



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

*Representation and data management from Mexican patients diagnosed with COVID-19 through knowledge graphs*

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

Universidad Autónoma Metropolitana, México

Recibido el 9 de octubre de 2022; aceptado el 26 de abril de 2023  
Disponible en Internet el: 3 de mayo de 2023

## 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

---

\* Autora para correspondencia

Correo electrónico: [mcbc@azc.uam.mx](mailto:mcbc@azc.uam.mx) (M. Bravo).

La revisión por pares es responsabilidad de la Universidad Nacional Autónoma de México.

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

0186- 1042/© 2019 Universidad Nacional Autónoma de México, Facultad de Contaduría y Administración. Este es un artículo Open Access bajo la licencia CC BY-NC-SA (<https://creativecommons.org/licenses/by-nc-sa/4.0/>)

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

## 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.

*JEL Code:* C63, D83, I00, L86

*Keywords:* Integration of knowledge graphs; reuse of graphs; biomedical knowledge graphs

---

## Introducción

Durante los últimos años se ha generado un enorme volumen de datos sobre pacientes que han padecido COVID-19, algunos resultan en complicaciones graves como es la intubación o el fallecimiento. El Gobierno de México a través de la Dirección General de Epidemiología (DGE)<sup>1</sup> realiza el registro de los datos de pacientes diagnosticados con COVID-19. Este registro consiste de un archivo CSV descargable con más de 7 millones de registros. Para realizar la gestión de grandes cantidades de datos y mejorar la administración de sistemas de salud, Khan y Yairi (2018) realizaron una revisión de tecnologías novedosas que permitan manejar adecuadamente la escalabilidad, así como la explotación de los datos para generar conocimiento derivado. De acuerdo con Kalra (2006), el registro, actualización, mantenimiento y explotación de datos de pacientes de forma automatizada es indispensable cuando se trabaja en el cuidado de la salud, para tener un mejor control y administración de la información. Los modelos de representación

---

<sup>1</sup> <https://datos.gob.mx/busca/dataset/informacion-referente-a-casos-covid-19-en-mexico>

basado en grafos de conocimiento permiten que se ejecuten razonadores que producen nuevas relaciones entre los datos mediante la definición de reglas y axiomas lógicos. En este artículo se describe la metodología de diseño, construcción e integración del grafo de conocimiento para la representación de datos de pacientes mexicanos con COVID-19. Como fuente de datos se utilizó el registro de casos asociados a COVID-19 proporcionados por la DGE. Mediante el empleo de herramientas de consulta e inferencia lógica sobre los grafos de conocimiento, se explotan los datos de pacientes con COVID-19 representados en el grafo de conocimiento.

La metodología de diseño y construcción del grafo de conocimiento se orientó a la extracción y reutilización de módulos de grafos de conocimiento preexistentes relacionados con el área médica. La representación basada en grafos tiene ventajas, ya que el manejo de tripletas <sujeito, predicado, objeto> ofrece una gran flexibilidad para el escalamiento de datos mediante la referenciación de IRI<sup>2</sup>s, permitiendo que se vinculen definiciones de otros grafos de conocimiento de forma sencilla. Para la evaluación del grafo de conocimiento que se describe en este artículo, se usaron tres enfoques: se evalúa la competencia del grafo, la calidad mediante principios de diseño y la usabilidad del grafo mediante la instanciación de un conjunto de datos de pacientes mexicanos con casos positivos de COVID-19.

Las principales contribuciones que se presentan en este artículo son:

- a) Metodología para la construcción del grafo de conocimiento considerando la extracción y reutilización de módulos de grafos de conocimiento en el área médica.
- b) Grafo de conocimiento integrado para la gestión de datos de pacientes mexicanos, el cual incluye los conceptos relacionados con el COVID-19, concretamente términos como vacunas, pruebas de laboratorio, medicamentos, enfermedades, síntomas y diagnósticos.
- c) Método de evaluación del grafo de conocimiento integrado basado en tres enfoques: evaluación de la usabilidad, evaluación basada en principios de diseño, y la evaluación de la competencia del grafo.

El resto del artículo está organizado de la siguiente manera. En la Sección Trabajos Relacionados se presenta una revisión de artículos de investigación relacionados con el desarrollo de métodos para la creación y evaluación de grafos de conocimiento para la gestión de datos clínicos de pacientes o para la gestión de datos de salud. En la Sección Metodología de Desarrollo del Grafo de Conocimiento se describen las etapas metodológicas que se definieron para la construcción e integración del Grafo de Conocimiento. En la Sección Evaluación del Grafo General se describen tres enfoques de evaluación que se aplicaron al grafo resultante. En la Sección Resultados se describe el grafo de conocimiento integrado y poblado con las instancias. Finalmente, se presentan las conclusiones.

---

<sup>2</sup> Identificador de recursos internacionalizado

## **Trabajos relacionados**

La representación y gestión de datos clínicos de pacientes es un tema de investigación que ha cobrado relevancia en los últimos años; esto se debe principalmente a que se busca resolver problemas relacionados con la interoperabilidad e intercambio de información médica especializada entre diversos usuarios de la salud. En esta sección se presenta una revisión de trabajos de investigación relacionados con los temas de este artículo.

En lo referente al registro electrónico de datos clínicos de pacientes, Farion et al. (2009), presentan el desarrollo y evaluación de un modelo de representación que sirve de apoyo en la toma de decisiones por parte de los médicos; este modelo se emplea en un entorno de atención aguda y de emergencias. Los autores describen la implementación de un nuevo diseño de protocolo de emergencia móvil 2, que proporciona un entorno unificador que puede manejar múltiples aplicaciones clínicas ejecutadas en múltiples plataformas, utilizando una base de conocimiento y los modelos derivados para representar componentes clave de un Sistema de Soporte para la Toma de Decisiones Clínicas (CDSS, por sus siglas en inglés Clinical Decision Support System).

En Köhler S. et al. (2009) los autores describen un sistema denominado Phenomizer, el cual no es un sistema experto sino más bien un sistema para expertos que pueden usar el sistema para apoyarse durante el proceso de diagnóstico diferencial en genética humana. El sistema Phenomizer puede indicar si las características de los datos clínicos ingresados por el médico son altamente sugestivas de un diagnóstico dado.

En Celi G. et al. (2012), los autores presentan una herramienta dinámica de representación y gestión de perfiles de pacientes, facilitando el apoyo en la toma de decisiones basada en datos empíricos. En este enfoque, el modelo probabilístico se realiza en subconjuntos de pacientes de la propia institución en lugar de poblaciones de pacientes heterogéneas de diferentes centros. En este modelo se adopta el término Experiencia Colectiva para este enfoque, ya que la información se extrae de la experiencia de varios médicos y ésta se almacena en el sistema de registros médicos electrónicos.

En Elizabeth D. Hermsen et al. (2012), los autores describen las ventajas de la representación de datos de pacientes mediante un CDSS, ya que éste permite mejorar la toma de decisiones sobre la administración de antimicrobianos mediante la disponibilidad de una combinación de datos y costos específicos del paciente. El uso de este sistema permite obtener varios beneficios, incluida la disminución de eventos adversos, la reducción de la duración de la estadía de un paciente en el centro de cuidado de la salud, la disminución de costos durante la estadía y el uso de antimicrobianos más apropiado para el paciente.

