



UNIVERSIDAD AUTÓNOMA METROPOLITANA
UNIDAD AZCAPOTZALCO

DIVISIÓN DE CIENCIAS BÁSICAS E INGENIERÍA
MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN

DESARROLLO E INTEGRACIÓN DE ONTOLOGÍAS PARA
LA REPRESENTACIÓN DE PACIENTES COVID-19

Idónea Comunicación de Resultados

para obtener el grado de:

Maestro en Ciencias de la Computación

PRESENTA:

Ing. Darinel González Villarreal

DIRECTORES:

Dra. Maricela Claudia Bravo Contreras

Dr. José Alejandro Reyes Ortiz



MAESTRÍA EN CIENCIAS
DE LA COMPUTACIÓN

Ciudad de México, Diciembre 2022

A las personas que son de gran apoyo siempre, y me lo han brindado en todo momento, con confianza y sobre todo por creer en mí, han sido mi motivación y ejemplo para seguir a lo largo de mi vida, gracias por su trabajo, sacrificio y amor. Gracias a ustedes he llegado hasta aquí y cumplido una meta más. Agradezco y me enorgullece ser su hijo.

A mis padres

Esas personas que me han ayudado en todo momento me han enseñado a seguir adelante a pesar de las adversidades, gracias por enseñarme que jamás hay que rendirse.

Familiares y amigos

Reconocimientos

Quiero agradecer a mis directores de tesis la Dra. Maricela Claudia Bravo Contreras y al Dr. José Alejandro Reyes Ortiz por su ayuda y sobre todo el conocimiento transmitido a lo largo de esta etapa, ya que me han guiado a través del proceso, gran parte del desarrollo de este proyecto fue desarrollado con ayuda de ellos. Espero siga dándose la oportunidad de seguir colaborando con ustedes en proyectos futuros.

De igual forma, quiero agradecer al Dr. Luis Fernando Hoyos Reyes y a la Maestría en Ciencias de la Computación por darme la oportunidad de formar parte y cursar este posgrado y siempre orientarme en momentos difíciles sobre continuar con el programa.

Agradezco a mis revisores Dra. Maricela Claudia Bravo Contreras, Dr. Leonardo Daniel Sánchez Martínez, Dra. Mireya Tovar Vidal, y el Dr. Luis Fernando Hoyos Reyes. Gracias por tomarse el tiempo en ayudarme a mejorar este trabajo.

Por último, quiero agradecer a la Universidad Autónoma Metropolitana – Unidad Azcapotzalco, por la formación que me ha brindado a lo largo de esta etapa.

Resumen

Desde que se identificó el primer caso conocido de síndrome respiratorio agudo severo coronavirus 2 (SARS-CoV-2) en Wuhan, China, en diciembre de 2019, la enfermedad se ha extendido por todo el mundo, dando lugar a la pandemia actual de COVID-19 que causa síntomas variables en las personas, pero a menudo incluyen fiebre, tos seca, dolor de cabeza, fatiga, dificultad para respirar, pérdida del olfato y pérdida del gusto, de aquellos que desarrollan síntomas lo suficientemente notables como para ser clasificados como pacientes, la mayoría desarrolla síntomas leves a moderados, mientras que el 14 % desarrolla síntomas graves y el 5 % sufre síntomas críticos, por lo que deben ser atendidos en centros especializados.

De modo que cada paciente reacciona de forma diferente ante la enfermedad, el saber cuáles son los factores que determinan la condición de un paciente es importante para conocer el comportamiento de la enfermedad, por ello se considera implementar bases de conocimientos que permitan generar nuevo conocimiento en base a los hechos establecidos de la información de pacientes, por ejemplo, el tratamiento de los pacientes.

Por tal motivo, en este proyecto de investigación se reporta el diseño y desarrollo un sistema de modelos ontológicos integrados para la representación y administración de perfiles de pacientes con COVID-19: diagnóstico y tratamiento, donde los resultados de las diferentes evaluaciones realizadas a dicho modelo, muestran la factibilidad de utilizar esta base de conocimiento integrada por ontologías para la representación de la información clínica del paciente y obtener nuevo conocimiento y servir como base para proyectos de investigación a futuro.

Abstract

Since the first known case of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was identified in Wuhan, China, in December 2019, the disease has spread throughout the world, giving rise to the current COVID pandemic. -19 which causes variable symptoms in people, but often include fever, dry cough, headache, fatigue, shortness of breath, loss of smell, and loss of taste, of those who develop symptoms notable enough to be classified as patients, the majority develop mild to moderate symptoms, while 14% develop severe symptoms and 5% suffer critical symptoms, so they must be treated in specialized centers.

So that each patient reacts differently to the disease, adequate knowledge of the factors that determine the condition of a patient is important to know the behavior of the disease, for this reason, it is considered to implement knowledge bases that allow generating new knowledge is based on established facts of patient information, for example, treatment of patients.

For this reason, this research project reports the design and development of a system of integrated ontological models for the representation and administration of profiles of patients with COVID-19: diagnosis and treatment, where the results of the different evaluations carried out on the said model, show the feasibility of using this knowledge base integrated by ontologies to represent the patient's clinical information and obtain new knowledge and serve as a basis for future research projects.

Índice general

Índice de figuras	XI
Índice de tablas	XIII
1. Introducción	1
1.1. Planteamiento del problema	2
1.2. Justificación	3
1.3. Objetivos	4
1.3.1. Objetivo general	4
1.3.2. Objetivos específicos	4
1.4. Metodología de Investigación	4
1.4.1. Diseño y construcción del modelo ontológico	5
1.4.2. Diseño y ejecución de una estrategia de evaluación del modelo ontológico	7
1.5. Contribuciones	7
1.6. Estructura de la tesis	8
2. Marco teórico	9
2.1. Ontología	9
2.2. Representación de perfiles médicos	10
2.3. Modelos ontológicos para la representación de perfiles de pacientes	11
2.4. Razonamiento e Inferencia lógica	11
3. Estado del arte	13
3.1. Revisión de trabajos relacionados	13
4. Diseño y construcción del modelo ontológico	19
4.1. Recopilación y análisis de información	19
4.2. Búsqueda de ontologías existentes	21
4.3. Diseño del modelo ontológico	22
4.4. Construcción del modelo ontológico	24
4.4.1. Ontología PatiendElectronicRecord	26
4.4.2. Ontología Vaccine	29

ÍNDICE GENERAL

4.4.3. Ontología Medicament	31
4.4.4. Ontología LaboratoryTest	34
4.4.5. Ontología Symptom	36
4.5. Extracción de módulos de ontologías existentes para su reutilización . .	37
4.5.1. Modelo ontológico Symptom	38
4.6. Integración de módulos en Ontología principal	42
4.7. Definición de axiomas y reglas de inferencia	46
5. Evaluación del modelo ontológico	47
5.1. Evaluación mediante el uso del modelo con instancias de datos	47
5.1.1. Caso de uso	51
5.2. Evaluación basada en principios de diseño de ontologías	52
5.3. Evaluación basada en preguntas de competencia	55
6. Conclusiones y trabajos a futuro	59
A. Publicaciones	65
A.1. Artículos científicos	65

Índice de figuras

1.1. Etapas de la metodología de desarrollo	5
4.1. Esquema general del modelo Ontológico	24
4.2. Modelado de la ontología PatientElectronicRecord en Protégé (OntoGraf)	29
4.3. Modelado de la ontología Vaccine en Protégé (OntoGraf)	30
4.4. Modelado de la ontología Medicament en Protégé (OntoGraf)	34
4.5. Modelado de la ontología LaboratoryTest en Protégé (OntoGraf)	35
4.6. Modelado de la ontología Symptom en Protégé (OntoGraf)	37
4.7. (a) Métricas de ontología Symp (con Protégé), (b) Jerarquía de clases de ontología Symp (con Protégé)	39
4.8. Clases java para exportar modulo a nueva ontología	40
4.9. Ontología Symptom para la representación de síntomas	41
4.10. Métricas de Ontología Symptom	42
4.11. Modelado de la ontología PatientElectronicRecord en Protégé (OntoGraf)	45
4.12. Fragmento de código OWL de la ontología PatientElectronicRecord	46
5.1. Etapas de metodología de evaluación	48
5.2. Fragmento de conjunto de datos sin procesar	48
5.3. Fragmento de conjunto de datos reducidos y procesados	49
5.4. Clases java para exportar datos de .xlsx al modelo ontológico	51
5.5. Caso de uso de la representación de perfil de un paciente	52
5.6. Verificación de consistencia con razonador	54
5.7. Diseño del modelo ontológico integrado	54
5.8. Métricas del modelo ontológico integrado	55
5.9. Tiempo de ejecución y tabla de resultado de 1ra consulta	56
5.10. Tiempo de ejecución y tabla de resultado de 3ra consulta	57
5.11. Tiempo de ejecución y tabla de resultado de 10ma consulta	58

Índice de tablas

3.1. Comparación entre los trabajos relacionados a este trabajo	18
4.1. Ontologías biomédicas consideradas para la reutilización parcial	22
4.2. Clases de la ontología PatientElectronicRecord	26
4.3. Propiedades de datos de la ontología PatientElectronicRecord	27
4.4. Propiedades de objetos de la ontología PatientElectronicRecord	28
4.5. Clases de la ontología Vaccine	29
4.6. Propiedades de datos de la ontología Vaccine	30
4.7. Clases de la ontología Medicament	31
4.8. Propiedades de datos de la ontología Medicament	32
4.9. Propiedades de objetos de la ontología Medicament	33
4.10. Clases de la ontología LaboratoryTest	34
4.11. Propiedades de datos de la ontología LaboratoryTest	35
4.12. Clases de la ontología Symptom	36
4.13. Propiedades de datos de la ontología Symptom	36
4.14. Ontologías resultantes del modelo integrado.	43
4.15. Propiedades de objetos resultantes del modelo integrado	44

Capítulo 1

Introducción

El término COVID-19 se refiere a la enfermedad infecciosa causada por el Síndrome respiratorio agudo severo coronavirus 2 (SARS-CoV-2) (WHO, 2022), este virus como la enfermedad que provoca eran desconocidos antes de que surgiera el primer brote en Wuhan, China en diciembre de 2019. La Organización Mundial de la Salud (OMS)¹ calificó la rápida propagación del virus como una pandemia, ya que afecta a muchos países de todo el mundo, debido a su fácil contagio, ya que se propaga principalmente a través de la vía aérea entre personas que se encuentran cerca físicamente. Las personas que están infectadas, pero no muestran síntomas (asintomáticos) también pueden transmitir el virus a otras personas.

Por ello, llevar el registro y el control de los pacientes que presentan casos positivos, es de vital importancia en el descubrimiento del comportamiento del virus en las personas, como por ejemplo saber: quienes son las personas más vulnerables, el rango de edad que tienen, tener alguna comorbilidad es un factor de alto riesgo y esta provoque la presencia de síntomas graves en los pacientes, e inclusive un alto porcentaje de deceso.

De acuerdo con, la Dirección General de Promoción de la Salud (DGPS)², de la secretaria de salud del gobierno de México, fomenta el supervisar y evaluar las acciones para promover la salud de la población mexicana, por medio de la elaboración y difusión de normas, modelos operativos, documentos técnicos y materiales de apoyo, son factores claves para el tratado y evaluación del impacto de la pandemia en el país.

Por otra parte, las bases de conocimiento como lo son: las ontologías son modelos útiles para organizar información y resolver problemas a partir de ella. Las ontologías se pueden definir como una colección de datos relacionados que permiten catalogar un conjunto de conceptos en un ámbito en específico como lo es el ámbito médico. Funcionan como una herramienta para el procesamiento de información, con la finalidad de obtener conocimiento nuevo de acuerdo con las premisas de datos relacionados en

¹<https://www.who.int/>

²<https://www.gob.mx/salud/documentos/dgps>

1. INTRODUCCIÓN

el modelo. Cabe mencionar, que existen ontologías que abordan conceptos médicos ya establecidos, por lo cual, el consultar o hacer uso de esos modelos (parcial o total), facilitan el desarrollo de nuevos modelos que ayuden a la integración de conceptos relacionados.

El gobierno de México, en conjunto con los datos abiertos de la Dirección General de Epidemiología (DGE)³, ha dispuesto a la población en general, la información referente a los casos asociados a COVID-19 con el propósito de facilitar a todos los usuarios que la requieran, el acceso, uso, reutilización y redistribución de esta. Esto permite la propuesta y desarrollo de nuevos proyectos de investigación que permitan explotar y obtener nuevo conocimiento en base a la información existente, como lo puede ser: el conocer la existencia de relaciones de entre conceptos mencionados en dicha información y que esto permita responder a preguntas que pueden esclarecer, ¿Qué conjunto de personas son más propensas a contraer dicha enfermedad? o ¿Cuáles son las comorbilidades en pacientes que pueden provocar un alto riesgo en los síntomas y estado del paciente al diagnosticar COVID-19?, entre otras preguntas de competencia en el entorno de la enfermedad.

Debido a lo antes mencionado, se plantea el desarrollo e integración de ontologías que permitan la representación de perfiles de pacientes mexicanos con COVID-19, en base a la reutilización parcial de ontologías que permitan abordar términos como los “síntomas” de las “enfermedades” (comorbilidades) que el paciente presenta, los “medicamentos” farmacológicos que pueden tratar a las enfermedades, las “pruebas de laboratorios” realizadas a los pacientes para confirmar la presencia del virus, así como la o las dosis de “vacuna” que el paciente contaba al momento de ser diagnosticado. Finalizando con la evaluación del modelo integrado, en base a la reutilización parcial de la información referente a los casos positivos de COVID-19 de pacientes mexicanos puesta a disposición por la DGE, sirviendo como base para posibles trabajos futuros de investigación como la implementación de sistemas de información capaces de consultar dicho modelo y servir para la administración remota de información de nuevos casos de COVID-19.

1.1. Planteamiento del problema

Los casos positivos de coronavirus en México han ido en aumento, debido a que dicho virus se propaga fácilmente a través de la vía aérea entre personas que se encuentran demasiado cerca o a través del contacto físico entre las mismas, provocando el fácil contagio a otras personas. Esto ha provocado muchos casos de contagios, hasta la fecha se contabilizan más de 6 millones de casos confirmados de acuerdo con el Informe General⁴ de los datos del Gobierno de México (de Salud, 2022), como promoción de

³<https://www.gob.mx/salud/documentos/datos-abiertos-152127>

⁴<https://datos.covid-19.conacyt.mx/>

información de salud.

Debido a dichos casos positivos se ha generado información acerca del entorno de salud de dichos pacientes, lo que la motivación de contar con una base de conocimiento como lo es una ontología para el procesamiento de dichos datos que han sido generados y publicados por el gobierno de México y una posterior explotación de estos a manera de pruebas.

Con lo establecido anteriormente, la implementación de una metodología para el desarrollo de una base de conocimiento que sea sustento para la importación de dichos datos para la representación de pacientes mexicanos con la enfermedad, es de gran ayuda para el descubrimiento de nuevo conocimiento en base a las relaciones establecidas en dicho modelo, para la explotación de la información establecida mediante consultas y la motivación a la comunidad en ciencias de la computación para el desarrollo de futuros sistemas ontológicos similares o incluso el uso del mismo para proyectos de investigación relacionados, haciendo uso de dicho modelo para la creación de sistemas de información capaces de administrar dicha información desde una interfaz gráfica para un uso remoto por personal médico.

1.2. Justificación

El primer caso de COVID-19 que se detectó en México fue el 27 de febrero de 2020. Desde ese día hasta la fecha, se han incrementado significativamente los contagios y los decesos, la estrategia de vigilancia centinela consisten en la recolección, integración, verificación y análisis de la información epidemiológica de un conjunto de Unidades de Salud Monitoras de Enfermedad Respiratoria viral (USMER)⁵. Por ello, el contar con un sistema de ontologías integradas que permitan la incorporación del conjunto de datos abiertos publicados por la DGE, es relevante para el descubrimiento del comportamiento de la enfermedad, así como en la representación de los datos de registros con la finalidad de tener nuevo conocimiento relacionado con los perfiles de pacientes.

Debido a lo anteriormente expuesto, la motivación de este trabajo de esta idónea comunicación de resultados presenta los alcances del proyecto de investigación, donde se plantea desde el diseño y desarrollo de un modelo integrado por ontologías reutilizadas parcialmente, que manejan conceptos médicos relacionados al COVID-19, hasta la implementación de evaluaciones del modelo. Evaluación de consistencia de datos, finalizando con la evaluación con un conjunto de datos abiertos con la finalidad de medir el nivel principio de diseño y consistencia del modelo ontológico.

De igual forma se plantea el aportar dicho modelo como base de conocimiento en inglés para la comunidad de científicos en computación, con la finalidad realizar proyec-

⁵Antes denominadas Unidades de Salud Monitoras de Influenza (USMI)

1. INTRODUCCIÓN

tos para explotación mediante la creación de herramientas o el uso de algunas existentes, para extraer nuevo conocimiento o inclusive motivar a realizar un modelo similar para la representación de perfiles de pacientes en un vocabulario medico en español.

Permitiendo tener una base de conocimientos para futuros proyectos de investigación, como lo puede ser el desarrollo de un sistema de apoyo en la toma de decisiones médicas, así como un sistema de información para la administración remota de perfiles de pacientes con COVID-19 de acuerdo con los datos que se solicitan en el modelo ontológico y la relación de estos.

1.3. Objetivos

En este proyecto de investigación se propone obtener los siguientes objetivos, de acuerdo con el siguiente orden de alcance.

1.3.1. Objetivo general

Desarrollar, integrar y evaluar un modelo basado en ontologías para la representación de pacientes mexicanos con COVID-19: diagnóstico y tratamiento.

1.3.2. Objetivos específicos

- a) Implementar una metodología para el diseño y construcción de modelo basado en ontologías que permita la representación de perfiles de pacientes mexicanos con COVID-19.
- b) Modularizar e integrar un conjunto de ontologías médicas para la representación de medicamentos, vacunas, enfermedades, síntomas, pruebas de laboratorio y perfiles de pacientes.
- c) Implementar un método para la población del modelo basado en ontologías con conjunto de datos de referencia del diagnóstico y tratamiento de pacientes mexicanos con COVID-19.
- d) Evaluar la integración del modelo ontológico para conocer los principios de diseño y la consistencia de este.

1.4. Metodología de Investigación

En esta sección se describe la metodología empleada para el requerimiento, diseño, desarrollo y evaluación del modelo ontológico en la representación de perfiles de pacientes mexicanos con COVID-19. Cada una de las etapas consiste en una secuencia de

tareas, las cuales se detallan en los siguientes párrafos.

El diseño y desarrollo del modelo basado en ontologías está compuesto por 2 etapas: Etapa 1 diseño y construcción del modelo ontológico, y la etapa 2 diseño y ejecución de una estrategia de evaluación del modelo ontológico, tal se describe y muestra a continuación en la Figura 1.1.

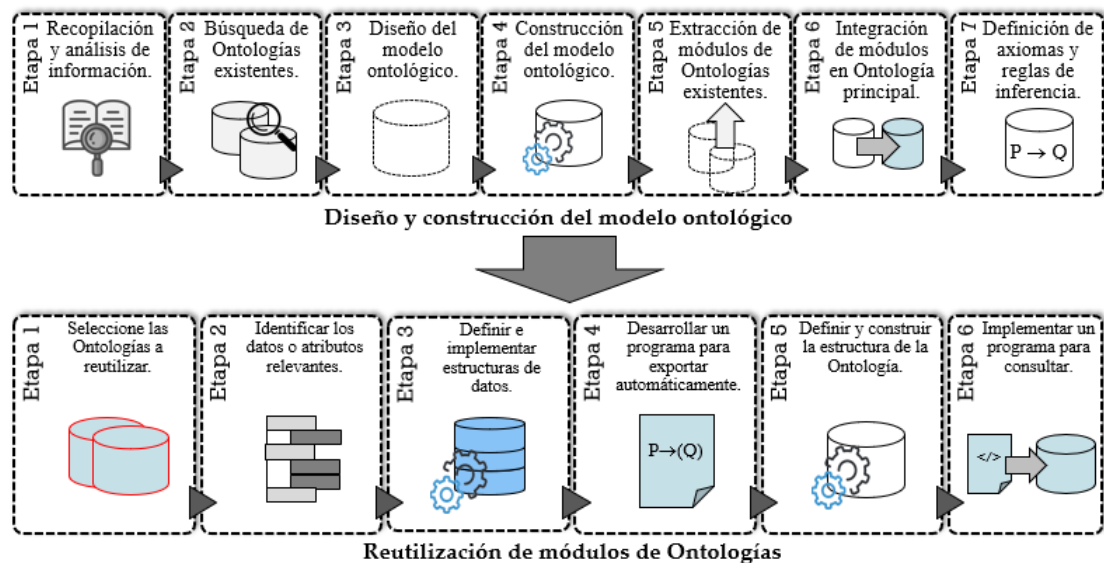


Figura 1.1: Etapas de la metodología de desarrollo.

1.4.1. Diseño y construcción del modelo ontológico

a) *Recopilación y análisis de información*

Esta etapa de la metodología consiste en la búsqueda, recopilación y análisis de información relacionada con normas nacionales e internacionales para el manejo de la información de pacientes con COVID-19. Se deben considerar fuentes de información válida, por ejemplo: la norma mexicana para el manejo de pacientes, norma mexicana para la gestión de expedientes clínicos, guías internacionales y manuales para el tratamiento y seguimiento de pacientes.

b) *Búsqueda de ontologías existentes*

La reutilización de ontologías existentes es una recomendación general que se incluye en la mayoría de las metodologías de diseño de ontologías. Por lo tanto, durante esta etapa se realiza la búsqueda y selección de vocabularios u ontologías con información médica relevante para el propósito del proyecto. Es importante determinar la relevancia de la ontología que se propone reutilizar con respecto

1. INTRODUCCIÓN

a los requerimientos del modelo ya que posiblemente se tenga que realizar un proceso de transformación de la ontología para que se reutilice solo una parte o un módulo de esta.

c) *Diseño del modelo ontológico*

Durante esta etapa se define el modelo principal, considerando los requerimientos y estableciendo las entidades conceptuales más importantes. El diseño se realiza de forma incremental estableciendo los atributos de cada concepto y las relaciones semánticas entre estos. Como resultado se debe generar un esquema general del modelo ontológico completo.

d) *Construcción del modelo ontológico*

Durante esta etapa se implementa el diseño del modelo utilizando el editor de ontologías y verificando que por cada propiedad de datos o propiedad entre objetos se definan correctamente el dominio y el rango. Asimismo, en caso de considerarse necesario se deben especificar axiomas en las definiciones, es decir, definir restricciones de pertenencia de los individuos a las clases. La construcción del modelo debe comenzar primero con las partes que son nuevas, es decir, que no impliquen la reutilización de otras ontologías. Posteriormente se debe realizar el proceso de reutilización, para lo cual se definirá una estrategia de extracción de módulos.

e) *Extracción de módulos de ontologías existentes para su reutilización*

La reutilización de ontologías comienza con un análisis profundo de las ontologías que se van a reutilizar. Se debe identificar cuáles conceptos e instancias son relevantes para cumplir con los requerimientos conceptuales del modelo. Se debe generar un modelo simple y más pequeño a partir del original, de tal forma que lo que se extraiga represente un módulo de la ontología original. Es poco recomendable que se realice la reutilización de una ontología completa mediante la importación, ya que esto genera modelos tan grandes que resulta prácticamente imposible realizar razonamiento lógico por la demanda de memoria.

f) *Integración de módulos de ontologías existentes para su reutilización*

Integración de ontologías se completa con la incorporación de las ontologías reutilizadas parcialmente, esto con ayuda de un programa editor de ontologías, permitiendo relacionar clases entre las ontologías, y la estructura final de una representación de los perfiles de pacientes.

g) *Definición de axiomas y reglas de inferencia*

Definir e integrar un conjunto de reglas de inferencia lógica y axiomas mediante los cuales se generen automáticamente nuevas relaciones semánticas y nuevas definiciones relacionadas con el dominio del modelo diseñado.

1.4.2. Diseño y ejecución de una estrategia de evaluación del modelo ontológico

a) *Evaluación mediante el uso del modelo con instancias de datos*

Esta forma de evaluación de un modelo ontológico consiste en el poblado del modelo con un conjunto de instancias (población) provenientes de datos reales para verificar la usabilidad y utilidad del modelo con respecto a posibles aplicaciones que harían uso de este.

b) *Evaluación basada en principios de diseño de ontologías*

Este tipo de evaluación consiste en revisar los principios de diseño descritos en la literatura especializada, de entre los cuales se deben seleccionar los más relevantes y aplicables al modelo construido. Con el conjunto de principios de diseño se debe realizar la evaluación del modelo uno por uno, ejecutando las herramientas de diagnóstico necesarias y presentando los resultados de tal forma que se pueda verificar el cumplimiento satisfactorio de cada principio.

c) *Evaluación basada en preguntas de competencia*

Este tipo de evaluación consiste en la definición de un conjunto de preguntas en lenguaje natural que se pretende que el modelo sea capaz de responder. Dichas preguntas se deben traducir a algún lenguaje de consultas, por ejemplo, SPARQL, SQWRL o alguno basado en Lógica Descriptiva (DL). Lo que establece este criterio de evaluación es determinar la competencia del modelo, de tal forma que un modelo se considera correctamente diseñado si es capaz de responder de forma correcta a todo el conjunto de preguntas definidas.

1.5. Contribuciones

De las principales contribuciones de este proyecto de investigación, es el aporte de un sistema ontologías integradas con los datos de pacientes mexicanos con casos positivos de COVID-19, permitiendo a la comunidad científica en ciencias de la computación contar con una base de conocimiento del ámbito medico acerca de conceptos médicos relacionados resulta de vital apoyo, si se trabaja con temas relacionados con la salud, llegando a obtener aquellas relaciones que existen en dicho ambiente, como lo son: los síntomas relacionados a una enfermedad, así como los tratamientos farmacológicos que existen o que se pueden emplear para tratamiento de una enfermedad, o conocer los efectos fisiológicos y mecanismo de acción de un medicamento en específico en los pacientes que presentan síntomas de esta enfermedad.

Contar con el modelo con dicha información permite el desarrollo de futuros proyectos en el ámbito de ciencias de la computación, que deseen implementar nuevas

herramientas como lo pueden ser para el Procesamiento del Lenguaje Natural (PLN), para el análisis de algunas relaciones semánticas del conjunto de datos de registro de pacientes mexicanos diagnosticados con COVID-19 considerados para dicho modelo ontológico, así como el desarrollo de sistemas de información en base a servicios para la administración de dicho modelo. También la explotación de los datos dicho modelo con la finalidad de conocer estadísticamente: el porcentaje de hombres y mujeres que fueron diagnosticados con dicha enfermedad, conocer la relación sí existe relación entre los pacientes que fuman y fueron más propensos a estar intubados en una unidad de cuidados intensivos, también se pudo conocer el rango de edad y estado federativo de residencia de los pacientes hospitalizados, entre otros tipos de proyectos que pueden ser considerados o basados en dicho modelo.

1.6. Estructura de la tesis

Esta idónea comunicación de resultado del proyecto de investigación está dividida en 6 capítulos, los cuales se explican a continuación.

Inicialmente el capítulo 1, se realizó una introducción sobre los temas y las razones que motivaron a realizar esta investigación, así como el planteamiento del problema a investigar, la justificación y los objetivos, la metodología de investigación a utilizar y la contribución de dicho trabajo.

En el capítulo 2, se presenta el marco teórico que describe los conceptos relacionados e importantes para tener una mejor comprensión en el desarrollo de este trabajo.

Seguido por el capítulo 3, se encuentra el estado del arte, donde se desarrolla un análisis de los trabajos relacionados con esta investigación, con la finalidad de obtener aquella investigación que existe hasta la fecha permitiendo compararla con la aportación de este trabajo.

En el capítulo 4, se desglosan las etapas y fases del desarrollo de la investigación, de acuerdo con la metodología propuesta para el diseño y construcción del modelo.

El capítulo 5, se comenta sobre los resultados de las evaluaciones realizadas al modelo ontológico para la representación de perfiles de pacientes.

Las conclusiones se abordan en el capítulo 6, explicando sí se ha cumplido con los objetivos de la investigación.

Finalmente, se enlistan las referencias bibliográficas consultadas en lo largo de este trabajo de investigación y los anexos.

Capítulo 2

Marco teórico

En este capítulo se describen los fundamentos teóricos relacionados con el termino de ontologías, así como el uso modelos ontológicos para la representación de conocimiento médico o de la salud, como lo es el caso de este proyecto: la representación de perfiles de pacientes con la enfermedad de COVID-19, también se aborda sobre el razonamiento y el uso del mismo, así como el uso de inferencia lógica en el descubrimiento de nuevo conocimiento basados a hechos establecidos, esto se describe a continuación.

2.1. Ontología

De acuerdo con el concepto de ontología que denomino con (Gruber, 1995) en 1993, una ontología se define como: “Una especificación explícita de una conceptualización”. En otras palabras, una ontología es un modelo de un dominio (conceptualización) que es descrita explícitamente (en forma de una especificación).

(Staab and Studer, 2010), explican los términos de esta definición de la siguiente manera: Una “conceptualización” se refiere a un modelo abstracto de algún fenómeno en el mundo al haber identificado los conceptos relevantes de ese fenómeno. ‘Explícito’ significa que el tipo de conceptos utilizados y las restricciones sobre su uso están definidos explícitamente. Por ejemplo, en los dominios médicos, los conceptos son enfermedades y síntomas, las relaciones entre ellos son causales y una restricción es que una enfermedad no puede causarse a sí misma. ‘Formal’ se refiere al hecho de que la ontología debe ser legible por máquina, lo que excluye el lenguaje natural. “Compartido” refleja la noción de que una ontología captura conocimiento consensuado, es decir, no es privado de algún individuo, sino aceptado por un grupo.

Por otra parte, de acuerdo con (Duran-Muñoz and Bautista-Zambrana, 2013), de igual forma se puede establecer que una ontología o un modelo de dominio, es la representación de la conceptualización explícita compartida de un dominio en particular, ya que se pueden utilizar para representar conceptos de una manera formal, entendible

para las personas y procesable por computadoras, precisando diversos componentes que sirvan en la representación de conocimiento, estos son: conceptos los cuales pueden ser clases de objetos, instancias e individuos de los conceptos, relaciones que pueden formar funciones, así como axiomas que sirven para realizar inferencias.

(Guarino, 1995), “ontología” se usa con diferentes sentidos en diferentes comunidades. La diferencia más radical es quizás entre el sentido filosófico que tiene una tradición bien establecida, y el sentido computacional, que surgió en los últimos años en la comunidad de ingeniería del conocimiento, a partir de una definición informal temprana de ontologías (computacionales) como “especificaciones explícitas de conceptualizaciones”. Establece una clasificación de las ontologías, donde se pueden dividir de acuerdo con su nivel de dependencia con una tarea en particular. La clasificación de ontologías es:

Ontología de un dominio. Representan el conocimiento especializado de un dominio o subdominio y son independiente de la aplicación.

Ontología de nivel superior. Estas describen conceptos generales y funcionales del conocimiento como las estructuras parte-todo, los tipos de objetos y los procesos.

Ontología de aplicación. Contienen todas las definiciones necesarias para modelar el conocimiento de una determinada aplicación, y generalmente son una especialización de ontologías de dominio.

Ontología de la tarea. Proporcionan un vocabulario sistemático de los términos utilizados para resolver problemas relacionados con las tareas que pueden o no pertenecer al mismo dominio.

Tal como es el caso de este proyecto de investigación se hace uso de ontología de aplicación ya que se definen conceptos y elementos para modelar el conocimiento de un dominio en particular, como lo es el dominio del cuidado de la salud mediante la representación de perfiles de pacientes.

2.2. Representación de perfiles médicos

En la actualidad contar con sistema para la representación de perfiles de pacientes es de gran utilidad para la administración de estos, por ello que el Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz (INPRF) (Cruz-Fuentes, 2011), establece que, para la representación de perfiles de pacientes como un expediente clínico, se debe realizar como un instrumento de gran relevancia para la materialización del derecho a la protección de la salud. Esto se trata del conjunto único de información y datos personales de un paciente, que puede estar integrado por documentos escritos, o electrónicos, mediante los cuales se hace constar en diferentes momentos del proceso de la atención

médica, las diversas intervenciones del personal del área de la salud, así como describir el estado de salud del paciente, antecedentes clínicos, resultados de laboratorios, etc.

Tal como la representación de los datos médicos que abarca el Expediente Clínico Electrónico (ECE)⁶, ya que esta herramienta que ofrece información sobre medicación, la historia clínica del paciente, así como los protocolos clínicos y las recomendaciones de estudios específicos, a su vez genera un incremento en la eficiencia en el rastreo de antecedentes clínicos y el cuidado preventivo, contribuyendo a reducir las complicaciones incluyendo los errores en la medicación.

2.3. Modelos ontológicos para la representación de perfiles de pacientes

En la última década el uso de ontologías para la representación de expedientes médicos de pacientes ha cobrado mayor relevancia y popularidad. Esto se debe a las ventajas que presentan las ontologías al diseñar y construir un sistema de información que permita la gestión del conocimiento de apoyo a los médicos en el control (Hayward, 2004) y seguimiento de tratamientos de pacientes, de este modo las ontologías proporcionan una gestión de conceptos en el sector de la salud.

Un ejemplo enfocado es el modelo OntoDiabetic a un sistema para la representación de perfiles de pacientes con la finalidad de funcionar como apoyo a la toma de decisiones, desarrollado por (Sherimon and Krishnan, 2016), en el cual desarrollan e implementan un modelo ontológico de procesos para la administración y seguimiento de perfiles de pacientes diabéticos, esto mediante el uso de reglas de inferencia donde se procesa la entrada de datos de acuerdo al conocimiento de guías clínicas almacenadas, llegando a una conclusión que muestra los porcentajes de riesgos así como las sugerencias de tratamientos que se pueden implementar a los pacientes, esto de acuerdo con los datos almacenados en la base de conocimiento, se comparan las condiciones del paciente mediante una base de reglas donde se proporcionan alertas y recomendaciones que se deben seguir en los tratamientos adecuados para los pacientes con diabetes.

2.4. Razonamiento e Inferencia lógica

Se entiende por razonamiento, como menciona (Aliseda, 1997), a la facultad que permite resolver problemas, extraer conclusiones y aprender de manera consciente de los hechos, estableciendo conexiones causales y lógicas necesarias entre ellos, del cual se puede obtener un nuevo conocimiento en base a la problemática resuelta. Se debe

⁶http://www.conamed.gob.mx/gobmx/boletin/pdf/boletin26/Besp26_10.pdf

2. MARCO TEÓRICO

contar con conocimiento previo de un dominio en específico, con la finalidad de usar motores computacionales en sistema inteligentes capaces de razonar en base a los hechos establecidos en un modelo de base de conocimiento.

Por otra parte, los hechos o premisas establecidos en las bases de conocimiento pueden llegar a generar nuevo conocimiento o conclusiones de un dominio establecido, sin embargo, se debe conocer que es la inferencia, tal como define (Neta, 2013), *“la inferencia es un acto mental, no un proceso computacional, pero algunos actos son procesos de hecho, el acto mental de computar es en sí mismo un proceso y, por lo tanto, sin alguna explicación de a qué se debe esta diferencia entre acto mental y proceso computacional”*.

