

Un grafo de conocimiento para la visualización de información del sector salud en un dominio específico y priorizado en Colombia, basado en los estándares HL7.

Tesis de Grado

**Luis Miguel Martinez Vargas
Obeimar Anibal Arias Suarez**

**Director tesis
Helder Yesid Castrillón Cobo**

**CORPORACIÓN UNIVERSITARIA COMFACAUCA - UNICOMFACAUCA
Facultad de Ingenierías
Ingeniería de Sistemas
Popayán - Cauca
2021**

Página de Aceptación

Página dedicatoria

A nuestros padres, por su amor, trabajo y sacrificio en todos estos años, gracias a ustedes hemos logrado llegar hasta aquí y convertirnos en lo que somos. Ha sido el orgullo y el privilegio de ser sus hijos, son los mejores padres.

A todas las personas que nos han apoyado y han hecho que el trabajo se realice con éxito en especial a aquellos que nos abrieron las puertas y compartieron sus conocimientos.

Página de Agradecimientos

A nuestros docentes y en especial a nuestro tutor por su ayuda, paciencia y dedicación. Agradecerles también a todas nuestras familias por darnos ánimo durante este proceso. A nuestros amigos de toda la vida que nos acompañan desde siempre.

Contenido

RESUMEN	10
ABSTRACT	10
1. CAPÍTULO 1 INTRODUCCIÓN	11
1.1. FORMULACIÓN DEL PROBLEMA	11
1.2. JUSTIFICACIÓN	14
1.3. OBJETIVOS	15
1.3.1. <i>Objetivo general</i>	15
1.3.2. <i>Objetivo específico</i>	16
1.4. METODOLOGÍA.	16
1.4.1. <i>Fases de la metodología investigación acción</i>	16
1.5. ORGANIZACIÓN DEL PROYECTO	18
2. CAPÍTULO 2 CONCEPTUALIZACIÓN Y PRIORIZACIÓN DEL DOMINIO DEL SECTOR SALUD EN COLOMBIA.	19
2.1. MARCO TEÓRICO	19
2.2. MARCO LEGAL	23
2.3. MAPEO SISTEMÁTICO	24
2.3.1. <i>Método</i>	24
2.3.2. <i>Preguntas de investigación</i>	24
2.3.3. <i>Palabras clave</i>	25
2.3.4. <i>Criterios de inclusión y exclusión</i>	26
2.3.5. <i>Búsquedas en bases de datos.</i>	26
2.3.6. <i>Resultados.</i>	34
2.3.7. <i>Discusión</i>	36
2.3.8. <i>Conclusión mapeo sistemático</i>	38
2.4. IDENTIFICACIÓN Y PRIORIZACIÓN DE ÁREAS Y DOMINIOS DEL SECTOR SALUD EN COLOMBIA	39
2.5. CARACTERIZACIÓN DE CONCEPTOS Y ESTÁNDARES PARA LA VACUNACIÓN COVID-19	43
2.6. REVISIÓN DE ESTÁNDARES HL7	45
2.6.1. <i>Identificación de conceptos y estándares de hl7</i>	45
2.6.2. <i>Adopción del estándar FHIR</i>	53
2.6.3. <i>Conclusiones</i>	54
3. CAPÍTULO 3 CONSTRUCCIÓN DE GRAFO DE CONOCIMIENTO	56
3.1. CONSTRUCCIÓN DE LA ONTOLOGÍA	57
3.1.1. <i>Definición del dominio y alcance de la ontología</i>	57
3.1.2. <i>Reutilización de ontologías existentes</i>	57
3.1.2.1. <i>Vo vaccine ontology</i>	58
3.1.2.2. <i>Schema.org</i>	58
3.1.3. <i>Identificación de términos importantes para la ontología</i>	59
3.1.4. <i>Definición de clases y la jerarquía de clases</i>	61
<i>Se definen las principales clases y su respectiva jerarquía, relacionadas con el dominio principal de Vacunación:</i>	61
3.1.5. <i>Definición propiedades de las clases: slots</i>	62
3.1.6. <i>Definir las facetas de los slots</i>	63
3.1.7. <i>Crear instancias</i>	67
3.2. TRANSFORMACIÓN DE DATOS ESTRUCTURADOS A FORMATO RDF-OWL	67