En Sherimon y Krishnan (2016), los autores describen OntoDiabetic un CDSS basado en una base de conocimiento para evaluar los factores de riesgo y generar sugerencias de tratamiento para pacientes que presentan diabetes. Los autores definen un conjunto de reglas de inferencia para obtener información sobre el estado de salud del paciente. Este modelo de ontología carece de información actualizada sobre fármacos, enfermedades y relaciones importantes entre ellos.

En Zhang, Gou, et al. (2017), los autores abordan las complejas interacciones entre los factores de riesgo, las enfermedades, las condiciones del paciente y las modalidades de tratamiento. Estas acciones incluyen estrategias de epidemiología y vigilancia para monitorear las tendencias y rastrear el progreso, políticas y enfoques ambientales para promover la salud.

En Ajami y Mcheick (2018), los autores presentan el desarrollo de un modelo que permite la representación de información de pacientes para crear entornos seguros para los pacientes con enfermedad pulmonar obstructiva crónica (EPOC). Este modelo está basado en la descripción formal de bases de conocimiento de un dominio relacionado con la salud, utiliza el lenguaje de reglas de la Web Semántica (SWRL). La base de conocimiento contiene todos los conceptos relevantes relacionados con EPOC, incluida la información personal del paciente, localización, actividad, síntomas, factores de riesgo, resultados de exámenes de laboratorio y plan de tratamiento.

En Oyelade et al. (2020), se presenta la construcción de una base de conocimiento para la representación de perfiles de pacientes, el marco propuesto se logra mediante la formalización del conocimiento. El resultado demostró un rendimiento interesante en comparación con estudios similares de razonamiento basado en casos (CBR) de última generación utilizando fuzzyCBR. La base de conocimiento es lo suficientemente general para su adopción y generalización a los problemas de diagnóstico asociados con otros fines médicos y enfermedades presentes en pacientes.

En Govindan, Mina, Alavi (2020), los autores presentan un sistema práctico para la representación de perfiles de pacientes como CDSS para clasificar a los miembros de la comunidad; y como consecuencia, gestionar la demanda y controlar los brotes epidémicos en la cadena de suministro de salud. En el enfoque propuesto, los usuarios se agrupan primero de acuerdo con dos criterios, rango de edad y enfermedades preexistentes (como diabetes, problemas cardíacos o presión arterial alta).

En Harry et al. (2020), se describe un sistema de representación de perfiles de pacientes para la toma de decisiones basadas en el Expediente Clínico Electrónico (EHR) que puede mejorar la prevención y la detección del cáncer en la atención primaria. Para ello se empleó el marco consolidado para la implementación de la investigación (CFIR, por sus siglas en inglés Consolidated Framework for Implementation Research).

De acuerdo con Bravo et al. (2020), la representación de perfiles de pacientes mediante un modelo de representación semántica ofrece beneficios para la inferencia y razonamiento automático, facilitando la identificación de casos de riesgo.

En Reyes-Peña et al. (2021) se describe una metodología para la construcción de una red de ontologías mediante la cual se realiza la representación de expedientes clínicos de pacientes con diabetes mellitus tipo 2.

Un enfoque similar al que se presenta en este artículo es el que presentan Mythili et al. (2022), quienes construyen un grafo de conocimiento a partir de registros de expedientes clínicos, identifican las relaciones entre las entidades de los pacientes, enfermedades y medicamentos. El grafo se construye a partir de diferentes fuentes de datos, y la información es extraída en forma de consultas. Los pasos que se implementaron para construir el grafo de conocimiento son: recolección de datos, reconocimiento de entidades nombradas, normalización de entidades, raqueo de entidades, y redes neuronales basadas en grafos.

Para una revisión más amplia sobre trabajos relacionados con la interoperabilidad semántica entre expedientes clínicos electrónicos, en De Mello et al. (2022) los autores describen las principales áreas de aplicación como son: búsqueda de la calidad en la salud, reducción de errores médicos, control y monitoreo de enfermedades y cuidado individualizado del paciente.

En este artículo se reporta una metodología que permite construir el grafo de conocimiento para la representación de perfiles de pacientes diabéticos tomando en consideración la existencia de terminologías médicas para su reutilización, así como conjuntos de datos de pacientes que padecieron COVID-19. Como resultado la metodología de construcción cubre necesidades específicas del sistema de salud en México, así como casos de pacientes reales.

## **Metodología de desarrollo del grafo de conocimiento**

El proceso de diseño, desarrollo e integración de un nuevo Grafo de Conocimiento implica la definición de una metodología que contemple desde la identificación de los conceptos clave que deben ser representados, la incorporación y reutilización de grafos existentes, hasta la evaluación del grafo resultante. En esta sección se describe la metodología implementada (ver Figura 1). Esta metodología contempla las siguientes etapas:

1. Recopilación y análisis de información relacionada con el dominio de representación.
2. Búsqueda y selección de grafos de conocimiento existentes relacionados con el dominio.
3. Diseño del modelo del Grafo de Conocimiento General.

4. Construcción del Grafo de Conocimiento General.
5. Extracción de módulos de grafos existentes para su reutilización.
6. Integración de módulos en el Grafo principal.
7. Definición de axiomas generales y reglas de inferencia.



Figura 1. Etapas de la metodología de desarrollo del Grafo de Conocimiento.

Fuente: Propia de los autores.

### *Etapa 1; recopilación y análisis de información*

El primer paso es contar con fuentes de información válidas y confiables. Esto resulta de mayor relevancia cuando la información que se representará en el grafo corresponde al dominio médico y de cuidado de la salud. Por lo anterior, lo primero que se realiza es la búsqueda de fuentes de información confiables acerca del coronavirus SARS-CoV-2, con la finalidad de conocer los términos o entidades conceptuales que son clave, así como las relaciones que pueden existir entre los conceptos.

Las fuentes de información que se consideraron son las siguientes:

- a) Página oficial con información acerca del COVID-19<sup>3</sup>, la cual es publicada por las autoridades de salud del Gobierno de México. En este portal se puede consultar la información oficial respecto a los programas de vacunación, el acceso a datos abiertos, y recomendaciones para la población, entre otros.
- b) Norma Oficial Mexicana NOM017 SSA2 2012<sup>4</sup>, la cual establece los criterios, especificaciones y directrices de operación del Sistema Nacional de Vigilancia Epidemiológica.
- c) Lineamiento Estandarizado para la Vigilancia Epidemiológica y por Laboratorio de la enfermedad respiratoria viral<sup>5</sup>.

<sup>3</sup> <https://coronavirus.gob.mx/covid-19/>

<sup>4</sup> [https://epidemiologia.salud.gob.mx/gobmx/salud/documentos/manuales/00\\_NOM-017-SSA2-2012\\_para\\_vig\\_epidemiologica.pdf](https://epidemiologia.salud.gob.mx/gobmx/salud/documentos/manuales/00_NOM-017-SSA2-2012_para_vig_epidemiologica.pdf)

<sup>5</sup> [https://www.gob.mx/cms/uploads/attachment/file/715464/Lineamiento\\_VE\\_y\\_Lab\\_Enf\\_Viral\\_05042022.pdf](https://www.gob.mx/cms/uploads/attachment/file/715464/Lineamiento_VE_y_Lab_Enf_Viral_05042022.pdf)

d) Guía clínica para el tratamiento de la COVID-19 en México<sup>6</sup>. Esta guía fue desarrollada con representantes de todas las instituciones públicas del sector salud. Esta guía describe los medicamentos que se pueden utilizar en el manejo de la COVID-19, el algoritmo de tratamiento para pacientes con COVID-19, entre otros aspectos y guías relevantes.

e) Guías internacionales y nacionales, manuales para el tratamiento y seguimiento de pacientes, las guías nacionales han sido publicadas por el Gobierno de México en conjunto con los datos abiertos de la DGE.

f) Grafos de Conocimiento descritos en los trabajos relacionados que fueron desarrollados para la representación de perfiles de pacientes en el dominio médico.