Por tal motivo, una inferencia es simplemente un razonamiento formal, en el sentido de que lo importante es la forma de las premisas y la conclusión, así como la relación entre ellas, no su contenido, en las bases de conocimiento esto se puede aplicar el mismo concepto de inferencia orientado a los sistemas computacionales, de esta forma se le denomina inferencia lógica, ya que en una sistema de ontológico se puede utilizar razonadores lógicos los cuales permiten crear nuevas clases, propiedades de objetos en base a los hechos establecidos en dicho modelo.

En este proyecto, se consideran eventos relacionados con la investigación, docencia y difusión de la cultura, los cuales serán detectados con el apoyo de redes de sensores en el espacio académico. Las ontologías serán utilizadas para la representación de eventos académicos junto con su espacio y tiempo, debido a las ventajas de interoperabilidad, comunicación e inferencia que ellas ofrecen.

Capítulo 3

Estado del arte

En esta sección, se presenta un análisis de revisión de trabajos relacionados y posteriormente una comparación con este proyecto de investigación, con la finalidad de conocer aquellos trabajos que abordan temas relacionados con esta investigación, como lo es: la representación de perfiles de pacientes en modelos ontológicos y el uso de reglas de inferencia para determinar un nuevo conocimiento en base a los hechos de los registros de los pacientes que presentan alguna enfermedad, y si esta se relaciona con el COVID-19, cabe mencionar que debido a que es una enfermedad relativamente nueva hay una limitación acerca de la trabajos similares sobre el tema abordado, por ello se consideró la revisión de artículos de la representación de perfiles de pacientes con comorbilidades relacionadas con esta enfermedad, ya que pueden provocar una mayor probabilidad de presentar casos graves de dicha enfermedad. La revisión de trabajos se hizo en el portal de Google Scholar, ya que alberga aquellos artículos más relevantes para su consulta.

3.1. Revisión de trabajos relacionados

La investigación descrita por los autores ([Farion et al., 2009](#)), **Clinical decision support system for point of care use: ontology driven design and software implementation**, se ocupa de respaldar el registro electrónico de perfiles de pacientes para la toma de decisiones por parte de los médicos en el punto de atención en un entorno de atención aguda y de emergencia. Implementando un nuevo diseño de protocolo de emergencia móvil 2 (MET2, por sus siglas en inglés Mobile Emergency Triage 2), que proporciona un entorno unificador que puede manejar múltiples aplicaciones clínicas ejecutadas en múltiples plataformas informáticas, utilizando una ontología y los modelos derivados para representar componentes clave de un Sistema de Apoyo a la Decisión Clínica (CDSS por sus siglas en inglés Clinical Decision Support System). Sin embargo, a diferencia de la mayoría de otros marcos de diseño, extiende la ontología para permitir la construcción de múltiples modelos de aplicaciones clínicas específicas no sólo en términos de datos y funcionalidad de soporte, sino también en términos de su

3. ESTADO DEL ARTE

interfaz de usuario y configuración de plataforma informática. El diseño MET2 separa los modelos de aplicación de los componentes de tiempo de ejecución e introduce el ciclo de solicitud-ejecución-purga de una aplicación. Esta solución asegura la extensibilidad del sistema MET2 y la reutilización de los componentes en tiempo de ejecución.

En este trabajo los autores (Köhler et al., 2009), **Clinical diagnostics in human genetics with semantic similarity searches in ontologies**, han implementado un método en un sistema web disponible gratuitamente, denominado Phenomizer. No pretende ser un sistema experto (software que intenta reproducir el desempeño de un experto humano) sino más bien un sistema para expertos, que pueden usar el sistema para ayudar a guiar el proceso de diagnóstico diferencial en genética humana. Al proporcionar una medida estadística de la importancia de los diagnósticos candidatos propuestos, el sistema puede proporcionar alguna indicación de si las características de representación y gestión de perfiles clínicos ingresadas por el médico son en sí mismas altamente sugestivas de un diagnóstico dado o, por otro lado, utilizando una estructura semántica del fenotipo humano (HPO, por sus siglas en inglés Human Phenotype Ontology), si no hay diagnóstico en el la base de datos coincide significativamente con los términos de la consulta.

En este trabajo los autores (Celi et al., 2012), **Database-driven Decision Support System: Customized Mortality Prediction**, presentan un marco de una herramienta dinámica de representación de perfiles de pacientes, para el apoyo a la decisión basada en datos empíricos, una desviación de los típicos sistemas expertos incorporados y derivados de grandes estudios publicados, generalmente multicéntricos, de intervención y de observación. 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, intercambiando la posibilidad de generalizar por precisión. Acuñan el término Experiencia Colectiva para este enfoque, ya que la información se extrae de la experiencia de varios médicos almacenada en el sistema de registros médicos electrónicos.

Los autores (Hermsen et al., 2012), **Implementation of a clinical decision support system for antimicrobial stewardship, infection control and hospital epidemiology**, describen en su investigación que la representación y la gestión de perfiles de pacientes en un CDSS, pueden mejorar la toma de decisiones sobre antimicrobianos mediante la disponibilidad de una combinación de datos y costos específicos del paciente. El uso de CDSS tiene varios beneficios, incluida la disminución de eventos adversos, la reducción de la duración de la estadía, la disminución de costos y el uso de antimicrobianos más apropiado, como lo plantearon en el desempeño de una auditoría prospectiva con intervención y retroalimentación para los programas de administración de antimicrobianos (ASPs, por sus siglas en inglés Antimicrobial Stewardship Programs) en el centro médico de Nebraska, un centro médico académico de 624 camas para cuidados intensivos.

OntoDiabetic CDSS, es un sistema para la representación de perfiles de pacientes con la finalidad de servir como un sistema de apoyo en la toma de decisiones de acuerdo con los hechos establecidos en el modelo, diseñado por los autores (Zhang et al., 2017), **OntoDiabetic: an ontology-based clinical decision support system for diabetic patients**, presentando el diseño y modelado de ontología, la implementación de la ontología de procesos. La ontología de procesos es una de las principales ontologías utilizadas en OntoDiabetic CDSS. Los conceptos de dominio y las directrices en sí se implementan en el lenguaje web de ontología 2 (OWL2, por sus siglas en inglés Ontology Web Language 2). El razonador, funciona procesando la entrada (perfil semántico) con el conocimiento almacenado (guías clínicas) para llegar a conclusiones correctas (puntajes de riesgo y sugerencias de tratamiento). Usando la inferencia de encadenamiento directo en el sistema. Los datos o hechos almacenados en la base de conocimientos se comparan con las condiciones de las reglas en la base de reglas. El sistema también proporciona alertas y recomendaciones, y sugiere tratamientos adecuados para los pacientes diabéticos que tienen tres complicaciones principales: antecedentes médicos posteriores a la enfermedad cardiovascular manifiestan (ECV), nefropatía diabética e hipertensión.

La prevención y el tratamiento de los pacientes con enfermedades crónicas requiere múltiples acciones coordinadas, los autores (Ajami and Mcheick, 2018), **An ontology-based approach to patient follow-up assessment for continuous and personalized chronic disease management**, abordan las complejas interacciones entre los factores de riesgo, las enfermedades, las condiciones del paciente y las modalidades de tratamiento, dichas estrategias se complementan con la vigilancia para monitorear las tendencias y rastrear el progreso, políticas y enfoques ambientales para promover la salud. Las contribuciones del trabajo incluyen: (1) modelos computarizados, explícitos y configurables para evaluaciones automáticas de pacientes durante las visitas de seguimiento de enfermedades crónicas; (2) estándar HL7 vMR adaptado para registrar datos de atención médica del paciente enriquecidos semánticamente; (3) compartir y reutilizar el conocimiento del dominio sobre el manejo de enfermedades crónicas; y (4) una arquitectura para la representación de perfiles de pacientes orientada al servicio para el manejo continuo y personalizado de pacientes con enfermedades crónicas.

En el trabajo de (Oyelade and Ezugwu, 2020), **Ontology-based model to support ubiquitous healthcare systems for COPD patients**, se presenta un modelo para la representación de información de pacientes para crear entornos seguros para los pacientes con enfermedad pulmonar obstructiva crónica (EPOC), el modelo está basado en la descripción formal ontológica de un dominio relacionado con la salud, utiliza el lenguaje de reglas de Web Semántica (SWRL, por sus siglas en inglés Semantic Network Rules Language). La ontología propuesta por los autores 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. Las reglas de SWRL se construyen a partir de

3. ESTADO DEL ARTE

las directrices médicas, la investigación y las opiniones de expertos independientes para estimar el riesgo de exacerbación de EPOC, la contribución novedosa del marco sugerido reside en el enfoque de monitoreo centrado en el paciente.

De acuerdo con la investigación de los autores (Govindan et al., 2020), **A case-based reasoning framework for early detection and diagnosis of novel coronavirus**, establecen la construcción de una base de conocimiento para la representación de perfiles de pacientes, el marco propuesto se logra mediante la formalización del conocimiento basada en ontologías. El resultado obtenido muestra que el marco propuesto demostró un rendimiento interesante en comparación con estudios similares de razonamiento basado en casos (CBR, por sus siglas en inglés Case-Based Reasoning) de última generación utilizando fuzzyCBR. Esto posiciona al marco basado en casos temporal orientado al procesamiento del lenguaje natural (NLP, por sus siglas en inglés Natural Language Processing) y la ontología para su adopción y generalización a los problemas de diagnósticos asociados con otros fines médicos y enfermedades presentes en pacientes.

En este trabajo los autores (Govindan et al., 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)**, presentan un sistema práctico para la representación de perfiles de pacientes como apoyo a la toma de decisiones para clasificar a los miembros de la comunidad, en 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). Una de las limitaciones del presente estudio es que los tres síntomas de fiebre, cansancio y tos seca se han considerado como los criterios para la evaluación de los residentes de la comunidad. Estos tres criterios figuran entre los síntomas más comunes de la infección por COVID-19, pero también se han observado en algunos pacientes otros síntomas, como diarrea, vómitos y similares.

Los autores (Harry et al., 2020), **Pre-implementation adaptation of primary care cancer prevention clinical decision support in a predominantly rural healthcare system**, presentan un sistema de representación de perfiles de pacientes para la toma de decisiones basadas en un Expediente Clínico Electrónico (EHR, por sus siglas en inglés Electronic Health Record) que pueden mejorar la prevención y la detección del cáncer en la atención primaria. Para ello se empleó el marco consolidado para la investigación de la implementación (CFIR, por sus siglas en inglés Consolidated Framework for Implementation Research) para describir los cambios previos a la implementación realizados en función de las barreras y los facilitadores encontrados al incorporar la prevención y el tratamiento del cáncer en un sistema (CDS, por sus siglas en inglés Clinical Decision Support) de gestión del riesgo cardiovascular actual para facilitar la implementación y maximizar el uso del sistema CDS integrado.

En la investigación desarrollada por los autores ([Bravo et al., 2020](#)), **Management of diabetic patient profiles using ontologies**, se describe el desarrollo de un sistema integrado de ontologías médicas, para la representación y administración de perfiles de pacientes con diabetes mellitus en México, que debido a la llegada de la pandemia del COVID-19 el presentar diabetes se ha convertido en un factor de riesgo de graves complicaciones, aumentando la tasa de mortalidad, por ello y tomando consideraciones de las normas mexicanas y características para el desarrollo de representación de perfiles de pacientes mexicanos, de esta manera el sistema de administración de perfiles está basado en modelos ontológicos que permiten la administración de dichos perfiles y a su vez sirve de apoyo en la toma de decisiones de los pacientes con diabetes mellitus y poder recomendar un tratamiento adecuado a los pacientes de con las condiciones que éstos presentan.

Se hizo una revisión de 25 artículos en total, cabe mencionar que del total de artículos revisados al final se seleccionaron aquellos que estuvieran más apegados a los temas mencionados, teniendo un total final de 11 artículos para realizar una comparación tal como se muestra en la Tabla [3.1](#), sobre la relación de los temas relacionados a este proyecto de investigación.

3. ESTADO DEL ARTE

Tabla 3.1: Comparación entre los trabajos relacionados a este trabajo.

Año	Autores	Representación de datos de pacientes	Uso de Ontologías	Lenguaje de Reglas de Web Semántica	Evaluación
2009	Farion, Michalowski, Wilk, et al.	Si	Si	No	Implementación de uso
2009	köhler, Schulz, Krawitz, et al.	Si	Si	Si	Implementación de uso
2012	Celi, Galvein, Davidzon, et al.	Si	No	No	Conjunto de datos
2012	Hermesen, VanSchooneveld, et al.	Si	No	No	Implementación de uso
2016	Sherimon, Krishnan	Si	Si	Si	Datos de monitoreo de pacientes
2017	Zhang, Guo, Zhuo, et al.	Si	Si	Si	Datos de monitoreo de pacientes
2018	Ajami, Mcheick, et al.	Si	No	No	Datos de monitoreo de pacientes
2020	Oyelade, Ezugwu	Si	Si	Si	Conjunto de datos
2020	Govindan, Mina, Alavi	Si	No	No	Conjunto de datos
2020	Harry, Saman, Truitt, et al.	Si	No	No	Implementación de uso
2020	Bravo, González, et al.	Si	Si	Si	Principios de diseño, Implementación de uso
-	Este trabajo.	Si	Si	Si	Principios de diseño, Conjunto de datos

Diseño y construcción del modelo ontológico

En este capítulo, se describe las fases realizadas de esta primera etapa de la metodología propuesta que permite el desarrollo del proyecto de investigación. La primera etapa denominada: “diseño y construcción del modelo ontológico”, está conformada por la recopilación y el análisis de información médica que tiene relación con la enfermedad del COVID-19, seguido de la búsqueda de ontologías existentes relacionadas con dicha enfermedad, para hacer una reutilización total o parcial de acuerdo a las necesidades del proyecto, continuando con el diseño y la construcción del modelo ontológico que será la ontología principal para la integración del modelo final, se procede con la extracción de módulos de ontologías existentes para su reutilización e integración con la ontología principal, para finalizar con la definición de axiomas y reglas de inferencia, cada fase de esta etapa se desglosa a continuación.

4.1. Recopilación y análisis de información

De acuerdo con la metodología propuesta para este proyecto de investigación, se procedió a consultar información existente acerca del coronavirus SARS-CoV-2, con la finalidad de conocer los términos y la relación del conocimiento médico, cabe mencionar que el virus y la enfermedad que este provoca no eran conocidos hasta principios del 2020, cuando se empezó a producir información acerca de estos, con la finalidad de difundir y conocer acerca del comportamiento del virus.

Es por esto qué, para el desarrollo del modelo ontológico que da soporte al sistema, se realizó una recopilación y análisis de información de diversas fuentes válidas, fuentes mexicanas. Las fuentes de información que se consideraron son las siguientes:

4. DISEÑO Y CONSTRUCCIÓN DEL MODELO ONTOLÓGICO

- a) Página oficial con información acerca del COVID-19⁷, 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 2012⁸, 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 viral⁹.
- d) Guía clínica para el tratamiento de la COVID-19 en México¹⁰. 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 internacionales¹¹ y nacionales¹², 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.

Por otra parte, se consultaron las ontologías descritas en los trabajos relacionados que fueron desarrollados para la representación de perfiles de pacientes en un dominio médico.

Con base en la información que se consideró y analizó, se realiza una identificación de conceptos relevantes para la representación de un perfil de pacientes, los términos de Paciente y Diagnóstico Clínico son importantes en el registro de información geográfica y demográfica, historial de antecedentes clínico de un paciente generando la relación entre estos, otro concepto identificado es Enfermedad que este se relaciona con el Síntoma, de igual forma se identificó que los enfermedades requieren de Prueba de Laboratorio, los conceptos mencionados anteriormente se relacionan al diagnóstico clínico del paciente, por otra parte, el concepto de Tratamiento Farmacológico está relacionado como un tratamiento del concepto Medicamento y este se relaciona con una enfermedad y a su

⁷<https://coronavirus.gob.mx/covid-19/>

⁸https://epidemiologia.salud.gob.mx/gobmx/salud/documentos/manuales/00_NOM-017-SSA2-2012_para_vig_epidemiologica.pdf

⁹https://www.gob.mx/cms/uploads/attachment/file/715464/Lineamiento_VE_y_Lab_Enf_Viral_05042022.pdf

¹⁰https://www.gob.mx/cms/uploads/attachment/file/659911/GuiaTx_COVID19_Consenso_2021.08.02_compressed.pdf

¹¹<https://apps.who.int/iris/bitstream/handle/10665/332638/WHO-2019-nCoV-clinical-2020.5-spa.pdf>

¹²https://coronavirus.gob.mx/wp-content/uploads/2020/04/Lineamiento_Clinico_COVID-19_CCINSHAE_14022020.pdf

vez está relacionado a un diagnóstico clínico de un paciente, de igual manera, el medicamento se relaciona con los conceptos de Mecanismo de Acción, Categoría Terapéutica, Efecto Fisiológico y Estructura Química, como propiedades principales de este, dichos conceptos se traducen al Inglés en la Ontología, de igual forma se traducen las relaciones definidas entre cada concepto.

4.2. Búsqueda de ontologías existentes

El objetivo de esta fase es hacer uso de aquellos portales y repositorios donde se almacenan y distribuyen de forma libre aquellas bases de conocimiento relacionadas al ámbito médico, es por ello por lo que se procedió a consultar uno de los repositorios más completos del mundo de ontologías biomédicas, denominado Bioportal¹³. La búsqueda y selección de vocabularios u ontologías de información médica válida y pública, ayuda a determinar si es conveniente el reutilizar total o parcialmente dichas ontologías, por ello, conocer las clases medicas relacionadas con la enfermedad del COVID-19 de acuerdo con la relevancia y nivel de consultas del repositorio, así como las propiedades que estas manejan, tal como menciona (Whetzel et al., 2011), son consideraciones importantes para cubrir las necesidades del desarrollo del modelo ontológico de este proyecto con la finalidad de obtener un buen modelado de conocimiento para la representación de perfiles de paciente, y poder integrar un conjunto de datos para la evaluación y consultas del modelo para obtener un nuevo conocimiento médico.

De acuerdo con los datos médicos considerados en base a la recopilación y análisis de fuentes mexicanas válidas para la representación de perfiles de pacientes, para el modelo de ontologías se consideró consultar ontologías de medicamentos, enfermedades, pruebas de laboratorio y vacunas ya que en algún momento se relacionan en la representación de un perfil para un paciente con COVID-19.

La primera ontología que se consultó fue *NDFRT* (National Drug File - Reference Terminology), ya que esta aborda conceptos relacionados de enfermedades con medicamentos, efectos fisiológicos de los medicamentos, categorías terapéuticas de los medicamentos, estructura química de los medicamentos, cuenta con un total aproximado de 36,200 clases hasta la fecha.

Por otra parte, el modelo *LOINC* (Logical Observation Identifier Names and Codes), es una ontología que aborda conceptos de todo tipo de pruebas de laboratorios, incluyendo aquellas pruebas de laboratorios para descartar o confirmar casos positivos de COVID-19, cuenta con 281,878 clases hasta la fecha.

Seguido del análisis de la ontología *SYMP* (Symptom Ontology), esta ontología contiene el conjunto de síntomas de manera general, síntomas ocasionados por la enfer-

¹³<https://bioportal.bioontology.org/>

4. DISEÑO Y CONSTRUCCIÓN DEL MODELO ONTOLÓGICO

medad, se diseñó de acuerdo con la relación del concepto síntoma: Un cambio percibido en la función, sensación o apariencia informado por un paciente indicativo de una enfermedadz cuenta con aproximadamente de 1000 clases hasta la fecha.

Por último, se consultó la ontología *NCIT* (National Cancer Institute Thesaurus), esta ontología aborda un amplio conjunto de conceptos biomédicos, desde procesos biológicos, la atención clínica, la investigación traslacional y básica, la información pública y las actividades administrativas de la biomedicina, y los materiales biomédicos de los que deriva el módulo de conjunto de vacunas contra dicha enfermedad, cuenta con un total de 174,278 clases hasta la fecha.

La Tabla 4.1 enlista la descripción y el identificador de concepto de cada una de las ontologías consultadas y descargadas para conocer el proceso de reutilización o transformación en un modelo más reducido y adaptado a las necesidades de este proyecto.

Tabla 4.1: Ontologías biomédicas consideradas para la reutilización parcial.

Ontología	Descripción	ID de Concepto
<i>NDFRT</i>	National Drug File - Reference Terminology	https://bioportal.bioontology.org/ontologies/NDFRT
<i>LOINC</i>	Logical Observation Identifier Names and Codes	https://bioportal.bioontology.org/ontologies/LOINC
<i>SYMP</i>	Symptom Ontology	https://bioportal.bioontology.org/ontologies/SYMP
<i>NCIT</i>	National Cancer Institute Thesaurus	https://bioportal.bioontology.org/ontologies/NCIT

4.3. Diseño del modelo ontológico

El diseño del esquema que da soporte al modelo ontológico integrado para la representación de perfiles médicos se llevó a cabo en base a las consideraciones de la recopilación y análisis de la información relacionada con normas nacionales e internacionales para el manejo de información (la representación de perfiles) de pacientes con COVID-19, con la finalidad de obtener un diseño de modelo para integrar un conjunto de datos médicos. Los datos para relacionar son: medicamentos, vacunas, enfermedades, síntomas, pruebas de laboratorio, de modo que se obtenga un modelo integrado de datos para la representación de perfiles de pacientes, a continuación, se describe una breve referencia al concepto del nombre de las ontologías propuestas para el modelo.

- a) Como ontología inicial se consideró el diseño de una ontología nueva que no dependa una reutilización y permita importar las demás ontologías, por ello se

denominó ***PatientElectronicRecord*** (Registro electrónico de paciente), en esta se describen las clases y propiedades para la representación de perfiles de pacientes, así como el diagnóstico de enfermedad del COVID-19, y las nuevas relaciones semánticas entre esta y las otras ontologías resultantes de la reutilización parcial.

- b) Por otra parte, la ontología nombrada ***Vaccine*** (Vacuna), fue diseñada para abordar los conceptos de clases de las vacunas relacionadas con la prevención del virus Sars-Cov-2, que han sido desarrolladas y aprobadas para su uso de emergencia por organizaciones nacionales e internacionales.
- c) Siguiendo un orden, se diseñó un modelo que contenga los conceptos de las enfermedades y los medicamentos que sirven para tratarlas, de modo que existe ese concepto de relación semántica que es importante para el desarrollo de este proyecto, aunque el modelo se nombró ***Medicament*** (Medicamento), está además de contener los medicamentos, también aborda los conceptos de las enfermedades y algunas clases relacionadas entre dichos conceptos que se describirán más adelante en la fase de construcción.
- d) De igual forma, se diseñó una ontología denominada ***LaboratoryTest*** (Prueba de laboratorio), esta aborda aquellos conceptos de los tipos de pruebas de laboratorios que existen para diagnosticar o descartar alguna enfermedad, así como las pruebas para determinar si existe un caso positivo en el contagio de COVID-19 como es el caso para este proyecto.
- e) Por último, se consideró el diseño de una ontología que abarque conceptos del conjunto de síntomas causados por enfermedades, por ello la ontología denominada ***Symptom*** (Síntoma) aborda aquellos conceptos mencionados para realizar una relación semántica entre enfermedades y los síntomas que esta provoca.

De acuerdo con lo anterior, se obtiene un total del diseño de 5 ontologías que forman el modelo integrado para la representación de perfiles de pacientes, en la Figura 4.1 se muestra el esquema general del modelo ontológico.

4. DISEÑO Y CONSTRUCCIÓN DEL MODELO ONTOLÓGICO

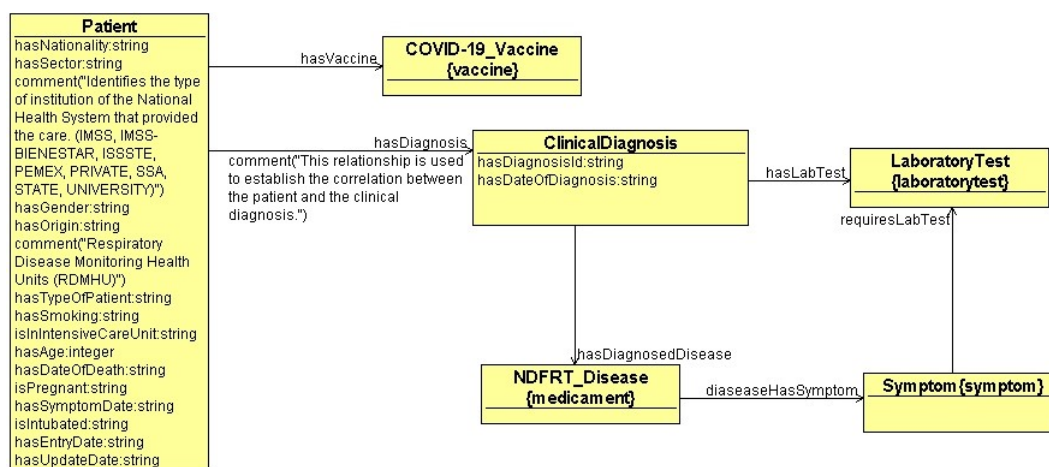


Figura 4.1: Esquema general del modelo Ontológico.

4.4. Construcción del modelo ontológico

Seguido de la fase anterior, se procedió a la construcción de cada una de las ontologías que dan soporte al modelo ontológico integrado para la representación de perfiles de pacientes, para ello se empezó por la construcción de aquellas ontologías nuevas que no implican realizar una reutilización y extracción de una ya existente, para ello se hizo uso de un programa local llamado Protégé¹⁴ en su versión 5.5.0 para la plataforma de Windows 10.

Es un programa gratuito y de código abierto, que permite crear y editar ontologías con la finalidad de crear sistemas inteligentes, fue desarrollado por el Centro Stanford para la Investigación en Informática Biomédica de la Facultad de Medicina¹⁵ de la Universidad de Stanford, con la finalidad de crear soluciones basadas en el conocimiento en áreas tan diversas como la biomedicina, el comercio electrónico y el modelado organizacional.

Para conocer cómo se constituye un modelo ontológico, se aborda la conceptualización semántica y las especificaciones explícitas, que permite la integración y consistencia en el desarrollo de las ontologías, los conceptos se definen a continuación:

Clases: Se trata de las representaciones concretas de conceptos, que se describen mediante declaraciones que indican los requerimientos para ser miembro de éstas.

¹⁴<https://protege.stanford.edu/about.php>

¹⁵<https://bmir.stanford.edu/>

Dominio: es el conjunto de individuos de una clase que se relaciona con el conjunto de individuos de otra clase.

Rango: es la propiedad del dominio que permite definir la clase a la que pertenece una propiedad de anotación, la clase a la que pertenece una propiedad de objeto o el tipo de dato a la que pertenece una propiedad de dato.

Propiedad de anotación: utilizadas para añadir información a clases, individuos y propiedades de objeto o dato, entre estas propiedades, destacan las *labels* y los *comments*.

Propiedades de datos: vinculan individuos y datos definidos mediante un rdf literal o utilizando un esquema de datos XML. Para establecer estas relaciones se utilizan las restricciones *tieneValor* (hasValue restrictions).

Propiedades de objetos: son relaciones binarias entre individuos, desde el dominio (Individuo) a un rango (otro Individuo), conformando una tripleta.

Individuos o Instancias: estos se definen como representación de objetos.

Axiomas: aserciones (incluyendo reglas) en una forma lógica que juntos incluyen toda la teoría que la ontología describe en su dominio de aplicación.

Funciones: complejas estructuras formadas de cierta relación que pueden ser usadas en lugar de un término individual en una declaración

Restricciones: establecen descripciones formales de lo que debe ser verdad con el objetivo de que alguna aserción pueda ser aceptada como entrada.

Reglas: son declaraciones con forma de oraciones si-entonces (antecedente-consecuente) que describen inferencias lógicas que pueden ser derivables de una aserción en una forma particular.

XSD (XML Schema Definition): es un lenguaje, también llamado simplemente XML Schema, que sirve para definir la estructura de un documento XML, permitiendo su validación.

OWL (Web Ontology Language): es un lenguaje basado en la lógica computacional tal que el conocimiento expresado en OWL puede ser explotado por programas de computadora.

En Base a los términos mencionados y descritos acerca del manejo y creación de ontologías, con el editor mencionado se realizó desde la creación de clases y subclases hasta

4. DISEÑO Y CONSTRUCCIÓN DEL MODELO ONTOLÓGICO

la creación de las propiedades de objetos si la relación entre clases se llegara a dar, cabe mencionar que quedan pendientes por realizar algunas relaciones de propiedades de objetos, ya que se realizan hasta que el modelo se haya integrado con las demás ontologías.

En la construcción de ontologías se abordaron conceptos ya mencionados, como: clases, dominio, rango, propiedades de datos y propiedades de objetos y la definición de axiomas, para la creación se abordó el siguiente orden: *PatientElectronicRecord*, *Vaccine*, *Medicament*, *LaboratoryTest*, *Symptom* y finalmente en la ontología *PatientElectronicRecord* se procede a importar los modelos de las demás ontologías creadas para su posterior extracción de módulos e integración.

4.4.1. Ontología PatientElectronicRecord

Para el desarrollo de esta ontología *PatientElectronicRecord*, se describió aquellas propiedades analizadas para la creación de un modelo de representación de perfiles de pacientes, por ello se definen las siguientes clases: *Patient*, *StateFederative*, *Municipality* y *ClinicalDiagnosis*, éstas clases presentadas en la Tabla 4.2, almacenarán aquellos datos del paciente que permiten representar un perfil médico.

Tabla 4.2: Clases de la ontología PatientElectronicRecord.

Clase	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.

Teniendo las clases definidas, se procede a incorporar aquellas propiedades de datos que permiten definir y restringir el tipo de individuo que pertenece a cada clase. Tal como se muestran y describen en la Tabla 4.3.

- En la clase de *Patient* se definen las propiedades de datos de: *hasOrigin*, *hasSector*, *hasGender*, *hasAge*, *hasTypeOfPatient*, *hasNationality*, *hasEntryDate*, *hasSymptomDate*, *hasSymptomUpdate*, *hasDateOfDeath*, *hasSmoking*, *isInIntensiveCareUnit*, *isPrenant* y *isIntubated*.
- Para la clase *StateFederative* se definieron las propiedades de datos de: *hasAbbreviationState*, *hasKeyState* y *hasNameState*.

- c) Y finalmente en la clase *Municipality* se definen las propiedades de datos de: *hasState*, *hasNameMunicipality* y *hasKeyMunicipality*.

Tabla 4.3: Propiedades de datos de la ontología PatientElectronicRecord.

Propiedad de dato	Dominio clase	Rango tipo de dato	Descripción
hasOrigin	Patient	integer	Origen de unidad de salud monitora de enfermedades respiratorias (USMER)
hasSector	Patient	string	Institución del Sistema Nacional de Salud que brindó la atención.
hasGender	Patient	string	Genero del paciente.
hasAge	Patient	integer	Edad del paciente.
hasTypeOfPatient	Patient	string	Paciente ambulatorio o 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.

4. DISEÑO Y CONSTRUCCIÓN DEL MODELO ONTOLÓGICO

Se ha definido las propiedades de datos de las clases, seguido de ellos se procede a definir las propiedades de objetos que permiten obtener las relaciones semánticas entre clases de la misma ontología *PatientElectronicRecord*, las cuales son: *belongsTo*, *hasDiagnosis*, *hasMunicipality*, *hasResidenceMunicipality*, *hasResidenceState*, *hasStateFederativeBirth* y *hasStateMedicalUnit*, tal como se muestran y describen en la Tabla 4.4.

Tabla 4.4: Propiedades de objetos de la ontología *PatientElectronicRecord*.

Propiedad de objetos	Dominio clase	Rango clase	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 del paciente.
hasResidenceState	Patient	StateFederative	Estado de residencia del paciente.
hasStateFederativeBirth	Patient	StateFederative	Estado de nacimiento del paciente.
hasStateMedicalUnit	Patient	StateFederative	Paciente tiene estado de unidad médica.

Una vez definida todas las clases, las propiedades de datos con el tipo de dato correspondiente y las propiedades de objetos de las clases que se relacionan en dicha ontología, el editor de ontologías, con ayuda de una herramienta de visualización llamada OntoGraf, se procede a visualizar el modelo principal, tal como se muestra en la Figura 4.2, de las clases relacionadas del modelo.

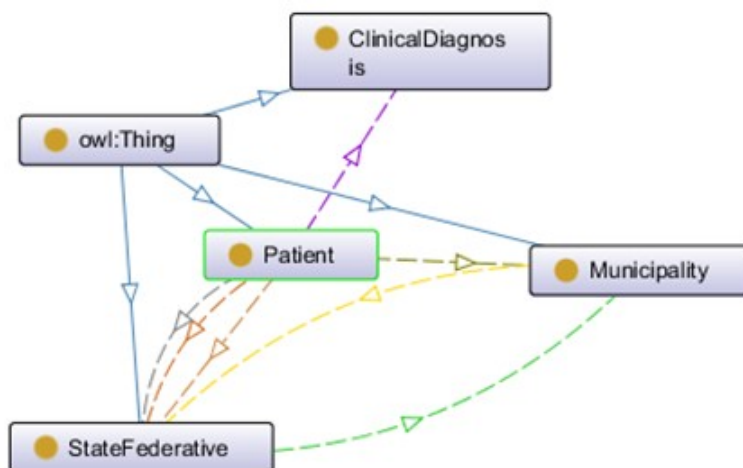


Figura 4.2: Modelado de la ontología PatientElectronicRecord en Protégé (OntoGraf).

4.4.2. Ontología Vaccine

Siguiendo con el desarrollo, en la ontología *Vaccine* se describió sobre aquellas propiedades consultadas y analizadas sobre los tipos de vacunas orientadas a prevenir la enfermedad del COVID-19¹⁶, por ello se definieron las siguientes clases y subclases presentadas en la Tabla 4.5 Clases y subclases que abordan los conceptos de vacunas: *COVID-19_Vaccine*, *mRNA_COVID-19_Vaccine* y *SARS- CoV-2_mRNA_Vaccine*.

Tabla 4.5: Clases de la ontología Vaccine.

Clase / Subclase	Descripción
COVID-19_Vaccine	Cualquier vacuna que pueda prevenir la enfermedad por COVID-19.
mRNA_COVID-19_Vaccine	Cualquier vacuna basada en ARNm que pueda prevenir la enfermedad por COVID -19.
SARS- CoV-2_mRNA_Vaccine	Una formulación que consiste en nanopartículas lipídicas (LNP) que encapsulan ARN mensajero (ARNm) que codifica la glicoproteína de punta (SP) del COVID-19, con posibles actividades inmunizantes y anti-COVID-19.

¹⁶<https://www.gob.mx/cofepris/acciones-y-programas/vacunas-covid-19-autorizadas>

4. DISEÑO Y CONSTRUCCIÓN DEL MODELO ONTOLÓGICO

Con las clases definidas, se procede a incorporar aquellas propiedades de datos que permiten definir y restringir el tipo de individuo que pertenece a cada clase. Tal como se muestran y describen en la Tabla 4.6.