3.3.	ELABORACIÓN DEL GRAFO DE CONOCIMIENTO A PARTIR DE DATOS ESTRUCTURADOS DEFINIDOS EN LA ONTOLOGÍA	69
3.4.	IMPORTANDO DATOS NO ESTRUCTURADOS AL GRAFO CREADO EN NEO4J	72
3.4.1.	<i>Recolección de datos (extracción de entidad)</i>	72
3.4.2.	<i>Importación de datos en neo4j</i>	73
3.5.	CONSULTAS AL GRAFO DE CONOCIMIENTO	75
3.5.1.	<i>Razonamiento del grafo</i>	77
3.6.	CONCLUSIONES	81
4.	CAPÍTULO 4: VALIDACIÓN DEL GRAFO DE CONOCIMIENTO	83
4.1.	DESARROLLO DEL PROTOTIPO DE VISUALIZACIÓN BASADO EN EL GRAFO	83
4.2.	ASPECTOS PARA LA VALIDACIÓN DE LA CALIDAD DE LOS DATOS EN EL GRAFO DE CONOCIMIENTO	95
4.3.	VALIDACIÓN DE LA CALIDAD Y EFECTIVIDAD DE LA VISUALIZACIÓN DE INFORMACIÓN	97
4.4.	RESULTADOS	99
4.4.1.	<i>Análisis de resultados encuesta de evaluación</i>	99
4.5.	DISCUSIÓN	110
4.6.	CONCLUSIONES Y TRABAJO FUTURO	111
	REFERENCIAS	113

LISTA DE FIGURAS

Imagen 1. Tendencia de Investigación de los últimos 5 años. Creación propia.....	36
Imagen 2 Estadísticas Vitales DANE. Defunciones Acumuladas 2020.	41
Imagen 3. Cobertura plan de vacunación de población para la inmunización del COVID-19	42
Imagen 4. Dimensión vida saludable y enfermedades transmisibles. Creación propia.	43
Imagen 5. Registro de la vacuna contra el COVID-19 en el carné de vacunación de adulto.....	44
Imagen 6. El estándar HL7 facilita la interoperabilidad.....	46
Imagen 7. Tecnologías emergentes año 2009 [70].....	50
Imagen 8. Gráfica de las tendencias en salud digital 2019 [73].	52
Imagen 9. Comparación de los estándares base de HL7 durante el tiempo. Creación Propia.	52
Imagen 10. Metodologías para la creación Grafos de Conocimiento. Creación Propia.....	56
Imagen 11. Jerarquía de clases. Creación Propia.....	62
Imagen 12. Interfaz principal de PROTÉGÉ.....	67
Imagen 13. Interfaz principal de PROTÉGÉ.....	68
Imagen 14. Importación de conceptos de ontologías externas.	69
Imagen 15. Grafo resultante de la importación de la ontología en Neo4j. Creación Propia	71
Imagen 16. Datos recolectados de encuesta desde EXCEL como archivo de formato CSV. Creación Propia	73
Imagen 17. Aplicación y uso de importación de datos y uso de lenguaje Cypher en Neo4j. Creación Propia.....	74
Imagen 18. Grafo resultante de toda la información importada. Creación Propia	75
Imagen 19. Consulta sobre efectos secundarios. Creación Propia	76
Imagen 20. Tabla de resultados de la consulta en Cypher. Creación Propia	76
Imagen 21. Consulta sobre creencias relacionadas a la vacuna de los vacunados. Creación Propia	77
Imagen 22. Etiquetas por cada grupo de nodos. Creación Propia	78
Imagen 23. Fórmula para el cálculo de la similitud de Jaccard.	79
Imagen 24. Ejemplo de cálculo de similitud de efectos secundarios de un vacunado específico con respecto a todos los vacunados. Creación Propia	80
Imagen 25. Ejemplo de cálculo de similitud de creencias de la vacuna de un vacunado con respecto a todos los vacunados. Creación Propia	80
Imagen 26 Comparación similaridades de Vacunados con respecto a Creencias. Creación Propia.	81
Imagen 27. Mockup de Interfaz Principal. Creación Propia	84
Imagen 28. Mockup prototipo de visualización. Creación Propia	84
Imagen 29. Mockup prototipo de visualización. Creación Propia	85
Imagen 30. Casos de uso. Creación Propia	86
Imagen 31. Diagrama de Componentes. Creación Propia.....	87
Imagen 32. Diagrama de Secuencia. Creación Propia.....	88
Imagen 33. Mockup prototipo de visualización. Creación Propia	89
Imagen 34. Interfaz de formulario. Creación Propia	89
Imagen 35. Interfaz de Visualización. Creación Propia	90