Con base en esta información se realizó la identificación de los siguientes términos relevantes para la representación y gestión de datos de pacientes: Paciente, Diagnóstico, Enfermedad, Medicamento, Síntoma, Prueba de Laboratorio y Vacuna. Adicionalmente se consideró incluir el registro de información del municipio y estado donde reside y dónde nació el paciente.

## *Etapa 2; búsqueda y selección de grafos de conocimiento existentes*

El objetivo de esta etapa es realizar la búsqueda de grafos de conocimiento para determinar si existen grafos que cuenten con los conceptos que se requieren en el grafo general que se va a integrar. Bioportal<sup>7</sup> es uno de los repositorios más completos de bases de conocimiento biomédicas. Una vez que se han identificado los grafos que cuentan con representaciones relacionadas, éstos se deben revisar y estudiar para conocer más detalladamente su estructura, de tal forma que se pueda decidir si es conveniente el reutilizarlos total o parcialmente. Por ello, es importante conocer las clases y conceptos médicos relacionados con la enfermedad del COVID-19, así como las propiedades que éstos manejan, tal como menciona Whetzel et al. (2011).

De acuerdo con los requerimientos conceptuales considerados, el modelo integral debería incluir grafos relacionados con los conceptos de: Enfermedad, Medicamento, Signo y Síntoma, Prueba de Laboratorio y Vacuna. El primer grafo que se consultó fue NDFRT (National Drug File - Reference Terminology), ya que este incluye conceptos relacionados con los medicamentos como son: los efectos fisiológicos, las categorías terapéuticas, la estructura química, las enfermedades que puede tratar, las enfermedades que puede prevenir, entre otros. NDFRT cuenta con 36,200 clases aproximadamente. Por otra parte, LOINC (Logical Observation Identifier Names and Codes), es un grafo que incluye conceptos

---

<sup>6</sup>[https://www.gob.mx/cms/uploads/attachment/file/659911/GuiaTx\\_COVID19\\_Consenso\\_2021.08.02\\_compressed.pdf](https://www.gob.mx/cms/uploads/attachment/file/659911/GuiaTx_COVID19_Consenso_2021.08.02_compressed.pdf)

<sup>7</sup> <https://bioportal.bioontology.org/>

de todo tipo de pruebas de laboratorio, incluyendo aquellas pruebas de laboratorio para descartar o confirmar casos positivos de COVID-19, cuenta con 281,878 clases aproximadamente. Otro grafo que se revisó es SYMP (Symptom), este grafo describe un conjunto amplio de síntomas de manera general, síntomas ocasionados por la enfermedad, cuenta con aproximadamente 1000 clases. Por último, se consultó el grafo NCIT (National Cancer Institute Thesaurus), este grafo incluye un amplio conjunto de conceptos biomédicos, desde la representación de procesos biológicos y los materiales biomédicos de los que se deriva el módulo de vacunas para COVID-19, cuenta con más de 174,278 clases.

La Tabla 1 muestra el nombre (o acrónimo), el significado y el Identificador de Recursos Internacionalizado (IRI) de cada uno de los grafos seleccionados para reutilización.

Tabla 1  
**Grafos de conocimiento biomédico considerados para la reutilización parcial.**

Grafo	Significado en inglés	IRI
NDFRT	National Drug File - Reference Terminology Logical Observation	<a href="https://bioportal.bioontology.org/ontologies/NDFRT">https://bioportal.bioontology.org/ontologies/NDFRT</a>
LOINC	Identifier Names and Codes	<a href="https://bioportal.bioontology.org/ontologies/LOINC">https://bioportal.bioontology.org/ontologies/LOINC</a>
SYMP	Symptom	<a href="https://bioportal.bioontology.org/ontologies/SYMP">https://bioportal.bioontology.org/ontologies/SYMP</a>
NCIT	National Cancer Institute Thesaurus	<a href="https://bioportal.bioontology.org/ontologies/NCIT">https://bioportal.bioontology.org/ontologies/NCIT</a>

Fuente: Propia de los autores.

### *Etapa 3; diseño del modelo general del grafo de conocimiento general*

El diseño que da soporte al Grafo de Conocimiento General se realizó tomando en consideración el resultado del análisis de la información relacionada con normas nacionales e internacionales para el manejo de información de pacientes con COVID-19. Mediante un proceso de elicitación de términos, se identificaron los conceptos principales que deben ser representados en el grafo de conocimiento general. En esta sección se describen los criterios del diseño cada uno de estos conceptos.

a) El grafo de conocimiento general se diseñó como un modelo basado en módulos, donde cada módulo representa un concepto principal, de tal forma que permita tanto la construcción de grafos desde cero, como la reutilización de otros grafos, este grafo general se denominó PatientElectronicRecord (Registro electrónico de paciente). En este grafo general se describen las clases y propiedades para la representación de datos de pacientes, así como el diagnóstico de la enfermedad del COVID-19, también se definen las meta-relaciones entre los distintos módulos o grafos de conocimiento a reutilizar.

b) Para la información sobre vacunas, se consideró reutilizar la información contenida en el grafo NCIT, el cual representa la información de las vacunas que han sido desarrolladas y aprobadas para uso de emergencia por organizaciones nacionales e internacionales. Se diseñó el módulo de grafo denominado Vaccine (Vacuna), el cual incluye los conceptos y relaciones para representar el conjunto de vacunas para la prevención del COVID-19.

c) Para la representación de medicamentos, se consideró reutilizar el grafo NDFRT. Este grafo incluye los atributos sobre los medicamentos, así como sus relaciones con las enfermedades, los componentes químicos, entre otros. Por ello se diseñó un módulo de grafo que incluye los conceptos de enfermedad, medicamento, efecto fisiológico y componente químico, entre otros. Asimismo, se definieron las relaciones semánticas entre estos conceptos, este módulo de grafo se nombró Medicament (Medicamento).

d) Para la representación de pruebas de laboratorio se consideró reutilizar la información del grafo LOINC. Por lo tanto, se diseñó un módulo de grafo denominado LaboratoryTest (Prueba de laboratorio), este representa los tipos de pruebas de laboratorio que existen para diagnosticar o descartar alguna enfermedad, así como las pruebas para determinar si existe un caso positivo por el contagio de COVID-19.

e) Por último, se diseñó un módulo de grafo denominado Symptom (Síntoma) para representar los síntomas asociados a las enfermedades, incluye la relación semántica entre síntomas y enfermedades.

En la Figura 2 se muestran los cinco subgrafos o módulos integrados en el grafo de conocimiento, así como las meta-relaciones entre estos grafos.