- a) a) La clase de *COVID-19_Vaccine*, es padre de las subclases *mRNA_COVID-19_Vaccine* y *SARS-CoV-2_mRNA_Vaccine*, por lo que asignar propiedades de datos a dicha clase hereda a las demás subclases, por ello se definen las propiedades de datos de: *hasVaccineID*, *hasVaccineDefinition* y *hasVaccineName*.

Tabla 4.6: Propiedades de datos de la ontología Vaccine.

Propiedad de dato	Descripción
hasVaccineID	Hace referencia al identificador de la vacuna en la ontología.
hasVaccineDefinition	Describe el compuesto y función de la vacuna.
hasVaccineName	Despliega el nombre de la vacuna anti-COVID-19.

Definida todas las clases y subclases, las propiedades de datos con el tipo de dato correspondiente y las propiedades de objetos de las clases y subclases que se relacionan en dicha ontología, con ayuda del OntoGraf se procede a visualizar las clases relacionadas del modelo, tal como se muestra en la Figura 4.3.

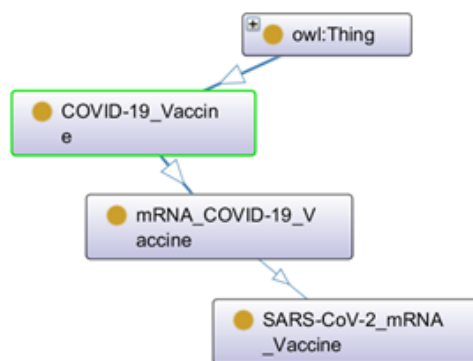


Figura 4.3: Modelado de la ontología Vaccine en Protégé (OntoGraf).

4.4.3. Ontología Medicament

En la ontología *Medicament* se describieron aquellas propiedades consideradas en el análisis de medicamentos y enfermedades, por ello se definen las siguientes clases: *ChemicalStructure*, *MechanismOfAction*, *Medicament*, *PharmacologicalTreatment*, *NDFRT_Disease*, *TherapeuticCategory* y *PhysiologicalEffect*, estas clases almacenarán aquellas propiedades de los medicamentos y las enfermedades, las clases se presentan y definen en la Tabla 4.7.

Tabla 4.7: Clases de la ontología Medicament.

Clase	Descripción
ChemicalStructure	Estructura química de los medicamentos.
MechanismOfAction	Mecanismo de acción de los medicamentos.
Medicament	Medicamentos.
PharmacologicalTreatment	Tratamiento farmacológico de enfermedades.
NDFRT_Disease	Enfermedades del Archivo Nacional de Drogas - Terminología de Referencia (NDF-RT).
TherapeuticCategory	Categoría terapéutica de los medicamentos.
PhysiologicalEffect	Efectos fisiológicos de los medicamentos.

Una vez definidas todas las clases de la ontología medicamentos, se procede a incorporar aquellas propiedades de datos que permiten definir y restringir el tipo de individuo que pertenece a cada clase. Tal como se muestra y describe en la Tabla 4.8.

- La clase de *ChemicalStructure*, en la que se define la propiedad de dato de: *hasChemicalName*.
- Para la clase de *PhysiologicalEffect*, en la que se define la propiedad de dato de: *hasEffectName*.
- En la clase de *MechanismOfAction*, en la que se define la propiedad de dato de: *hasMechanismName*.
- Seguido de la clase de *NDFRT_Disease*, en la que se define la propiedad de dato de: *hasNDFRT_DiseaseName*.
- En la clase de *Medicament*, en la que se define la propiedad de dato de: *hasDisplayName*, *hasNUI*, *hasRxNormCUI* y *hasSynonym*.

4. DISEÑO Y CONSTRUCCIÓN DEL MODELO ONTOLÓGICO

- f) Por último, la clase de *TherapeuticCategory*, en la que se define la propiedad de dato de: *hasTherapeuticCategoryName*.

Tabla 4.8: Propiedades de datos de la ontología Medicament.

Propiedad de dato	Dominio clase	Rango tipo de dato	Descripción
hasChemicalName	ChemicalStructure	string	Nombre de estructura química.
hasEffectName	PhysiologicalEffect	string	Nombre del efecto fisiológico.
hasMechanismName	MechanismOfAction	string	Nombre del mecanismo de acción.
hasNDFRT_DiseaseName	NDFRT_Disease	string	Nombre de enfermedad.
hasDisplayName	Medicament	string	Nombre del medicamento.
hasNUI	Medicament	string	Identificador único numérico.
hasRxNormCUI	Medicament	string	Identificador único numérico (normalizado).
hasSynonym	Medicament	string	Sinónimo del nombre del medicamento.
hasTherapeuticCategoryName	TherapeuticCategory	string	Nombre de la categoría terapéutica.

Siguiendo en el proceso de desarrollo del modelo ontológico, se procede a realizar las propiedades de objetos que permiten obtener las relaciones semánticas entre clases de la misma ontología definiendo los dominios y rangos, las cuales son: *hasChemicalStructure*, *hasMechanismOfAction*, *hasMedication*, *hasPhysiologicalEffect*, *hasTherapeuticCategory*, *mayDiagnose*, *mayPrevent* y *mayTreat*, tal como se muestran y describen en la Tabla 4.9.

Tabla 4.9: Propiedades de objetos de la ontología Medicament.

Propiedad de objeto	Dominio clase	Rango clase	Descripción
hasChemicalStructure	Medicament	ChemicalStructure	Estructura química de medicamentos.
hasMechanismOfAction	Medicament	MechanismOfAction	Mecanismos de acción en medicamentos.
hasMedication	PharmacologicalTreatment	Medicament	Tratamiento medicamento tiene medicación.
hasPhysiologicalEffect	Medicament	PhysiologicalEffect	Efectos fisiológicos de medicamentos.
hasTherapeuticCategory	Medicament	TherapeuticCategory	Categoría terapéutica de medicamentos.
mayDiagnose	Medicament	NDFRT_Disease	Medicamento puede diagnosticar enfermedad.
mayPrevent	Medicament	NDFRT_Disease	Medicamento puede prevenir enfermedad.
mayTreat	Medicament	NDFRT_Disease	Medicamento puede tratar enfermedad.

Por último, después de haber definido todas las clases y subclases, las propiedades de datos con el tipo de dato correspondiente y las propiedades de objetos de las clases que se relacionan en dicha ontología, con ayuda del OntoGraf se procede a visualizar las clases relacionadas del modelo, tal como se muestra en la Figura 4.4.

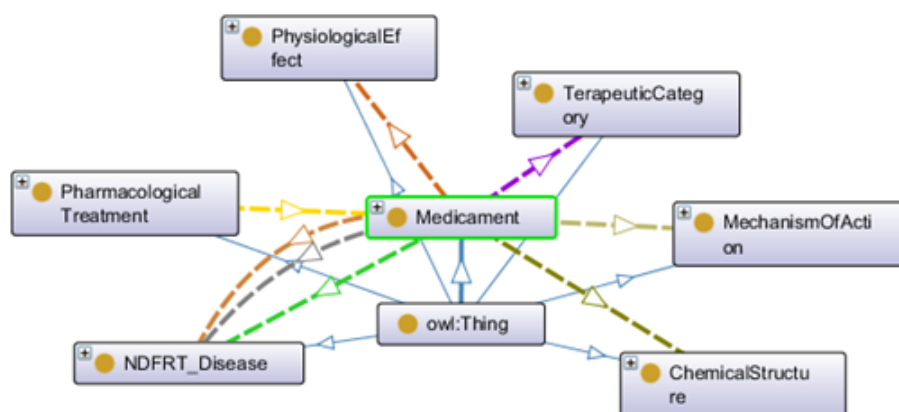


Figura 4.4: Modelado de la ontología Medicament en Protégé (OntoGraf).

4.4.4. Ontología LaboratoryTest

Para el desarrollo de la ontología *LaboratoryTest*, se describieron aquellas propiedades consultadas para la creación de un modelo ontológico de pruebas de laboratorios, específicamente la consulta de aquellas pruebas de laboratorio para detectar el COVID-19. De acuerdo con el análisis de la información consultada, menciona que ante la emergencia sanitaria por el virus SARS-CoV-2¹⁷, la Comisión Federal para la Protección contra Riesgos Sanitarios (Cofepris), estableció un protocolo de evaluación en conjunto con el Instituto Nacional de Ciencias Médicas y Nutrición “Salvador Zubirán” y TecSalud del Tecnológico de Monterrey, por ello se definen las siguientes clases: *LaboratoryTest*, que abarca 73 subclases de los tipos de pruebas de laboratorio en general, por ello en la Tabla 4.10 solo se presenta la clase *PCR_test*, que es la subclase de las pruebas PCR.

Tabla 4.10: Clases de la ontología LaboratoryTest.

Clase / Subclase	Descripción
LaboratoryTest	Cualquier prueba de laboratorio clínico.
PCR_test	Prueba de reacción en cadena de la polimerasa (PCR).

Teniendo las clases definidas, se procede a incorporar aquellas propiedades de datos que permiten definir y restringir el tipo de individuo que pertenece a cada clase. Tal

¹⁷<https://www.gob.mx/cofepris/acciones-y-programas/pruebas-autorizadas-para-covid-19>

4.4.5. Ontología Symptom

Por último, en la ontología *Symptom* se describen aquellas propiedades consultadas acerca de los síntomas que se presentan cuando se padece una enfermedad, esto de acuerdo con el Centers of Disease Control and Prevention¹⁸, para ello se define la clase: *Symptom*, presentada en la Tabla 4.12.

Tabla 4.12: Clases de la ontología Symptom.

Clase	Descripción
Symptom	Síntomas de enfermedades

Teniendo las clases definidas, se procede a incorporar aquellas propiedades de datos que permiten definir y restringir el tipo de individuo que pertenece a cada clase. Tal como se muestra y describe en la Tabla 4.13.

En la clase de *Symptom* se definen las propiedades de datos de: *hasSymptomDefinition*, *hasSymptomID* y *hasSymptomName*.

Tabla 4.13: Propiedades de datos de la ontología Symptom.

Propiedad de dato	Dominio clase	Rango tipo de dato	Descripción
hasSymptomDefinition	Symptom	string	Definición del síntoma.
hasSymptomID	Symptom	string	Identificador del síntoma.
hasSymptomName	Symptom	string	Nombre del síntoma.

De igual forma, una vez definida la clase *Symptom*, las propiedades de datos con el tipo de dato correspondiente, se procedió a visualizar las clases relacionas del modelo con ayuda de OntoGraf, tal como se muestra en la Figura 4.6.

¹⁸<https://www.cdc.gov/>

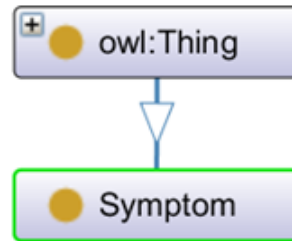


Figura 4.6: Modelado de la ontología Symptom en Protégé (OntoGraf).

4.5. Extracción de módulos de ontologías existentes para su reutilización

En esta fase de la primera etapa, con ayuda del programa editor de ontologías Protégé, se procedió a visualizar aquellos modelos ontológicos consultados en la etapa de búsqueda de ontologías existentes. Seguido de ello, se procedió a desarrollar e implementar un programa de consola en Java, de igual forma se hace uso de OWL API 4.5.9¹⁹, que es una API que permite trabajar con ontologías en formatos RDF o OWL (cómo lo es este caso).

De acuerdo con (Schlicht and Stuckenschmidt, 2006), la reutilización parcial o de módulos de ontologías consiste en utilizar métodos para separar modelos en módulos que optimizan la integridad del razonamiento ya sea de forma local o distribuido, ha sido bien aceptado por la comunidad de web semántica, ya que generar un modelo más simplificado al original es de gran ayuda empleando una serie de criterios.

Tal como se ha consultado en la fase 4.2 sobre la búsqueda de ontologías existentes, y se ha seleccionado aquellas ontologías a reutilizar de forma parcial, ya que, de acuerdo con los criterios establecidos en (Cuenca Grau, 2007), el hacer importación grandes modelos demandan la importación de todos los axiomas del modelo original y esto al final demanda mucha memoria para hacer razonamiento del modelo completo.

Debido a esto, se consideró definir e implementar un enfoque diferente, utilizar un método basado en modularización para la reutilización de las ontologías seleccionadas, este método consiste en la transformación de parte de las definiciones ya sean por símbolos que definen las ontologías que son relevantes para el modelo específico que se está integrando, reduciendo el costo de cómputo cuando se importan dichos módulos,

¹⁹<http://owlcs.github.io/owlapi/>

tal como describen las fases del método a utilizar, a continuación:

1. Seleccione las ontologías a reutilizar, esto debe ser descargado y analizado localmente con ayuda del editor de ontologías.
2. Identificar los datos o atributos relevantes a reutilizar para completar la conceptualización de la ontología que se está integrando.
3. Definir e implementar estructuras de datos usando el paradigma de Programación Orientada a Objetos para permitir 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.
5. Definir y construir la estructura de la ontología que se utilizará para contener los conceptos y atributos de interés.
6. Desarrollar un programa para exportar automáticamente la ontología con la lista de conceptos obtenidos.

En las siguientes secciones, se describe un ejemplo de caso de uso de la aplicación de este método para los conceptos requeridos de la ontología *Symptom*, ya que dicho método se aplicó para todos los modelos ontológicos seleccionados anteriormente.

4.5.1. Modelo ontológico Symptom

De acuerdo con el método para la reutilización de módulos de ontologías, se procedió a implementar dicho método, tal como se describe a continuación.

1. Selección de la ontología **SYMP** del portal, para la representación de síntomas de las enfermedades, para su reutilización e integración con otras ontologías.

Dentro del vocabulario medico orientado a la representación de perfiles de pacientes con COVID-19, se considera el concepto de síntoma debido a que este se relaciona con las enfermedades causantes, dicha ontología aborda vocabulario médico y propiedades para las diferentes clases y subclases que constituyen dicho modelo ontológico, en la primera parte de la Figura 4.7, se puede visualizar las métricas de dicha ontología descargada y abierta desde el Protégé, solo tiene una propiedad de objetos, 21 propiedad de anotaciones, cuenta con un total de 6271 axiomas, siendo de estos 889 axiomas lógicos, 1015 axiomas declarados, cuenta con un conjunto de 993 clases de los tipos de síntomas, y un total de 889 subclases.

De igual forma, en la parte derecha se puede visualizar el identificador **SYMP:0000613** de la subclase seleccionada, este hace referencia al síntoma de “fever” (fiebre), que a su vez es una subclase de la clase “neurological and

psychological symptom” (síntoma neurológico y psicológico), de igual forma cuenta con anotaciones y propiedades que son útiles para conocer los detalles específicos cuando se selecciona la parte del modelo a exportar. Dichas subclases serán procesadas como instancias de datos del nuevo modelo, esto para tener un mejor modelo donde una clase de síntomas contengan todos aquellos síntomas como individuos y no como subclases.

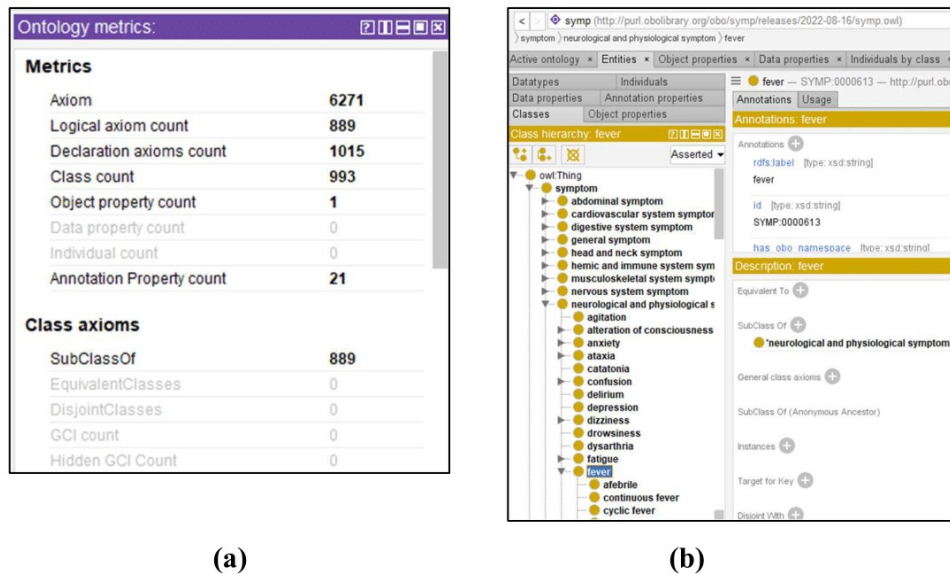


Figura 4.7: (a) Métricas de ontología Symp (con Protégé), (b) Jerarquía de clases de ontología Symp (con Protégé).

- Identificación de los datos y atributos a importar, en esta fase de revisión se procesó y guardo la ontología en formato Turtle²⁰ para visualizar dicha sintaxis del formato. En la visualización de sintaxis se seleccionó la subclase neurological and psychological symptom, específicamente el concepto de **SYMP:0000613**, donde la sintaxis muestra las propiedades de dicho concepto, tal como se muestra en el código a continuación.

²⁰[https://dbpedia.org/page/Turtle_\(syntax\)](https://dbpedia.org/page/Turtle_(syntax))

4. DISEÑO Y CONSTRUCCIÓN DEL MODELO ONTOLÓGICO

```

### 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 ;
  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 dicha ontología 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 contemplaron y adaptaron las propiedades de dato: *hasSymptomDefinition*, *hasSymptomID*, *hasSymptomName* y sobre todo la IRI de dicha clase.

3. Para la extracción y representación de síntomas en objetos, se desarrolló y utilizó una aplicación de clases de Java, la clase *Symptom*, *SymptomOntologyCreation* y *OntologyUtils*, tal como se muestra en la Figura 4.8.

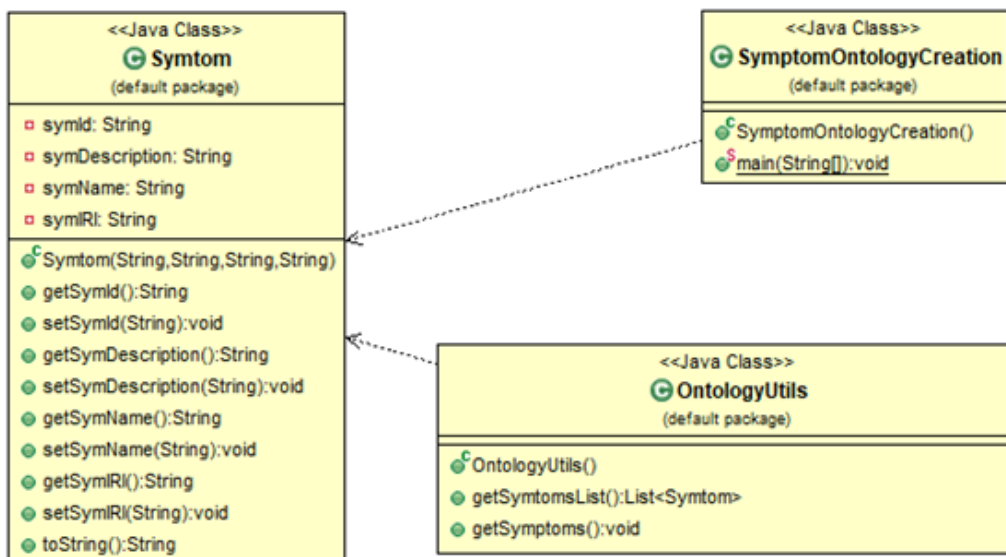


Figura 4.8: Clases java para exportar modulo a nueva ontología.

4. En esta fase, se implementó dicha clase de *OntologyUtils* mediante el uso de la API RDF4J²¹, con la finalidad de implementar una consulta **SPARQL** (tal como se muestra debajo), para generar la lista de objetos de los síntomas y guardarlas en una nueva ontología llamada *Symptom.owl*.

```

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 .
}

```

5. En la construcción de la ontología fue necesario diseñar el nuevo modelo en el cual se importan los conceptos de clase y definiciones como individuos, de esta forma se tiene en cuenta el tipo de dato de la consulta devuelta, con la finalidad de adaptarlo a las propiedades de datos establecidas en el nuevo modelo, como se muestra en la Figura 4.9. En este nuevo modelo se simplifican las propiedades de datos y los conceptos de clases se definen como instancias de los síntomas, con la finalidad que dicho modelo sea más ligero y se útil para la integración con la ontología principal para la representación de perfiles de pacientes con COVID-19.

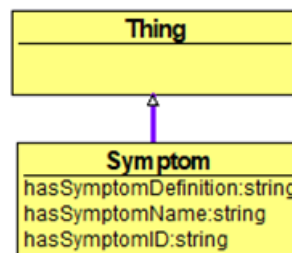


Figura 4.9: Ontología Symptom para la representación de síntomas.

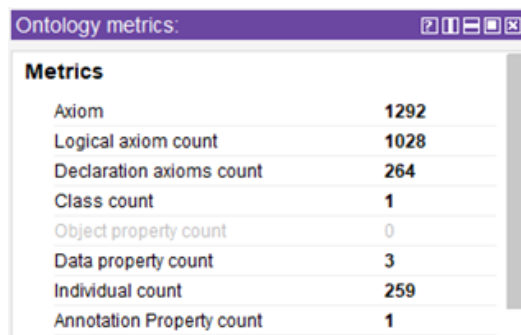
6. Finalmente, en la clase *SymptomOntologyCreation*, se procedió a ejecutar el método principal del programa de Java para la población de datos, para la modularización de ontología mediante consultas SPARQL, el código de la consulta se muestra a continuación, donde se procede a instanciar el método *OntologyUtils*

²¹<https://rdf4j.org>

que contiene la consulta SPAQRL y posteriormente este devuelve la lista de síntomas y se envía a través del método *populateSymptomOntology* al nuevo modelo de la ontología *Symptom*.

```
public class SymptomOntologyCreation {
    public static void main(String[] args) {
        OntologyUtils util = new OntologyUtils();
        List<Symptom> symptomList = util.getSymptomsList();
        for(int i = 0; i < symptomList.size(); i++) {
            System.out.println("=====");
            System.out.println("Symptom id " + symptomList.get(i).toString());
        }
        util.populateSymptomsOntology("Ontologies/Symptom.owl", symptomList);
    }
}
```

Como resultado de la implementación del método para la reutilización parcial, se obtiene un nuevo módulo de ontología de *Symptom*, esta ontología consta de 1 clase, 3 propiedades de datos, 259 individuos y un total de 1292 axiomas, tal como se muestra en la Figura 4.10, cabe mencionar que este método se realizó para el resto de las ontologías seleccionadas para la reutilización parcial.



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

Figura 4.10: Métricas de Ontología Symptom.

4.6. Integración de módulos en Ontología principal

En definitiva, se obtiene los 4 modelos ontológicos que han sido reutilizados de forma modular, seguido de ello, con ayuda de una herramienta del editor de ontologías Protégé se procede a importar dichos modelos a la ontología principal *PatientElectronicRecord*, que funciona como ontología de soporte de para la representación de perfiles de pacientes con COVID-19, con la finalidad de obtener los modelos ontológicos correspondientes de acuerdo con el diseño inicial se tienen las siguientes ontologías: *LaboratoryTest*, *Medicament*, *Symptom*, *Vaccine*, *PatientElectronicRecord*, dichos modelos se describen en la Tabla 4.14.

Tabla 4.14: Ontologías resultantes del modelo integrado.

Ontología	Descripción	ID de Concepto
<i>LaboratoryTest</i>	Lab's test	http://www.semanticweb.org/darinelstein/ontologies/2021/8/LaboratoryTest
<i>Medicament</i>	Medicaments	http://www.semanticweb.org/darinelstein/ontologies/2021/8/Medicament
<i>Symptom</i>	Symptoms	http://www.semanticweb.org/darinelstein/ontologies/2021/8/Symptom
<i>Vaccine</i>	Vaccines	http://www.semanticweb.org/darinelstein/ontologies/2021/8/Vaccine
<i>PatientElectronicRecord</i>	Patient Clinical Information	http://www.semanticweb.org/darinelstein/ontologies/2021/8/PatientElectronicRecord

De igual forma, en la integración final de las ontologías reutilizadas y la ontología principal, se deben de generar aquellas relaciones de propiedad de objetos entre clases de dichas ontologías. Una vez se ha integrado el modelo ontológico final, se procede a conceptualizar las propiedades de objetos que deben resultar al relacionar clases de diferentes ontologías integradas, como lo son: *diseaseHasSymptom*, *hasDiagnosedDisease*, *hasLabTest*, *hasPharmalogicalTreatment*, *hasVaccine*, *hasVaccine*, tal como se muestra y describen las relaciones de objetos en la Tabla 4.15.

4. DISEÑO Y CONSTRUCCIÓN DEL MODELO ONTOLÓGICO

Tabla 4.15: 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	NDFRT_Disease	LaboratoryTest	Síntoma requiere prueba de laboratorio.

Con ayuda de la herramienta OntoGraf²² del editor de ontologías, se procedió a visualizar el modelo ontológico integrado, de modo que se puede ver las relaciones de objetos entre clases de las 5 ontologías total integradas, tal como se muestra en la Figura 4.11. De igual forma en la Figura 4.12, se muestra un fragmento del código OWL del modelo ontológico integrado *PatientElectronicRecord*, se muestra en sintaxis Manchester

footnote[23]<https://www.w3.org/2007/OWL/wiki/ManchesterSyntax> para tener una mejor comprensión de dicho código, en la parte del encabezado se muestra los principales prefijos de las ontologías importadas seguido de las propiedades de anotaciones y las propiedades de datos, de igual forma se puede apreciar una propiedad de objeto que relaciona dos clases del modelo integrado.

²²<https://protegewiki.stanford.edu/wiki/OntoGraf>

4.6 Integración de módulos en Ontología principal

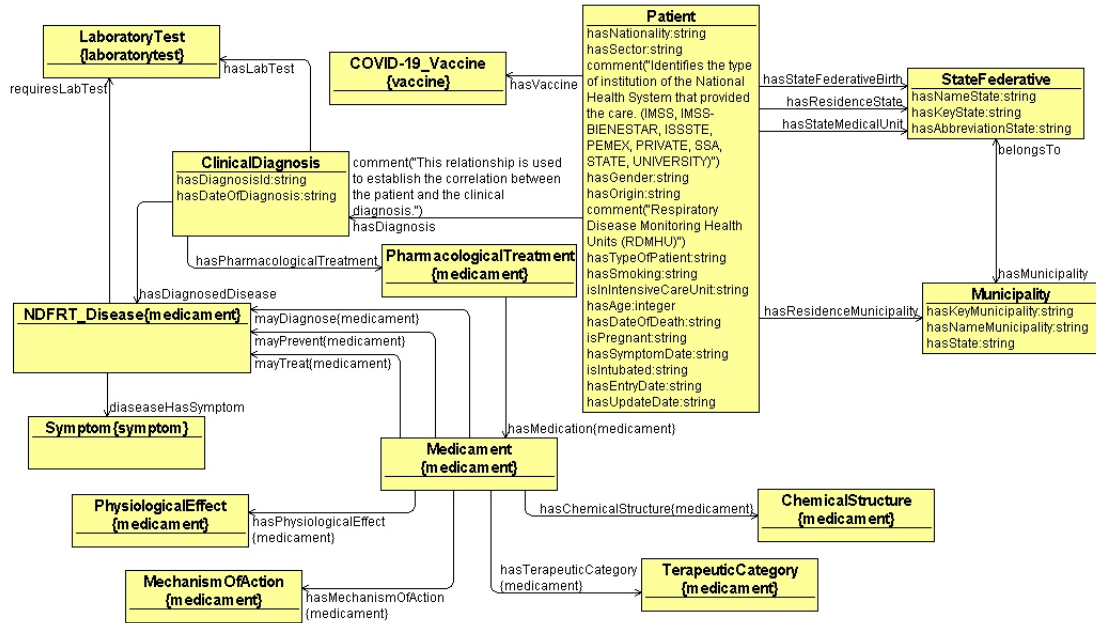


Figura 4.11: Modelado de la ontología PatientElectronicRecord en Protégé (OntoGraf).

4. DISEÑO Y CONSTRUCCIÓN DEL MODELO ONTOLÓGICO

```
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 4.12: Fragmento de código OWL de la ontología PatientElectronicRecord.

4.7. Definición de axiomas y reglas de inferencia

Posteriormente, en la parte final de la primera etapa de esta metodología, se procedió a definir los axiomas que integran el modelo ontológico, así como las reglas de inferencia que permiten generar automáticamente nuevas relaciones semánticas y nuevas definiciones relacionadas con el dominio del modelo ontológico, esto mediante el uso de un lenguaje de consulta integrado en el editor de ontologías antes mencionado y utilizado, de acuerdo a los criterios de axiomas para que el modelo ontológico cuenta con los principios de diseño necesarios con la finalidad de contar con un nivel de expresividad y coherencia en relación a los conceptos establecidos en el dominio.

De igual forma, se hizo uso de la herramienta antes mencionada, se para el procesamiento de ontologías, concretamente haciendo uso de un módulo para la creación y la ejecución de reglas de inferencia en un lenguaje de web semántica, como define (McGuinness et al., 2004), con la finalidad de procesar los datos que se importen a dicho modelo para obtener nuevo conocimiento en base a hechos establecidos en dicho modelo.

Evaluación del modelo ontológico

En esta sección se presenta la segunda parte de la metodología propuesta para este proyecto, esta etapa aborda las evaluaciones realizadas al modelo ontológico, las evaluaciones se realizan desde la población del modelo ontológico con un conjunto de datos de instancias los cuales son procedentes de una fuente validada del gobierno de México. Seguido de la evaluación basada en preguntas de competencias hacia el modelo con instancias de datos, las cuales se plantean en lenguaje natural y posteriormente se traducen esas preguntas a un lenguaje de consultas con la finalidad de emplear dichas consultas al modelo ontológico y poder responder a esas preguntas con conocimiento inferido y determinar el nivel de competencia del modelo. Finalmente, como evaluación se revisan los principios de diseño de acuerdo con literatura especializada por expertos en el tema, de tal forma que se seleccionen los principios más relevantes y se evalúa el modelo integrado.

5.1. Evaluación mediante el uso del modelo con instancias de datos

En la fase de evaluación del modelo con instancias de datos, se dio a la tarea de buscar aquellos portales de conjuntos de datos de pacientes COVID-19 ya sean nacionales o internacionales, con la finalidad de poblar el modelo ontológico y evaluar su funcionamiento de acuerdo con las necesidades planteadas en el modelo, de esta forma se llegó al portal del gobierno de México, concretamente a los Datos Abiertos Dirección General de Epidemiología proporcionados por la Dirección General de Epidemiología (DGE) conforme al Decreto publicado en el diario Oficial de la Federación el 20 de Febrero del 2015, que establece la regulación en materia de datos abiertos, en base en los ordenamientos aplicables en dicha materia, pone a disposición de la población en general los datos referente 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, de acuerdo a la metodología de la Figura 5.1.

5. EVALUACIÓN DEL MODELO ONTOLÓGICO

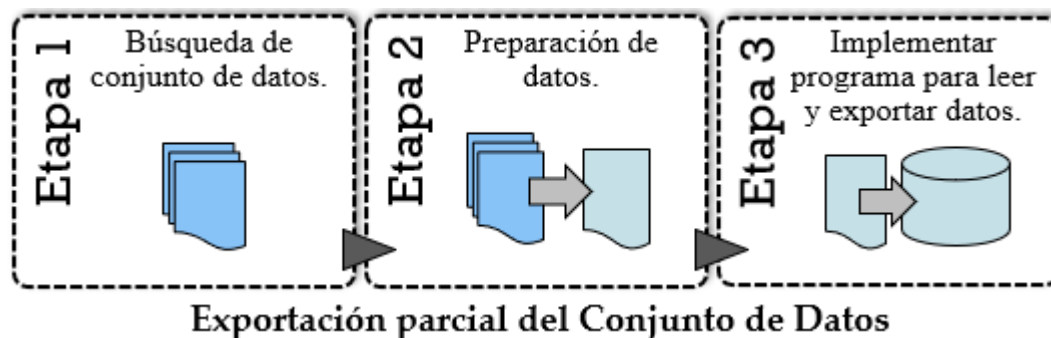


Figura 5.1: Etapas de metodología de evaluación.

Debido a lo anterior, se procedió a descargar el archivo separado por comas .csv (Comma Separated Values) del conjunto de datos de pacientes con COVID-19 llamado *220320COVID19MEXICO.csv*, el archivo hasta la fecha del desarrollo de este proyecto de investigación cuenta con un conjunto de datos de más de 6000000 de registros de pacientes mexicanos y extranjeros atendidos en las Unidades de Salud Monitoras de Enfermedad Respiratoria viral (USMER), no obstante, para fines de pruebas y debido a las limitantes de equipo de cómputo, se trabajó con solo 500 registros de pacientes, filtrados por nacionalidad, de igual forma, se adaptó el conjunto de datos de registro de pacientes de acuerdo con las propiedades consideradas para el modelo ontológico, ya que el archivo de origen cuenta con 40 campos de los datos del paciente, tal como se muestra en la Figura 5.2.

```

FECHA_ACTUALIZACION,ID_REGISTRO,ORIGEN,SECTOR,ENTIDAD_UM,SEXO,ENTIDAD_NAC,ENTIDAD_RES,MUNICIPIO_RES,TIPO_PACIENTE,FECHA_INGRESO,FECHA_SINTOMAS
04/08/2022,z5c20f,1,12,31,2,31,31,102,1,25/05/2020,22/05/2020,9999-99-99,97,2,40,1,97,1,1,2,2,2,2,2,2,2,2,2,2,1,1,1,1,3,99,México,97,97
04/08/2022,z2a1ad,2,12,4,2,4,4,2,1,09/06/2020,05/06/2020,9999-99-99,97,2,47,1,97,2,2,2,2,2,2,2,2,2,2,2,2,1,1,1,1,1,1,3,99,México,97,97
04/08/2022,z3bf80,2,12,8,2,8,8,37,1,28/07/2020,20/07/2020,9999-99-99,97,2,35,1,97,2,2,2,2,2,2,2,2,2,2,2,2,2,2,1,1,1,1,3,99,México,97,97
04/08/2022,z54912,1,12,31,1,31,31,79,1,12/06/2020,10/06/2020,9999-99-99,97,2,56,1,2,2,2,1,2,2,2,2,2,2,2,2,2,2,1,1,1,1,3,99,México,97,97
04/08/2022,z552ac,1,12,1,1,9,1,1,1,02/06/2020,30/05/2020,9999-99-99,97,2,58,1,2,2,2,1,2,2,2,2,2,2,2,2,2,1,1,1,1,3,99,México,97,97
04/08/2022,z59345,1,12,31,2,31,31,89,1,01/07/2020,30/06/2020,9999-99-99,97,2,37,1,97,2,2,2,2,2,2,2,2,2,2,2,2,2,2,1,1,1,1,3,99,México,97,97
04/08/2022,z4887b,2,12,8,1,10,8,37,1,07/08/2020,05/08/2020,9999-99-99,97,2,71,1,2,2,2,2,2,2,2,2,2,2,2,2,2,1,1,1,1,3,99,México,97,97
04/08/2022,z13788,1,12,9,1,9,9,4,1,23/09/2020,23/09/2020,9999-99-99,97,2,67,1,2,2,2,1,2,2,2,2,2,2,2,2,2,2,1,1,1,1,3,99,México,97,97
04/08/2022,z3d1e2,2,12,9,1,9,9,5,1,22/04/2020,20/04/2020,9999-99-99,97,2,45,1,2,2,2,2,2,2,2,2,2,2,2,2,2,1,1,1,1,3,99,México,97,97
04/08/2022,z21f6f,2,12,7,1,7,7,9,1,27/04/2020,24/04/2020,9999-99-99,97,2,40,1,2,2,2,2,1,2,2,2,2,2,2,2,2,2,1,1,1,1,3,99,México,97,97
04/08/2022,z99040,1,12,19,1,19,19,45,1,06/09/2020,03/09/2020,9999-99-99,97,2,57,1,2,2,2,1,2,2,2,2,2,2,2,2,2,2,1,1,1,1,3,99,México,97,97
04/08/2022,z21c58,2,12,19,1,19,19,26,1,10/07/2020,06/07/2020,9999-99-99,97,2,49,1,2,2,2,2,2,2,2,2,2,2,2,2,2,1,1,1,1,3,99,México,97,97

```

Figura 5.2: Fragmento de conjunto de datos sin procesar.