Imagen 36. Clase de Conexión al grafo de conocimiento. Creación Propia	91
Imagen 37. Parámetros de conexión a neo4j. Creación Propia	91
Imagen 38. Función Nodo, crea nodos en el grafo de conocimiento. Creación Propia.....	92
Imagen 39. Transacción que pasa datos como parámetros a la función nodo. Creación Propia	92
Imagen 40. Función de Similaridad. Creación Propia	92
Imagen 41. Función de similaridad en el grafo de conocimiento. Creación Propia	93
Imagen 42. Transacción de similaridad. Creación Propia.....	93
Imagen 43. Mapeo de los vacunados existentes en la visualización. Creación Propia.....	94
Imagen 44. Función de solicitud de información al grafo de conocimiento. Creación Propia	94
Imagen 45. Mapeo de datos seleccionado por el usuario en pantalla. Creación Propia	94
Imagen 46. Resultado Prototipo Pregunta 1. Creación Propia.	100
Imagen 47. Resultado Mi Vacuna Pregunta 1. Creación Propia.....	100
Imagen 48. Resultado Prototipo Pregunta 2. Creación Propia	101
Imagen 49. Resultado Mi Vacuna Pregunta 2. Creación Propia.....	101
Imagen 50. Resultado Prototipo Pregunta 3. Creación Propia	102
Imagen 51. Resultado Mi Vacuna Pregunta 3. Creación Propia.....	102
Imagen 52. Resultado Prototipo Pregunta 4. Creación Propia	102
Imagen 53. Resultado Mi Vacuna Pregunta 4. Creación Propia.....	103
Imagen 54. Resultado Prototipo Pregunta 5. Creación Propia	103
Imagen 55. Resultado Mi Vacuna Pregunta 5. Creación Propia.....	103
Imagen 56. Resultado Prototipo Pregunta 6. Creación Propia	104
Imagen 57. Resultado Mi Vacuna Pregunta 6. Creación Propia.....	104
Imagen 58. Resultado Prototipo Pregunta 7. Creación Propia	105
Imagen 59. Resultado Mi Vacuna Pregunta 7. Creación Propia.....	105
Imagen 60. Resultado Prototipo Pregunta 8. Creación Propia	105
Imagen 61. Resultado Mi Vacuna Pregunta 8. Creación Propia.....	106
Imagen 62. Resultado Prototipo Pregunta 9. Creación Propia	106
Imagen 63. Resultado Mi Vacuna Pregunta 9. Creación Propia.....	107
Imagen 64. Resultado Prototipo Pregunta 10. Creación Propia	107
Imagen 65. Resultado Mi Vacuna Pregunta 10. Creación Propia.....	107
Imagen 66. Resultado Prototipo Pregunta 11. Creación Propia	108
Imagen 67. Resultado Mi Vacuna Pregunta 11. Creación Propia.....	108
Imagen 68. Resultado Prototipo Pregunta 12. Creación Propia	109
Imagen 69. Resultado Mi Vacuna Pregunta 12. Creación Propia.....	109
Imagen 70. Resultado Prototipo Pregunta 13. Creación Propia	110
Imagen 71. Resultado Mi Vacuna Pregunta 13. Creación Propia.....	110

LISTA DE TABLAS

Tabla 1. Criterios de Inclusión y Exclusión.....	26
Tabla 2. Cadena de búsqueda para la pregunta 1.....	27
Tabla 3. Cadena de búsqueda para pregunta 2.	29
Tabla 4. Cadena de búsqueda para pregunta 3.	32
Tabla 5. Recursos de HL7 con algunos atributos.....	54
Tabla 6. Conceptos importados de VO Vaccine Ontology.	58
Tabla 7. Conceptos importados de Schema.org.....	59
Tabla 8. Términos usados para la construcción de Ontología.	59
Tabla 9. Clases y Jerarquía de Clases.....	61
Tabla 10. Propiedades de clases de ontología.	62
Tabla 11. Facetas de los Slots.	63

RESUMEN

Los grafos de conocimiento son usados para mejorar los resultados de búsqueda de la información recopilada de todo tipo de fuente, ofreciéndonos la información más importante respecto a una palabra dada o las relaciones que se pueden encontrar en una base de conocimiento, por tal motivo en este proyecto se propone un grafo de conocimiento para la visualización de registros de salud en un dominio específico priorizado en Colombia, basado definiciones del estándar HL7 FHIR.

ABSTRACT

Knowledge graphs are used to improve the search results of the information collected from all kinds of sources, offering us the most important information regarding a given word or the relationships that can be found in a knowledge base, for this reason in this This project proposes a knowledge graph for the visualization of health records in a specific domain prioritized in Colombia, based on definitions of the HL7 FHIR standard.

1. CAPÍTULO 1 INTRODUCCIÓN

1.1. Formulación del problema

El desarrollo actual de las tecnologías de la información y las telecomunicaciones (TIC), ha logrado el intercambio de información a una mayor velocidad a la que se conocía anteriormente. Además, recordemos que la portabilidad es la propiedad de los programas o aplicaciones informáticos que permite a sus usuarios acceder a ellos desde diferentes sistemas [1], de una forma más **técnica** según la ISO/IEC 25012 define la **portabilidad** como el *“Grado en el que los datos tienen atributos que les permiten ser instalados, reemplazados o eliminados de un sistema a otro, preservando el nivel de calidad en un contexto de uso específico.”* [2]. Con lo mencionado, el uso de la portabilidad se puede multiplicar el rendimiento en procedimientos de intercambio de datos, mejorando un acceso a sistemas y aplicaciones, también puede lograr con éxito que la participación de los sistemas de información sea más amplia.