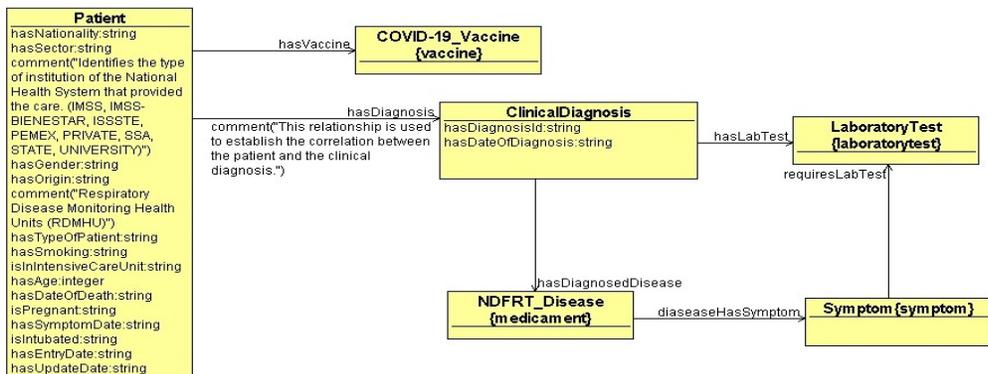


Figura 2. Modelo general de los módulos que integran el grafo de conocimiento PatientElectronicRecord.

Fuente: Propia de los autores.

#### *Etapa 4; construcción del grafo de conocimiento general*

El grafo PatientElectronicRecord, incluye los conceptos que se muestran en la Tabla 2.

Tabla 2

Conceptos definidos en el grafo de conocimiento PatientElectronicRecord.

Concepto	Descripción
ClinicalDiagnosis	Expediente del diagnóstico clínico del paciente.
Patient	Datos del registro del paciente.
StateFederative	Estado federativo de la república mexicana.
Municipality	Municipio de estado.

Fuente: Propia de los autores.

Para cada uno de los conceptos de la Tabla 2 se definieron las propiedades de datos que caracterizan a los individuos que pertenece a cada concepto (ver Tabla 3).

Tabla 3

Propiedades de datos del grafo de conocimiento PatientElectronicRecord.

Propiedad de dato	Dominio	Rango	Descripción
hasOrigin	Patient	integer	USMER <sup>8</sup> de origen del paciente.
hasSector	Patient	string	Institución del Sistema de Salud que brindó la atención.
hasGender	Patient	string	Género del paciente.
hasAge	Patient	integer	Edad del paciente.
hasTypeOfPatient	Patient	string	Paciente ambulatorio u hospitalizado.
hasNationality	Patient	string	Nacionalidad del paciente.
hasEntryDate	Patient	string	Fecha de ingreso del paciente a USMER.
hasSymptomDate	Patient	string	Fecha de síntomas del paciente.
hasSymptomUpdate	Patient	string	Fecha de actualización de síntomas del paciente.
hasDateOfDeath	Patient	string	Fecha de deceso del paciente.
hasSmoking	Patient	string	El paciente fuma.
isInIntensiveCareUnit	Patient	string	Se encuentra en unidad de cuidados intensivos.
isPregnant	Patient	string	La paciente se encuentra embarazada.
isIntubated	Patient	string	El paciente esta intubado.
hasAbbreviationState	StateFederative	string	Abreviación del nombre de estado.
hasKeyState	StateFederative	string	Clave de estado.
hasNameState	StateFederative	string	Nombre de estado.
hasState	Municipality	string	Municipio pertenece a estado.
hasNameMunicipality	Municipality	string	Nombre del municipio.
hasKeyMunicipality	Municipality	string	Clave del municipio.

Fuente: Propia de los autores.

<sup>8</sup> Unidad de Salud Monitora de Enfermedad Respiratoria viral

En la Tabla 4 se muestran las propiedades entre conceptos (meta-relaciones) del grafo de conocimiento general PatientElectronicRecord.

Tabla 4

Propiedades entre conceptos del grafo de conocimiento PatientElectronicRecord.

Propiedad	Dominio	Rango	Descripción
belongsTo	Municipality	StateFederative	Municipio pertenece a estado.
hasDiagnosis	Patient	ClinicalDiagnosis	Paciente tiene diagnóstico clínico.
hasMunicipality	StateFederative	Municipality	Estado federativo tiene municipio.
hasResidenceMunicipality	Patient	Municipality	Municipio de residencia.
hasResidenceState	Patient	StateFederative	Estado de residencia del paciente.
hasStateFederativeBirth	Patient	StateFederative	Estado de nacimiento del paciente.
hasStateMedicalUnit	Patient	StateFederative	Paciente tiene unidad médica.

Fuente: Propia de los autores.

### *Etapa 5; extracción de módulos de grafos existentes para su reutilización*

La reutilización basada en módulos ha sido bien aceptada por la comunidad de la Web Semántica, ya que generar un modelo más simplificado al original es de gran ayuda. Como se señaló en la Etapa 3 de esta metodología, se identificaron un conjunto de grafos para ser reutilizados; por lo tanto, se definió un método de reutilización basado en la extracción de módulos a partir de los grafos que se listan en la Tabla 1. Este método consiste en la extracción de uno o varios módulos de un grafo de conocimiento, el diseño y generación automática de un modelo simplificado y la reutilización del modelo simplificado mediante su importación en el grafo general, tal como se muestra en la Figura 3.

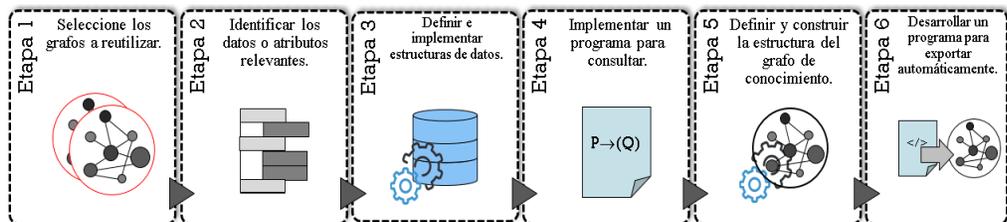


Figura 3. Etapas de método para reutilización de grafos.

Fuente: Propia de los autores.

- 1) Seleccionar el grafo a reutilizar, el cual debe ser descargado y analizado localmente con ayuda del editor de ontologías.

- 2) Identificar los atributos relevantes a reutilizar para completar la conceptualización del grafo que se está integrando.
- 3) Definir e implementar estructuras de datos usando el paradigma de Programación Orientada a Objetos para el manejo más claro y eficiente de los conceptos de interés.
- 4) Implementar un programa para consultar las ontologías usando SPARQL y obtener automáticamente la lista de conceptos de interés y almacenarla en la estructura de datos.
- 5) Definir y construir el T-Box del módulo de grafo de conocimiento que se utilizará para representar los conceptos y propiedades de interés.
- 6) Desarrollar un programa que genere automáticamente el grafo de conocimiento con la lista de conceptos obtenidos.

Para ilustrar la ejecución de este método de reutilización se describe la construcción del módulo del grafo de conocimiento Symptom. Una de las terminologías médicas importantes que fueron requeridas para construir el grafo de conocimiento general PatientElectronicRecord es el de los síntomas de las enfermedades. Los síntomas son cruciales para el diagnóstico certero de la enfermedad por COVID-19.

### *Seleccionar el grafo a reutilizar*

Se seleccionó y descargó el grafo SYMPT<sup>9</sup>, el cual define un síntoma como: “un cambio percibido en la función, sensación o apariencia, el cual es informado por el paciente y que puede ser indicativo de una enfermedad”. Esta grafo considera la relación que existe entre el concepto de signo y síntoma, por lo que se incluyen también los signos y se considera que algún término pueda ser un signo o un síntoma. En la parte izquierda de la Figura 4 se muestran las métricas de la ontología SYMPT, la cual cuenta con un total de 6271 axiomas y 993 clases de los tipos de síntomas, y un total de 889 subclases. De igual forma, en la parte derecha se puede visualizar la clase SYMP:0000613 la cual se refiere al síntoma de “fever” o fiebre, que a su vez es una subclase de la clase SYMP\_0000410 “neurological and psychological symptom”. Todas las clases en este grafo cuentan con anotaciones y propiedades que son útiles para conocer los detalles específicos cuando se selecciona la parte del modelo a exportar.