Con ayuda del programa de hoja de cálculo de Excel de la paquetería de Office, que es una suite de ofimática, se filtró y acoto el total de registro de pacientes mexicanos del archivo .csv, de igual forma se filtró el número de campos a utilizar para el modelo, obteniendo un total de 500 registros de datos procesados, con una simplificación de 28 campos de cada registro, para la representación de perfil de un paciente con dicha enfermedad, tal como se muestra en la Figura 5.3. Finalmente, se guardó los datos de registro de pacientes, en un archivo llamado *patinets.xlsx*. Con la finalidad de procesarlos a través de una aplicación de consola desarrollada mediante el lenguaje de programación

de Java²⁴.

```
ID_REGISTRO,ORIGEN,SECTOR,EDAD,SEXO,TIPO_PACIENTE,INTUBADO,UCI,EMBARAZO,TABAQUISMO,PAIS_NACIONALIDAD,FECHA_INGRESO,FECHA_SINTOMAS,FECHA_DEF,FECHA_ACTUALIZACION,ENTIDAD,
23bF80,OUTSIDE_RDMHU,SSA,35,MALE,AMBULATORY,NOT APPLICABLE,NOT APPLICABLE,NOT APPLICABLE,NO,MEXICO,07/28/2020,07/28/2020,00/00/0000,03/20/2022,08,08,08,037,2,2,2,2,2,2,2,
21e370,RDMHU,SSA,42,FEMALE,AMBULATORY,NOT APPLICABLE,NOT APPLICABLE,NO,NO,MEXICO,04/22/2020,04/18/2020,00/00/0000,03/20/2022,14,14,14,085,2,2,2,2,2,2,2,1,7
z2e974,RDMHU,ISSSTE,34,FEMALE,AMBULATORY,NOT APPLICABLE,NOT APPLICABLE,NO,NO,MEXICO,02/28/2021,02/28/2021,00/00/0000,03/20/2022,24,24,24,035,99,2,2,2,2,2,2,2,1,7
z27067,RDMHU,SSA,51,MALE,AMBULATORY,NOT APPLICABLE,NOT APPLICABLE,NOT APPLICABLE,NO,MEXICO,08/18/2020,08/17/2020,00/00/0000,03/20/2022,09,09,09,007,2,2,2,2,2,2,2,1,7
21da1e,RDMHU,SSA,30,MALE,AMBULATORY,NOT APPLICABLE,NOT APPLICABLE,NOT APPLICABLE,NO,MEXICO,03/09/2020,03/05/2020,00/00/0000,03/20/2022,01,01,01,001,99,2,2,2,2,2,2,2,1,7
2393a3,RDMHU,SSA,47,FEMALE,AMBULATORY,NOT APPLICABLE,NOT APPLICABLE,NO,NO,MEXICO,12/28/2020,12/28/2020,00/00/0000,03/20/2022,09,09,09,017,2,2,2,2,2,2,2,2,7
259da,RDMHU,SSA,47,MALE,AMBULATORY,NOT APPLICABLE,NOT APPLICABLE,NOT APPLICABLE,NO,MEXICO,06/28/2020,06/24/2020,00/00/0000,03/20/2022,07,07,07,078,2,2,2,2,2,2,2,2,6
25ba5b,OUTSIDE_RDMHU,SSA,38,MALE,AMBULATORY,NOT APPLICABLE,NOT APPLICABLE,NOT APPLICABLE,NO,MEXICO,07/31/2020,07/28/2020,00/00/0000,03/20/2022,08,10,08,037,2,2,2,2,2,2,2,
21f605,OUTSIDE_RDMHU,SSA,42,MALE,AMBULATORY,NOT APPLICABLE,NOT APPLICABLE,NOT APPLICABLE,NO,MEXICO,08/26/2020,08/20/2020,00/00/0000,03/20/2022,07,07,07,024,99,2,2,2,2,2,2,2,
24985e,OUTSIDE_RDMHU,SSA,26,MALE,AMBULATORY,NOT APPLICABLE,NOT APPLICABLE,NOT APPLICABLE,NO,MEXICO,05/21/2020,05/21/2020,00/00/0000,03/20/2022,19,19,19,039,2,2,2,2,2,2,2,
2eace,RDMHU,STATE,7,FEMALE,HOSPITALIZED,NO,NO,NO,MEXICO,09/23/2020,09/20/2020,00/00/0000,03/20/2022,15,15,15,106,2,2,2,2,2,2,2,1,2,1,6
238de4,RDMHU,SSA,7,FEMALE,AMBULATORY,NOT APPLICABLE,NOT APPLICABLE,NO,NO,MEXICO,05/23/2020,05/20/2020,00/00/0000,03/20/2022,07,07,07,101,2,2,2,2,2,2,2,2,2,6
2579ac,RDMHU,SSA,33,MALE,AMBULATORY,NOT APPLICABLE,NOT APPLICABLE,NOT APPLICABLE,NO,MEXICO,10/14/2020,10/10/2020,00/00/0000,03/20/2022,15,15,15,060,2,2,2,2,2,2,2,2,1,6
22669f,OUTSIDE_RDMHU,SSA,43,MALE,AMBULATORY,NOT APPLICABLE,NOT APPLICABLE,NOT APPLICABLE,YES,MEXICO,06/18/2020,06/16/2020,00/00/0000,03/20/2022,14,14,14,039,2,2,2,2,1,2,2,
```

Figura 5.3: Fragmento de conjunto de datos reducidos y procesados.

El programa de Java está constituido con clases que permiten la definición e instancia de objetos, comenzando con la clase *PatientClass* que permiten la definición de un paciente de acuerdo las propiedades de datos de dicho individuo, así como la lectura del archivo .xlsx con ayuda de la Api Apache POI Project²⁵, es una API de Java que permite manipular varios formatos de archivo basados en los estándares Office Open XML (OOXML) y el formato de documento compuesto de OLE 2 de Microsoft (OLE2), en resumen, puede leer y escribir archivos de MS Excel, Word y Powerpoint desde un programa de Java, seguido de la clase principal que instancia todos los métodos para la población del modelo, como se describen a continuación.

Clase *PatientClass*. En base al programa de Java para exportar los datos de pacientes al modelo integrado, se desarrolló la clase de pacientes con la finalidad de definir las propiedades de datos que integran a la representación del perfil del paciente, de modo que la clase *Patinets.class* perteneciente al paquete *populateOntology* tiene las siguientes atributos privados: *ID*, *Origin*, *Sector*, *Age*, *Gender*, *PatientsType*, *Intubated*, *ICU*, *Pregnant*, *Smoking*, *Nationality*, *EntryDate*, *SymptomDate*, *UpdateDate*, *MedicalUnit*, *FederativeBirth*, *FederativeResi*, *MunicipalityResi*, *Pneumonia*, *Diabetes*, *COPD*, *Astma*, *Hypertension*, *Infartion*, *Obesity*, *CKD* y *Lab*, con sus respectivos métodos de *setters* y *getters* para acceder a dichos atributos desde un método externo.

Clase *OntoUtils*. En esta clase perteneciente al paquete *crudOntology* se define las definen todos los métodos que permiten acceder al directorio del equipo en el que se trabaja, de modo que se puedan leer archivos y cargarlos en memoria para su posterior procesamiento con Java, por ello con ayuda de dicho programa se dio a la tarea de cargar el archivo *patients.xslx* resultante de los registros y posteriormente imprimir dichos datos en consola para saber si los datos eran procesados correctamente, de igual forma se hizo uso de la API mencionada anteriormente OWL API, para la manipulación del modelo ontológico desde un programa de Java con la finalidad de cargar las clases de dicho modelo de acuerdo a los identificadores de cada ontología y escribir nuevas ins-

²⁴<https://www.java.com/es/>

²⁵<https://poi.apache.org/>

5. EVALUACIÓN DEL MODELO ONTOLÓGICO

tancias en las clases establecidas y así mismo generar las relaciones de propiedades de objetos de dichas instancias, por ello se desarrolló un método para establecer los datos cargados en variables que sean enviadas al método de exportación en la misma clase, mediante ciclos que permiten iterar cada registro del archivo .xlsx e ir registrando en el modelo ontológico, esto con la ejecución de la clase principal.

Clase *Main*. En esta clase principal se procede a instanciar los métodos de objetos sin parámetros de variables, ya que como clase principal perteneciente al paquete *crudOntology* cuenta como el método *Main()* que permite ejecutar el método para cargar ontologías en memoria para su posterior consulta y/o actualización, así mismo se ejecuta el método que permite leer archivo .xlsx y cargarlo en memoria para posterior en el mismo método exportar los datos de registro de pacientes al modelo ontológico integrado.

Las clases involucradas mencionadas anteriormente de dicho programa se muestra en el diagrama de clases en la Figura [5.4](#).

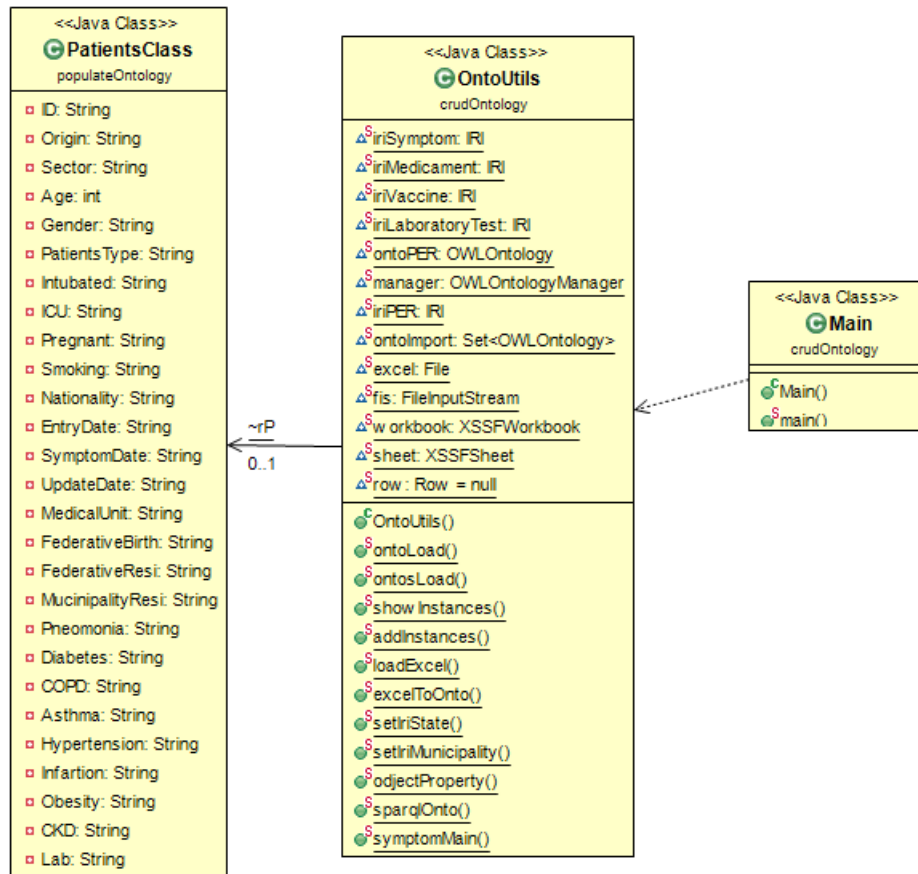


Figura 5.4: Clases java para exportar datos de .xlsx al modelo ontológico.

5.1.1. Caso de uso

Siguiendo con el primer el proceso de evaluación, a continuación, se describe el registro de un paciente con caso positivo COVID-19 a manera de caso de uso. El paciente mexicano *pat01e35a* cuenta con: 56 años de edad, género femenino, originaria de la Ciudad de México de la alcaldía Coyoacán, atendida de igual forma en la Ciudad de México en el sector del Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado (ISSSTE) con fecha de inicios de síntomas del 08 de enero del 2020 y fecha de registro de caso 09 de enero del 2020, sin embargo no se procedió a intubarla ni ser hospitalizada en la USMER, aunque los antecedentes médicos establecen que el paciente tiene comorbilidades de diabetes mellitus y obesidad, las cuales pueden provocar un nivel alto de complicaciones al presentar dicha enfermedad, debido a lo anterior se establece que el paciente puede estar tratándose con algunos de los medicamentos farmacológicos establecidos para cada enfermedad que la paciente presenta, tal como se muestra en la Figura 5.5 del diagrama del registro del perfil de dicha paciente, donde

5. EVALUACIÓN DEL MODELO ONTOLÓGICO

se relacionan aquellas instancias involucradas en dicho registro.

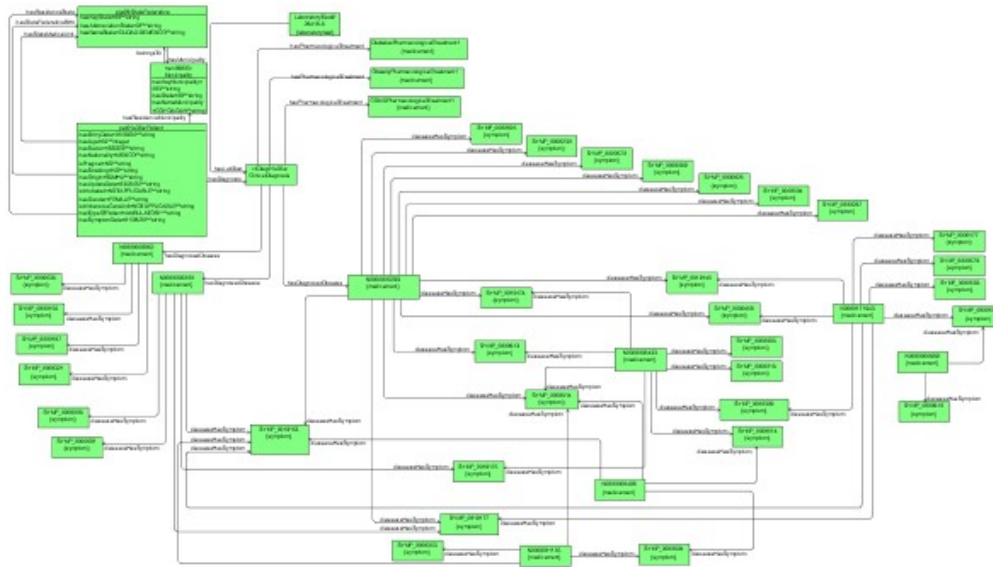


Figura 5.5: Caso de uso de la representación de perfil de un paciente.

5.2. Evaluación basada en principios de diseño de ontologías

Finalmente, se procedió a evaluar dicho modelo de acuerdo con las métricas establecidas por diferentes autores (Gómez-Pérez et al., 1996), (Duque-Ramos et al., 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. En las métricas de evaluación de ontología, de acuerdo con (Gruber, 1995), define un listado de principios de diseño en una ontología, los cuales se mencionan a continuación:

- Claridad.
- Coherencia.
- Extensibilidad.
- Sesgo de codificación mínimo:
- Compromiso ontológico mínimo.

De acuerdo con las métricas presentada se procede a evaluar el modelo ontológico, tal como se describen a continuación:

- a) *Claridad*: Siendo un modelo para la representación de perfiles de pacientes con COVID-19, de acuerdo con las consideraciones para dicho modelo se abarcaron conceptos para dicha representación de axiomas lógicos, concepto de vacunas contra la enfermedad causada por el SARS-Cov-2, enfermedades incluyendo al COVID-19, y morbilidades asociadas que pueden hacer más vulnerable y complicar el estado de salud del paciente, medicamentos para dichas enfermedades, síntomas provocados por estas y las pruebas de laboratorios para destacar o confirmar alguna enfermedad específicamente la enfermedad del COVID-19.
- b) *Coherencia*: Los conceptos establecidos en el modelo permiten definir formalmente el dominio para la representación de perfiles de pacientes en base a los hechos establecidos de los datos de pacientes con la enfermedad, de modo que una oración establecida en dicho modelo no permita contradecir una definición establecida, tal como que un paciente tenga relación directa con el tratamiento sin ser diagnosticado antes con alguna enfermedad.
- c) *Extensibilidad*: Dicho modelo ontológico permite la anticipación de la integración de nuevos términos o conceptos de nuevas ontologías que formen parte de este dominio con la intención de expandir dicho modelo para abarcar y establecer nuevos axiomas para la representación de otros módulos basados en el vocabulario existente en dicho modelo, sin requerir la revisión de todo el modelo para entender su estructura y nivel de formalidad.
- d) *Sesgo de codificación mínimo*: El nivel de la conceptualización se especificó de acuerdo con el nivel de conocimiento sin depender de una codificación en particular de un nivel de símbolo o de notación, los sesgos de codificación se minimizan debido a que los agentes puedan intercambiar el conocimiento al implementar dicha conceptualización en diferentes sistemas o diferentes estilos de representación.
- e) *Compromiso ontológico mínimo*: Dicho modelo requiere un mínimo de compromiso ontológico con la finalidad de soportar tareas de intercambio de conocimiento, así como realizar la menor cantidad posible de afirmaciones sobre el ámbito que se está modelando, permitiendo a las partes comprometidas con la ontología, la libertad de especializarse e instanciar la ontología según sea necesario.

Del conjunto de principios de diseño se debe realizar la evaluación del modelo uno por uno, ejecutando herramientas de diagnóstico necesarias y presentando los resultados de tal forma que se pueda verificar el cumplimiento satisfactorio como lo es la coherencia, para ello se utilizó la herramienta de razonador en Protégé, específicamente el razonador Pellet²⁶, tal como se muestra en la Figura 5.6, así también se usó una

²⁶<https://www.w3.org/2001/sw/wiki/Pellet>

5. EVALUACIÓN DEL MODELO ONTOLÓGICO

herramienta Online para verificar errores de consistencia, denominado OOPS! (Poveda-Villal'ón et al., 2014).

```

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 5.6: Verificación de consistencia con razonador.

Por otra parte la evaluación de diseño del modelo ontológico debe abarcar el punto de elicitación de un listado de terminos médicos de este modo al ser un modelo orientado al dominio para la representación de perfiles de pacientes con COVID-19, con la finalidad de que las personas que hagan uso del modelo sean capaces de identificar el dominio al que está orientado y sobre todo la identificación de modularización y la formalización en la representación de ontologías de manera individual, tal como se visualiza en la Figura 5.7.

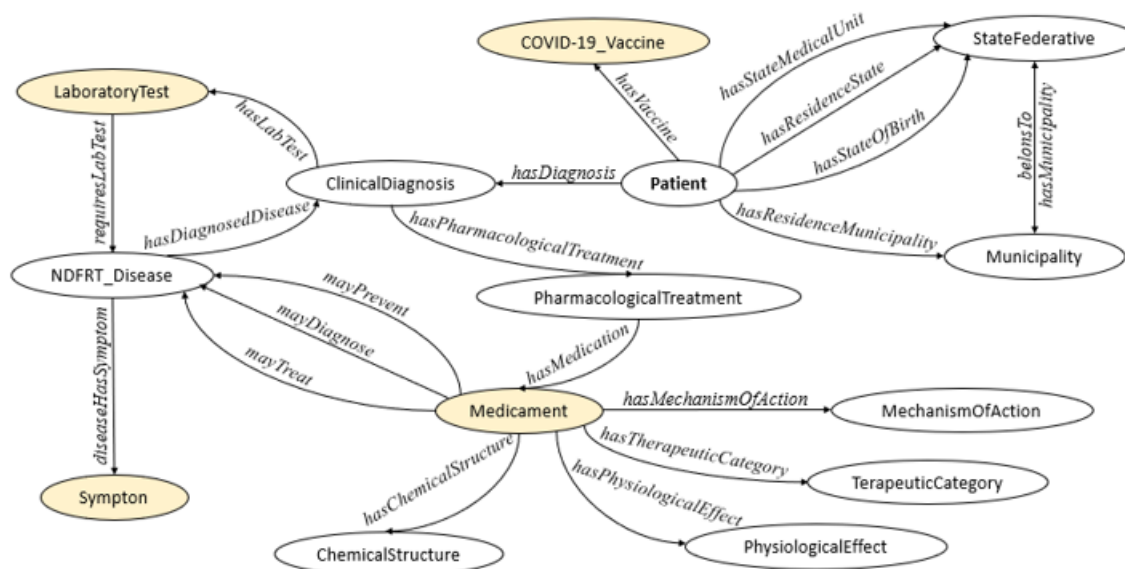



Figura 5.7: Diseño del modelo ontológico integrado.

Dicho modelo integrado y poblado cuenta con las siguientes métricas, un total de 36594 axiomas, 31687 axiomas lógicos, 17 clases, 21 propiedad de objetos, 40 propiedades de datos y 4768 individuos, tal como se muestran las métricas en la Figura 5.8.



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

Figura 5.8: Métricas del modelo ontológico integrado.

5.3. Evaluación basada en preguntas de competencia

Siguiendo con el proceso de evaluación, de acuerdo con la recopilación de información y la revisión de trabajos relacionados de modelos ontológicos para la representación de perfiles de pacientes con alguna comorbilidad, se plantearon las siguientes preguntas de competencias de acuerdo con las necesidades que se presentan y analizan en las instancias de datos del modelo integrado, con la finalidad de obtener conocimiento, como por ejemplo: conocer acerca de los principales síntomas de la enfermedad del COVID-19, el rango de edad de las personas que presentan dicha enfermedad, así como conocer las características de las personas que han sido hospitalizadas, están intubados, si el paciente es fumador activo y debido a ello presenta condiciones graves de la enfermedad, tal como se expresan en el siguiente listados de preguntas.

- a) *What are the main symptoms of patients with COVID-19?*
- b) *What is the average age of intubated patients?*
- c) *What is the gender and age of the intubated patients?*
- d) *What is the gender and age of the patients who die?*
- e) *How much time is there between admission to the hospital and the date of death?*
- f) *Are there pregnant patients who were diagnosed with COVID-19?*
- g) *Are there patients who smoke and who were diagnosed with COVID-19?*

5. EVALUACIÓN DEL MODELO ONTOLÓGICO

- h) *Is there any correlation between patients being intubated if they smoke?*
- i) *What are the characteristics of hospitalized patients who worsen and are intubated?*
- j) *What are the characteristics of hospitalized patients who die?*

Teniendo el modelo ontológico poblado (con instancias de datos), se procede a consultar las preguntas de competencia planteadas, para obtener aquellos datos que permitan conocer acerca del comportamiento de la enfermedad en las personas, para ello se implementa un lenguaje de consulta que permite trabajar con las instancias de dicho modelo ontológico, de acuerdo a (O'Connor and Das, 2009), el trabajar con el lenguaje de consulta Semantic Query-enhanced Web Rule Language (por la siglas en inglés SQWRL) en ontologías en formato RDF o OWL es un lenguaje formal para trabajar con dichos estándares de los conceptos del modelo, por ello a continuación se presenta ejemplos de la definición de la consulta y los resultados en formato de tabla de los registros encontrados en el modelo ontológico.

Pregunta 1: What are the main symptoms of patients with COVID-19?, a continuación, se muestra la consulta 1 en lenguaje SQWRL, y posteriormente se muestra en la Figura 5.9 el tiempo de ejecución y la tabla de resultado de dicha consulta, de los principales síntomas en pacientes con COVID-19.

```
medicament:NDFRT_Disease(medicament:N0000003799) ^
medicament:hasNDFRT_DiseaseName(medicament:N0000003799, ?disease) ^
patientelectronicrecord:diseaseHasSymptom(medicament:N0000003799,
?idSymptom) ^ symptom:hasSymptomName(?idSymptom, ?symptomName) ->
sqwrl:select(?disease, ?idSymptom, ?symptomName)
```

See the S1 tab to review results of the SQWRL query.
The query took 5300 milliseconds. 13 rows were returned.

disease	idSymptom	symptomName
Coronavirus Infections [Disease/Finding]	symptom:SYMP_0000614	Cough
Coronavirus Infections [Disease/Finding]	symptom:SYMP_0019145	Vomiting
Coronavirus Infections [Disease/Finding]	symptom:SYMP_0019153	Dyspnea
Coronavirus Infections [Disease/Finding]	symptom:SYMP_0019174	Chills
Coronavirus Infections [Disease/Finding]	symptom:SYMP_0019177	Fatigue
Coronavirus Infections [Disease/Finding]	symptom:SYMP_0000613	Fever
Coronavirus Infections [Disease/Finding]	symptom:SYMP_0000504	Headache
Coronavirus Infections [Disease/Finding]	symptom:SYMP_0000505	Throat_Pain
Coronavirus Infections [Disease/Finding]	symptom:SYMP_0000570	Diarrhea
Coronavirus Infections [Disease/Finding]	symptom:SYMP_0000458	Nausea
Coronavirus Infections [Disease/Finding]	symptom:SYMP_0000237	Congestion
Coronavirus Infections [Disease/Finding]	symptom:SYMP_0000230	Body_Ache
Coronavirus Infections [Disease/Finding]	symptom:SYMP_0000099	Pain

Figura 5.9: Tiempo de ejecución y tabla de resultado de 1ra consulta.

Pregunta 3: What is the gender and age of the intubated patients?, a continuación, se muestra la consulta 3 en lenguaje SQWRL, y posteriormente se muestra

en la Figura 5.10 el tiempo de ejecución y la tabla de resultado de dicha consulta, sobre el género y edad de los pacientes con dicha enfermedad que se encuentran intubados.

```

patischelectronicrecord:hasGender(?patient, ?gender) ^
patischelectronicrecord:isIntubated(?patient, "YES") ^
patischelectronicrecord:hasAge(?patient, ?age) ^
patischelectronicrecord: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
patischelectronicrecord.pat1135b2	YES	FEMALE	91
patischelectronicrecord.pat49a69	YES	FEMALE	65
patischelectronicrecord.pat477ab	YES	MALE	14
patischelectronicrecord.pat4e838	YES	MALE	60
patischelectronicrecord.pat12d70a	YES	MALE	58

Figura 5.10: Tiempo de ejecución y tabla de resultado de 3ra consulta.

Pregunta 10: What are the characteristics of hospitalized patients who die?, a continuación, se muestra la consulta 10 en lenguaje SQWRL, y posteriormente se muestra en la Figura 5.11 el tiempo de ejecución y la tabla de resultado de dicha consulta, de las características de los pacientes que fallecieron debido a la enfermedad.

```

patischelectronicrecord:Patient(?patient) ^
patischelectronicrecord:hasDateOfDeath(?patient, ?dateOfDeath) ^
patischelectronicrecord:hasGender(?patient, ?gender) ^
patischelectronicrecord:hasAge(?patient, ?age) ^
patischelectronicrecord:hasTypeOfPatient(?patient, ?tipoOfPatient)
^ patischelectronicrecord:hasSmoking(?patient, ?smoking) ^
patischelectronicrecord:isIntubated(?patient, ?intubated) ->
sqwrl:select(?patient, ?dateOfDeath, ?gender, ?age,
?tipoOfPatient, ?smoking, ?intubated)

```

5. EVALUACIÓN DEL MODELO ONTOLÓGICO

See the S10 tab to review results of the SQWRL query.
The query took 5365 milliseconds. 18 rows were returned.

patient	dateOfDeath	gender	age	tipoOfPatient	smoking	intubated
patientelectronicrecord.pat:49a69	07/21/20	FEMALE	66	HOSPITALIZED	NO	YES
patientelectronicrecord.pat12914d	12/15/20	FEMALE	30	HOSPITALIZED	NO	NO
patientelectronicrecord.pat18cc57	01/19/20	MALE	57	HOSPITALIZED	NO	NOT SPECIFIED
patientelectronicrecord.pat19fac0	01/27/20	MALE	44	HOSPITALIZED	NO	NOT SPECIFIED
patientelectronicrecord.pat:4e838	05/11/20	MALE	60	HOSPITALIZED	YES	YES
patientelectronicrecord.pat158f2e	01/13/20	FEMALE	77	HOSPITALIZED	NO	NOT SPECIFIED
patientelectronicrecord.pat14c8f5	01/14/20	MALE	69	HOSPITALIZED	YES	NOT SPECIFIED
patientelectronicrecord.pat0a5909	01/13/20	FEMALE	47	HOSPITALIZED	NO	NOT SPECIFIED
patientelectronicrecord.pats62292	01/12/20	MALE	84	HOSPITALIZED	NO	NOT SPECIFIED
patientelectronicrecord.pat10f594	01/10/20	MALE	74	HOSPITALIZED	NO	NOT SPECIFIED
patientelectronicrecord.pat0759ce	01/18/20	MALE	30	HOSPITALIZED	NO	NO
patientelectronicrecord.pat17603a	06/30/20	MALE	56	HOSPITALIZED	YES	NO
patientelectronicrecord.pat05c673	01/09/20	MALE	6	HOSPITALIZED	NO	NOT SPECIFIED
patientelectronicrecord.pat0a7fa2	06/28/20	MALE	53	HOSPITALIZED	YES	NO

Figura 5.11: Tiempo de ejecución y tabla de resultado de 10ma consulta.

Capítulo 6

Conclusiones y trabajos a futuro

En este trabajo se presentó el desarrollo de un modelo integrado para la representación de perfiles de pacientes con casos positivos COVID-19, abordando las vacunas, pruebas de laboratorio, medicamentos, enfermedades, síntomas entre otros conceptos médicos, de acuerdo con la metodología empleada para este proyecto, se abordó desde la etapa de recopilación y a análisis de la información acerca de la enfermedad del COVID-19, sobre la representación de perfiles de pacientes de acuerdo con las normas y manuales establecidos, de igual forma se consultó aquellos trabajos relacionados con el desarrollo de este proyecto de investigación. En la etapa de búsqueda de ontologías se consultó un portal de ontologías existentes del dominio médico que permitiera la representación de perfiles de pacientes, con la finalidad de reutilizar las ontologías parcialmente. El diseño de dicho modelo que permite la representación de perfiles de pacientes se basó en principios de diseños establecidos y estandarizados en proyecto de investigaciones referentes a este proyecto. Posteriormente, con ayuda de un editor de ontologías se construyó cada una de las ontologías (Symptom, Vaccine, Medicament, LaboratoryTest y PatientElectronicRecord) que integran el modelo final. Aplicando métodos y programas para la exportación de módulos parciales. Finalmente se implementó la definición de axiomas y definición de reglas de inferencia en el modelo integrado.

Dicho modelo ontológico fue puesto a prueba, mediante los tipos de evaluaciones definidas para dicho modelo, donde una de ellas fue la población del modelo ontológico con instancias de datos de referencia de acuerdo con el diseño y la definición de conceptos médicos para la representación de perfiles de pacientes con COVID-19. De igual forma, se procedió a evaluar dicho modelo, aplicando consultas de preguntas de competencia (en lenguaje natural) a dichas instancias para responder (en lenguaje de consultas) de forma clara y precisa con conocimiento nuevo de las instancias de datos. Finalmente, se procedió a evaluar dicho modelo mediante los principios de diseño establecidos por autores que han desarrollado estrategias de evaluación a lo largo de los años.

De acuerdo con el conjunto de datos procesados en el modelo ontológico integrado

6. CONCLUSIONES Y TRABAJOS A FUTURO

de pacientes mexicanos, del conjunto de datos considerados y la evaluación mediante consultas a dicho modelo, los resultados obtenidos muestran el número de casos positivos en pacientes mexicanos. El 53.7% de los casos son de pacientes del sexo femenino, y el 1.4% de las pacientes del sexo femenino se encuentran embarazadas, del total de datos de pacientes mexicanos, el 0.4% de los pacientes se encuentran en unidad de cuidados intensivos y a su vez están intubados, el 3.6% de los pacientes con casos positivos han fallecido de acuerdo a la actualización de la fecha el cual se analizó el conjunto de datos de pacientes mexicanos, la edad promedio de los pacientes que fallecieron por COVID-19 es de 36 años, a su vez 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 a una unidad de salud monitorea enfermedades respiratorias hasta el fallecimiento del paciente en la misma, oscilando entre 2 y 18 días.

Con el desarrollo de este proyecto de investigación contribuye: con un modelo ontológico integrado que permite la representación de perfiles de pacientes con COVID-19, como investigación y modelo de referencia para la comunidad de ciencias de la computación que deseen el reutilizar o basarse en la aplicación del método para la integración de ontologías. De igual forma, se aporta una base de conocimiento con datos de pacientes con dicha enfermedad.

Por otra parte, como trabajos de investigación a futuro se puede partir del modelo integrado con conceptos médicos relacionados, mediante la implementación de un sistema de información en base a servicios para la administración de dicho modelo. También una posible implementación de herramientas para explotar dichos datos, como lo puede ser para el Procesamiento del Lenguaje Natural (PLN), con la finalidad de obtener un análisis o descubrimiento y extracción de relaciones semánticas, reconocimiento de entidades nombradas, etc., en dicho modelo ontológico integrado. Así también la integración de nuevos conceptos médicos a este modelo o la orientación del modelo referente a otra enfermedad de la cual se requiere un perfil del paciente más adecuado.

