph should effectively communicate the intended meaning of the defined terms. Definitions should be objective; formalization is the means to this end. Complete definitions are preferable to primitive definitions. All definitions should be documented in natural language. Being a graph for the representation of patient data with COVID-19 and according to the principle of clarity, the concepts for the representation were defined using logical axioms. Moreover, all defined concepts and relations were documented using labels.

b) Coherence: The concepts in the model are defined by logical axioms based on the facts established by the patient data so that an axiom established in the model should not contradict an established definition. The Pellet reasoner[14] was executed from Protégé to verify coherence, showing that the graph is consistent and therefore coherent, as shown in Figure 13.

```
INFO  15:13:43  ------------------------------- Running Reasoner
INFO  15:14:18  Pre-computing inferences:
INFO  15:14:18      - class hierarchy
INFO  15:14:18      - object property hierarchy
INFO  15:14:18      - data property hierarchy
INFO  15:14:18      - class assertions
INFO  15:14:18      - object property assertions
INFO  15:14:18      - same individuals
INFO  15:14:34  Ontologies processed in 50585 ms by Pellet
```

Figure 13. Verification of logical consistency with the reasoner
Source: created by the authors.

c) Extensibility: The design of the knowledge graph should consider in advance the integration of new concepts to expand the graph for the representation of other modules without implying the revision of the whole graph or its affectation.

d) Minimum ontological commitment: Only those concepts that are necessary should be included without going overboard by including definitions that are not strictly indispensable. Consequently, the knowledge graph developed presents a minimum ontological commitment; that is, it can support the required knowledge exchange tasks but with as few assertions as possible about the domain being modeled, allowing the parties committed to the graph the freedom to specialize and instantiate the graph as needed.

[14] https://www.w3.org/2001/sw/wiki/Pellet

Evaluation of the competence of the graph

The competence of a knowledge graph can be defined as the set of questions that the graph can answer according to the knowledge included. This section presents the set of competency questions in natural language, which are translated into the SQWRL query language to execute such queries.

Based on the above, the following competency questions were posed:

1. What are the main symptoms of patients with COVID-19?
2. What is the average age of intubated patients?
3. What is the gender and age of the intubated patients?
4. What is the gender and age of the patients who die?
5. How much time is there between admission to the hospital and the date of death?
6. Are there pregnant patients who were diagnosed with COVID-19?
7. Are there patients who smoke and who were diagnosed with COVID-19?
8. Is there any correlation between patients being intubated if they smoke?
9. What are the characteristics of hospitalized patients who worsen and are intubated?
10. What are the characteristics of hospitalized patients who die?

Figure 14 shows an example of the execution of a query and the execution time. The result is shown in table format with all the records found in the graph. As an example, the execution of question 3 is shown: What is the gender and age of the intubated patients? Translated into SQWRL language as follows:

patientelectronicrecord:hasGender(?patient, ?gender) ^ patientelectronicrecord:isIntubated(?patient, “YES”) ^ patientelectronicrecord:hasAge(?patient, ?age) ^ patientelectronicrecord:Patient(?patient) -> sqwrl:select(?patient, “YES”, ?gender, ?age)

See the S3 tab to review results of the SQWRL query.
The query took 11939 milliseconds. 5 rows were returned.

patient	C1	gender	age
patientelectronicrecord:pat1135b2	YES	FEMALE	91
patientelectronicrecord:patz49a69	YES	FEMALE	66
patientelectronicrecord:patz477ab	YES	MALE	14
patientelectronicrecord:patz4e838	YES	MALE	60
patientelectronicrecord:pat12d70a	YES	MALE	58

Figure 14. Execution time and table of results of 3rd query
Source: created by the authors.

Results

As a result, the graph PatientElectronicRecord, whose structure can be seen in Figure 15, is obtained. This graph includes the modules that were reused (Medicament, COVID-19_Vaccine, LaboratoryTest, and Symptom), which leads to the conclusion that the module extraction method is efficient for the reuse of biomedical graphs, avoiding the import of large amounts of data.