Una de las transformaciones importantes que se implementaron durante este proceso de modularización fue el de utilizar las subclases de síntomas como instancias (o individuos) en el nuevo modelo. Con el propósito de contar con un modelo donde se tenga una jerarquía de las clases principales

---

<sup>9</sup> <https://bioportal.bioontology.org/ontologies/SYMP>

de síntomas, dentro de las cuales se incluyan todos los síntomas específicos como individuos y no como subclases, facilitando de esta manera establecer relaciones entre enfermedades y sus signos o síntomas.

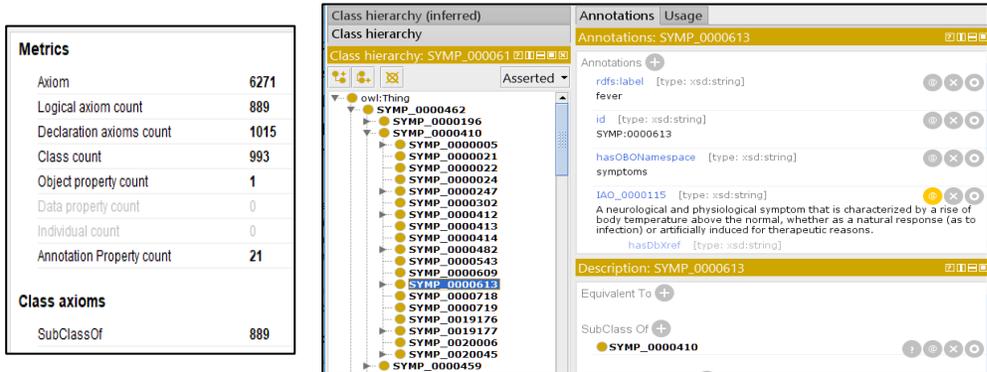


Figura 4. Visualización de la ontología SYMPT (con Protégé).  
 Fuente: Propia de los autores.

### Identificación de los datos y atributos relevantes

En esta etapa se realiza un análisis sobre la ontología SYMPT para identificar los conceptos, atributos (propiedades) y tipos de datos que se requieren para ser importados y reutilizados en el grafo de conocimiento general PatientElectronicRecord. Para facilitar la lectura por parte del analista, se recomienda guardar la ontología SYMPT utilizando la sintaxis Turtle<sup>10</sup>. Para muestra de los datos y atributos de esta ontología, en la parte de abajo se muestra un extracto de la categoría de síntomas neurológicos y psicológicos, específicamente el concepto de SYMPT:0000613, que se corresponde con el síntoma de la fiebre.

```
### http://purl.obolibrary.org/obo/SYMP_0000613
obo:SYMPT_0000613    rdf:type owl:Class ;
                    rdfs:subClassOf obo:SYMPT_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
```

<sup>10</sup> [https://dbpedia.org/page/Turtle\\_\(syntax\)](https://dbpedia.org/page/Turtle_(syntax))

```
"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 ;
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 ]
```

Las propiedades que existen en dicho grafo son: 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. De las cuales se incluyeron y adaptaron como propiedades de dato: hasSymptomDefinition, hasSymptomID, hasSymptomName, además de la IRI de la clase.

### *Definir e implementar las estructuras de datos*

Para la extracción y representación de síntomas se desarrolló un programa en Java basado en la librería OWL API. Este programa consiste en las clases Symptom, SymptomOntologyCreation y OntologyUtils, tal como se muestra en el diagrama de clases de la Figura 5.

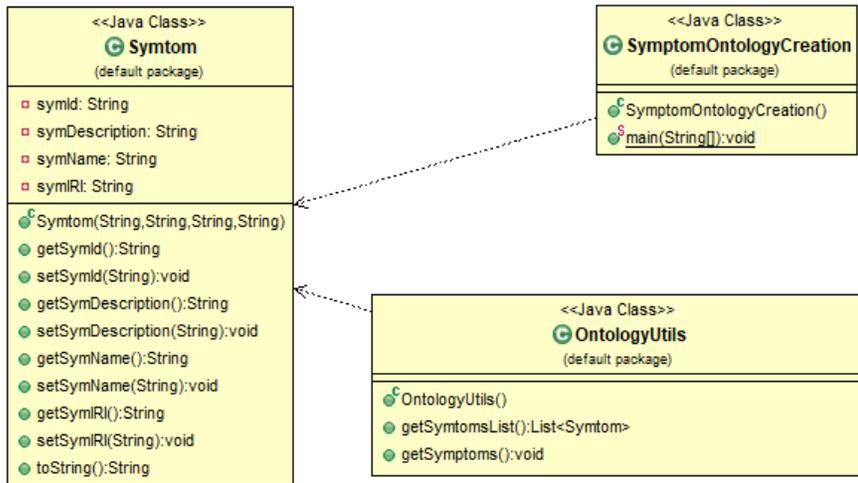


Figura 5. Diagrama de clases del programa que extrae los síntomas y los registra en el nuevo módulo.  
 Fuente: Propia de los autores.

### Implementar un programa para consultar la ontología SYMPT

Durante esta etapa, se implementaron un conjunto de métodos en la clase OntologyUtils mediante el uso de la API RDF4J, con la finalidad de ejecutar consultas en el lenguaje SPARQL, de tal forma que se automatice la extracción de la información de los síntomas y se genere la lista de objetos de los síntomas, y posteriormente guardar esta información en un nuevo grafo llamado Symptom.owl. La consulta en SPARQL que se empleó se muestra a continuación:

```

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

### *Definir y construir el modelo del T-Box del módulo del grafo*

Se diseñó el nuevo modelo en el cual se registran las instancias de síntomas, la definición de la clase Symptom se apega al tipo de dato manejado por el programa que extrae la lista de síntomas. En este nuevo modelo de grafo se simplifican las propiedades de datos, los conceptos de los síntomas que se encuentran registrados como clases y subclases en la ontología original se registran como instancias o individuos en el nuevo grafo, con la finalidad que dicho módulo sea más ligero y fácil de integrar en el grafo de conocimiento general PatientElectronicRecord.

### *Desarrollar un programa que genere automáticamente el módulo de grafo de conocimiento*

Finalmente, en la clase OntologyUtils, se integraron los métodos para la extracción de información de síntomas a partir del grafo (getSymptomsList), y para la generación automática del módulo del grafo que será reutilizado (populateSymptomOntology). Estos métodos son ejecutados desde una aplicación principal (SymptomOntologyCreation) como se muestra en el siguiente código en Java:

```
public class SymptomOntologyCreation {
    public static void main(String[] args) {
        OntologyUtils util = new OntologyUtils();
        //Extracción de la lista de síntomas a través de SPARQL
        List<Symptom> symptomList = util.getSymptomsList();
        //Registro de la lista de síntomas en el nuevo módulo del grafo
        util.populateSymptomsOntology("Ontologies/Symptom.owl",
            symptomList);    }}

```

Como resultado de la implementación del método para la reutilización, se obtiene un nuevo módulo del grafo de conocimiento llamado Symptom.owl, este consta de 1 clase, 3 propiedades de datos, 259 individuos y un total de 1292 axiomas, tal como se muestra en la Figura 6.

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

Figura 6. Métricas del módulo del grafo Symptom.  
Fuente: Propia de los autores.

El método de reutilización descrito anteriormente (tal como se muestra en la Figura 3) se implementó para la extracción de módulos de los siguientes grafos: NDFRT, National Drug File - Reference Terminology; LOINC, Logical Observation Identifier Names and Codes; NCIT, National Cancer Institute Thesaurus. Como resultado se obtuvieron los siguientes grafos de conocimiento (GCs):

a) GC Medicament. En este módulo de grafo se incluyen las definiciones de las propiedades de datos y propiedades de objeto identificadas como relevantes a partir del análisis del grafo de conocimiento NDF-RT. Como resultado en este grafo de Medicament se definieron las siguientes clases: Medicament, ChemicalStructure, MechanismOfAction, PharmacologicalTreatment, NDFRT\_Disease, TherapeuticCategory y PhysiologicalEffect. En estas clases se registran los individuos que tienen relación con los medicamentos y las enfermedades.

b) GC LaboratoryTest. En este grafo se incluyen las propiedades que se identificaron como relevantes en el grafo de conocimiento original LOINC; en particular se incluyen las pruebas de laboratorio para detectar el COVID-19. De acuerdo con el análisis de la información consultada en el grafo original se define la clase LaboratoryTest que consiste en 73 subclases de los tipos de pruebas de laboratorio en general.

c) GC Vaccine. Este grafo incluye las definiciones de clases, de propiedades de datos y de objetos consultadas en el grafo de conocimiento original NCIT. Específicamente se revisaron e incluyeron las definiciones conceptuales para incorporar los diversos tipos de vacunas orientadas a

prevenir la enfermedad del COVID-19<sup>11</sup>. Como resultado el grafo Vaccine incluye las siguientes clases de vacunas: COVID-19\_Vaccine, mRNA\_COVID-19\_Vaccine y SARS- CoV-2\_mRNA\_Vaccine.

### *Etapa 6; integración de módulos en el grafo de conocimiento principal*

Una vez finalizada la etapa de extracción, se realiza el proceso de integración importando los módulos en el grafo de conocimiento general PatientElectronicRecord; y definiendo meta-relaciones entre los conceptos de los grafos importados de acuerdo con los requerimientos del modelo general. Como resultado de esta integración se cuenta con un modelo general que permitirá más adelante representar casos de pacientes mexicanos.

La definición de meta-relaciones es una tarea importante para la integración de los conceptos externos, estas meta-relaciones se pueden definir mediante propiedades entre objetos o instancias de las clases. Es importante recordar que las propiedades entre objetos establecen relaciones binarias, es decir, por cada propiedad se debe definir el dominio y el rango. Durante la integración del grafo general PatientElectronicRecord se definieron los nombres de las propiedades, sus dominios y rangos que se muestran en la Tabla 5.

Tabla 5  
 Propiedades de objetos resultantes del modelo integrado.

Propiedad de objetos	Dominio clase	Rango clase	Descripción
diseaseHasSymptom	NDFRT_Disease	Symptom	Enfermedad tiene síntoma.
hasDiagnosedDisease	ClinicalDiagnosis	NDFRT_Disease	Diagnóstico clínico tiene enfermedad.
hasLabTest	ClinicalDiagnosis	LaboratoryTest	Diagnóstico clínico tiene prueba de laboratorio.
hasPharmalogicalTreatment	ClinicalDiagnosis	PharmalogicalTreatment	Diagnóstico clínico tiene tratamiento clínico.
hasVaccine	Patient	COVID-19_Vaccine	Paciente tiene vacuna aplicada.
requiresLabTest	NDF-RT_Disease	LaboratoryTest	Síntoma requiere prueba de laboratorio.

Fuente: Propia de los autores.

Para visualizar el modelo general integrado (ver Figura 7). De igual forma en la Figura 8, se muestra un fragmento del código del grafo de conocimiento general PatientElectronicRecord notación

<sup>11</sup> <https://www.gob.mx/cofepris/acciones-y-programas/vacunas-covid-19-autorizadas>

Manchester<sup>12</sup>. En el encabezado se pueden observar los prefijos de los grafos importados seguido de las propiedades de anotaciones y las propiedades de datos, entre otras.

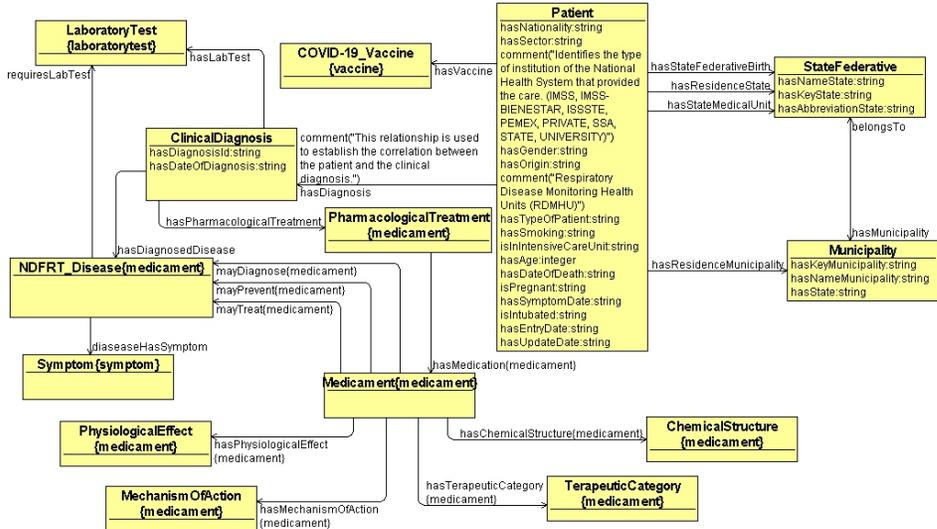


Figura 7. Grafo de conocimiento integrado PatientElectronicRecord.

Fuente: Propia de los autores.

<sup>12</sup> <https://www.w3.org/2007/OWL/wiki/ManchesterSyntax>