Desde el 31 de enero del 2020 el gobierno nacional colombiano dirigido por el presidente Iván Duque aprobó la ley 215 de 2020 [3], medio por el cual se solicita la creación de una historia clínica electrónica interoperable, que tiene como objeto la estandarización de la historias clínicas electrónicas para permitir el registro integral y cronológico de las condiciones de salud de un paciente, además de ser administradas por sistemas de información y software con la capacidad de comunicarse e intercambiar información para brindar las herramientas necesarias para la utilización de estos datos; así a través de este medio se busca agilizar y garantizar, el acceso a los derechos a la salud y a la información de los usuarios Colombianos. Cabe resaltar que el área de salud en Colombia cuenta con muchas

subdivisiones que manejan grandes cantidades de datos, por lo cual se hace necesario la selección de un dominio en específico que permita hacer énfasis en el análisis de la información.

Por otra parte, con la implementación de la ley 2015 de 2020 se hace necesario la utilización de estándares con el cual se pretende lograr la optimización de los recursos, reducir costos en varios campos, y generar muchas ventajas competitivas que permitan desarrollarse con mayor facilidad en cualquier ámbito que se desee. El estándar **(HL7)** o **Health Level Seven** [4] reconocido a nivel mundial son un conjunto de estándares dirigidos a facilitar el intercambio electrónico de información clínica, y con visión para asegurar el acceso y utilización de los datos recolectados de forma adecuada, segura, cuando y donde se necesiten. También cabe mencionar que el HL7 incluye estrategias como: el desarrollo de nuevos estándares y sus metodologías, educar el sector de salud, promover su uso, estimular y colaborar con otras organizaciones que desarrollan estándares para un uso de mayor compatibilidad entre ellos mismo estando muy a favor de su aplicación en la ley mencionada anteriormente [5].

Los **grafos de conocimientos (Knowledge Graph)** son una base de conocimiento usada para mejorar todo resultado de búsqueda que los usuarios lleven a cabo en el sistema, con la información recopilada de todo tipo de fuente, ofreciéndonos la información más importante respecto a una palabra dada o las relaciones que se pueden encontrar con esta.

Debido a la falta de visualización de los registros médicos de forma portable con la implementación de la ley 2020 en los sistemas de diferentes hospitales, clínicas, centros de salud y EPS. Con lo mencionado anteriormente el problema se divide en tres enfoques; **a. Uso de diferentes formatos y estándares en las historias**

clínicas: Las diferentes entidades de salud en el país aún manejan formatos únicos y propios de forma física para almacenar las historias clínicas, e inclusive varios documentos relacionados al paciente. Cada entidad de salud maneja diferentes estándares y variables para su especificación de documentos clínicos, esto afecta mucho a los usuarios, un ejemplo es hacer el cambio de una historia clínica de una entidad de salud a otra, ya que para hacerlo tendría que crearla de nuevo en la entidad a la que esté interesado en ingresar.

b. La heterogeneidad de los datos dificulta su visualización: La cantidad de formatos y estándares hacen que la visualización para los usuarios o los mismos empleados de salud no sea la más óptima para su respectivo análisis. Es por ello que no hay una serie de variables y estándares que se maneje de forma global para todas las entidades sin afectar la visualización y análisis de los datos. Ejemplo, cuando los usuarios solicitan algún formato de un documento clínico en cualquier entidad de salud, diferencian que otras entidades generan el mismo documento con diferentes formatos y estándares.

c. La falta de visualización de los registros médicos de forma portable, esto sucede por las diferentes formas de almacenamiento que se pueden evidenciar en cada una de los entes de salud, debido a que los datos como historiales clínicos, radiografías, exámenes médicos, etc. Son almacenados de forma física y no de forma electrónica, imposibilitando que haya una visualización adecuada de toda la información manejada y perjudicando al análisis de la información. Concluyendo de forma más general al problema, se evidencia que existe una falta de visualización de las grandes cantidades de datos manejados por los entes médicos sin tener presente algún estándar o base de datos en común.

Este trabajo se basa en proponer el uso de un grafo de conocimiento, que se apoya en información recolectada para lograr visualizar registros electrónicos en salud apoyado en definiciones del estándar HL7 y en la experiencia dirigida a usuarios en general, con lo anterior se propone la siguiente pregunta de investigación:

¿Cómo mejorar la visualización de información para el sector salud, utilizando como base grafos de conocimiento y los estándares HL7?