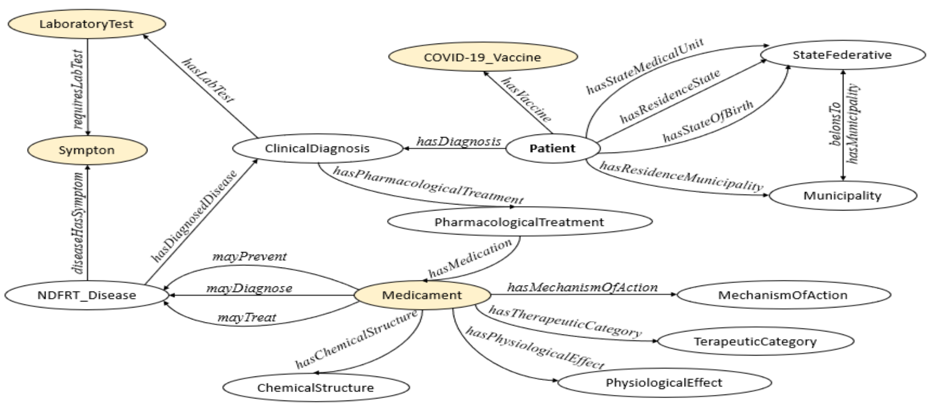

Figure 15. Integrated knowledge graph with entities and relations
Source: created by the authors.

The resulting knowledge graph shows a smaller number of axioms considering the source graphs that were reused. This reduction helps reduce the execution time of reasoners, avoiding the excessive demand of computational resources. The PatientElectronicRecord graph has the metrics shown in Figure 16: 36,594 axioms: 31,687 logical axioms, 17 classes, 21 object properties, 40 data properties, and 4,768 individuals.

Metrics

Axiom	36594
Logical axiom count	31687
Declaration axioms count	4850
Class count	17
Object property count	21
Data property count	40
Individual count	4768
Annotation Property count	4

Figure 16. Metrics of the PatientElectronicRecord graph
Source: created by the authors.

Some general statistics can be obtained based on the data set instantiated in the PatientElectronicRecord graph. For example, 53.7% of the positive cases are female, and 1.4% are pregnant. Considering the total patient data, 0.4% are in the intensive care unit and intubated, and 3.6% of the patients with positive cases are deceased. The average age of patients who died of COVID-19 is 36 years. 8.4% of the patients with positive cases were considered smokers and had more acute complications. The results also show that the average number of days from the patient's admission to a USMER to the patient's death ranged from 2 to 18 days.

Conclusions

One of the areas of research that has gained enormous relevance in recent years is the representation, management, and handling of clinical data and medical specialty data, as well as the efficient handling of large volumes of data. This is mainly due to the COVID-19 pandemic, for which several research institutions around the world have undertaken the task of investigating multiple aspects of COVID-19. This has led to the publication of numerous knowledge bases presented as ontologies, knowledge graphs, or medical vocabularies.

This paper describes the methodology for the development, construction, and integration of a comprehensive and relevant knowledge graph for the representation and management of data on patients diagnosed with COVID-19 in order to integrate and take advantage of all these public and available knowledge resources. In particular, the aim was to design and adapt this graph considering the standards established by the Mexican government and using a set of records published by the DGE through the open data policy. As a result, a general knowledge graph called PatientElectronicRecord was obtained, incorporating concepts and specialized and updated information on medications, diseases, laboratory tests, signs and symptoms, vaccines, and a set of relations between concepts. The general knowledge graph was evaluated considering usability, design principles and competence. The result of these evaluations makes it possible to determine that the development of the graph has been adequate according to the established standards and that it fulfills its purpose of representing knowledge of a domain.

Future work will incorporate the FHIR ontology, a standard to facilitate health data exchange and semantic interoperability between health-related organizations and institutions.

References

Ajami, H., & Mcheick, H. (2018). Ontology-based model to support ubiquitous healthcare systems for COPD patients. Electronics, 7(12), 371. https://doi.org/10.3390/electronics7120371

Bravo, M., González, D., Ortiz, J. A. R., & Sánchez, L. (2020). Management of diabetic patient profiles using ontologies. Contaduría y administración, 65(5), 12. http://dx.doi.org/10.22201/fca.24488410e.2020.3050

Bravo, M., Hoyos Reyes, L. F., & Reyes Ortiz, J. A. (2019). Methodology for ontology design and construction. Contaduría y administración, 64(4). https://doi.org/10.22201/fca.24488410e.2020.2368

Celi, L.A.; Galvin, S.; Davidzon, G.; Lee, J.; Scott, D.; Mark, R. A Database-driven Decision Support System: Customized Mortality Prediction. J. Pers. Med. 2012, 2, 138-148. https://doi.org/10.3390/jpm2040138

de Mello, B. H., Rigo, S. J., da Costa, C. A., da Rosa Righi, R., Donida, B., Bez, M. R., & Schunke, L. C. (2022). Semantic interoperability in health records standards: a systematic literature review. Health and Technology, 12(2), 255-272. https://doi.org/10.1007/s12553-022-00639-w