```
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
```

Figura 8. Fragmento de código del grafo de conocimiento general PatientElectronicRecord.

Fuente: Propia de los autores.

## *Etapas 7; definición de axiomas generales y reglas de inferencia*

Los grafos de conocimiento se implementan mediante la definición de axiomas lógicos, en un sentido estricto un axioma es una verdad inobjetable acerca de un concepto o un hecho. Vistos desde un enfoque práctico los axiomas son definiciones formales de los conceptos, es decir, son el conjunto de definiciones formales que restringen la pertenencia de los individuos a las clases. Un axioma lógico visto como una restricción establece las condiciones necesarias y/o suficientes para que un individuo sea miembro o pertenezca a una clase. Por lo anterior, la definición de axiomas permite implementar la caracterización fina y detallada de los conceptos. Durante esta etapa del desarrollo del grafo de conocimiento se procedió a definir los axiomas que integran el modelo, así como un conjunto de reglas de inferencia que permiten generar automáticamente nuevas relaciones semánticas y nuevas definiciones.

## Evaluación del grafo general PatientElectronicRecord

En esta sección se describe la implementación de tres enfoques de evaluación: la evaluación de la usabilidad del grafo, la evaluación de los principios de diseño y la evaluación de la competencia del grafo.

### *Evaluación de la usabilidad del grafo de conocimiento*

La usabilidad de un grafo de conocimiento tiene como propósito medir la facilidad que tiene el usuario para aprovechar el modelo mediante la instanciación de datos del mundo real. Mediante la usabilidad se conoce que “tan adecuado” es el diseño de un modelo para representar casos del mundo real. En esta sección se presenta el método implementado para la representación de un conjunto de datos “reales” de pacientes COVID-19, tal como se muestra en la Figura 9. La DGE del gobierno de México pone a disposición de la población en general los datos referentes a los casos asociados a COVID-19, con el propósito de facilitar a todos los usuarios que requieran el acceso, uso, reutilización y redistribución de los mismos.



Figura 9. Método para la representación de datos de pacientes COVID-19 usando el grafo PatientElectronicRecord.  
Fuente: Propia de los autores.

### *Búsqueda de conjunto de datos*

En la etapa de evaluación de la usabilidad del modelo se buscaron conjuntos de datos de pacientes COVID-19 públicos para instanciarlos el grafo general y evaluar su usabilidad de acuerdo con las necesidades planteadas inicialmente. Los datos que se utilizaron se obtuvieron de la Dirección General de Epidemiología (DGE).



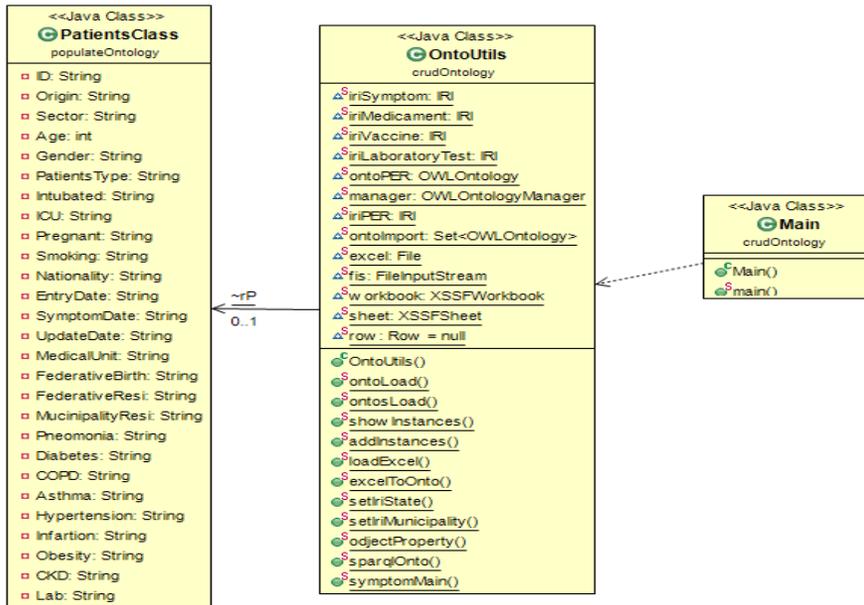


Figura 11. Diagrama de clases para la lectura y registro de datos de pacientes en el grafo general.  
 Fuente: Propia de los autores.

En la Figura 12 se muestra un ejemplo de un paciente registrado en el grafo PatientElectronicRecord. Se trata de una paciente mexicana con el identificador pat01e35a, tiene 56 años, género femenino, originaria de la Ciudad de México de la alcaldía Coyoacán, atendida en la Ciudad de México en el Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado (ISSSTE), la fecha de inicio de síntomas es el 08 de enero del 2020 y la fecha de registro del caso es el 09 de enero del 2020. Su registro refiere que no fue necesaria la intubación, ni la hospitalización en la USMER, aunque los antecedentes médicos establecen que el paciente tiene comorbilidades de diabetes mellitus y obesidad, las cuales representan un elevado riesgo de complicaciones. En su registro se estableció que la paciente debería ser tratada con los medicamentos indicados por el médico.

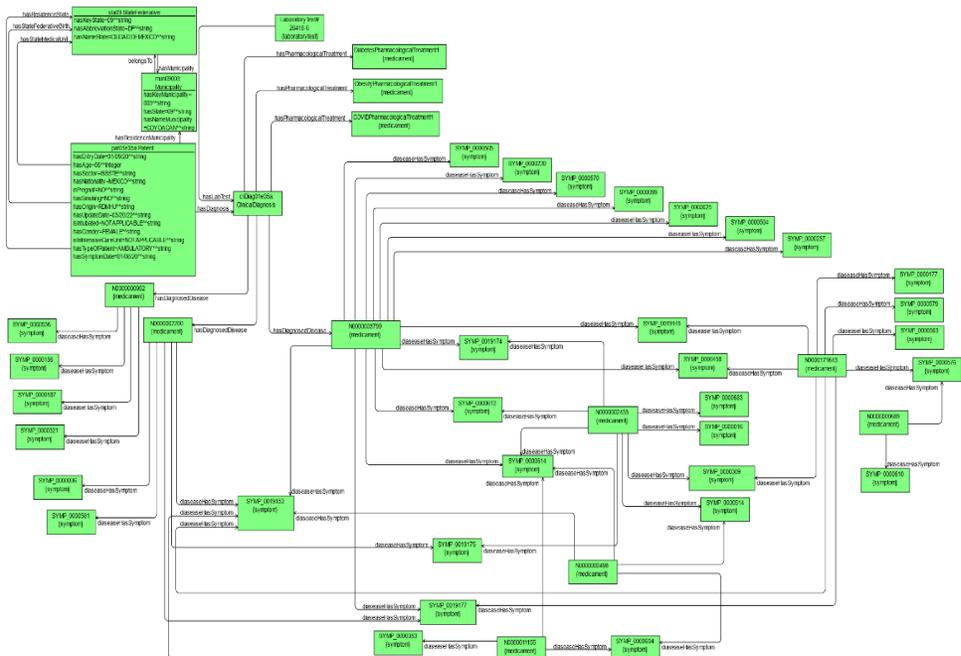


Figura 12. Ejemplo del resultado del registro de los datos de una paciente mexicana con COVID-19.  
 Fuente: Propia de los autores.

### Evaluación basada en principios de diseño

De acuerdo con Tom Gruber (1995) los principios de diseño son guías que orientan el diseño y construcción de modelos ontológicos. En esta sección se revisan los principios de diseño aplicados en la evaluación del grafo PatientElectronicRecord.

Se seleccionaron los principios de diseño propuestos por Gómez-Pérez (1996), Duque-Ramos (2014), tal como mencionan Bravo et al. (2019), la construcción de ontologías abarca métodos, técnicas y principios de diseño que son el soporte de un diseño y construcción de ontologías eficientes, de esta manera los principios de diseño y métodos se consideraron en el desarrollo de este proyecto del modelo ontológico integrado.

### *Principios de diseño aplicados en la evaluación del grafo de conocimiento*

- a) Claridad. Un grafo de conocimiento debería comunicar efectivamente el significado intencionado de los términos definidos. Las definiciones deberían ser objetivas, la formalización es el medio para lograr este fin. Las definiciones completas son preferibles a las definiciones primitivas. Todas las definiciones deberían estar documentadas en lenguaje natural. Siendo un grafo para la representación de datos de pacientes con COVID-19, de acuerdo con el principio de claridad, se definieron los conceptos para la representación mediante axiomas lógicos. Asimismo, se documentaron todos los conceptos y relaciones definidas mediante etiquetas.
- b) Coherencia: Los conceptos en el modelo se definen mediante axiomas lógicos con base en los hechos establecidos por los datos de los pacientes, de modo que un axioma establecido en dicho modelo no debe contradecir una definición establecida. Para verificar la coherencia se ejecutó el razonador Pellet<sup>14</sup> desde Protégé, el cual mostró que el grafo es consistente y por lo tanto coherente tal como se muestra en la Figura 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
```

Figura 13. Verificación de consistencia lógica con el razonador.  
Fuente: Propia de los autores.