Bibliografía

- Ajami, H. and Mcheick, H. (2018). Ontology-based model to support ubiquitous health-care systems for copd patients. *Electronics*, 7(12):371. 15
- Aliseda, A. (1997). Seeking explanations: Abduction in logic. *Philosophy of Science and Artificial Intelligence. Ph. D. Thesis, Institute for Logic, Language and Computation, Amsterdam, The Netherlands*. 11
- Bravo, M., Gonz'lez, D., Ortiz, J. A. R., and S'anchez, L. (2020). Management of diabetic patient profiles using ontologies. *Contadur'ia y administraci'ón*, 65(5):12. 17
- Bravo, M., Hoyos Reyes, L. F., and Reyes Ortiz, J. A. (2019). Methodology for ontology design and construction. *Contadur'ia y administraci'ón*, 64(4). 52
- Celi, L. A., Galvin, S., Davidzon, G., Lee, J., Scott, D., and Mark, R. (2012). A database-driven decision support system: customized mortality prediction. *Journal of personalized medicine*, 2(4):138–148. 14
- Cruz-Fuentes, C. S. (2011). Instituto nacional de psiquiatría “ram'ón de la fuente muñiz”. *Revista de Investigaci'ón Cl'nica*, 63(3):219–222. 10
- Cuenca Grau, B. (2007). A logical framework for modularity of ontologies. 37
- de Salud, S. (2022). Dirección general de promoción de la salud. <https://www.gob.mx/salud/documentos/direccion-general-de-promocion-de-la-salud>. Accessed: 2022-10-21. 2
- Duque-Ramos, A., Boeker, M., Jansen, L., Schulz, S., Iniesta, M., and Fern'andez-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. 52

- Duran-Muñoz, I. and Bautista-Zambrana, M. R. (2013). Applying ontologies to terminology: Advantages and disadvantages. *Hermes-Journal of Language and Communication in Business*, (51):65–77. [9](#)
- Farion, K., Michalowski, W., Wilk, S., O’Sullivan, D., Rubin, S., and Weiss, D. (2009). Clinical decision support system for point of care use. *Methods of information in medicine*, 48(04):381–390. [13](#)
- G’omez-P’erez, A., Fernández, M., and Vicente, A. d. (1996). Towards a method to conceptualize domain ontologies. [52](#)
- Govindan, K., Mina, H., and 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. [16](#)
- 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. [9](#), [52](#)
- Guarino, N. (1995). Formal ontology, conceptual analysis and knowledge representation. *International journal of human-computer studies*, 43(5-6):625–640. [10](#)
- Harry, M. L., Saman, D. M., Truitt, A. R., Allen, C. I., Walton, K. M., O’Connor, P. J., Ekstrom, H. L., Sperl-Hillen, J. M., Bianco, J. A., and Elliott, T. E. (2020). Pre-implementation adaptation of primary care cancer prevention clinical decision support in a predominantly rural healthcare system. *BMC Medical Informatics and Decision Making*, 20(1):1–15. [16](#)
- Hayward, R. (2004). Clinical decision support tools: Do they support clinicians? *Canadian Medical Association. Journal*, 170(10):FP66. [11](#)
- Hermesen, E. D., VanSchooneveld, T. C., Sayles, H., and Rupp, M. E. (2012). Implementation of a clinical decision support system for antimicrobial stewardship. *Infection Control & Hospital Epidemiology*, 33(4):412–415. [14](#)
- Köhler, S., Schulz, M. H., Krawitz, P., Bauer, S., Dölken, S., Ott, C. E., Mundlos, C., Horn, D., Mundlos, S., and 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. [14](#)
- McGuinness, D. L., Van Harmelen, F., et al. (2004). Owl web ontology language overview. *W3C recommendation*, 10(10):2004. [46](#)
- Neta, R. (2013). What is an inference? *Philosophical issues*, 23:388–407. [12](#)
- O’Connor, M. J. and Das, A. K. (2009). Sqwrl: a query language for owl. In *OWLED*, volume 529. [56](#)

- Oyelade, O. N. and Ezugwu, A. E. (2020). A case-based reasoning framework for early detection and diagnosis of novel coronavirus. *Informatics in Medicine Unlocked*, 20:100395. 15
- Poveda-Villal’ón, M., G’ómez-P’erez, A., and Su’ rez-Figueroa, M. C. (2014). Oops!(ontology pitfall scanner!): An on-line tool for ontology evaluation. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 10(2):7–34. 54
- Schlicht, A. and Stuckenschmidt, H. (2006). Towards structural criteria for ontology modularization. In *WoMO*. Citeseer. 37
- Sherimon, P. and Krishnan, R. (2016). Ontodiabetic: an ontology-based clinical decision support system for diabetic patients. *Arabian Journal for Science and Engineering*, 41(3):1145–1160. 11
- Staab, S. and Studer, R. (2010). *Handbook on ontologies*. Springer Science & Business Media. 9
- Whetzel, P. L., Noy, N. F., Shah, N. H., Alexander, P. R., Nyulas, C., Tudorache, T., and Musen, M. A. (2011). Bioportal: enhanced functionality via new web services from the national center for biomedical ontology to access and use ontologies in software applications. *Nucleic acids research*, 39(suppl.2):W541–W545. 21
- WHO (2022). Coronavirus disease (COVID-19). <https://www.who.int/health-topics/coronavirus>. Accessed: 2022-10-21. 1
- Zhang, Y.-f., Gou, L., Zhou, T.-s., Lin, D.-n., Zheng, J., Li, Y., and 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. 15

Apéndice A

Publicaciones

A.1. Artículos científicos

En este apartado se anexan las publicaciones derivadas en base a este proyecto de investigación, las cuales se anexan en orden de mención.

Los artículos: *"Management of diabetic patient profiles using ontologies"* en *Contaduría y Administración*, ISSN 0186-1042, Vol. 65, N^o. Extra 5, 2020 (Ejemplar dedicado a: Especial COVID-19).

"Modularization Method to Reuse Medical Knowledge Graphs" en *Pre-prints*, MATHEMATICS & COMPUTER SCIENCE, Information Technology & Data Management.

Por otro lado, el artículo: *Representación y gestión de datos de pacientes mexicanos con COVID-19 mediante Grafos de Conocimiento*, se encuentra proceso de revisión en la revista *Contaduría y Administración* hasta la fecha de impresión de esta Tesis.



Management of diabetic patient profiles using ontologies

Administración de perfiles de pacientes diabéticos usando ontologías

Maricela Bravo*, Darinel González, José A. Reyes-Ortiz,
Leonardo Sánchez

Systems Department, Autonomous Metropolitan University, Mexico

*Received September 14, 2020; accepted November 26, 2020
Available online November 27, 2020*

Abstract

In 2019 the International Diabetes Federation estimated that 12.8 million Mexicans had diabetes. The diabetes epidemic ranked second in causes of death in Mexico, a situation that was severely complicated during the second quarter of 2020 with the COVID-19 pandemic. Studies carried out by the Ministry of Public Health showed that the comorbidity of diabetes with COVID-19 has become a risk factor for serious complications, increasing the mortality rate. For this reason, it is necessary to develop personalized information management systems to support medical decision-making considering the specific characteristics of patients in Mexico. Information management of the diabetic patient profile begins with the investigation and registration of the relevant clinical data, data used by the physician to make the diagnosis and determine a personalized treatment. This article reports the development and integration of an ontology model for the management of diabetic patient profiles, incorporating medical ontologies. The results of the evaluation show the feasibility of using this integrated ontology for the management of diabetic patient presenting comorbidities. Likewise, a consistent ontological model is achieved, which complies with extensibility and reusability quality characteristics.

JEL code: I18, D83, C88, C63, C55

Keywords: Medical information management; Ontology modeling; Diabetic patient profile; Decision making for healthcare

* 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.2020.3050>

0186- 1042/©2019 Universidad Nacional Autónoma de México, Facultad de Contaduría y Administración. This is an open access article under the CC BY-NC-SA (<https://creativecommons.org/licenses/by-nc-sa/4.0/>)

Resumen

En 2019, la Federación Internacional de Diabetes estimó que 12.8 millones de mexicanos tenían diabetes. La epidemia de diabetes ocupó el segundo lugar en causas de muerte en México, una situación que se complicó gravemente durante el segundo trimestre de 2020 con la pandemia de COVID-19. Estudios realizados por la Secretaría de Salud Pública mostraron que la comorbilidad de la diabetes con COVID-19 se ha convertido en un factor de riesgo de graves complicaciones, aumentando la tasa de mortalidad. Por ello, es necesario desarrollar sistemas personalizados de administración de la información que apoyen la toma de decisiones médicas considerando las características específicas de los pacientes en México. La administración de la información del perfil del paciente diabético comienza con la investigación y registro de los datos clínicos relevantes, datos que utiliza el médico para realizar el diagnóstico y determinar un tratamiento personalizado. Este artículo reporta el desarrollo e integración de un modelo ontológico para la administración de información de perfiles de pacientes diabéticos, incorporando ontologías médicas. Los resultados de la evaluación muestran la viabilidad de utilizar esta ontología integrada para el manejo del paciente diabético con comorbilidades. Asimismo, se logra un modelo ontológico consistente, que cumple con las características de calidad de extensibilidad y reutilización.

Código JEL: I18, D83, C88, C63, C55

Palabras clave: Administración de Información Médica; Modelado de ontologías; Perfil de paciente diabético; Toma de decisiones para el cuidado de la salud

Introduction

The first case of COVID-19 pandemic was reported in Wuhan China on December 19th, 2019. COVID-19 is a disease caused by the SARS-CoV-2 virus. During the first weeks of the pandemic, *Diabetes Mellitus* rapidly emerged as the main comorbidity of patients who developed serious complications or death. According with the World Health Organization (WHO), diabetes is a chronic disease that appears when the pancreas does not produce enough insulin or when the body does not effectively use the insulin it produces. The Standards of Medical Care in Diabetes (2019) state that the effect of uncontrolled diabetes is hyperglycemia. Type 2 diabetes has its origin in the body's inability to effectively use insulin, which is often the result of excess weight or physical inactivity. Regarding the data about diabetes in Mexico, in 2019 the International Diabetes Federation¹ estimated that 12.8 million Mexicans had diabetes (IDF, 2019). Regarding causes of death, data reported by INEGI in 2018 showed that diabetes mellitus represents the second cause of mortality in Mexico with 106,525 cases. Similarly, the National Institute of Public Health (INSP) notes that more than 74,000 people die every year due to diabetes and its complications in Mexico. Another important risk factor is *obesity*. INSP indicates that overweight and obesity are health problems precursors of diabetes; in Mexico, 68% of the population over 20 years of age is overweight and obese. According to (Denova-Gutiérrez, 2020), 17.4% of Mexicans diagnosed with COVID-19 had obesity, 14.5% diabetes, 18.9% hypertension and 2.8% had cardiovascular disease.

This research project focuses on Type 2 Diabetes Mellitus (T2DM), which has several therapeutic options. The purpose of this project is to develop an ontological model for the management of patient profiles with T2DM. This ontological model will serve as a broad knowledge base of the clinical information related with the diagnosis and medical treatment of diabetes mellitus and other diseases. The most relevant information considered for the diagnosis and personalized medical treatment is the patient data, physical exercise, body mass index, basal metabolism rate, symptoms, laboratory tests, medication and diseases. The risks faced in the treatment of diabetes derived from the selection of medications are: hyperglycemia, hypoglycemia, kidney disease, retinopathy, gastroparesis, and sexual dysfunction.

With regard to treatment, not all patients with T2DM will respond equally to the same treatment, the reason could be that the doctor does not consider all the specific characteristics of the patient. The treatment that works best for most patients is selected first, even if this treatment is not effective for "some" patients. According to (Lasierra, 2013) there are no identical patients, each requires their own treatment according to their chronic conditions. In fact, it is very

¹ <https://idf.org/>

common that patients who have a chronic disease suffer from another chronic condition (comorbidity). Personalized medical treatment is a relevant and emerging approach that involves the design of strategies to prevent, detect, treat and monitor patients individually according to their complete medical profile (Inzucchi, 2012) (Hempo, 2015) (El-Sappagh, 2018).

The use of ontologies for modelling profiles of patients with T2DM will allow the detailed characterization of all the factors involved in diagnosis and management, achieving a high degree of semantic expressiveness by using description logics, and the incorporation of inference rules based on medical knowledge. A revision of existing ontologies and related sources of information was conducted to define the most relevant patient profile characteristics, attributes and semantic relations that need to be included in the ontological model. Additionally, data required for diagnosis and treatment will be included: patient height, weight, age, BMI, Basal Metabolic Rate, daily caloric intake, physical exercise, symptoms, laboratory tests, medications, and disease. In order to evaluate the usefulness and applicability of the ontology model two study cases are presented and the quality design criteria are explained.

The main contribution of this research is the diabetic patient profile ontology, which incorporates knowledge bases from valid medical resources and quality design criteria. The semantic approach derived from ontologies can be effectively exploited in personalized management services according to the patient, since it confers an inference capacity on the information, making use of logic-based languages. The rest of the paper is organized as follows: Section 2 presents a revision of related ontologies with diabetes mellitus, Section 3 describes the ontology development methodology, Section 4 presents the global ontology integration, Section 5 describes the evaluation, and finally in Section 6 conclusions are presented.

Revision of Related Ontologies

The use of ontologies to support diagnosis and treatment of diabetes mellitus is not a new issue, many researchers have investigated the benefits of logic-based reasoning approaches to automate diagnosis and support Clinical Decision Support Systems (CDSS). In this section we present a revision of ontologies that address the representation of health care support for diabetes mellitus.

In (Paganelli, 2007) authors described an ontology-based context model, and a context management middleware to monitor and assist patients at home. The model consists of four ontologies: Patient Personal Domain Ontology, Home Domain Ontology, Alarm Domain Ontology, and Social Context Ontology. Additionally, authors presented a Web application for manual input of biomedical parameters: heart rate frequency, pulse oximetry, systolic and diastolic blood pressure, body temperature, and glycemia.

In (Buranarach, 2011) presented an ontology for Type II diabetes and a clinical support system. The clinical support system consists of an information system and a reminding system. The ontology concepts included are: patient card, person data, family history, signs, symptoms, and health status (diabetes and diabetes complications).

In (Chen, 2012) authors presented a recommendation system for diabetes medication; this recommendation system was built using WSRL rules and Jess Rule engine. The kind of recommendations are related with specific medication treatment depending of various input data: HbA1c levels, safety, tolerance, among others. They also included information about patient tests. However, this recommendation system lacks of other important information for drug recommendation, for instance, disease history, familiar disease history, physical activity, and diet.

(Lasierra, 2013) authors described an ontology-based solution to provide personalized care to chronic patients at home. They implemented a method in three stages: ontology design and implementation, ontology application study, and software prototype implementation. Of particular interest is the incorporation of physician's knowledge and clinical guidelines to represent patient profiles, which is a similar approach to the work reported in this paper. The ontology

incorporated biological measurements: weight, height, glucose level, peripheral capillary oxygen saturation (SpO₂), amount of air forced from lungs per second (FEV₁), body water, body fat, blood pressure, and pulse rate.

In (Rahimi, 2014) authors presented a validation of the Diabetes Mellitus Ontology (DMO) using real world Electronic Health Records (EHR) data. They evaluated the accuracy of the ontology to execute inferences regarding the detection of the DM2. The sources of information utilized for the experimentation were taken from a literature review, the Australian National Guidelines for Type 2 Diabetes Mellitus, and 908 real world Electronic Health Records (HER).

(Hempo, 2015) authors presented a personalized care recommendation approach for diabetic patients, using an ontology model that consists of three ontologies: one that describes the particularities of patient profiles, a second ontology for the description of complications of diabetes, and a third ontology that describes self-care practices focusing on food and exercise recommendations.

In (El-Sappagh, 2016) authors described the Diabetes mellitus Diagnosis Ontology (DDO), which was constructed following the principles of the Open Biomedical Ontologies (OBO) Foundry; authors also specialized the Basic Formal Ontology (BFO) upper level ontology; and reused the Ontology for General Medical Science (OGMS). The main classes included in DDO are: diabetic complication (disease), drug, laboratory test, physical examination, and diabetes symptom. This ontology model does not consider treatment plans or medicaments, and family history.

Sherimon and Krishnan (2016) describe OntoDiabetic, an ontology-based decision support system to assess the risk factors and generate treatment suggestions for diabetic patients. Authors define a set of inference rules to derive information about the health status of the patient. This ontology model lacks up-to-date information on drugs, diseases, and important relationships between them.

In (El-Sappagh, 2018) authors described the development of Diabetes Mellitus Treatment Ontology (DMTO) following the design principles of the OBO Foundry, reusing the OGMS ontology, and incorporating the DDO previously defined concepts. The novel inclusions in DMTO are: interactions between drugs, interactions between drugs and complications, interactions between drugs and food, and interactions between drugs and exercise. The main goal of DMTO is to support the automation of diabetes treatment process and provide an intelligent and distributed clinical decision support system to be integrated in an Electronic Health Record (EHR):

In (Chen, 2019) authors describe an Ontology-based Model for Diabetic Patients (OMDP). In OMDP authors specialized the concepts from the BFO and reused the OGMS ontology concepts. OMDP consists of the following main definitions: Gene, Laboratory test, Diagnosis, Treatment, Diet, Exercise, Drug, Complications, Physical examination, Symptoms, and Patient information.

The review of related ontologies is done with the main objective of reusing one or more models, either complete or partial. Therefore, the breadth of concepts they cover regarding the diagnosis and treatment of diabetes mellitus must be observed first. From this perspective, we consider that the ontologies DMTO and OMDP are the most complete because they include treatment plans. However, the reuse of a complete model must be carefully decided. DMTO and OMDP are complex models that extend and specialize from BFO and OGMS ontologies. Therefore, these ontologies import several models and handle many naming conventions that require the ontology designer to become familiar with these rules and regulatory principles of the Open Biomedical Ontology (OBO) Foundry.

The main difficulties found in reusing of DMTO and OMDP ontologies is that there are no reports of current applications that use them. This is an indicator of the following design problems: lack of intelligibility, very large and complex models, performance problems related to memory resources for reasoning and inference execution when dealing with hundreds of patient profiles. Literature reporting these ontologies mention nothing about performance measures regarding large volumes of patient data. Therefore, in this work we have decided to carry out a hybrid ontology design which consists of two phases: design and develop from scratch the patient profile model; and select, modularize and

integrate existing medical ontologies that are usable, clear and valid. The integration procedure of existing resources consists of extracting modules in the form of vocabularies and integrating them as lighter modules, thus avoiding performance issues.

Ontology Design Method

Taking as a reference the methodology reported in (Bravo, 2019) we designed the ontology model that is reported in this section. The process of designing an ontology begins by establishing the domain and competence of the ontology.

Ontology Requirements Specification

This ontology model aims at representing the knowledge base for DMT2 diagnosis, and treatment. DMT2 diagnosis and treatment depend mainly on the particular characteristics of the patient, therefore this ontology will consider the acquisition of the patient data. In order to specify the initial requirements of the ontology model, it is necessary to define a list of competency questions that the ontology model should be capable of answering.

Regarding the patient profile, the following competency questions should be answered:

- a) *What is the basic data that characterizes the diabetic patient profile?*
- b) *What is the basal metabolism of the patient?*
- c) *What is the body mass index?*
- d) *How many calories does the patient consume daily?*
- e) *What is the BMI classification of the patient?*
- f) *What type of physical exercise does the patient practice?*
- g) *How many calories does the patient burns per day?*

Regarding diabetes diagnosis, the following competency questions should be answered

- a) *What are the symptoms associated to T2DM?*
- b) *What are the laboratory tests required to confirm a T2DM diagnosis?*
- c) *What are the T2DM complications?*

Regarding diabetes *Clinical Diagnostic*, the following questions should be answered:

- a) *What are the drugs utilized in T2DM treatment?*
- b) *What are the main elements of a treatment plan?*
- c) *What type of exercise can a diabetic treatment include considering the patient characteristics?*
- d) *What type of diet, food and meals can a diabetic treatment include considering the patient preferences?*

There are other important questions that an ontology model of DMT2 should answer; however, in this article we concentrate on the representation of the diabetic patient profile, since this represents a crucial entry point for the entire process of diagnosis and subsequent treatment.

Ontology Design

The process of designing an ontology consists of making decisions about the concepts to be included into the model and the relationships between them. The type of relationships that are to be implemented are hierarchical relations (is-a) and semantic relations (object and data properties). Based on the competency questions specified above, and the revision of related ontologies, the following terms were identified as the core concepts involved in diagnosis and treatment of DMT2:

Patient Profile, Demographic Data, Physical Activity, BMI Classification, Disease, Drug (Medication), Lab Test, and Clinical Diagnosis among others. Figure 1 shows the main concepts included in the ontology model.

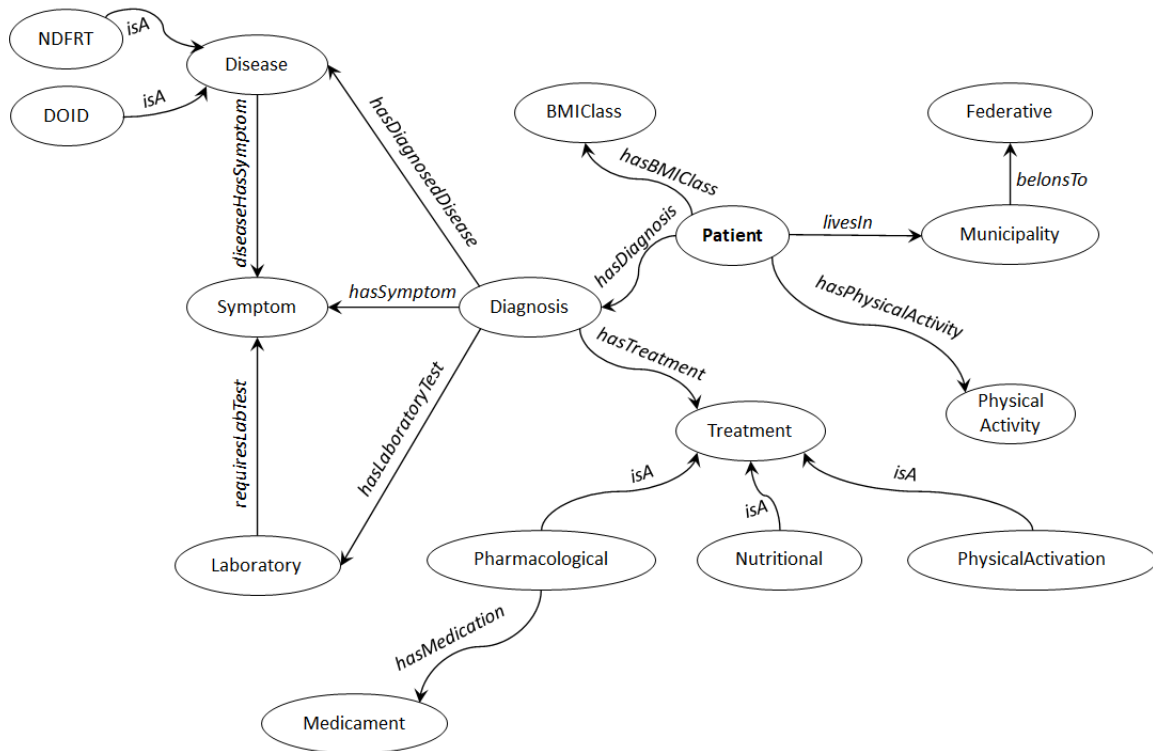


Figure 1. Ontology model to represent *Patient* profiles.
 Source: Author's own

Patient concept was designed considering the important data (without private personal data such as name, address or phone number) to integrate a basic clinical record of the patient. The personal data of the patients is not stored, only the data necessary for the establishment of the state of health are obtained with the express authorization of the participants, the identity of the patients is never known. However, it is necessary to have a way to identify the patient and his / her medical and biomedical data, therefore a username and password are requested as identifiers to access to the system. During the process of deciding which were the most important characteristics to represent a patient, the need to include the data that allowed calculating the metabolism data was observed. It is relevant for the development of this project to calculate the basal metabolic rate and the body mass index.

Patient concept consists of the following data type properties: *hasWeight*, *hasHeight*, *hasGender*, *hasBirthDate*, *hasNumberOfChildren*, *hasBloodType*, *bodyMassIndex*, *numberCigarettesPerDay*, *basalMetabolism*, where *basalMetabolism* and *bodyMassIndex* are two data properties which are calculated using inference rules. Figure 2 and Table 1 details the data type properties of the *Patient* class.

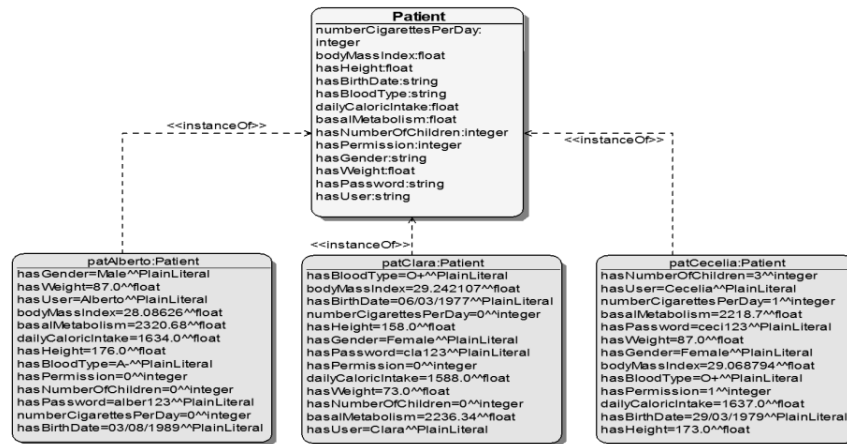


Figure 2. Ontology model to represent *Patient* profiles.
 Source: Author's own

Table 1
 Details of the data properties of the *Patient* concept

Data property	Domain	Range	Description
<i>hasUser</i>	Patient	xsd:string	Represents a user identification to access the system. This data attribute is generated when the patient registers for the first time.
<i>hasPassword</i>	Patient	xsd:string	This is an access password created by the user when registered for first time.
<i>hasWeight</i>	Patient	xsd:float	This data type is used to specify the weight in kilograms of the patient during registration.
<i>hasHeight</i>	Patient	xsd:float	This data type is used to specify the height in meters of the patient during registration.
<i>hasGender</i>	Patient	xsd:string	This data type is used to specify the gender of the patient.
<i>hasBirthDate</i>	Patient	xsd:string	This data type is used to specify the date of birth of the patient.
<i>hasBloodType</i>	Patient	xsd:string	This data type is used to specify the type of blood of the patient.
<i>hasNumberOfChildren</i>	Patient	xsd:integer	This data type is used to specify the number of children of the patient.
<i>numberCigaretesPerDay</i>	Patient	xsd:integer	This data type is used to specify the number of cigarettes smoked per day of the patient.
<i>dailyCaloricIntake</i>	Patient	xsd:float	This type of data is used to establish the number of calories that the patient ingests in total per day.
<i>bodyMassIndex</i>	Patient	xsd:float	This data type is calculated using inference rules to represent the weight of a person in kilograms divided by the square of height in meters.
<i>basalMetabolism</i>	Patient	xsd:float	This data type represents the rate of energy required for the body.

Source: Author's own

According with the Center for Disease Control and Prevention (CDC)² the Body Mass Index (BMI) represents the weight of a person in kilograms divided by the square of height in meters. Likewise, the CDC indicates that a high BMI can be an indicator of high body fatness. Therefore, BMI calculation can be used to filter out categories of body weight that may lead to possible health problems. In (Pelley, 2012) the Basal Metabolic Rate (BMR) is defined as the rate of energy expenditure of a person at rest; it eliminates the variable effect of physical activity.

The **BMIClassification** concept was defined based on the World Health Organization (WHO)³ definition, with the objective of identifying the patient BMI class of body weight. **BMIClassification** is defined with three data type attributes: *hasMinValue*, *hasMaxValue* and *hasNameBMI*. Table 2 shows the particular ranges of each individual. Figure 3 shows the semantic relationship that enables the correlation of each *Patient* individual with its weight classification.

Table 2
 Classification of the weight status of patients according to the BMI

hasNameBMI	hasMinValue	hasMaxValue	WHO classification
<i>UnderWeight</i>		18.49f	Under weight
<i>Normal</i>	18.5f	24.9f	Normal weight
<i>OverWeight1</i>	25.0f	26.9f	Over weight
<i>OverWeight2</i>	27.0f	29.9f	Over weight
<i>Obesity1</i>	30.0f	34.9f	Class I obesity
<i>Obesity2</i>	35.0f	39.9f	Any BMI >= 35 is severe obesity
<i>Obesity3</i>	40.0f	49.9f	Any BMI >= 40 is morbid obesity
<i>Obesity4</i>	50.0f		Any BMI >= 50 is super obese

Source: Author's own

During *Patient* registration, the *basalMetabolism* and *bodyMassIndex* values are calculated, also a semantic relation is defined using the *hasBMIClassification* object property to associate every patient with his weight class.

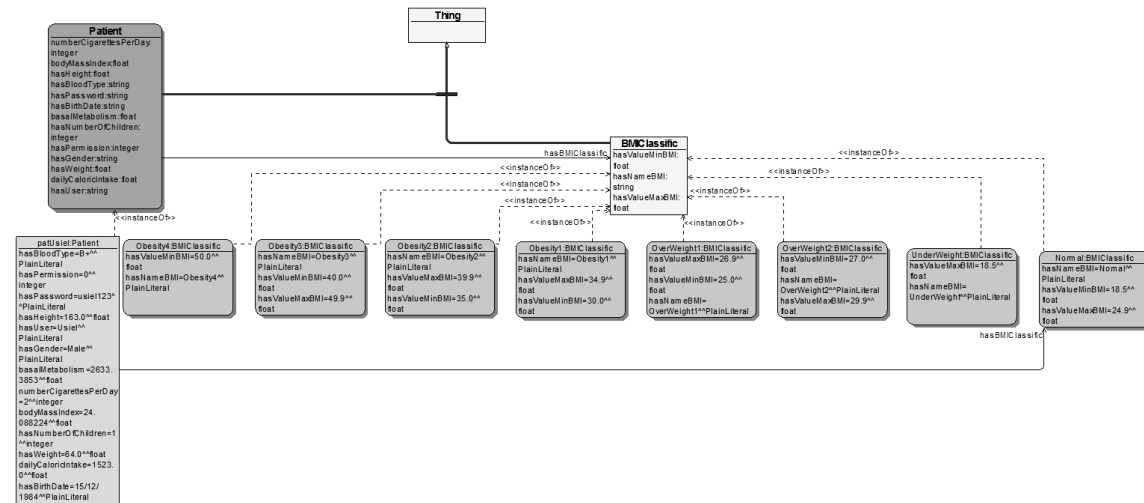


Figure 3. Semantic relationships between *Patient* and *BMIClassification*
 Source: Author's own

According with a position statement of the American Diabetes Association (Colberg et al., 2016) adoption and maintenance of physical activity is a relevant aspect for the prevention and treatment of diabetes mellitus. Therefore, it is important to obtain this information from the patient during registration. For the definition of the concept **PhysicalActivity** the following data were considered: the frequency and type of exercise that is performed. For the type of exercise we

² <https://www.cdc.gov/healthyweight/assessing/bmi/index.html>

³ https://www.who.int/health-topics/obesity#tab=tab_1

defined a specific list of physical activities that the patient can perform routinely, and the approximate number of calories per hour associated, so that it is possible to make an approximate calculation of the number of calories burned. Figure 4 shows the semantic relationship used to define the type of physical activity that the patient performs on a periodical basis. The physical activity also defines a numerical value to represent the number of calories that are burn per hour.

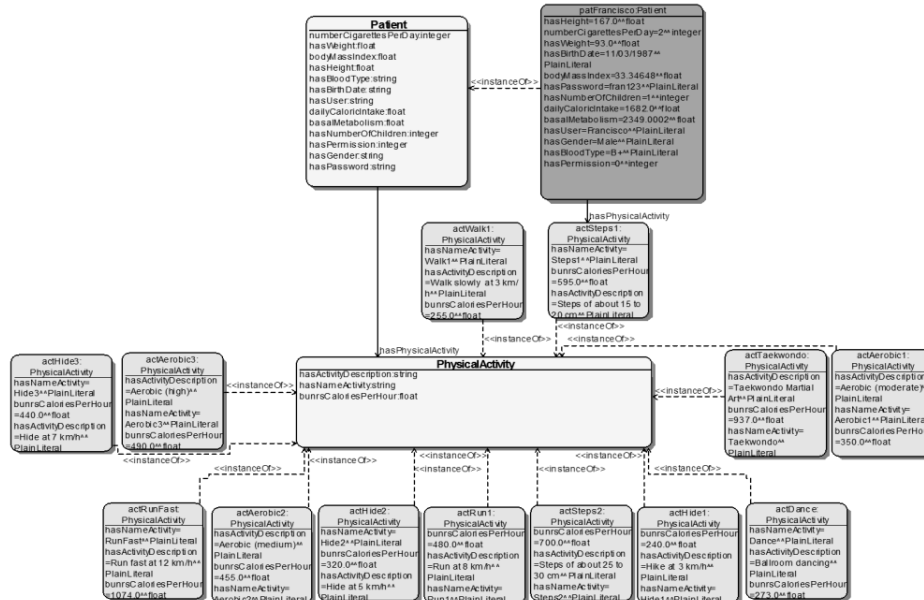


Figure 4. Semantic relationships between *Patient* and *PhysicalActivity*
 Source: Author's own

The purpose of this research project is to develop a comprehensive infrastructure to facilitate the integration and development of applications for remote monitoring of diabetic patients. In this sense, it is important to study other factors that affect DMT2. In (Ansari, 2019) authors presented a conceptual model of the aspects for the self-management of diabetes, which includes: socio-demographic characteristics, such as age, gender, education; behavioral and psychological characteristics: monitoring of blood sugar, adherence to diet, and physical activity; barriers to self-management: lack of knowledge, self-confidence, financial, and family support; and cultural characteristics: cultural beliefs, and dietary preferences. This article describes a prototype that does not cover exhaustively all the aforementioned aspects; however, the model is possible to be augmented easily. For now, the information regarding the state, province or municipality where the patient was born and lives are included. Figure 5 shows the semantic relationship between the Patient and the Municipality.

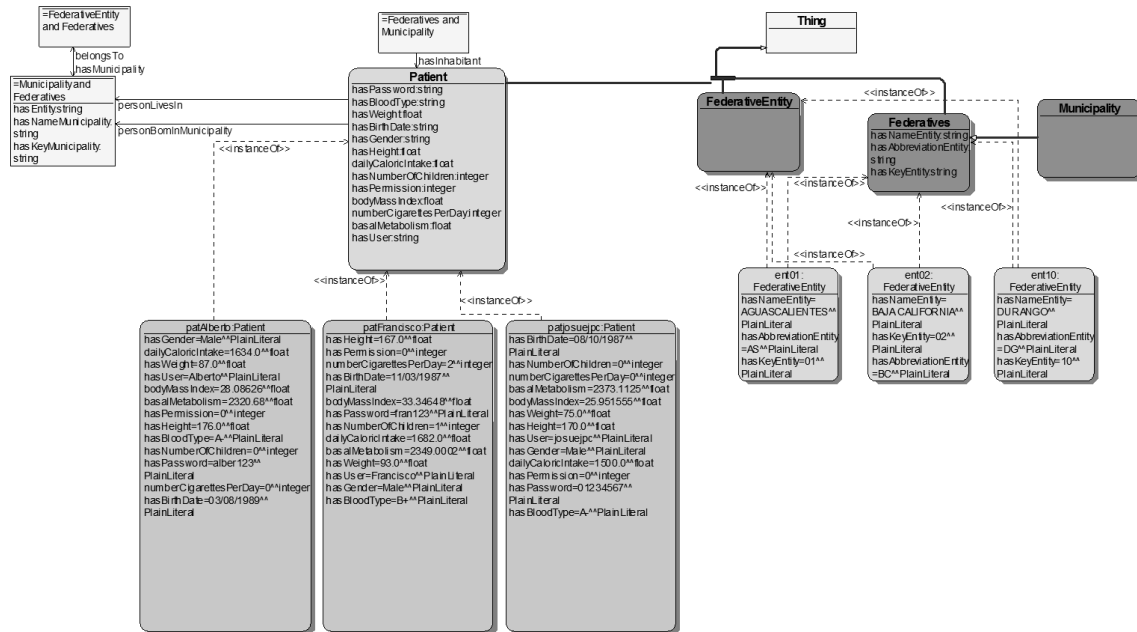


Figure 5. Semantic relationships between *Patient* and *Municipality*
 Source: Author's own

The semantic relationship of *personBornInMunicipality* is used to describe the specific municipality where the patient was born, also *personLivesIn* is used to indicate the current address of the patient. This information will be used to correlate the patient profile with his social context.