1.2. Justificación

El número de clínicas que se disponen en el país aproximadamente es de 1800 tanto en el sector público como privado [6], cada una de las cuales manejan historias clínicas para los pacientes, algunas manejan historias clínicas almacenadas físicamente afectando la eficiencia y calidad en la atención a pacientes, por otro lado se manejan formatos digitales donde es necesario que se puedan compartir y comunicar diferentes formatos entre ellos que faciliten el acceso al personal médico (o administrativo) en cualquier lugar del país sin demoras en consultas y entregas de resultados clínicos, pero esto conlleva mucho tiempo para su implantación y requiere de diferentes métodos y/o estándares que faciliten la comunicación de datos. La portabilidad se conoce como el “Grado en el que los datos tienen atributos que les permiten ser instalados, reemplazados o eliminados de un sistema a otro, preservando el nivel de calidad en un contexto de uso específico.” [2], esta es una manera eficiente de hacer que los sistemas de información de los diferentes actores de salud puedan visualizar sus registros electrónicos en salud.

Es en ese punto donde la historia clínica electrónica entra a jugar un papel importante en la implantación de la interoperabilidad en el sistema de salud, la cual según la ley 2015 del 31 de enero del 2020 la define como “el registro integral y cronológico de las condiciones de salud del paciente, que se encuentra contenido en sistemas de información y aplicaciones de software con capacidad de comunicarse, intercambiar datos y brindar herramientas para la utilización de la información refrendada con firma digital del profesional tratante.” [3]. Siendo esta una opción muy eficiente en el momento de acortar esperas innecesarias y eliminar barreras geográficas. Pero aún falta agregar un elemento más para apoyar la portabilidad, y es el uso de grafos de conocimiento apoyados en librerías que sirven de insumo para la visualización de registros electrónicos de salud. Un grafo de conocimiento, es una forma sencilla de representar relaciones entre entidades, que permite establecer los vínculos semánticos entre datos y metadatos, y realizar razonamientos a través de reglas que recorren los nodos según relaciones semánticas que se determinen [7]. Esto permite que los datos sean representados más fácilmente y que se puedan integrar de forma que su consulta sea más coherente, se puede apoyar con estándares de salud como el HL7 en su versión 3, en el marco de su derivado FHIR en el uso de sus definiciones sobre salud.

1.3. Objetivos

1.3.1. Objetivo general

Proponer un grafo de conocimiento para la visualización de información del sector salud para un dominio específico y priorizado en Colombia, basado en los estándares de HL7.

1.3.2. Objetivo específico

- Caracterizar los conceptos y estándares de HL7 requeridos para la visualización de información del sector salud de un dominio específico y priorizado en Colombia.
- Construir un grafo de conocimiento para la visualización de información del sector salud bajo los conceptos previamente caracterizados.
- Validar la efectividad y calidad del grafo de conocimiento construido para la visualización de información, basada en la experimentación con un software utilizado en el sector salud.

1.4. Metodología.

Para el desarrollo del proyecto se siguieron las fases de acuerdo a la metodología investigación acción, con el fin de proponer una idea como solución al problema planteado, logrando obtener un efecto real sobre la situación [8] .

1.4.1. Fases de la metodología investigación acción

- **FASE DIAGNÓSTICO:** en esta fase, se analiza el origen y se delimita el problema, se identifica los principales objetivos y enfoques encontrando una relación entre ellos para obtener la posible solución. Entre las tareas que se desarrollaron: mapeo sistemático, en el cual se desarrollan las preguntas de investigación, para identificar que se ha realizado con respecto al problema a abordar, revisar el estado del arte del tema de investigación, lo que permitirá la selección y clasificación de estudios.

- **FASE DE PLANIFICACIÓN:** definiendo los objetivos y haber conocido el contexto del problema, en esta fase se llevan a cabo las actividades programadas para cumplir con los objetivos delimitados para implementar. Entre las tareas definidas se identificaron las áreas y dominios del sector salud, así como la revisión de los estándares HL7, para posteriormente adoptar términos del estándar. Otra de las tareas fue obtener ontologías que ya se encuentren especificadas y crear una ontología con los términos definidos del dominio de salud priorizado.
- **FASE DE ACCIÓN:** en esta fase se lleva a cabo la puesta en marcha el desarrollo del grafo de conocimiento. Entre las tareas definidas se realizó la construcción del grafo de conocimiento definido por la ontología previamente creada, en donde se especificó las herramientas para su desarrollo, la importación de la ontología y los datos necesarios para poblar el grafo de conocimiento.
- **FASE DE REFLEXIÓN:** en esta fase se interpreta y se obtiene conclusiones de los resultados generados por las actividades de acuerdo con los objetivos planteados en la fase de planificación. En esta fase se pretende validar el grafo de conocimiento propuesto con el desarrollo de un prototipo de visualización de datos, el cual haga uso del grafo de conocimiento. Con el prototipo desarrollado se valida la calidad del grafo de conocimiento, así como su efectividad, con la aplicación de un instrumento de validación en el cual se compara con otra aplicación del dominio en específico. Al final se recopilan los resultados obtenidos y se obtiene el análisis para cada uno de los requisitos de la validación.