Duque-Ramos, A., Boeker, M., Jansen, L., Schulz, S., Iniesta, M., & Fernández-Breis, J. T. (2014). Evaluating the good ontology design guideline (GoodOD) with the ontology quality requirements and evaluation method and metrics (OQuaRE). PloS one, 9(8), e104463. https://doi.org/10.1371/journal.pone.0104463

Farion, K., Michalowski, W., Wilk, S., O'Sullivan, D. M., Rubin, S., & Weiss, D. (2009). Clinical decisión support system for point of care use: ontology driven design and software implementation. Methods of information in medicine, 48(4), 381-390. https://doi.org/10.3414/ME0574

Gómez-Pérez, A., Fernández, M., & Vicente, A. D. (1996). Towards a method to conceptualize domain ontologies. Disponible en: https://oa.upm.es/7228/ y Consultado: 08/10/2022

Govindan, K., Mina, H., & Alavi, B. (2020). A decision support system for demand management in healthcare supply chains considering the epidemic outbreaks: A case study of coronavirus disease 2019 (COVID-19). Transportation Research Part E: Logistics and Transportation Review, 138, 101967. https://doi.org/10.1016/j.tre.2020.101967

Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing? International journal of human-computer studies, 43(5-6), 907-928. https://doi.org/10.1006/ijhc.1995.1081

Harry, M.L., Saman, D.M., Truitt, A.R. et al. Pre-implementation adaptation of primary care cancer prevention clinical decision support in a predominantly rural healthcare system. BMC Med Inform Decis Mak 20, 117 (2020). https://doi.org/10.1186/s12911-020-01136-8

Hermsen, E. D., VanSchooneveld, T. C., Sayles, H., & Rupp, M. E. (2012). Implementation of a clinical decision support system for antimicrobial stewardship. infection control and hospital epidemiology, 33(4), 412. https://doi.org/10.1086/664762

Kalra, D. (2006). Electronic health record standards. Yearbook of medical informatics, 15(01), 136-144.

Khan, S., & Yairi, T. (2018). A review on the application of deep learning in system health management. Mechanical Systems and Signal Processing, 107, 241-265. https://doi.org/10.1016/j.ymssp.2017.11.024.

Köhler, S., Schulz, M. H., Krawitz, P., Bauer, S., Dölken, S., Ott, C. E., ... & Robinson, P. N. (2009). Clinical diagnostics in human genetics with semantic similarity searches in ontologies. The American Journal of Human Genetics, 85(4), 457-464. https://doi.org/10.1016/j.ajhg.2009.09.003

McGuinness, D. L., & Van Harmelen, F. (2004). OWL web ontology language overview. W3C recommendation, 10(10), 2004. http://www. w3. org/TR/2004/REC-owl-features-20040210/. Mythili, R., Parthiban, N., & Kavitha, V. (2022). Construction of heterogeneous medical knowledge graph from electronic health records. Journal of Discrete Mathematical Sciences and Cryptography, 25(4), 921-930. https://doi.org/10.1080/09720529.2022.2068604

Oyelade, O. N., & Ezugwu, A. E. (2020). A case-based reasoning framework for early detection and diagnosis of novel coronavirus. Informatics in Medicine Unlocked, 20, 100395. https://doi.org/10.1016/j.imu.2020.100395

Reyes-Peña, C., Tovar, M., Bravo, M., & Motz, R. (2021). An ontology network for Diabetes Mellitus in Mexico. Journal of Biomedical Semantics, 12(1), 1-18. https://doi.org/10.1186/s13326-021-00252-2

Sherimon, P. C., & Krishnan, R. (2016). OntoDiabetic: an ontology-based clinical decision support system for diabetic patients. Arabian Journal for Science and Engineering, 41(3), 1145-1160. https://doi.org/10.1007/s13369-015-1959-4

Whetzel PL, Noy NF, Shah NH, Alexander PR, Nyulas C, Tudorache T, Musen MA. BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. Nucleic Acids Res. 2011 Jul;39(Web Server issue):W541-5. Epub 2011 Jun 14. https://doi.org/10.1093/nar/gkr469

Zhang, Y. F., Gou, L., Zhou, T. S., Lin, D. N., Zheng, J., Li, Y., & Li, J. S. (2017). An ontology-based approach to patient follow-up assessment for continuous and personalized chronic disease management. Journal of biomedical informatics, 72, 45-59. https://doi.org/10.1016/j.jbi.2017.06.021