Integrate Medical Ontologies

Based on the list of required concepts related with the diagnosis and treatment of diabetes, a search was made for existing ontological resources to be reused in this model. Specifically, the following concepts were searched: disease, drug or medication, symptom, and laboratory tests. A good repository of biomedical ontologies is *BioPortal*⁴, which provides search mechanisms to find and explore among a large collection of biomedical ontologies. We executed the search of the afore-mentioned concepts, Table 3 shows the list of ontologies found for each concept.

⁴ <http://bioportal.bioontology.org/>

Table 3
 List of ontologies found using the *BioPortal*

Ontology	General Description	URL	Institutions involved
SNOMEDCT	Consist of a large collection of cross referred medical vocabularies aiming at supporting the clinical recording of patient data.	https://bioportal.bioontology.org/ontologies/SNOMEDCT	Published by the College of American Pathologists (CAP). Currently the intellectual property of SNOMED belongs to International Health Terminology Standards Development Organisation (IHTSDO).
DOID	Is a standardized ontology which provides descriptions of human disease terms.	https://bioportal.bioontology.org/ontologies/DOID	Developed at the Institute for Genome Sciences and the University of Maryland School of Medicine.
NDF-RT	Is the National Drug file reference terminology	https://bioportal.bioontology.org/ontologies/NDFRT	Produced by the U.S. Department of Veterans Affairs, Veterans Health Administration (VHA).
DRON	The Drug Ontology.	https://bioportal.bioontology.org/ontologies/DRON	Division of Biomedical Informatics, University of Arkansas for Medical Sciences.
SYMP	The Symptom ontology.	https://bioportal.bioontology.org/ontologies/SYMP	Institute for Genome Sciences (IGS) at the University of Maryland.
CSSO	An ontology for describing clinical signs and symptoms.	https://bioportal.bioontology.org/ontologies/CSSO	The National Bioscience Database Center (NBDC) is the organization for life sciences databases in Japan. To promote the creation of new knowledge, we conduct research and development and provide services that contribute to data sharing and integration.
LOINC	The international standard for identifying health measurements, observations, and documents.	https://bioportal.bioontology.org/ontologies/LOINC	The LOINC vocabulary is developed and maintained by the Regenstrief Institute. The LOINC project was initiated in 1994.
LABO	Is an ontology that describes laboratory tests prescriptions and reporting documents.	https://bioportal.bioontology.org/ontologies/LABO	Institut de Recherche en Informatique de Toulouse (IRIT), CNRS, France. Groupe de Recherche Interdisciplinaire en Informatique de la Santé (GRIIS), Université de Sherbrooke, Quebec, Canada.

Source: Author's own

Search results returned a large list of ontologies that matched the concepts; therefore the selection of the ontologies was made based on the intelligibility and usability criteria. That is, that the ontologies that were clear, legible and usable were selected. In this section we describe the specific ontologies selected and integrated into the Patient Profile ontology model.

a) The National Drug File – Reference Terminology (NDF-RT) is a controlled medical terminology developed by the Department of Veterans Affairs Veterans Health Administration (VHA). The NDF-RT provides a formal model to describe *medications*, includes the FDA Established Pharmacologic Class(EPC), a set of Mechanisms of Action (MoA), Physiological Effects, Therapeutic Categories, and cross references to other important vocabularies such as the RxNorm⁵ provided by the National Library of Medicine (see Figure 6).

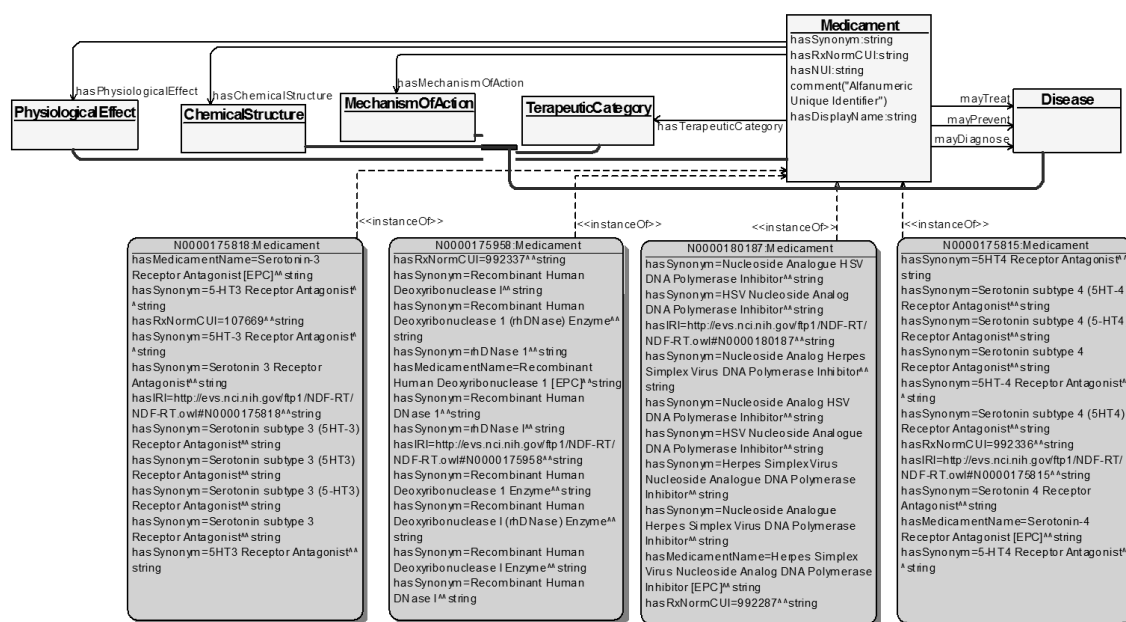


Figure 6. The National Drug File – Reference Terminology (NDF-RT) ontology reference.
 Source: Author's own

NDF-RT ontology was selected because it provides a revised and valid vocabulary of medicaments with relevant information such as the physiological effect, the chemical ingredients, the mechanism of action, the diseases that each medicament may prevent, may diagnose or may treat. This latter information is crucial to correlate pharmacological treatments with diseases.

b) The Human Disease Ontology (DOID)⁶ is a standardized reference that provides human disease terms, phenotype characteristics and medical vocabulary. DOID integrates mappings with important medical vocabularies: MeSH, ICD, NCI's thesaurus, SNOMED and OMIM. The DOID ontology was selected because it represents disease terms correlated with several important vocabularies.

c) SYMPT⁷. The symptom ontology (SYMP, 2020) is a collaborative effort between researchers from Colorado State University and Northwestern University, Chicago, IL. This ontology represents general classes of symptoms according with the human anatomy systems. The Symptom Ontology was developed as part of the Gemina project starting in 2005. Currently this ontology is maintained on the project at the Institute for Genome Sciences (IGS) at the University of Maryland.

d) The Logical Observation Identifiers Names and Codes (LOINC)⁸ system. The LOINC ontology provides standardized vocabularies for laboratory terms. LOINC is currently maintained by the Regenstrief Institute,

⁵ <https://www.nlm.nih.gov/research/umls/rxnorm/index.html>

⁶ <https://disease-ontology.org/>

⁷ <https://bioportal.bioontology.org/ontologies/SYMP/>

⁸ <https://bioportal.bioontology.org/ontologies/LOINC>

which in 1994 initiated the LOINC project to address the lack of standardization with multiple laboratories that used different codes for different test observations. LOINC provides a standard way of identifying observations using approximately 41,000 observation terms. 31,000 of these terms are used for laboratory testing.

To determine the ontologies that could be reused and imported into the model, we considered that the importance of each ontology or vocabulary lies mainly in the validity of the knowledge represented. Therefore, we have selected the vocabularies from institutions and groups of researchers that meet the requirements of the OBO Foundry to publish their ontologies in the BioPortal. Additionally, the intelligibility and ease of use of each vocabulary were considered for reutilization and integration.

Global Ontology Integration

Once the design and implementation of the core concepts have been completed, and having modularized the ontologies for reuse, the integration process consisted of the following steps: integrate core concepts of the Patient Profile in a single ontology file named *PatientElectronicRecord.owl*; select and import modularized ontologies and concepts to fulfill the requirements for the Patient Profile modeling; define the global semantic relationships (object properties) between the core concepts and the imported concepts that are necessary to meet the requirements of the ontology; and evaluate the resulting integrated ontology by representing case studies, and executing the logical consistency.

Definition of Global Semantic Relations

The list of global semantic relations is shown in Table 4, specifying the domain and range for each.

Table 4
 Details of the object properties defined for the Patient profile ontology

Object property	Domain	Range	Description
<i>personLivesIn</i>	Patient	Federative or Municipality	This relationship was created to indicate the particular municipality and federative entity where the patient lives.
<i>hasBMIClassification</i>	Patient	BMIClassification	This relationship was established to classify the patient's health status with respect to his weight.
<i>hasChemicalStructure</i>	Medicament	ChemicalStructure	This relationship is used to indicate the chemical structure of each drug.
<i>hasMechanismOfAction</i>	Medicament	MechanismOfAction	This relationship is used to describe how a medicament produces an effect in the body. For example, a drug's mechanism of action could be how it affects a specific target in a cell, such as an enzyme, or a cell function, such as cell growth.
<i>hasPhysiologicalEffect</i>	Medicament	PhysiologicalEffect	It is used to establish the biochemical and physiological effects of drugs and their mechanisms of action and the relationship between the concentration of the drug and its effect on an organism.
<i>diseaseHasSymptom</i>	Disease	Symptom	This relationship was established to indicate the symptoms associated to each disease.
<i>requiresLaboratoryTest</i>	Disease	LaboratoryTest	This relationship is used to specify the laboratory tests required for disease confirmation.
<i>hasLaboratoryTest</i>	ClinicalDiagnosis	LaboratoryTest	This relationship is used to indicate the set of laboratory tests necessary for each diagnosis.
<i>hasSymptom</i>	ClinicalDiagnosis	Symptom	It is used to establish the symptoms associated with the clinical diagnosis.
<i>hasDiagnosedDisease</i>	ClinicalDiagnosis	DOID_Disease or NDFRT_Disease	This relationship is used to indicate the diagnosed disease.
<i>hasDiagnosis</i>	Patient	ClinicalDiagnosis	This relationship is used to establish the correlation between the patient and the clinical diagnosis.
<i>hasTreatment</i>	ClinicalDiagnosis	Treatment	This semantic relationship is used to indicate that every diagnosis has an associated treatment.
<i>hasMedication</i>	PharmacologicalTreatment	Medicament	This semantic relationship is used to indicate that all pharmacological treatment consists of the prescription of drugs.

Source: Author's own

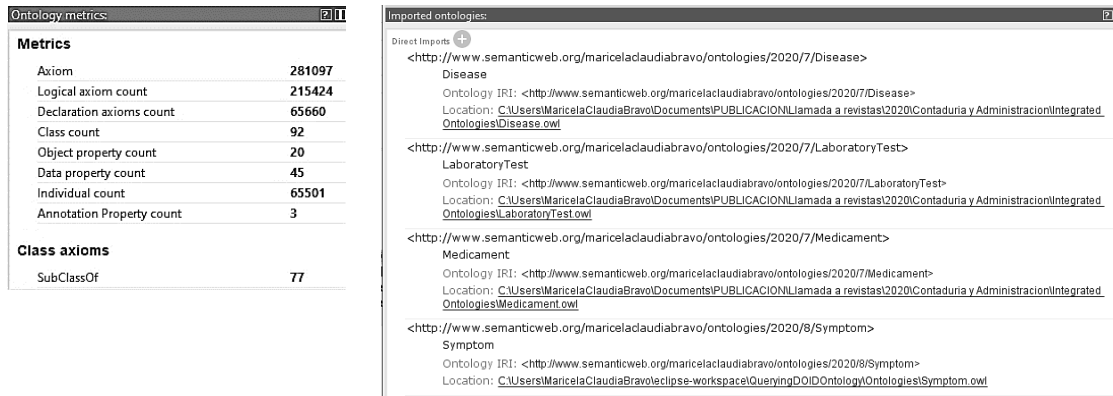


Figure 8. Metrics of the ontology model.
 Source: Author's own

Study Cases

Three study cases were implemented using the ontology model (Figure 9 shows the details). The first case study corresponds to a 57-year-old female patient, who was born and lives in Irapuato, Guanajuato. She is 1.58 meters tall, weighs 68 kilos, doesn't smoke, has A + blood type. Her body mass index is 27.23, which is why her BMI rating is overweight. Her physical activity is walking, her base metabolism is 1827.37 and her daily caloric intake is 1482. He has a diagnosis of Type 2 Diabetes Mellitus.

The second case study corresponds to a 55-year-old male patient, who was born in Cozumel, Quintana Roo and currently lives in Santiago Yaveo, Oaxaca. The patient weighs 68 kilos, measures 1.73 meters, his blood type is AB-, he does not smoke. His body mass index is 22.72 and his basal metabolic rate is 1579. This patient has a diagnosis of COVID.

The third case study corresponds to a 44-year-old male patient, who was born in Tuxpan, Jalisco and currently lives in Xalisco, Nayarit. The patient weighs 75 kilos, measures 1.71 meters, his blood type is B+, he smokes two cigarettes per day. His body mass index is 25.64 and his basal metabolic rate is 2856. This patient has a diagnosis of Hypertension.

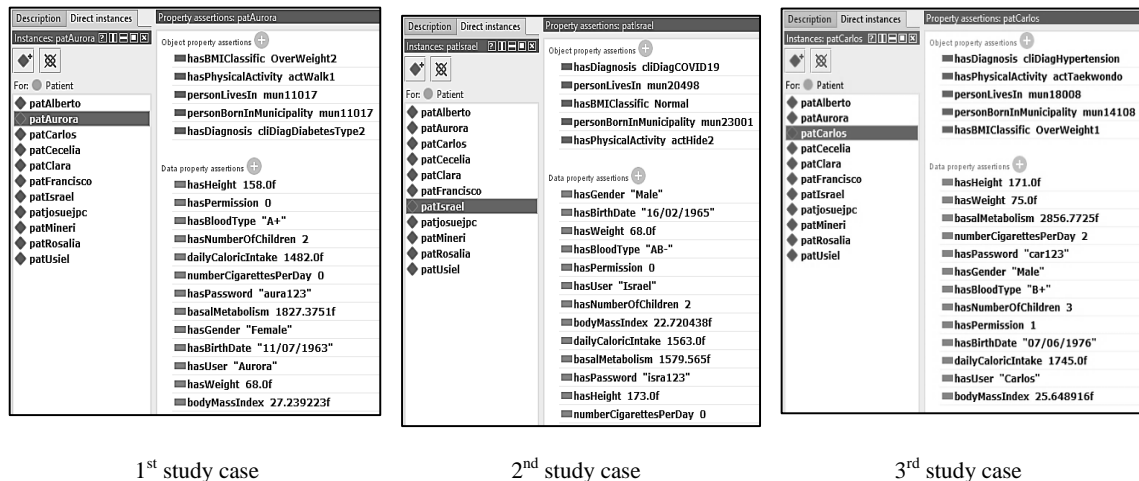


Figure 9. Representation of three patient profiles using the integrated ontology model.
 Source: Author's own

The three cases presented show the feasibility of representing profiles of patients with chronic diseases such as COVID, diabetes and hypertension. It is worth mentioning that this ontology imports a long vocabulary of diseases, which enables the representation of clinical records of patients with varied and multiple diagnoses and histories of suffered diseases.

Ontology Consistency Checking

Ontology development and evaluation require the verification of Consistency of the model. Reasoning with inconsistent ontologies may lead to erroneous conclusions. The verification of consistency of an ontology is the task carried out by the reasoner program that contrasts what is formally defined in the T-Box with respect to what is asserted in the A-Box. If inconsistencies are found, the reasoner will generate explanations of the inconsistencies. In the case of the ontological model presented in this article, the verification of logical consistency is of particular interest, since modules from external ontologies and vocabularies are imported, which can unintentionally cause the occurrence of logical inconsistencies. Ontology consistency is an important evaluation criteria, because it states that the entire integrated ontology is consistent and has the possibility to execute inference rules to find new concepts and execute queries over the concepts.

Figure 10 shows the result of executing the reasoner over the integrated model. During the execution of the reasoner, Pellet executes a tableaux reasoner that has only one functionality: checking the consistency of an ontology. According with (Sirin, 2007) an ontology is consistent if there is an interpretation that satisfies all the facts and axioms in the ontology. The tableaux algorithm constructs a graph-based representation of the A-Box, where each node is associated with its corresponding type, all property-value assertions are represented as directed edges between nodes. The reasoner repeatedly applies expansion rules until a contradiction is detected. All other reasoning tasks are defined in terms of consistency checking. Therefore, inference about the class hierarchy, the object property hierarchy, and the data property hierarchy derives all non-explicit superclass-subclass relationships which exist in these hierarchies. Regarding the inference about assertions of classes, object properties, and individuals, the reasoner determines if there are no inconsistencies between what is established as class axioms and what is instantiated in the A-Box. From a system management perspective, the results of the reasoner shown in Figure 10 means that user applications or software agents will be able to ask questions and obtain logically correct answers, as well as produce inferences about the facts established in the ontology, facilitating decision making process.

```
INFO 08:04:36 ----- Running Reasoner -----
INFO 08:04:39 Pre-computing inferences:
INFO 08:04:39   - class hierarchy
INFO 08:04:39   - object property hierarchy
INFO 08:04:39   - data property hierarchy
INFO 08:04:39   - class assertions
INFO 08:04:39   - object property assertions
INFO 08:04:39   - same individuals
INFO 08:06:53 Ontologies processed in 136965 ms by Pellet
INFO 08:06:53
```

Figure 10. Ontology Consistency Checking.
Source: Author's own

Gruber (1983) introduced the definition of Coherence design principle as follows: “an ontology should sanction inferences that are consistent with the definitions”. Based on the results obtained by the reasoner, after integration of ontological models, and some case scenarios it is possible to state the resulting ontology is *Coherent*.

Conclusions

In this work we have presented an ontological model that integrates two fundamental concepts: the patient's profile and the clinical diagnosis. The model was designed and built considering relevant information requirements for medical decision making. With this ontological model, it was possible to integrate a broad knowledge base that includes medical information from existing valid resources such as DOID, NDF-RT, SYMP, and LOINC ontology references.

The ontology model has been designed following important quality criteria, such as extendibility and reusability. Extensibility was achieved by designing several separate ontologies. For example, the disease ontology is integrated in a single separate owl file, within this ontology the keys and references of the diseases obtained from the original DOID ontology are incorporated. The same criteria were followed for the ontologies of drugs, symptoms and laboratory tests. Thus achieving a modular design, which allows each ontology to be updated separately without affecting the main design of the ontology that integrates the patient's profile and the clinical diagnosis. The reusability of the patient profile is achieved as a consequence of the modular and lightweight design. The patient profile can be reused in different scenarios and applications. Likewise, patient data can be used to perform data mining and find patterns that allow recognizing risk groups by categories, be it gender, age, social context, etc.

Study cases showed the feasibility of the model to be used as a broad knowledge base to support the management of patient profiles. The resulting ontology will be augmented by including more aspects such as management of diabetic patient based on physical activation and treatment plans based on an analysis of the mental and psychological state of the patient. Likewise, an extension will be made of the patients' data on their family history. The long-range purpose is to develop remote patient follow-up and recommendation applications.

References

- Ansari, R. M., Hosseinzadeh, H., Harris, M., & Zwar, N. (2019). Self-management experiences among middle-aged population of rural area of Pakistan with type 2 diabetes: a qualitative analysis. *Clinical Epidemiology and Global Health*, 7(2), 177-183. DOI: <https://doi.org/10.1016/j.cegh.2018.04.003>
- Chen, Yanshu; Tan, Dixin; Xu, Yihua; Wang, Bin; Li, Xuhui; Cai, Xiaonan; Li, Mengyu; Tang, Cong; Wu, Yun; Shu, Wen; Zhang, Guoliang; Huang, Jiayu; Zhang, Ying; Yan, Yaqiong; Liang, Xunchang; Yu, Songlin. (2020). In Patient Education and Counseling. 103(2):328-335 Language: English. DOI: <https://doi.org/10.1016/j.pec.2019.09.007>.
- Bravo, M., Hoyos Reyes, L. F., & Reyes Ortiz, J. A. (2019). Methodology for ontology design and construction. *Contaduría y Administración*, 64(4), 1–24. DOI: <https://doi.org/10.22201/fca.24488410e.2020.2368>
- Buranarach, M., Chalortham, N., Thein, Y. M., & Supnithi, T. (2011). Design and implementation of an ontology-based clinical reminder system to support chronic disease healthcare. *IEICE transactions on information and systems*, 94(3), 432-439. DOI: <https://doi.org/10.1587/transinf.e94.d.432>
- Chen, R. C., Huang, Y. H., Bau, C. T., & Chen, S. M. (2012). A recommendation system based on domain ontology and SWRL for anti-diabetic drugs selection. *Expert Systems with Applications*, 39(4), 3995-4006. DOI: <https://doi.org/10.1016/j.eswa.2011.09.061>
- Colberg, S. R., Sigal, R. J., Yardley, J. E., Riddell, M. C., Dunstan, D. W., Dempsey, P. C., ... & Tate, D. F. (2016). Physical activity/exercise and diabetes: a position statement of the American Diabetes Association. *Diabetes care*, 39(11), 2065-2079. DOI: <https://doi.org/10.2337/dc16-1728>
- Denova-Gutiérrez, E., Lopez-Gatell, H., Alomia-Zegarra, J. L., López-Ridauro, R., Zaragoza-Jimenez, C. A., Dyer-Leal, D. D., Cortés-Alcala, R., Villa-Reyes, T., Gutiérrez-Vargas, R., Rodríguez-González, K., Escondrillas-Maya, C., Barrientos-Gutiérrez, T., Rivera, J. A., & Barquera, S. (2020). The Association of Obesity, Type 2 Diabetes, and Hypertension with Severe Coronavirus Disease 2019 on Admission Among Mexican Patients. *Obesity (Silver Spring, Md.)*. DOI: <https://doi.org/10.1002/oby.22946>
- El-Sappagh, S., & Ali, F. (2016, December). DDO: a diabetes mellitus diagnosis ontology. In *Applied Informatics* (Vol. 3, No. 1, p. 5). Springer Berlin Heidelberg. DOI: <https://doi.org/10.1186/s40535-016-0021-2>
- Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2), 199-220. DOI: <https://doi.org/10.1006/knac.1993.1008>
- Hempo, B., Arch-int, N., Arch-int, S., & Pattarapongsin, C. (2015). Personalized care recommendation approach for diabetes patients using ontology and SWRL. In *Information science and applications* (pp. 959-966). Springer, Berlin, Heidelberg. DOI: https://doi.org/10.1007/978-3-662-46578-3_114

- International Diabetes Federation. IDF Diabetes Atlas, 9th ed. Brussels, Belgium: 2019. Available at: <https://www.diabetesatlas.org>, and last Accessed: 03/12/2020.
- Inzucchi, S. E., Bergenstal, R. M., Buse, J. B., Diamant, M., Ferrannini, E., Nauck, M., Peters, A. L., Tsapas, A., Wender, R., & Matthews, D. R. (2012). Management of hyperglycemia in type 2 diabetes: a patient-centered approach: position statement of the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetes Care*, 35(6), 1364–1379. DOI: <https://doi.org/10.2337/dc12-0413>.
- John W. Pelley, in *Elsevier's Integrated Review Biochemistry (Second Edition)*, 2012.
- Lasierra, N., Alesanco, A., Guillén, S., & García, J. (2013). A three stage ontology-driven solution to provide personalized care to chronic patients at home. *Journal of biomedical informatics*, 46(3), 516-529. DOI: <https://doi.org/10.1016/j.jbi.2013.03.006>.
- Li Chen, Dongxin Lu, Menghao Zhu, Muhammad Muzammal, Oluwarotimi Williams Samuel, Guixin Huang, Weinan Li, & Hongyan Wu. (2019). OMDP: An ontology-based model for diagnosis and treatment of diabetes patients in remote healthcare systems. *International Journal of Distributed Sensor Networks*, 15. DOI: <https://doi.org/10.1177/1550147719847112>
- Paganelli, F., & Giuli, D. (2007, May). An ontology-based context model for home health monitoring and alerting in chronic patient care networks. In *21st International Conference on Advanced Information Networking and Applications Workshops (AINAW'07) (Vol. 2, pp. 838-845)*. IEEE. DOI: <https://doi.org/10.1109/ainaw.2007.90>
- Sirin, E., Parsia, B., Grau, B. C., Kalyanpur, A., & Katz, Y. (2007). Pellet: A practical owl-dl reasoner. *Journal of Web Semantics*, 5(2), 51-53. DOI: <https://doi.org/10.1016/j.websem.2007.03.004>
- Rahimi, A., Liaw, S. T., Taggart, J., Ray, P., & Yu, H. (2014). Validating an ontology-based algorithm to identify patients with type 2 diabetes mellitus in electronic health records. *International journal of medical informatics*, 83(10), 768-778. DOI: <https://doi.org/10.1016/j.ijmedinf.2014.06.002>
- Shaker El-Sappagh, Daehan Kwak, Farman Ali, & Kyung-Sup Kwak. (2018). DMTO: a realistic ontology for standard diabetes mellitus treatment. *Journal of Biomedical Semantics*, 9(1), 1–30. DOI: <https://doi.org/10.1186/s13326-018-0176-y>.
- 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. DOI: <https://doi.org/10.1007/s13369-015-1959-4>
- American Diabetes Association. (2020). Standards of medical care in diabetes—2020 abridged for primary care providers. *Clinical diabetes: a publication of the American Diabetes Association*, 38(1), 10. DOI: <https://doi.org/10.2337/cd20-as01>
- SYMP; Symptom Ontology; DOI: <https://doi.org/10.25504/FAIRsharing.ay74mj>; Last edited: July 9, 2019, 10:34 a.m.; Available at: <https://biportal.bioontology.org/ontologies/SYMP>, and last Accessed: 03/12/2020.

Article

Modularization Method to Reuse Medical Knowledge Graphs

Maricela Bravo ^{1,*}, Darinel González-Villarreal ¹, José-A. Reyes-Ortiz ¹, and Leonardo-D. Sánchez-Martínez ¹

¹ Universidad Autónoma Metropolitana; mcbc@azc.uam.mx, darinelramstein@gmail.com, jaro@azc.uam.mx, ldsm@azc.uam.mx

* Correspondence: mcbc@azc.uam.mx

Abstract: During the creation and integration of a health care system based on medical knowledge graphs, it is necessary to review and select the vocabularies and definitions that best fit the information requirements of the system being developed. This implies the reuse of medical knowledge graphs; however, full importation of knowledge graphs is not a tractable solution in terms of memory requirements. In this paper we present a modularization-based method for knowledge graph reuse. A case study of graph reuse is presented by transforming the original model into a lighter one.

Keywords: Medical knowledge graphs, knowledge graphs reuse, ontology modularization

1. Introduction

Knowledge graphs in the medical area have become enormously relevant to support medical research development and to facilitate exchange of clinical data and scientific advances. That is the case of the well-known NCBO BioPortal, a medical ontology repository, where multiple health institutions publish their ontologies or knowledge graphs. In this article we describe a modularization-based method for reutilization of portions of medical knowledge graphs into a general model maintaining reasoning efficiency.

Reutilization of knowledge graphs occurs during the design and integration of multiple semantic representation models (ontologies or knowledge graphs) with the aim of generating a complete representation model to solve the requirements of intelligent information systems or expert systems. Reutilization of a knowledge graph or ontology is the task of completely or partially importing a model into another. The reuse of a knowledge graph may involve carrying out a process of modularization and adaptation or mapping from one model to another.

The process of reusing a knowledge graph begins with the search and selection of the graphs or ontologies that are considered appropriate to complete the definitions required by the system. Once the knowledge graphs to be reused have been identified, it is necessary to review in detail how this integration will be carried out. To address this problem, the following questions should be considered:

Is it necessary to import the entire model or only a part of the model is required?

Is it necessary to adjust or make adaptations between the imported model and the host model?

What is the most suitable modularization approach??

What is the computational cost of each option?

This article describes a general method for generating and reusing knowledge graph modules. This method arises from the need to integrate a general knowledge graph for managing electronic patient records. During the process of designing and integrating the general knowledge graph, a great disadvantage was observed in reusing some biomedical ontologies. This is mainly because most of these ontologies are very large and complex. Therefore, it was necessary to develop a method that would allow obtaining useful information for the general graph by extracting modules.

As a result, four modules of knowledge graphs were obtained that were imported into the general knowledge graph. The resulting knowledge graph functioned adequately for the information needs required for the project.

The rest of the paper is organized as follows: in Section 2 a review on modularization of knowledge graphs is presented, in Section 3 the method to generate and reuse medical knowledge graphs modules, in Section 4 the application of the method to generate the Medicament knowledge graph module is presented, in Section 5 the application of the method to generate the Disease knowledge graph module is described, in Section 6 the integrated knowledge graph modularization system is presented, in Section 7 two evaluation approaches are described, and finally in Section 8 conclusions are presented.

2. Modularization of Knowledge Graphs

One of the first attempts to address modularization of representation models was reported by Alan Rector in 2003 [1], who stated the importance of modularity as a key requirement for ontologies to achieve reutilization, maintainability, and evolution. Modularization of ontologies was defined as the task of decomposing an ontology into independent disjoint skeleton taxonomies.

In 2006 Schlicht and Stuckenschmidt [2] described a modularization approach aiming at supporting the distribution of knowledge in a P2P network. Authors describe the requirements of reasoning efficiency, robustness and maintainability; and propose the following structural criteria: connectedness of modules, stating that the efficiency of distributed computation is closely related with the degree of interconnectedness of the generated modules; size and number of modules, this criteria has a strong impact on the robustness; and redundancy of representation, the use of redundant representations will improve robustness at the cost of increasing maintenance. The goal of this approach is to apply modularization on a given ontology with the purpose of splitting it considering numerical aspects, such as: size, number of modules and modules interconnectedness.

From a logic-based perspective a series of contributions have been reported by a research group led by Bernardo Cuenca and Ian Horrocks. In 2006 [3] Cuenca et al. described the notion of a module from a model-theoretic perspective as a self-contained unit within the ontology. They stressed additional module requirements such as: the scope of the module, its size, and the correct interpretation of the module. Later in 2007 [4] Cuenca et al. presented an approach to extract modular fragments from ontologies preserving the minimal conditions. They defined a module with respect to a signature S , aiming at reducing the cost of importing external ontologies they proposed to import only a fragment of given ontologies preserving the meaning of the terms imported.

Based on the afore mentioned references a module can be defined as a self-contained knowledge unit within the knowledge graph, where a module has a clearly defined scope, its size, and a correct interpretation of the module.

Whereas modularization can be defined as the task of decomposing a knowledge graph into independent disjoint modules, splitting it considering modules size, number of modules and modules interconnectedness.

Accordingly, knowledge graph modularization should address the following criteria:

- a) Facilitate reutilization, and evolution of knowledge graphs.
- b) Maintain reasoning efficiency.
- c) Keep the balance between robustness and maintainability.
- d) Enable distribution of knowledge in open networks.
- e) Support connectedness of modules.

The concepts of module and modularization are closely related to the process of reusing knowledge graphs, since the purpose is to extract modules from knowledge graphs so that they can be imported into another model, ensuring that the semantic definitions and relations are not lost while maintaining the efficiency of reasoning.

Therefore, in the rest of this article we will use the concept of Knowledge Graph Module (KGM) as the resulting product of modularization process.

In 2009 Pathak et al. [5] presented a survey on modular ontologies techniques that are based on logical formalisms and graph theories. They state that modular ontologies techniques are crucial for the biomedical domain, since most popular ontologies are large and complex. Therefore, the development of tools for managing multiple distributed ontologies will benefit reasoning performance. Authors establish that an ontology module should be self-contained and logically consistent. Also defined the goals of ontology modularization: partial reuse, complexity, ownership and customization, efficient reasoning, and tooling support. Finally, authors outlined evaluation criteria and requirements: localized semantics, correct reasoning, transitivity, safe reuse and decidability.

In Courtot M. et al. [6] authors presented the Minimum Information to Reference an External Ontology Term (MIREOT) guidelines to aid the development of the Ontology of Biomedical Investigations (OBI). The purpose of this guide is to import the minimum required reference from an external ontology, for this the authors propose three cases, depending on what is required as a reference, all of them generally require that the identifier or URI (currently IRI) of the class or term reference be included. The method described in this article agrees to keep the unique identifier of each concept or class, but includes more relevant details of each concept, avoiding an overload in the memory of the reasoner.

One of the main trade offs of ontology modularization is to determine the level of expressiveness that you want to maintain from the original knowledge graph. However, the higher the level of expressiveness, the lower the ability to reason with large knowledge graphs and the efficiency will be very poor. Regarding this problem, Algergawy et al. [7] introduced the Ontology Analysis and Partitioning Tool (OAPT), a framework for analyzing and partitioning ontologies. However, the main approach used by the authors is to partition an ontology into modules and then apply evaluation criteria. Instead, in this article we focus on the extraction of specific information or knowledge that needs to be included in another knowledge graph of a particular purpose. The original graphs are not partitioned into modules.

In 2012 Alan Rector et al. [8] presented use cases for modular development of ontologies by using the import mechanism of OWL. According to the authors, ontology modularization concerns two topics: module extraction, which consists of separating existing ontologies into modules; and modular development, which resembles modular software engineering. In this article we present a modularization method that combines aspects of both cases, on the one hand, the general design of the knowledge graph is done with a modular development approach from the beginning; and on the other hand we present a semi-automatic method to perform the extraction of modules for reuse.

In 2021 Shimizu, Hammar & Hitzler [9] described the reasons why ontologies are not frequently reused as follows: differing representational granularity, lack of conceptual clarity in many ontologies, lack and difficulty of adherence to good modeling principles, lack of support in ontology engineering tools.

Another relevant issue that we deal with in the modularization method described in this paper is related to the way biomedical ontologies are developed, most of which do not have A-Boxes, they are mainly designed with T-Boxes. This represents a difficulty for their reuse, in case of establishing an *is_A* relationship between an object and a class, the interpretation is that the object is a member of that class, which generates misleading interpretations.

Although there are various approaches and studies presented for the reuse of ontologies or knowledge graphs, there are still difficulties for this practice to become a reality. In this paper we present a module extraction approach that depends on a careful design of the concepts and relationships that are required to be reused.

3. Method to Generate and Reuse Medical KGM

To describe the analysis and decision-making process that was carried out, this section first presents the system that required a method for reusing knowledge graphs. Specifically, the analysis of information requirements of the system required the search, selection, and reuse of various external knowledge graphs. This system is based on the specification of a comprehensive knowledge graph to support the representation and management of Electronic Health Records (EHR) of patients. We start from a base model that was designed with the aim of representing patient profiles. This base model was first reported in [10].