1.5. Organización del proyecto

El siguiente documento se organiza de la siguiente forma:

- En el capítulo 1 se hace la introducción al tema a abordar al igual, la justificación del problema y se especifican los objetivos a alcanzar en la realización del proyecto.
- En el capítulo 2 se hizo la conceptualización del dominio específico en el sector salud en Colombia, también se realizó la investigación del estado del arte teniendo en cuenta el dominio específico y se definió los conceptos del estándar HL7 a adoptar
- En el capítulo 3 está dirigido al desarrollo de la ontología necesaria para la construcción del grafo de conocimiento, al igual que los requerimientos necesarios para la construcción del grafo de conocimiento, el uso de un algoritmo de grafos para el cálculo de la similitud.
- En el capítulo 4 se realizó la validación del grafo de conocimiento con el desarrollo de un prototipo, en cuanto a su calidad y efectividad se realizó mediante una encuesta además de la explicación de cómo se llevó a cabo el desarrollo para así obtener resultados y conclusiones.

2. CAPÍTULO 2 CONCEPTUALIZACIÓN Y PRIORIZACIÓN DEL DOMINIO DEL SECTOR SALUD EN COLOMBIA.

En este capítulo se espera abordar los conceptos teóricos relacionados con la investigación, también el estado actual de los trabajos relacionados con la representación de información basada en grafos de conocimiento, utilizados en el sector salud. Finalmente se hace la priorización de conceptos y estándares necesarios para abordar la problemática del proyecto.

2.1. Marco teórico

Historias clínicas electrónicas: con historias clínicas electrónicas se puede lograr un registro automático de los datos sociales, preventivos y médicos de un paciente, de lo cual se obtiene de forma directa o indirecta [9].

Portabilidad: Grado en el que los datos tienen atributos que les permiten ser instalados, reemplazados o eliminados de un sistema a otro, preservando el nivel de calidad en un contexto de uso específico [10].

Widget: los widgets son pequeñas aplicaciones cuyo objetivo es dotar de información visual y facilitar el acceso a las funciones que se utilizan de forma frecuente [11].

Visualización de Datos: La visualización de datos es la representación gráfica de información y datos. Al utilizar elementos visuales como cuadros, gráficos y mapas, las herramientas de visualización de datos proporcionan una manera accesible de ver y comprender tendencias, valores atípicos y patrones en los datos [12].

Estándar: es aquello que puede tomarse como referencia, patrón o modelo para la construcción y guía de objetos de la misma especie o categoría [13].

Health Level Seven (HL7): “HL7 es el acrónimo de Health Level Seven Internacional. Una organización sin ánimo de lucro dedicada a proporcionar estándares para la interoperabilidad en el ámbito de salud” [14].

HL7 v2: es un estándar de mensajería, el cual es el caballo de batalla del intercambio electrónico de datos en el ámbito clínico y posiblemente el estándar más implementado para la atención médica en el mundo. Este estándar de mensajería permite el intercambio de datos clínicos entre sistemas. Está diseñado para admitir un sistema central de atención al paciente, así como un entorno más distribuido donde los datos residen en sistemas departamentales [15].

HL7 V3 RIM: son un conjunto de especificaciones basadas en el modelo de información de referencia (RIM) de HL7, proporciona una fuente única que permite a los implementadores de las especificaciones V3 trabajar con el conjunto completo de mensajes, tipos de datos y terminologías. necesario para construir una implementación completa. Representa un nuevo enfoque para el intercambio de información clínica basado en una metodología impulsada por modelos que produce mensajes y documentos electrónicos expresados en sintaxis XML [16].

HL7 CDA: es un estándar de marcado de documentos que especifica la estructura y la semántica de los "documentos clínicos" con el fin de intercambiarlos entre los proveedores de atención médica y los pacientes. Un CDA puede contener cualquier tipo de contenido clínico; los documentos de CDA típicos serían un resumen de alta, un informe de imágenes, un informe de admisión y examen físico, de patología y más. El uso más popular es para el intercambio de información entre empresas [17].

HL7 FHIR (Fast Healthcare Interoperability Resources- Recursos de Interoperabilidad Sanitaria Rápida): es un estándar de interoperabilidad

destinado a facilitar el intercambio de información de atención médica entre proveedores de atención médica, pacientes, cuidadores, pagadores, investigadores y cualquier otra persona involucrada en el ecosistema de atención médica. Consta de 2 partes principales: un modelo de contenido en forma de "recursos" y una especificación para el intercambio de estos recursos en forma de interfaces RESTful en tiempo real, así como mensajes y documentos [18].