- c) Extensibilidad: El diseño del grafo de conocimiento debe considerar de manera anticipada la integración de nuevos conceptos con la intención de expandir el grafo para la representación de otros módulos, sin que esto implique la revisión de todo el grafo o su afectación.
- d) Compromiso ontológico mínimo: Se deben incluir solo los conceptos que sean necesarios, sin excederse incluyendo definiciones que no sean estrictamente indispensables. En este sentido el

---

<sup>14</sup> <https://www.w3.org/2001/sw/wiki/Pellet>

grafo de conocimiento desarrollado presenta un mínimo de compromiso ontológico, esto es que puede soportar las tareas de intercambio de conocimiento requeridas, pero con la menor cantidad posible de afirmaciones sobre el ámbito que se está modelando, permitiendo a las partes comprometidas con el grafo, la libertad de especializarse e instanciar el grafo según sea necesario.

### *Evaluación de la competencia del grafo*

La competencia de un grafo de conocimiento puede definirse como el conjunto de preguntas que el grafo es capaz de responder de acuerdo con el conocimiento incluido. En esta sección se presenta el conjunto de preguntas de competencia en lenguaje natural, las cuales son traducidas al lenguaje de consulta SQWRL con la finalidad de ejecutar dichas consultas.

Por lo anterior, se plantearon las siguientes preguntas de competencia:

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?

En la Figura 14 se presenta un ejemplo de la ejecución de una consulta, así como el tiempo de ejecución. El resultado se muestra en formato de tabla con todos los registros encontrados en el grafo. Como ejemplo se muestra la ejecución de la pregunta 3: What is the gender and age of the intubated patients? Traducida al lenguaje SQWRL de la siguiente forma:

```
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:patz4e638	YES	MALE	60
patientelectronicrecord:pat12d70a	YES	MALE	58

Figura 14. Tiempo de ejecución y tabla de resultado de 3ra consulta.  
 Fuente: Propia de los autores.

## Resultados

Como resultado se obtiene el grafo PatientElectronicRecord cuya estructura se puede ver en la Figura 15. Este grafo incluye los módulos que fueron reutilizados (Medicament, COVID-19\_Vaccine, LaboratoryTest, y Symptom), lo que permite concluir que el método de extracción de módulos es eficiente para la reutilización de grafos biomédicos, evitando la importación de grandes cantidades de datos.

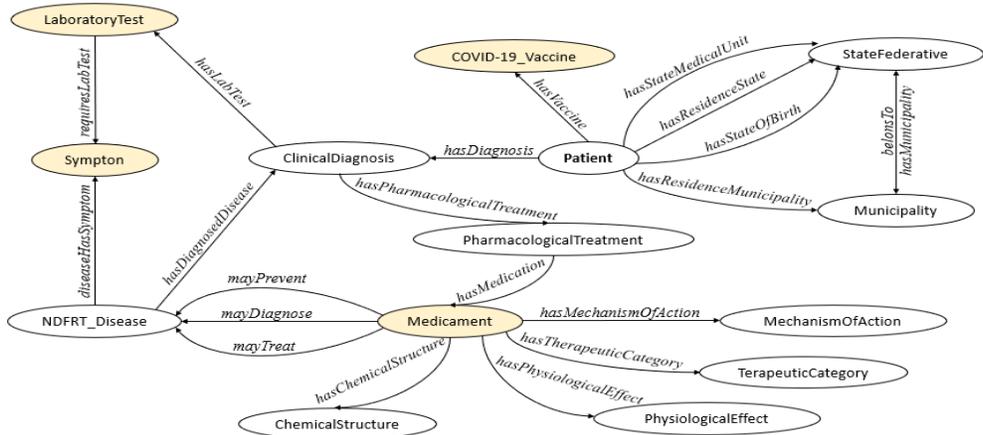


Figura 15. Grafo de conocimiento integrado con entidades y relaciones.  
 Fuente: Propia de los autores.

El grafo de conocimiento resultante muestra un menor número de axiomas considerando los grafos de origen que fueron reutilizados. Esta reducción permite disminuir el tiempo de ejecución de razonadores, evitando la excesiva demanda de recursos de cómputo. El grafo PatientElectronicRecord cuenta con las métricas que se muestran en la Figura 16: un total de 36594 axiomas, 31687 axiomas lógicos, 17 clases, 21 propiedad de objetos, 40 propiedades de datos y 4768 individuos.

<b>Metrics</b>	
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

Figura 16. Métricas del grafo PatientElectronicRecord.  
Fuente: Propia de los autores.

Con base en el conjunto de datos instanciados en el grafo PatientElectronicRecord, se pueden obtener algunas estadísticas generales. Por ejemplo, el 53.7% de los casos positivos son de pacientes del sexo femenino, y el 1.4% de las pacientes del sexo femenino se encuentran embarazadas. Considerando el total de datos de pacientes, el 0.4% se encuentran en unidad de cuidados intensivos y a su vez están intubados, el 3.6% de los pacientes con casos positivos han fallecido. La edad promedio de los pacientes que fallecieron por COVID-19 es de 36 años. El 8.4% de los pacientes con casos positivos se consideraron fumadores y tuvieron más complicaciones agudas, los resultados también muestran que el promedio de días desde el ingreso del paciente en una USMER hasta el fallecimiento del paciente oscila entre 2 y 18 días.

## Conclusiones

Una de las áreas de investigación que ha cobrado enorme relevancia en los últimos años, es el de la representación, gestión y manejo de datos clínicos y de datos de especialidades médicas, así como el manejo eficiente de grandes volúmenes de datos. Esto se debe principalmente a la pandemia por COVID-19, por la cual diversas instituciones de investigación alrededor del mundo se han dado a la tarea de investigar múltiples aspectos del COVID-19. Lo que ha generado que se publiquen numerosas bases de conocimiento presentadas como ontologías, grafos de conocimiento o vocabularios médicos. Para poder integrar y aprovechar todos estos recursos de conocimiento públicos y disponibles, en este artículo se describió la metodología de desarrollo, construcción e integración de un grafo de conocimiento integral y pertinente para la representación y gestión de datos de pacientes diagnosticados con COVID-19. En particular, se buscó diseñar y adaptar este grafo considerando las normas establecidas por el gobierno

mexicano, además de utilizar un conjunto de registros publicados por la DGE mediante la política de datos abiertos. Como resultado se obtuvo un grafo de conocimiento general llamado PatientElectronicRecord, el cual incorpora conceptos e información especializada y actualizada sobre: medicamentos, enfermedades, pruebas de laboratorio, signos y síntomas, vacunas y un conjunto de relaciones entre los conceptos. El grafo de conocimiento general fue evaluado considerando la usabilidad, los principios de diseño y la competencia. El resultado de estas evaluaciones permite determinar que el desarrollo del grafo ha sido adecuado de acuerdo con los estándares establecidos y que permiten cumplir con su finalidad de representar conocimiento de un dominio.

Como trabajo futuro se incorporará la ontología FHIR, el cual es un estándar para facilitar el intercambio y la interoperabilidad semántica de datos de salud entre organismos e instituciones relacionadas con la salud.

## Referencias

- 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 decision

- 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. <https://oa.upm.es/7228/>
- 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>
- Hermesen, 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>