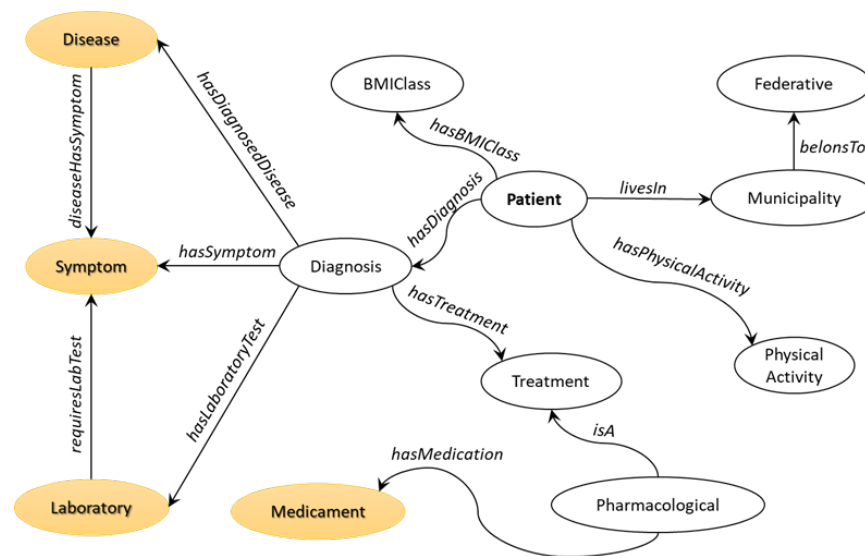


Figure 1. Knowledge graph for the representation and management of Electronic Health Records (EHR).

Figure 1 shows the general EHR model, which depicts in highlighted color the concepts that require reutilization of medical vocabularies: Medicament, Disease, Symptom, and Laboratory.

As a result of the analysis of conceptual requirements of the initial representation model, it was necessary to find reliable and scientifically valid sources of information for the inclusion of concepts about medical knowledge. Therefore, to complete this model, we decided to reuse vocabularies and definitions from external references. A series of searches were executed in BioPortal, resulting in the list of ontologies shown in Table 1.

Table 1. NCBO Ontologies identified for reuse.

Acronym	Description	Ontology
SYMPT	The Symptom Ontology has been developed as a standardized ontology for symptoms of human diseases.	http://purl.obolibrary.org/obo/symp.owl
LOINC	Logical Observation Identifier Names and Codes (LOINC)	http://purl.bioontology.org/ontology/LINC/
NDF-RT	National Drug File - Reference Terminology Public Inferred Edition	http://evs.nci.nih.gov/ftp1/NDF-RT/NDF-RT.owl
DOID	A hierarchical controlled vocabulary for human disease representation	http://purl.obolibrary.org/obo/doid.owl

¹ Tables may have a footer.

To reuse each of these knowledge graphs various methods exist, for instance, importing full models into a general knowledge graph. However, importing a complete model is not a good alternative, this is because the number of axioms that are imported are added, making the integrated representation model intractable in terms of the memory resources that would be required to make use of it. Therefore, a different approach should be defined and implemented.

We considered using a modularization-based method for reutilization of knowledge graphs, this method consists in the transformation of part of the knowledge graph definitions that are relevant for the specific model being integrated. For each of the general concepts highlighted in Figure 1, we defined and implemented the following procedure (method) that allowed the modularization and reutilization of knowledge graphs.

1. Select the knowledge graph or ontology that will be reused. The knowledge graph must be downloaded and analyzed locally.
2. Identify the relevant data or attributes that will be reused to complete the conceptualization into the knowledge graph being integrated.
3. Define and implement data structures using the Object-Oriented Programming paradigm to allow the clearest and most efficient handling of the concepts of interest.
4. Implement a program to query the knowledge graph using SPARQL and automatically obtain the list of concepts of interest.
5. Define and construct the model of the knowledge graph that will be used to contain the concepts and attributes of interest.
6. Develop a program to automatically populate the knowledge graph with the list of concepts obtained.

The application of this method for each of the required concepts is described in the following sections.

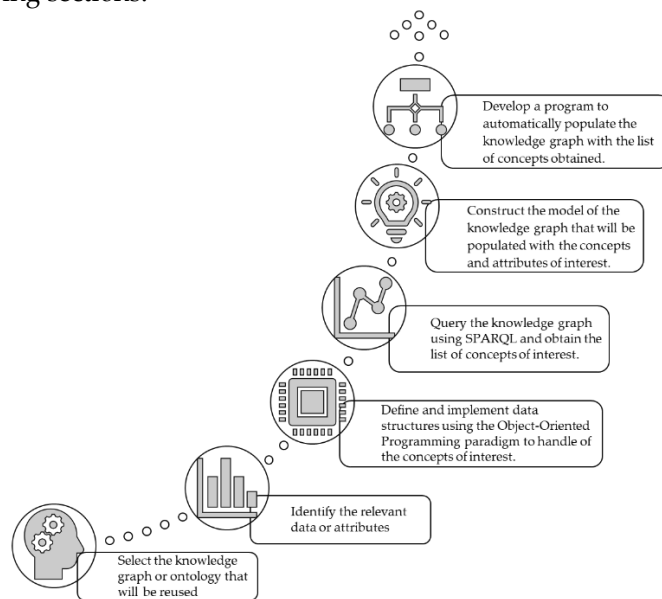


Figure 2. Method to generate and reuse medical KGM.

4. Medicament KGM

For the representation of an EHR it is necessary to specify the medications that a patient may be taking, or that will be indicated as part of his pharmacological treatment. Therefore, it is necessary to use a medical vocabulary of drugs.

To obtain a Medicament KGM, we implemented the method to generate and reuse medical KGM as follows:

4.1. Selection of the knowledge graph

For the representation of medicaments concepts and relations we selected the NDF-RT knowledge graph. NDF-RT is the National Drug File - Reference Terminology produced by the Veterans Health Administration. Figure 3 shows a snapshot of the NDF-RT model as visualized in Protégé editor.

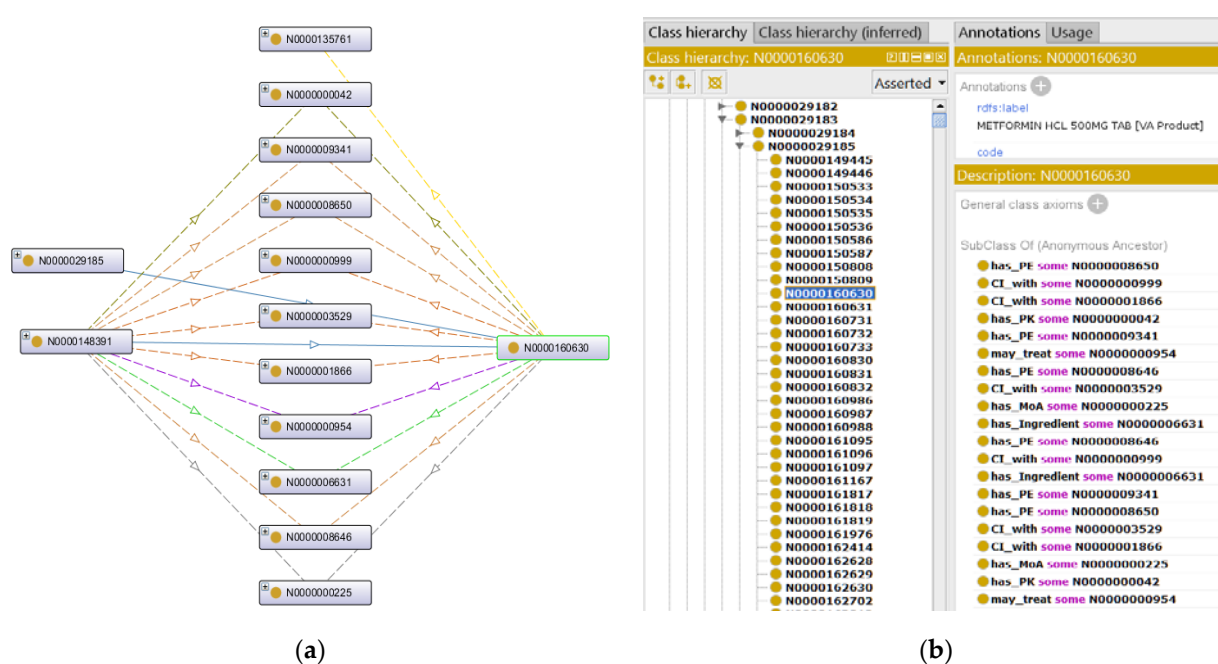


Figure 3. This figure shows a partial view of the NDF-RT ontology, specifically the N0000160630 class is highlighted which represents “Metformin hydrochloride 500 MG Oral Tablet”. (a) Presents the relations of N0000160630 with other classes; (b) Presents the details of the relations or properties of the N0000160630 concept.

As can be seen in Figure 3, the NDF-RT knowledge graph is very complex and is not an intuitive model to understand. This is in part because the medical information it represents is highly specialized and technical. However, the structure of the ontological model is also difficult to understand. For example, class naming is based on a nomenclature established in the design principles indicated by the NCBO for the publication of ontologies in BioPortal. As a result of this naming convention, class names do not provide any clue of which drugs they represent, it is necessary to expand the classes to identify the class that groups the entire hierarchy of drugs. Additionally, it is necessary to explore all the annotations to know the specific details of each concept.

Another problem that is observed with this knowledge graph is the fact that all the represented concepts are organized mainly using class definitions, that is, there are no individuals as member of the classes, which represents a difficulty in establishing a relationship between objects according to what is established in the base model shown in Figure 1.

Despite the difficulties that this model presents, NDF-RT is quite useful to be reused within the EHR because it defines concepts that relate drugs to diseases, for example, may-treat and may-prevent relationships, which we consider relevant.

4.2. Identification of the data and attributes to be imported

In order to review and analyze the NDF-RT knowledge graph was saved it in Turtle syntax. For the extraction of medicament products, we selected only the “Drug Products by VA Class”, which represent clinical pharmaceutical products that can be ordered. As an example, the code shown in Figure 4 describes the concept N0000160630, that identifies

the drug product "Metformin hydrochloride 500 MG Oral Tablet". Based on this code we can identify the important attributes that should be considered, for instance: the RxNorm_Name, the RxNorm CUI, units, strength, and NUI, among others.

```
NDF-RT:N0000160630 NDF-RT:RxNorm_Name "Metformin hydrochloride 500 MG Oral Tablet" ;
NDF-RT:code "C45762" ;
NDF-RT:Level "VA Product" ;
NDF-RT:Units "MG" ;
NDF-RT:Product_Component NDF-RT:N0000148391 ;
NDF-RT:Status "Active" ;
NDF-RT:NUI "N0000160630"^^xsd:string ;
NDF-RT:RxNorm_CUI "861007" ;
NDF-RT:VANDF_Record "
  <VANDF_Record>50.68^12369^</VANDF_Record>
  <VA_File>50.68</VA_File>
  <VA_IEN>12369</VA_IEN>" ;
NDF-RT:Print_Name "METFORMIN HCL 500MG TAB" ;
NDF-RT:UMLS_CUI "C0978483" ;
NDF-RT:Display_Name "METFORMIN HCL 500MG TAB" ;
NDF-RT:Product_Component_Qualifier "
  <Product_Component>METFORMIN HYDROCHLORIDE</Product_Component>
  <Strength>500</Strength>
  <Unit>MG</Unit><VA.IEN>3556</VA.IEN>" ;
rdfs:label "METFORMIN HCL 500MG TAB [VA Product]" ;
metadata:prefixIRI "N0000160630"^^xsd:string ;
NDF-RT:VA_National_Formulary_Name "METFORMIN TAB,ORAL" ;
NDF-RT:VUID "4012735" ;
NDF-RT:Strength "500" .
```

Figure 4. Excerpt from the NDF-RT representation of the N0000160630 class, that identifies the drug product "Metformin hydrochloride 500 MG Oral Tablet".

The set of relations or roles that NDF-RT uses can be extracted from a different part of the definitions, the code shown in Figure 5 presents the set of drug relationships with other concepts, for example, mechanism of action, physiological effects, clinical kinetics, among others.

```
### http://evs.nci.nih.gov/ftp1/NDF-RT/NDF-RT.owl#N0000160630
NDF-RT:N0000160630 rdf:type owl:Class ;
                    rdfs:subClassOf NDF-RT:N0000029185 ,
                                    NDF-RT:N0000148391 ,
                                    [ rdf:type owl:Restriction ;
                                      owl:onProperty NDF-RT: ;
                                      owl:someValuesFrom NDF-RT:N0000135761
                                    ] ,
                                    [ rdf:type owl:Restriction ;
                                      owl:onProperty NDF-RT:CI_with ;
                                      owl:someValuesFrom NDF-RT:N000000999
                                    ] ,
                                    [ rdf:type owl:Restriction ;
                                      owl:onProperty NDF-RT:CI_with ;
                                      owl:someValuesFrom NDF-RT:N000001866
                                    ] ,
                                    [ rdf:type owl:Restriction ;
                                      owl:onProperty NDF-RT:CI_with ;
                                      owl:someValuesFrom NDF-RT:N000003529
                                    ] ,
                                    [ rdf:type owl:Restriction ;
                                      owl:onProperty NDF-RT:has_Ingredient ;
                                      owl:someValuesFrom NDF-RT:N0000006631
                                    ] ,
                                    [ rdf:type owl:Restriction ;
                                      owl:onProperty NDF-RT:has_MoA ;
                                      owl:someValuesFrom NDF-RT:N000000225
                                    ] ,
                                    [ rdf:type owl:Restriction ;
                                      owl:onProperty NDF-RT:has_PE ;
                                      owl:someValuesFrom NDF-RT:N0000008646
                                    ] ,
```

Figure 5. Excerpt from the NDF-RT representation of, that describes the list of properties of the N0000160630 class.

4.3. Define and implement data structures

To extract and represent medicament information using an object-oriented paradigm, the JavaTM classes shown in Figure 6 were implemented. The Medicament class includes the attributes that were selected for their relevance. It should be noted that a lighter extraction of the concept attributes is performed. On the other hand, the Property class is used as a helper class in extracting all the important relationships that drugs can have with other concepts. The OntologyUtils class is an important class in which all the information extraction methods have been implemented, as well as the ontology population methods. The DrugOntologyCreation class represents the main program in which the data extraction, transformation, and module generation tasks are executed in sequence.

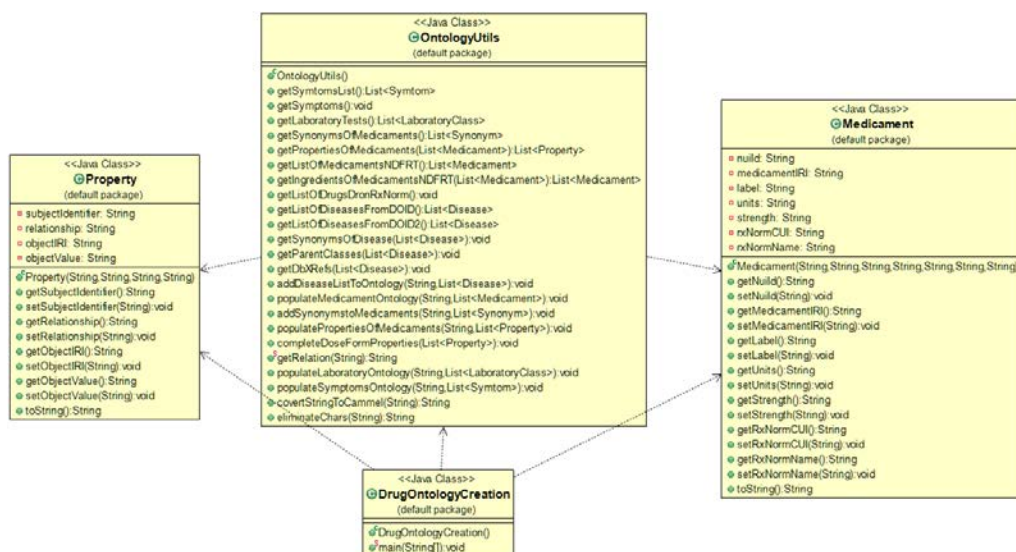


Figure 6. Class diagram of the Medicament KGM creation.

4.4. Implement a program to query the knowledge graph

To obtain the information from the NDF-RT knowledge graph, we implemented a JavaTM program using the API RDF4J that allowed issuing SPARQL queries through which a list of objects with the previously defined types were generated. The purpose of the first query in Table 2 is to obtain the list of medications and their basic attributes, the second query obtains the entire set of relationships that are all characterized as properties.

Table 2. SPARQL queries executed to retrieve the list of Medicament objects from NDF-RT.

SPARQL Query	Typed Collection Returned
PREFIX NDF-RT: <http://evs.nci.nih.gov/ftp1/NDF-RT/NDF-RT.owl#> SELECT distinct ?medicament ?label ?units ?strength ?nui ?cui ?name WHERE { ?medicament rdf:type owl:NamedIndividual . ?medicament rdfs:label ?label . ?medicament NDF-RT:Level "VA Product" . ?medicament NDF-RT:Units ?units . ?medicament NDF-RT:Strength ?strength . ?medicament NDF-RT:NUI ?nui . }	List<Medicament> listMed;

```

?medicament NDF-RT:RxNorm_CUI ?cui .
?medicament NDF-RT:RxNorm_Name ?name .      }
PREFIX NDF-RT: <http://evs.nci.nih.gov/ftp1/NDF-RT/NDF-RT.owl#>
SELECT ?property ?value ?valueLabel
WHERE {
  NDF-RT:" + medId + " rdf:type owl:Class .
  NDF-RT:" + medId + " rdfs:subClassOf ?s1 .
  ?s1 owl:onProperty ?property; owl:someValuesFrom ?value
  ?value rdfs:label ?valueLabel .      }

```

¹ SPARQL queries that are embedded into java code that produce a list of Medicament objects.

4.5. Define and construct the model of the knowledge graph

For the construction of the Medicament KGM that represents the data about drugs, it is necessary to design and implement the conceptual model of the target module graph, which is, the T-Box of the ontology. Therefore, and considering the types of data (or objects) returned by queries, the ontological model presented in Figure 7 was designed. This model represents a different view than the original NDF-RT knowledge graph. The resulting KGM is a simplified and much lighter version of the original, this model has more intuitive named concepts and relationships that facilitate reuse in another knowledge graph. For instance, it is clearer that a given instance of the Medicament concept may treat, or may diagnose, or may prevent any instance of the Disease concept.

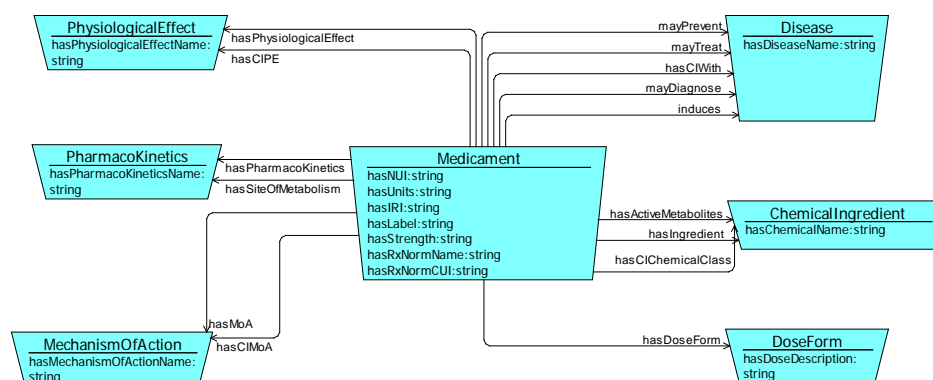


Figure 7. Model of Medicament KGM based on the NDF-RT ontology model.

According with the NDF-RT documentation , the following classes (concepts) are defined as follows:

- Medicament. This concept is at the core of the hierarchy in NDF-RT. It includes classifications of medications, generic ingredient preparations used in medications, and orderable drug products.
- Clinical Kinetics. Represents a collection of concepts describing the absorption, distribution, and elimination of drug active ingredients.
- Chemical Ingredient. Represents chemicals or other drug ingredients, organized into a chemical structure classification hierarchy.
- Mechanism of Action. Represents molecular, subcellular, or cellular effects of drug generic ingredients, organized into a chemical function classification hierarchy.
- Physiological Effect. Concerns tissue, organ, or organ system effects of drug generic ingredients, organized into an organ system classification hierarchy.
- Dose Form. Represents specific hierarchy of administered medication dose.

g) Disease. Represents pathophysiologic as well as certain non-disease physiologic states that are treated, prevented, or diagnosed by an ingredient or drug product. May also be used to describe contraindications.

The following semantic relations (object properties) are defined as follows:

- a) CI Chemical class. This relation is used to specify that a medicament is contraindicated with a chemical ingredient.
- b) Has ingredient. Relation used to indicate that a medicament has a chemical ingredient.
- c) Has active metabolites. Relation used to indicate that a medicament has active metabolites.
- d) Has PE. This relation is used to specify that a medicament has a physiological effect.
- e) CI PE. Relation used to indicate that a medicament is contraindicated with a physiological effect.
- f) Has dose form. This relation correlates the medicament with its form of dose.
- g) Has MoA. This relation is used to specify that a medicament has a mechanism of action.
- h) CI MoA. Relation used to indicate that a medicament is contraindicated with a mechanism of action.
- i) Induces. Relation used to indicate that a medicament induces a disease.
- j) May diagnose. Relation used to indicate that a medicament may diagnose a given disease.
- k) May prevent. Relation used to indicate that a medicament may prevent a disease.
- l) May treat. Relation used to indicate that a medicament may treat a disease.
- m) CI with. This relation is used to specify that a medicament is contraindicated with a given disease.
- n) Has PK. Relation that indicates that a medicament has a clinical kinetics (absorption, distribution, and elimination of drug active ingredients).
- o) Site of metabolism. Relation used to indicate that a medicament has a site of metabolism.

4.6. Develop a program to automatically populate the knowledge graph

Finally, a general JavaTM program was developed for the construction of the Medicament KGM, which integrates the information retrieval programs and the ontology population methods. The general procedure consists of the following steps: read the list of medicaments from the NDF-RT knowledge graph, recover the properties of the medicaments, populate the Medicament KGM with the list of medicaments, and finally for each medicament instance, populate the list of properties and update the Medicament KGM.

```
public class DrugOntologyCreation
{
    public static void main(String[] args)
    {
        OntologyUtils util = new OntologyUtils();

        //Step 1. Read the list of medicaments and register them into the ontology
        List<Medicament> listMed = util.getListOfMedicamentsNDFRT();

        //Step 2. Obtain the relationships for each medicament
        List<Property> listProps = util.getPropertiesOfMedicaments(listMed);
        util.completeDoseFormProperties(listProps);

        //Step 3. Write medicament objects into the Ontology
        util.populateMedicamentOntology("Ontologies/Medicament.owl", listMed);

        //Step 4. Write properties into the ontology
        util.populatePropertiesOfMedicaments("Ontologies/Medicament.owl", listProps);
    }
}
```

Figure 8. Main program in which the data extraction, transformation, and module generation tasks are executed in sequence.

5. Disease KGM

One of the most important concepts of an EHR is that of diseases. There are various definitions of disease in dictionaries or in the medical literature, here we present some general definitions:

“Disease, any harmful deviation from the normal structural or functional state of an organism, generally associated with certain signs and symptoms and differing in nature from physical injury.” (Encyclopedia Britannica)

“(an) illness of people, animals, plants, etc., caused by infection or a failure of health rather than by an accident.” (Cambridge Dictionary)

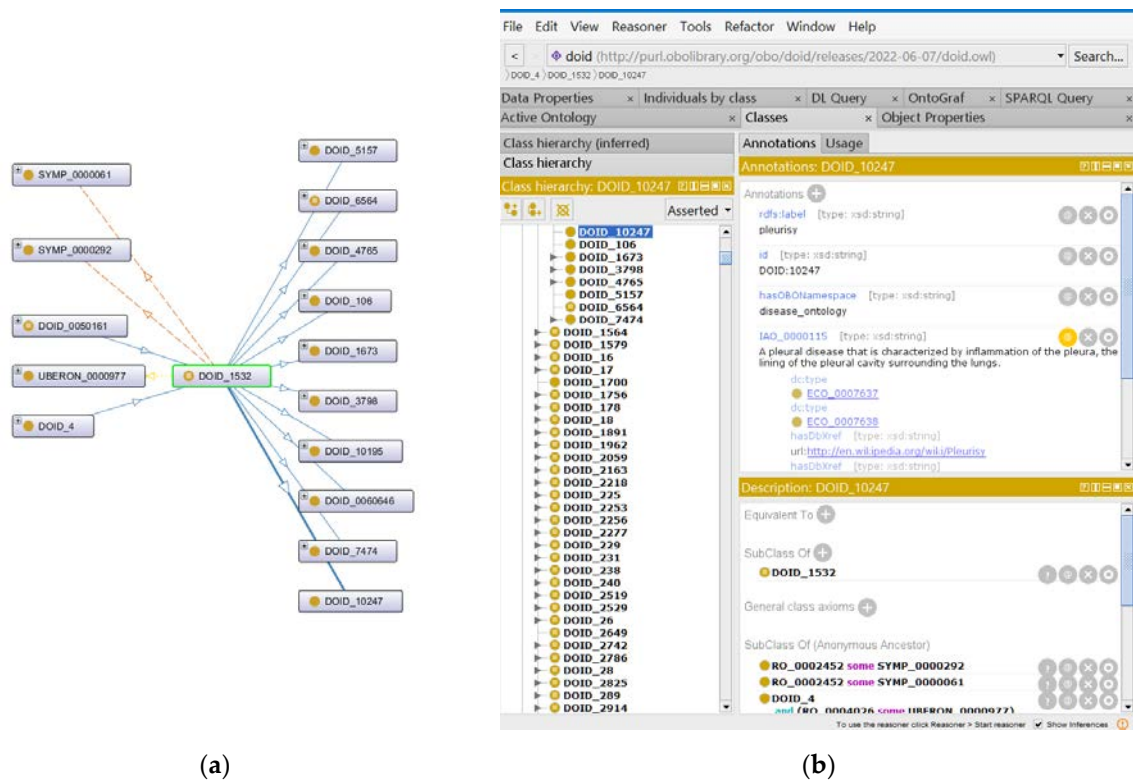
“A disease is an illness which affects people, animals, or plants, for example one which is caused by bacteria or infection.” (Collins Dictionary)

To obtain a Disease KGM we implemented the method to generate and reuse medical KGM as follows:

5.1. Selection of the knowledge graph

Considering the management of information for an EHR, a disease is associated with the person or patient to whom the file refers; likewise, a disease is associated with a set of signs and symptoms. Therefore, we have decided to reuse the DOID knowledge graph for this purpose. The Human Disease Ontology (DOID) was developed with the collaboration between biomedical researchers coordinated by the University of Maryland School of Medicine, Institute for Genome Sciences.

Figure 9 shows a partial view of the DOID ontology, as can be seen, this model presents a complex class structure, and the names are not intuitive either. If there is a requirement for obtaining the information of a particular disease, it is necessary to know the specific id with which it has been classified to do a search and read all the tags added as annotations of the concept. Of course, with a well-developed search tool, the use of this type of model is feasible, but it still requires knowledge and experience dealing with this kind of vocabularies.



(a)

(b)

Figure 9. This figure shows a partial view of the DOID ontology, specifically the DOID_10247 class which represents “pleurisy” a disease. (a) Presents the relations of DOID_10247 with other classes; (b) Presents the details of the relations or properties of the DOID_10247 concept.

5.2. Identification of the data and attributes to be imported

To execute this revision the DOID knowledge graph was saved in Turtle syntax. The code shown in Figure 10 describes the concept **DOID_9352**, which corresponds to the diabetes mellitus type 2 disease. Based on this code we can identify important attributes and relations, such as: `obo:IAO_0000115` which is an annotation property used for defining and explaining the meaning of a class or property; `oboInOwl:hasDbXref` is a qualifier that allows cross referencing other data bases, enabling the recovery of information about the disease by querying the included references; `oboInOwl:hasExactSynonym` is a property used to include synonyms of the disease; and `rdfs:label` which is used to describe the disease.

```

### http://purl.obolibrary.org/obo/DOID_9352
obo:DOID_9352 rdf:type owl:Class ;
  rdfs:subClassOf obo:DOID_9351 ;
  obo:IAO_0000115 "A diabetes mellitus that is characterized by high blood sugar,
  insulin resistance, and relative lack of insulin."^^xsd:string ;
  oboInOwl:hasDbXref "EFO:0001360"^^xsd:string ,
    "ICD10CM:E11"^^xsd:string ,
    "KEGG:04930"^^xsd:string ,
    "MESH:D003924"^^xsd:string ,
    "NCI:C26747"^^xsd:string ,
    "OMIM:125853"^^xsd:string ,
    "OMIM:601283"^^xsd:string ,
    "OMIM:601407"^^xsd:string ,
    "OMIM:603694"^^xsd:string ,
    "OMIM:608036"^^xsd:string ,
    "SNOMEDCT_US_2021_09_01:44054006"^^xsd:string ,
    "UMLS_CUI:C0011860"^^xsd:string ;
  oboInOwl:hasExactSynonym "NIDDM"@en ,
    "insulin resistance"^^xsd:string ,
    "non-insulin-dependent diabetes mellitus"@en ,
    "type 2 diabetes"@en ,
    "type II diabetes mellitus"@en ;
  oboInOwl:hasOBONamespace "disease_ontology"^^xsd:string ;
  oboInOwl:id "DOID:9352"^^xsd:string ;
  oboInOwl:inSubset doid:DO_FlyBase_slim ,
    doid:NCIthesaurus ;
  rdfs:comment ""Xref MGI.
  OMIM mapping confirmed by DO. [SN].""^^xsd:string ;
  rdfs:label "type 2 diabetes mellitus"^^xsd:string .

```

Figure 10. Excerpt from the DOID ontology describing the DOID_9352 concept which corresponds to the diabetes mellitus disease.

5.3. Define and implement data structures

To extract and represent disease information as a list of objects the JavaTM classes shown in Figure 11 were implemented.



Figure 11. Class diagram for the extraction of disease concepts from the DOID ontology, and ontology population of the Disease KGM.

5.4. Implement a program to query the knowledge graph

To obtain the information from the DOID knowledge graph, we implemented various JavaTM programs using the API RDF4J, these programs execute the SPARQL queries shown in Table 3. The first query retrieves the data of diseases from DOID ontology, the rest of the queries are used to complete the definitions with other important data such as the synonyms of each disease, the parent classes, and the set of data base cross references for each disease. It is important to note that the information retrieved and included in the new ontology contains IRI references to enable further queries in case that additional information of a particular disease is required.

Table 3. SPARQL queries executed to retrieve the list of Disease objects from DOID.

SPARQL Query	Typed Collection Returned
<pre> PREFIX oboIn: <http://www.geneontology.org/formats/oboInOwl#> PREFIX obo: <http://purl.obolibrary.org/obo/> SELECT DISTINCT ?s ?label ?def ?synonym ?id WHERE { ?s rdf:type owl:Class . ?s rdfs:label ?label . ?s obo:IAO_0000115 ?def . ?s oboIn:hasExactSynonym ?synonym . ?s oboIn:id ?id . } </pre>	List<Disease> listDis;
<pre> PREFIX oboIn: <http://www.geneontology.org/formats/oboInOwl#> PREFIX obo: <http://purl.obolibrary.org/obo/> SELECT ?syn WHERE { obo:" + id + " oboIn:hasExactSynonym ?syn . } </pre>	List<Disease> listDis;
<pre> PREFIX oboIn: <http://www.geneontology.org/formats/oboInOwl#> PREFIX obo: <http://purl.obolibrary.org/obo/> SELECT ?parent WHERE { obo:" + id + " rdfs:subClassOf ?parent . } </pre>	List<Disease> listDis;
<pre> PREFIX oboIn: <http://www.geneontology.org/formats/oboInOwl#> PREFIX obo: <http://purl.obolibrary.org/obo/> SELECT ?ref WHERE { obo:" + id + " oboIn:hasDbXref ?ref . } </pre>	List<Disease> listDis;

¹ SPARQL queries that are embedded into java code to obtain the list of Disease objects.

5.5. Define and construct the model of the knowledge graph

For the construction of the Disease knowledge graph it is necessary to design the conceptual model, which is, the T-Box of the ontology. Considering the structure of the Disease type returned by SPARQL queries, the ontological model depicted in Figure 12 was developed. This model represents mainly the Disease concepts with the most important references and definitions to other medical data bases, also is a simplified and lighter representation resource compared with the original DOID knowledge graph. The most important difference is the elimination of direct and indirect imported ontologies. The extraction and construction of an ontological representation model without the use of imported ontologies offers benefits in terms of managing memory resources to perform logical inference. Specifically, the representation model developed has the basic attributes that allow diseases to be identified without the need to include a burden of reference models. If required, the concepts can be expanded through queries using the IRIs and the referenced databases.

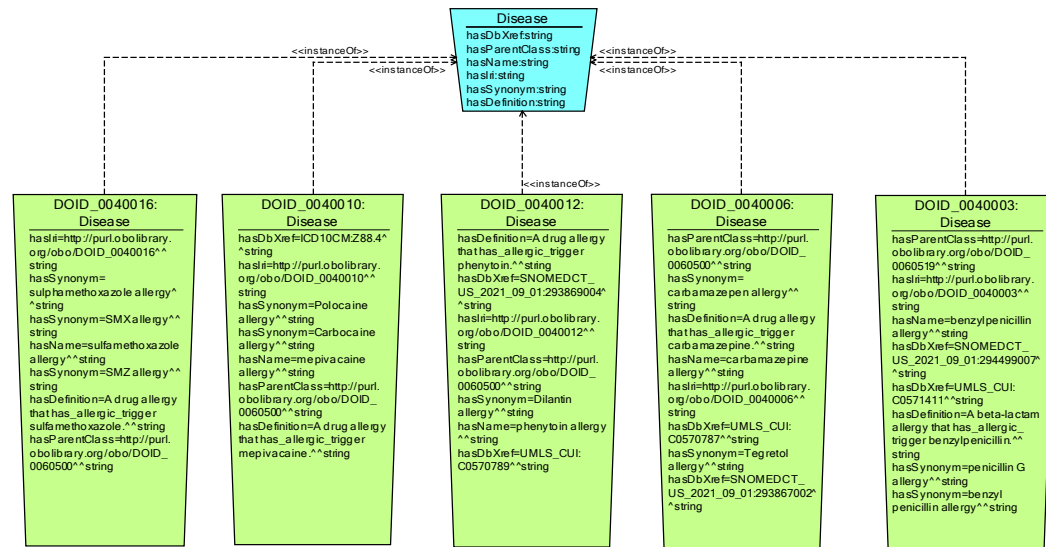


Figure 12. Ontology model for the representation of Disease information.