Mapa de conocimiento: base de conocimiento que de manera estructurada recibe, maneja, organiza y utiliza información o conceptos en el mundo físico y sus interrelaciones de forma semántica [19].

Entidad: "algo cuya existencia es reconocida por un sistema ontológico" [20].

Relación: "una relación es una correspondencia o conexión entre algo o alguien con otra cosa u otra persona"[21].

Datos semiestructurados: "No tienen un esquema definido. No encajan en un formato de tablas/filas/columnas, sino que se organizan mediante etiquetas o "tags" que permiten agruparlos y crear jerarquías. También se les conoce como no relacionales o NoSQL" [22].

Datos no estructurados: "generalmente son datos binarios que no tienen estructura interna identificable. Es un conglomerado masivo y desorganizado de varios objetos que no tienen valor hasta que se identifican y almacenan de manera organizada" [23].

RDF: "es un lenguaje para la representación de información sobre recursos que debe ser procesada por máquinas sin que esto suponga una pérdida de significado" [24] .

Arquitectura de Software: “consiste en un conjunto de patrones y abstracciones coherentes que proporcionan un marco definido y claro para interactuar con el código fuente del software [25].

Extracción de información: es un campo de trabajo perteneciente al ámbito del Procesamiento de lenguaje natural (Natural Language Processing (NLP), en inglés), que busca localizar cierta información en texto libre en un dominio determinado, ignorando otra información irrelevante. A saber: quién hizo qué a quién, dónde, cómo y cuándo. La Extracción de información se centra en derivar información estructurada a partir de un texto no estructurado [26].

NER: hace referencia a reconocimiento de entidad con nombre por sus siglas en inglés, también conocido por la extracción de entidades, es una tarea de extracción de información que busca localizar y clasificar en categorías, como personas, organizaciones, lugares, expresiones de tiempo y cantidades, las entidades nombradas encontradas en un texto [27].

Semántica: “es el estudio del significado de las palabras. A diferencia de la etimología, la semántica comprende un campo de relación entre palabras, formas y símbolos con el material real o abstracto que trata la elaboración de una frase” [28].

Ontología: es una termino que asimila y codifica el conocimiento, definiendo las relaciones existentes entre los diferentes conceptos de una determinada categoría o dominio [29].

Tuplas: “Una tupla es una secuencia de valores agrupados. Una tupla sirve para agrupar, como si fueran un único valor, varios valores que, por su naturaleza, deben ir juntos” [30].

Sistemas expertos: son los sistemas que almacenan conocimiento en un formato computable y que las usan para realizar operaciones en ellos [31].

Base de conocimiento: es un complejo de conocimientos, información y datos, en poder de un sistema experto [32].

2.2. Marco legal

Ley 2064 del 9 de diciembre del 2020: La presente ley tiene por objeto declarar de interés general la estrategia para la inmunización de la población colombiana contra la Covid-19 y establecer medidas administrativas y tributarias para la financiación y la gestión de los asuntos relacionados con la inmunización contra la Covid-19 y otras pandemias [33].

Decreto 109 del 29 de enero del 2021 “Por el cual se adopta el Plan Nacional de Vacunación contra el COVID -19 y se dictan otras disposiciones [34].

Ley estatutaria 1751 de 2015: La presente ley tiene por objeto garantizar el derecho fundamental a la salud, regularlo y establecer sus mecanismos de protección [35].

Ley 1581 de 2012: La presente ley tiene por objeto desarrollar el derecho constitucional que tienen todas las personas a conocer, actualizar y rectificar las informaciones que se hayan recogido sobre ellas en bases de datos o archivos, y los demás derechos, libertades y garantías constitucionales a que se refiere el artículo 15 de la Constitución Política; así como el derecho a la información consagrado en el artículo 20 de la misma [36].

Ley 2015 de 2020: La presente ley tiene por objeto regular la Interoperabilidad de la Historia Clínica Electrónica - IHCE, a través de la cual se intercambiarán los elementos de datos clínicos relevantes, así como los documentos y expedientes

clínicos del curso de vida de cada persona. A través de la Historia Clínica Electrónica se facilitará, agilizará y garantizará el acceso y ejercicio de los derechos a la salud y a la información de las personas, respetando el Hábeas Data y la reserva de la misma [3].

2.3. Mapeo sistemático

2.3.1. Método

Basado en lo propuesto por [37] y teniendo presente el objeto de estudio se realizó la formulación de tres preguntas de investigación (PI).

2.3.2. Preguntas de investigación

P1. ¿Como los grafos de conocimiento se han utilizado para la visualización de información?

P2. ¿Cómo visualizar información electrónica en salud de forma más usable?