5.6. Develop a program to automatically populate the knowledge graph

Figure 13 shows the general Java program that was implemented for the construction of the Disease KGM, which starts by executing the program that obtains the list of disease definitions from DOID ontology, then updates the list of disease objects obtaining the synonyms, the parent class, and the data base cross references of each disease. Finally, the program writes the list of disease objects into the Disease ontology file.

```

public class DiseaseOntologyCreation
{
    //Generate ontology Disease module from DOID ontology
    public static void main(String[] args)
    {
        //Step1. Extract the list of diseases from DOID ontology
        OntologyUtils util = new OntologyUtils();
        List <Disease> listDis = util.getListOfDiseasesFromDOID2();
        //Step 2. Obtain the synonyms of each disease and update the list
        util.getSynonymsOfDisease(listDis);
        //Step 3. Get the parent classes IRIs of each disease and update the list
        util.getParentClasses(listDis);
        //Step 4. Get the data base cross references of each disease and update the list
        util.getDbXRefs(listDis);
        //Step 5. Write the disease individuals into the ontology
        util.addDiseaseListToOntology("Ontologies/Disease.owl", listDis);
    }
}

```

Figure 13. Main program in which the data extraction, transformation, and module generation tasks are executed in sequence.

6. Knowledge Graph Modularization System

We integrated all the programs developed in an modularization system. This system design is based on the object-oriented programming paradigm which makes it easy to carry out updates or adaptations in any of the tasks of the modularization process. Figure 14 shows the class diagram of the modularization system. The objective of this system is to extract the concepts of interest from diverse medical knowledge graphs, represent these concepts using a object oriented paradigm, and generate specific KGMs that will be reused and usable in the target EHR knowledge graph.

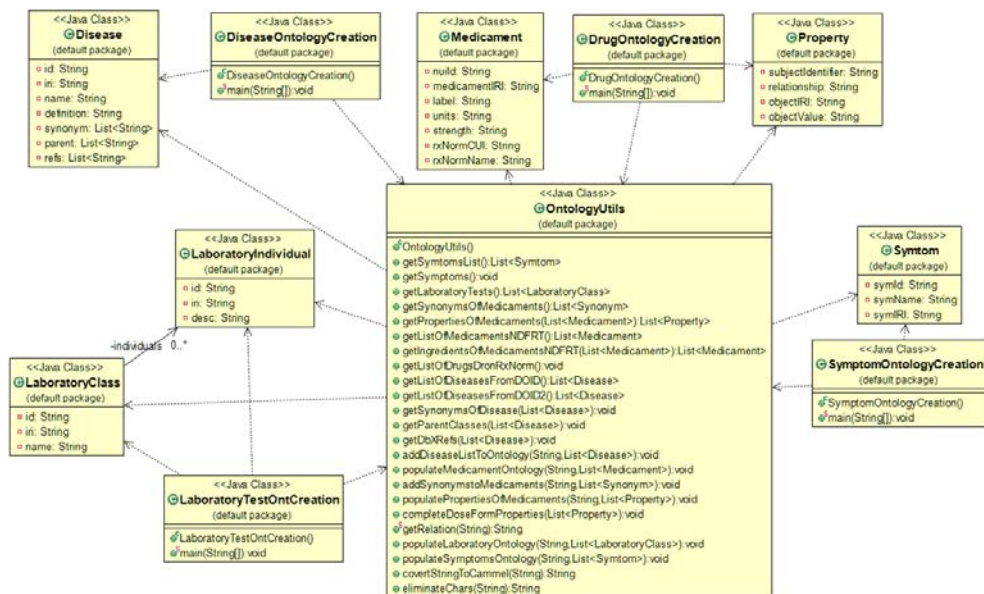


Figure 14. Class diagram of the Knowledge Graph Modularization System.

Applying the described method, the *LaboratoryTest*, *Symptom*, and *Vaccine* KGMs were also generated. As can be seen, the method for generating and reusing medical KGM is semi-automatic. This is because it is necessary to know in detail the knowledge graphs from which the modules will be extracted, in addition to the fact that important decisions must be made regarding what attributes and relationships are required in the target graph.

Another relevant aspect that must be taken into consideration is that the medical knowledge graphs are constantly being updated, for which it is essential to review the changes in the original model and eventually make the pertinent updates in the system.

As a result of the implementation of the method for the generation and reuse of KGMs, four modules were obtained, which are described below.

7. Evaluation

Aiming at evaluating the method reported in this paper, in this section we describe two different approaches to evaluate the resulting modules.

7.1 Evaluation of the usefulness of the knowledge graph modules

According with Duque-Ramos et al. [11] “the quality of an ontology module can be defined as the degree of conformance to functional and non-functional requirements”. In the case of the knowledge graph modules generated, they were intended to be used as part of an integral knowledge graph to support management of EHR. With this goal in mind, we review the general knowledge graph and its usefulness to represent patient data with relationships to these graphs. Figure 15 shows the EHR general graph diagram, which presents in light blue the predefined concepts and attributes: *Patient*, *ClinicalDiagnosis*, *StateFederative*, and *Municipality*; this diagram also shows in dark blue the concepts that come from the imported graph modules: *LaboratoryTest*, *NDFRT_Disease*, *PharmacologicalTreatment*, and *COVID-19 Vaccine*.

It is important to mention that this general knowledge graph for the management of EHRs was populated with 500 clinical records of patients. To show the usefulness of the integrated knowledge graph, Figure 16 shows an instance of real data from a female patient who was diagnosed with Coronavirus, this diagnosis is related to the *NDFRT_Disease* graph.

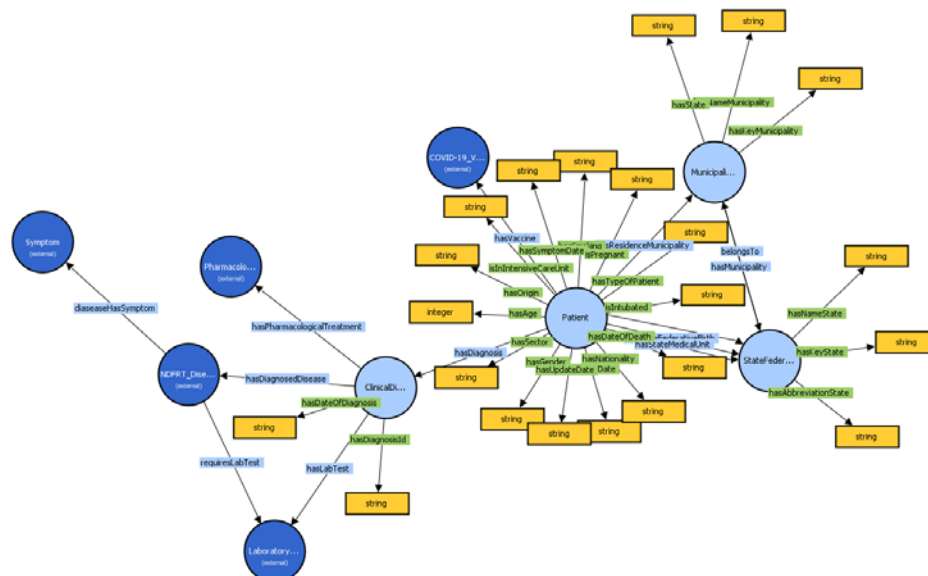


Figure 15. Integrated general knowledge graph.

Figure 16. Example of an instance of a patient registered in the general graph.

Figure 17 shows the specific diagnosis related to the disease (NDFRT_Disease imported graph), with the laboratory test required for the diagnosis (LaboratoryTest imported graph) and with the pharmacological treatment.

Figure 17. Example of an instance of a diagnosis related with NDFRT_Disease and LaboratoryTest.

Figure 18 shows an example of pharmacological treatment for COVID-19, which is related to the medications (Medicament imported graph) indicated by the doctor.

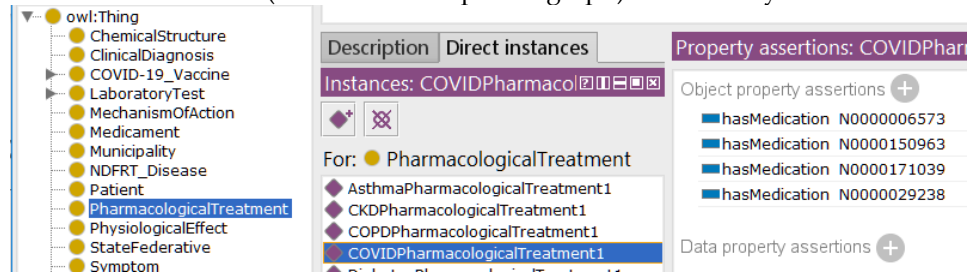


Figure 18. Example of a Pharmacological treatment related with specific medications.

Figure 19 shows the coronavirus disease related signs and symptoms from the Symptom knowledge graph.

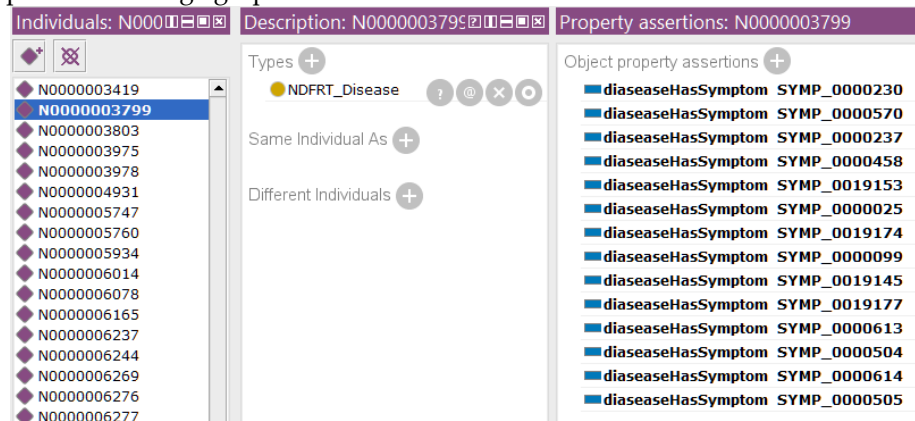


Figure 19. Integrated general knowledge graph.

Based on the initial requirements of the EHR management system, it can be determined that the imported graph modules satisfactorily meet the objective of the model. Therefore, from the application and utility point of view, the graph modules generated by the method described in this article are correctly designed and meet the required information needs.

7.2 Evaluation based on graph metrics

To carry out a quantitative evaluation, a comparison of the metrics of the original knowledge graphs versus the generated knowledge graph modules is described in this subsection. The set of metrics used to make the comparison are those generated by the Protégé ontology editor. Among all the metrics presented, the most important is the number of axioms, which despite being very general, offers a good measure of the reduction achieved with modularization., the general idea is to have an estimation of the size of the module extracted and the reduction achieved. Figure 20 shows on the left side the original NDF-RT metrics, whereas on the right side the metrics of the Medicament KGM generated. As can be seen the extracted Medicament KGM represents the 20.33% of the complete NDF-RT. Other important difference is the number of classes used un the Medicament KGM which is 7, while the original has 46047 classes. This is mainly because the original graph does not define individuals, it handles everything with classes.

NDF-RT original KG metrics	Medicament KGM metrics
----------------------------	------------------------

Ontology metrics:		Ontology metrics:	
Metrics			
Axiom	991831	Axiom	201647
Logical axiom count	238553	Logical axiom count	188795
Declaration axioms count	70984	Declaration axioms count	12852
Class count	46047	Class count	7
Object property count	26	Object property count	16
Data property count	29	Data property count	14
Individual count	24830	Individual count	12815
Annotation Property count	56	Annotation Property count	0

Figure 20. Comparison of NDF-RT metrics with the Medicament KGM obtained

Figure 21 on the left shows the metrics of the original DOID knowledge graph, and on the right the metrics of the Disease KGM obtained from the modularization process. Disease KGM metrics show 56% reduction in the number of axioms of the original knowledge graph.

DOID original KG metrics	Disease KGM metrics		
Metrics			
Axiom	179536	Axiom	78874
Logical axiom count	29282	Logical axiom count	72418
Declaration axioms count	17994	Declaration axioms count	6456
Class count	17898	Class count	1
Object property count	45	Object property count	0
Data property count	0	Data property count	6
Individual count	0	Individual count	6449
Annotation Property count	52	Annotation Property count	0
Class axioms			
SubClassOf	28496	SubClassOf	0
EquivalentClasses	722	EquivalentClasses	0
DisjointClasses	26	DisjointClasses	0

Figure 21. Comparison of DOID metrics with the Disease KGM obtained

8. Conclusions

One of the difficult problems to solve during the reuse of medical knowledge graphs is that they directly or indirectly import other ontologies or graphs. This generates an overload of models in RAM memory, which prevents reasoning and inference from being carried out more efficiently.

In this article we have described a modularization method applicable to large knowledge graphs; in particular, in graphs from the biomedical area that, in addition to being large in volume, are highly complex for the user. The described modularization method differs from methods that seek to divide a graph into parts, each of which is self-contained. It is also different from fully automatic methods that extract a module considering logically complete modules.

The modularization method described is semi-automatic and is more oriented towards having support tools to solve very specific needs of information or specialized knowledge that must be incorporated into larger and complex knowledge graphs.

The examples of application of the method allow to observe and analyze the specific decision-making that must be done during the design of the knowledge modules, the

selection of the necessary attributes to meet the requirements, and that once the design is complete, a mechanism for automatic population is programmed.

It should be noted that the reuse of knowledge graphs entails many challenges, one of the most complicated being that the way in which the structure of the original graph may not be the most appropriate for the destination graph. Therefore, in the examples shown in this article, we have made the decision to make changes in the way of representing the classes and class hierarchies. We have chosen to handle many concepts or classes as individuals in the destination graph.

Author Contributions: Conceptualization, Maricela Bravo and Darinel González-Villarreal; Formal analysis, Maricela Bravo and José-A. Reyes-Ortiz; Investigation, Maricela Bravo, José-A. Reyes-Ortiz and Leonardo-D. Sánchez-Martínez; Methodology, Maricela Bravo; Software, Darinel González-Villarreal; Validation, Maricela Bravo, José-A. Reyes-Ortiz and Leonardo-D. Sánchez-Martínez; Writing – original draft, Maricela Bravo, José-A. Reyes-Ortiz and Leonardo-D. Sánchez-Martínez; Writing – review & editing, Maricela Bravo, Darinel González-Villarreal, José-A. Reyes-Ortiz and Leonardo-D. Sánchez-Martínez.

Funding: “This research received no external funding”.

Data Availability Statement: In this section, please provide details regarding where data supporting reported results can be found, including links to publicly archived datasets analyzed or generated during the study. Please refer to suggested Data Availability Statements in section “MDPI Research Data Policies” at <https://www.mdpi.com/ethics>. If the study did not report any data, you might add “Not applicable” here.

Conflicts of Interest: “The authors declare no conflict of interest.” “The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results”.

References

1. Rector, A., L. Modularisation of Domain Ontologies Implemented in Description Logics and related formalisms including OWL. Proceedings of the 2nd International Conference on Knowledge Capture, pp. 121-128, 2003.
2. Schlicht, A., Stuckenschmidt, H. Towards Structural Criteria for Ontology Modularization. Proceedings of the ISWC Workshop on Modular Ontologies, 2006.
3. Cuenca, B., Parsia, P., Sirin, E., Kalyanpur, A. Modularity and Web Ontologies. Proceedings of the Tenth International Conference on Principles of Knowledge Representation and Reasoning, Lake District of the United Kingdom, June 2-5, 2006.
4. Grau, Bernardo Cuenca, et al. "A Logical Framework for Modularity of Ontologies." IJCAI. Vol. 2007.
5. Pathak, J., Johnson, T. M., & Chute, C. G. (2009). Survey of modular ontology techniques and their applications in the biomedical domain. *Integrated computer-aided engineering*, 16(3), 225-242.
6. Courtot, M., Gibson, F., Lister, A. L., Malone, J., Schober, D., Brinkman, R. R., & Ruttenberg, A. (2011). MIREOT: The minimum information to reference an external ontology term. *Applied Ontology*, 6(1), 23-33.
7. Algergawy, A., Babalou, S., Klan, F., & König-Ries, B. (2020). Ontology modularization with OAPT. *Journal on Data Semantics*, 9(2), 53-83.
8. Rector, A., Brandt, S., Drummond, N., Horridge, M., Pulestin, C., & Stevens, R. (2012). Engineering use cases for modular development of ontologies in OWL. *Applied Ontology*, 7(2), 113-132.
9. Shimizu, C., Hammar, K., & Hitzler, P. (2021). Modular ontology modeling. *Semantic Web*, (Preprint), 1-31.
10. Bravo, M., González, D., Ortiz, J. A. R., & Sánchez, L. "Management of diabetic patient profiles using ontologies". *Contaduría y administración*, 65(5), 12 (2020).
11. Duque-Ramos, A., Fernández-Breis, J. T., Iniesta, M., Dumontier, M., Aranguren, M. E., Schulz, S., ... & Stevens, R. (2013). Evaluation of the OQuaRE framework for ontology quality. *Expert Systems with Applications*, 40(7), 2696-2703.



www.cya.unam.mx/index.php/cya
Contaduría y Administración, 2022, 1-12

Representación y gestión de datos de pacientes mexicanos con COVID-19 mediante Grafos de Conocimiento

Darinel González¹, Maricela Bravo¹, José-A. Reyes-Ortiz¹,
Luis Fernando Hoyos-Reyes¹

¹ *Universidad Autónoma Metropolitana, Av. San Pablo Xalpa 180, Reynosa Tamaulipas,
Azcapotzalco, México City, México*

*Received INSERTAR FECHA RECIBIDO.; accepted INSERTAR FECHA ACEPTADO.
Available online fecha.*

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. Por otra parte, el empleo de grafos de conocimiento favorece la integración y expansión del conocimiento mediante la vinculación con otros grafos. En este artículo 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. 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.

JEL code: C63, D83, I00, L86

Palabras clave: Integración de grafos de conocimiento, reutilización de grafos, grafos de conocimiento biomédico.

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)¹ 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 esta cantidad de datos se requiere contar con 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 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 <sujeeto, predicado, objeto> ofrece una gran flexibilidad para el escalamiento de datos mediante la referenciación de IRIs, 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:

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

- 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.

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).

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 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.

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.

Figura 1.1 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-19², la cual es publicada por las autoridades de salud del Gobierno de México. En este portal se puede consultar la

² <https://coronavirus.gob.mx/covid-19/>

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 2012³, 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 viral⁴.
- d) Guía clínica para el tratamiento de la COVID-19 en México⁵. 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.

Etapas 2. Búsqueda y selección de grafos de conocimiento existentes

El objetivo de 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. Bioportal⁶ 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

³ https://epidemiologia.salud.gob.mx/gobmx/salud/documentos/manuales/00_NOM-017-SSA2-2012_para_vig_epidemiologica.pdf

⁴ https://www.gob.mx/cms/uploads/attachment/file/715464/Lineamiento_VE_y_Lab_Enf_Viral_05042022.pdf

⁵ https://www.gob.mx/cms/uploads/attachment/file/659911/GuiaTx_COVID19_Consenso_2021.08.02_compressed.pdf

⁶ <https://bioportal.bioontology.org/>

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

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.

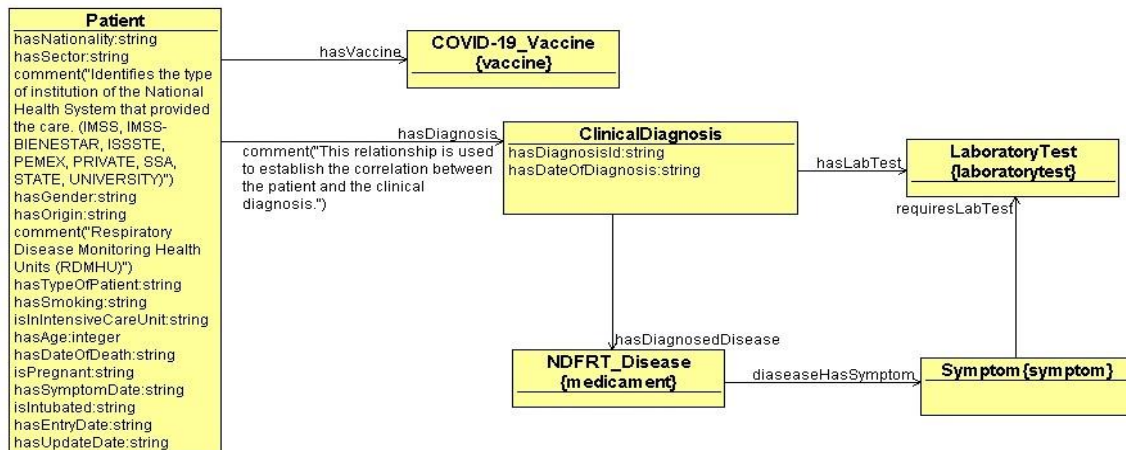


Figura 2. Grafos de conocimiento o módulos que integran el grafo de conocimiento *PatientElectronicRecord*. Fuente: Propia de los autores.

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
<i>ClinicalDiagnosis</i>	Expediente del diagnóstico clínico del paciente.
<i>Patient</i>	Datos del registro del paciente.
<i>StateFederative</i>	Estado federativo de la república mexicana.
<i>Municipality</i>	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
<i>hasOrigin</i>	Patient	integer	USMER de origen del paciente.
<i>hasSector</i>	Patient	string	Institución del Sistema de Salud que brindó la atención.
<i>hasGender</i>	Patient	string	Género del paciente.
<i>hasAge</i>	Patient	integer	Edad del paciente.

<i>hasTypeOfPatient</i>	Patient	string	Paciente ambulatorio u hospitalizado.
<i>hasNationality</i>	Patient	string	Nacionalidad del paciente.
<i>hasEntryDate</i>	Patient	string	Fecha de ingreso del paciente a USMER.
<i>hasSymptomDate</i>	Patient	string	Fecha de síntomas del paciente.
<i>hasSymptomUpdate</i>	Patient	string	Fecha de actualización de síntomas del paciente.
<i>hasDateOfDeath</i>	Patient	string	Fecha de deceso del paciente.
<i>hasSmoking</i>	Patient	string	El paciente fuma.
<i>isInIntensiveCareUnit</i>	Patient	string	Se encuentra en unidad de cuidados intensivos.
<i>isPregnant</i>	Patient	string	La paciente se encuentra embarazada.
<i>isIntubated</i>	Patient	string	El paciente esta intubado.
<i>hasAbbreviationState</i>	StateFederative	<u>string</u>	Abreviación del nombre de estado.
<i>hasKeyState</i>	StateFederative	string	Clave de estado.
<i>hasNameState</i>	StateFederative	string	Nombre de estado.
<i>hasState</i>	Municipality	string	Municipio pertenece a estado.
<i>hasNameMunicipality</i>	Municipality	string	Nombre del municipio.
<i>hasKeyMunicipality</i>	Municipality	string	Clave del municipio.

Fuente: Propia de los autores.

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
<i>belongsTo</i>	Municipality	StateFederative	Municipio pertenece a estado.
<i>hasDiagnosis</i>	Patient	ClinicalDiagnosis	Paciente tiene diagnóstico clínico.
<i>hasMunicipality</i>	StateFederative	Municipality	Estado federativo tiene municipio.
<i>hasResidenceMunicipality</i>	Patient	Municipality	Municipio de residencia.
<i>hasResidenceState</i>	Patient	StateFederative	Estado de residencia del paciente.
<i>hasStateFederativeBirth</i>	Patient	StateFederative	Estado de nacimiento del paciente.
<i>hasStateMedicalUnit</i>	Patient	StateFederative	Paciente tiene unidad médica.

Fuente: Propia de los autores.

Etapas 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.

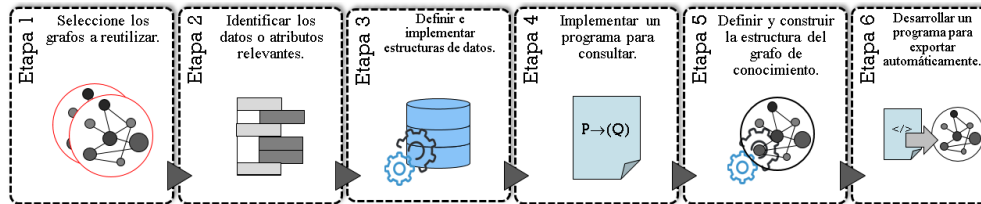


Figura 3. Etapas de método para reutilización de grafos.

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.

5.1 Seleccionar el grafo a reutilizar

Se seleccionó y descargó el grafo SYMPT⁷, 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”. Este 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

⁷ <https://bioportal.bioontology.org/ontologies/SYMP>

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 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.

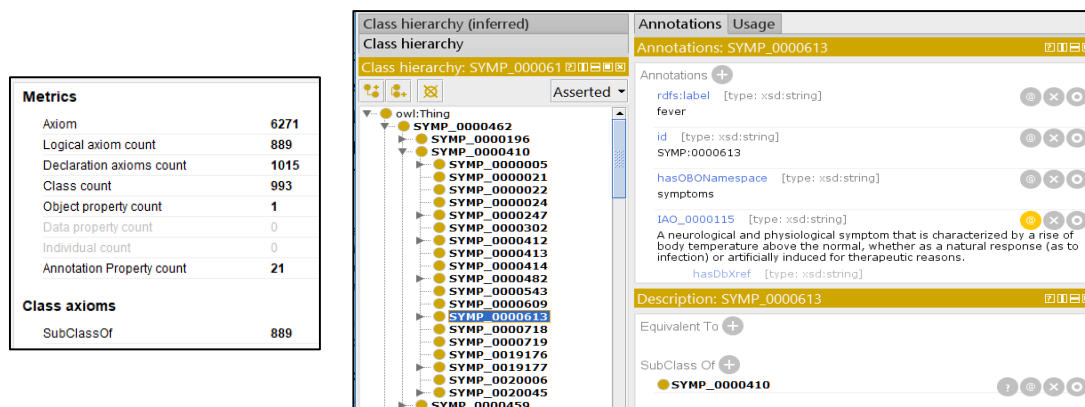


Figura 4. Visualización de la ontología SYMPT (con Protégé).
Fuente: Propia de los autores.

5.2 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 Turtle⁸. Para muestra de los datos y atributos de esta ontología, en la parte de abajo se muestra un

⁸ [https://dbpedia.org/page/Turtle_\(syntax\)](https://dbpedia.org/page/Turtle_(syntax))

extracto de la categoría de síntomas neurológicos y psicológicos, específicamente el concepto de **SYMP:0000613**, que se corresponde con el síntoma de la fiebre.

```
### 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 ;
  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/mwmedn1m?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.

5.3 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.

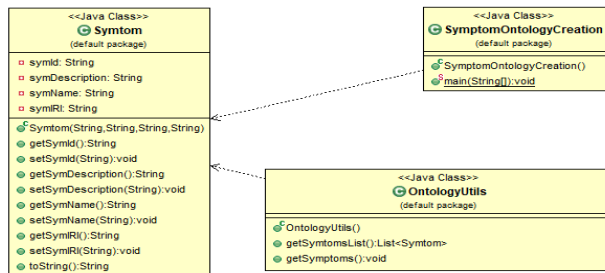


Figura 5. Diagrama de clases del programa que extrae los síntomas y los registra en el nuevo módulo.
 Fuente: Propia de los autores.

5.4 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 rdf:type owl:Class .
    ?s rdfs:subClassOf obo:SYMP_0000462 .
    ?s obo:IAO_0000115 ?description .
    ?s rdfs:label ?label .
    ?s oboInOwl:id ?id .}
    
```

5.5 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*.

5.6 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.

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

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*, *TerapeuticCategory* 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-19⁹. Como resultado el grafo *Vaccine* incluye las siguientes clases de vacunas: *COVID-19_Vaccine*, *mRNA_COVID-19_Vaccine* y *SARS-CoV-2_mRNA_Vaccine*.

Etapas 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

⁹ <https://www.gob.mx/cofepris/acciones-y-programas/vacunas-covid-19-autorizadas>

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
<i>diseaseHasSymptom</i>	NDFRT_Disease	Symptom	Enfermedad tiene síntoma.
<i>hasDiagnosedDisease</i>	ClinicalDiagnosis	NDFRT_Disease	Diagnóstico clínico tiene enfermedad.
<i>hasLabTest</i>	ClinicalDiagnosis	LaboratoryTest	Diagnóstico clínico tiene prueba de laboratorio.
<i>hasPharmalogicalTreatment</i>	ClinicalDiagnosis	PharmalogicalTreatment	Diagnóstico clínico tiene tratamiento clínico.
<i>hasVaccine</i>	Patient	COVID-19_Vaccine	Paciente tiene vacuna aplicada.
<i>requiresLabTest</i>	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 Manchester¹⁰. 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.

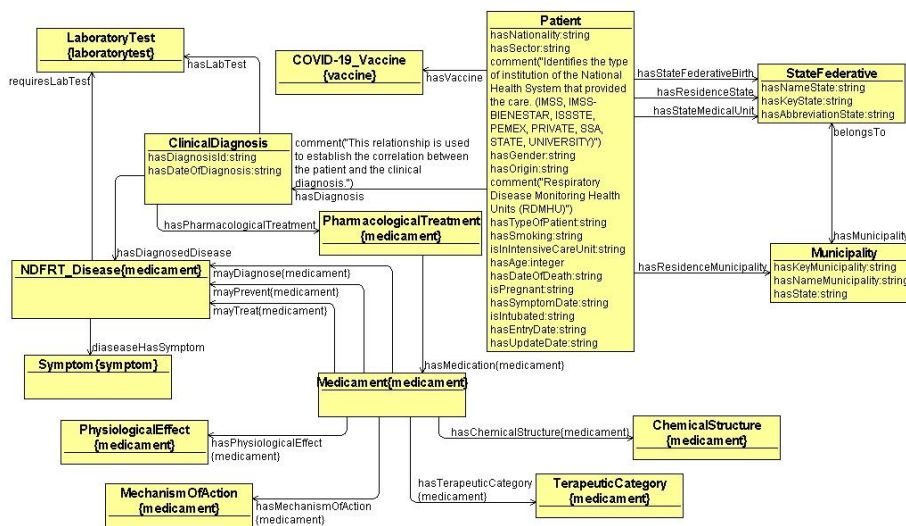


Figura 7. Grafo de conocimiento integrado *PatientElectronicRecord*.

Fuente: Propia de los autores.

¹⁰ <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.

Etapa 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).

Preparación de datos

Se descargó el archivo en formato CSV con los datos de pacientes. Este archivo cuenta con más de 7,000,000 de registros de pacientes atendidos en las Unidades de Salud Monitoras de Enfermedad Respiratoria viral (USMER). Con el propósito de realizar una prueba de la usabilidad del modelo, se utilizaron 500 registros, seleccionando aquellos de nacionalidad mexicana. Para llevar a cabo el registro se seleccionaron los atributos de los registros de pacientes que coinciden con los atributos definidos en el grafo general. El archivo de origen se muestra en la Figura 10.

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.

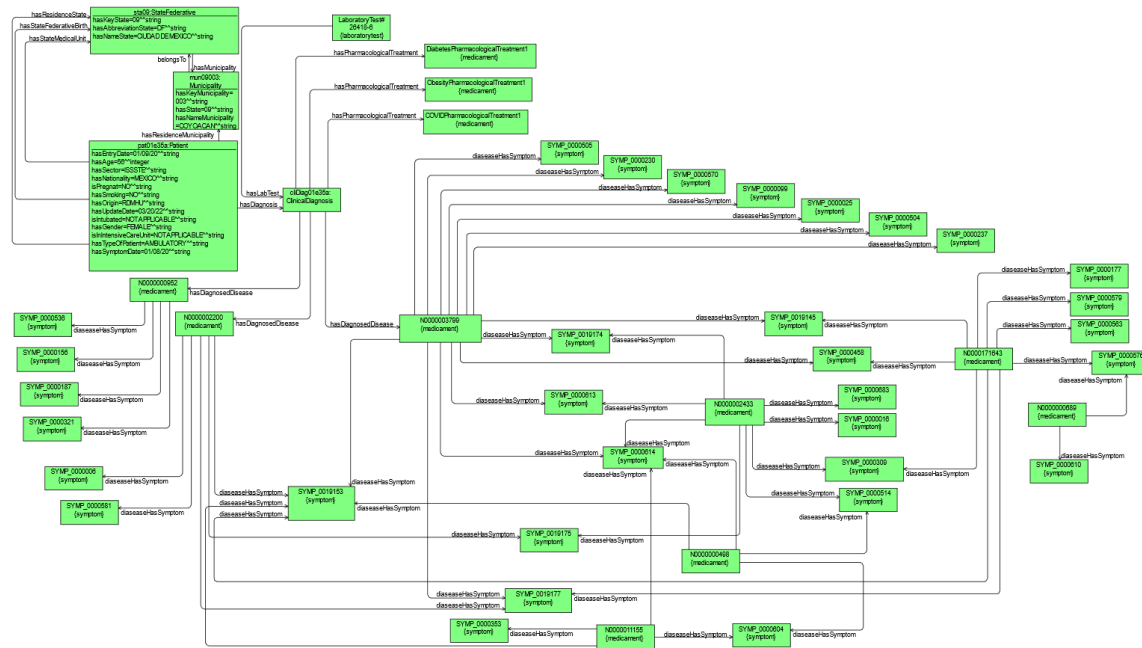


Figura 12. Ejemplo del resultado del registro de los datos de una paciente mexicana con COVID-19. Fuente: Propia de los autores.

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 Pellet¹² 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 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

¹² <https://www.w3.org/2001/sw/wiki/Pellet>

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:patz4e838	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, 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.

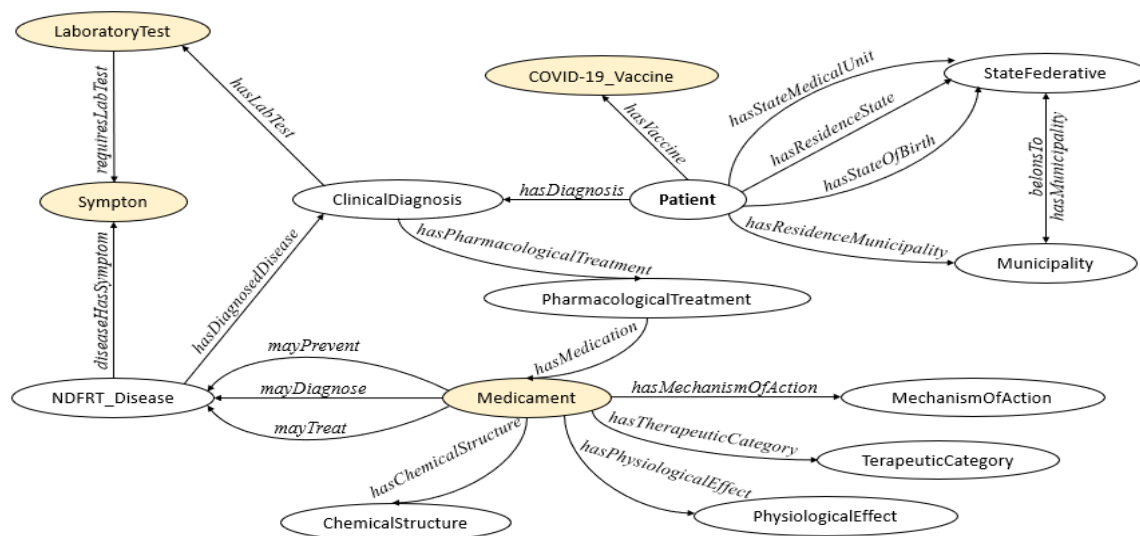


Figura 15. Grafo de conocimiento integrado con entidades y relaciones.
 Fuente: Propia de los autores.

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.

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

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.

Referencias Bibliográficas

- 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>
- 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.
- 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.
- 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>
- 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.
- 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>