P3. ¿Como se han usado los grafos de conocimiento para la representación de información en salud?

Con **P1** se busca identificar los distintos métodos con los cuales se han utilizado los grafos de conocimiento en el apoyo de la visualización de datos.

Con **P2** se pretende identificar los tipos de visualización más eficaces para la interpretación de información electrónica en salud. En **P3** se pretende identificar el uso de grafos de conocimiento en la representación de información en salud.

A partir de las preguntas de investigación se siguió la estrategia PICO (Population, Intervention, Comparison and Outcomes) sugerida por [38] para identificar palabras clave y formular cadenas de búsqueda a partir de preguntas de investigación.

Población: En nuestro contexto, la población es la información de los pacientes relacionada con el tratamiento de EHR, incluye los diferentes pacientes vinculados a una entidad promotora de salud (EPS).

Intervención: La intervención está dirigida a la metodología, herramienta, tecnología o procedimiento que aborda el problema. En el contexto de este estudio, los procesos para visualización de información, estándares para manejo de información electrónica en salud y construcción de grafos de conocimientos son el objeto de investigación.

Comparación: Se revisan los diferentes modelos que se han usado en el ámbito de los EHR, con el fin de analizar las diferentes estrategias que se han implementado, la forma de representación de los datos, el contenido de estos y la forma de ver los resultados.

Resultados: Los resultados se relacionan con factores de importancia para la población involucrada. En nuestro caso como resultado tenemos la representación de información médica en salud de una forma visualizada y de fácil acceso para sus usuarios.

2.3.3. Palabras clave

Se identificaron y agruparon las palabras clave teniendo en cuenta sinónimos y abreviaturas:

- Población: Electronic Health Record OR "electronic health record" OR "public health record" OR "personal health record" OR health information OR "Clinical knowledge graph"
- Intervención: visualization OR knowledge graph OR hl7 OR knowledge graph application OR linked data OR ontology
- Comparación: Render OR Rendering OR Content OR representation
- Resultados: widget OR "web app" OR application* OR build OR "web application"

2.3.4. Criterios de inclusión y exclusión

A continuación, se presentan los criterios que fueron tenidos en cuenta para incluir y excluir documentos.

Tabla 1. Criterios de Inclusión y Exclusión.

Inclusión	Exclusión
Últimos 5 años	Libros y Literatura Gris.
Artículos, revisiones	No se relaciona con Registros clínicos electrónicos (EHR). No se relaciona con grafos de conocimientos.
Estudios en campos de la ingeniería, medicina, biomedicina y computación.	Estudios no presentados en inglés

2.3.5. Búsquedas en bases de datos.

Las palabras identificadas fueron consideradas para construir tres cadenas de búsqueda, y posteriormente fueron incluidas en los buscadores de las bases de datos bibliográficas Scopus y Google Scholar, obteniendo los siguientes resultados.

Tabla 2. Cadena de búsqueda para la pregunta 1.

Base de datos	Búsqueda	Nº Resultados
Scopus	TITLE-ABS-KEY ("knowledge graph") AND TITLE-ABS-KEY (visualization OR render OR rendering OR content OR widget OR "web app" OR application* OR build) AND PUBYEAR > 2016 AND PUBYEAR < 2022 AND (LIMIT-TO (OA , "all")) AND (LIMIT-TO (SUBJAREA , "COMP") OR LIMIT-TO (SUBJAREA , "ENGI") OR LIMIT-TO (SUBJAREA , "MEDI") OR LIMIT-TO (SUBJAREA , "BIOC"))	393

De acuerdo con el mapeo sistemático se evidenció que, para la **P1**, se encontraron 416 artículos de los cuales por medio de criterios de inclusión se seleccionaron 11 artículos. Según los diferentes autores, se han utilizado los grafos de conocimiento (KG) para permitir la visualización de información con el uso de interfaces, apoyando la extracción de información ayudando a usuarios a mejorar la interpretación de los datos consultados [39]. Con el apoyo de los KG a través de los diferentes tipos de interfaces manejados (*Plantillas o Widgets*), se pueden realizar investigaciones, análisis, consultas, etc. mediante la interpretación de información obtenida, beneficiando a médicos, usuarios e investigadores [40].

En [41] los grafos de conocimientos son construidos utilizando una ontología que permite describir conceptos y relaciones con respecto a la infraestructura de detección, la procedencia, las actividades de adquisición de datos, los indicadores y las propias entidades de la ciudad, permitiendo serializar la información en formato CSV y así generar paneles (widget) para el análisis de los metadatos recolectados de diferentes zonas de una ciudad.

En [42] se implementa un prototipo de aplicación web, con el uso de ALOHA,

el cual es un grafo de conocimiento que ayuda a los usuarios a buscar información relacionada con suplementos dietarios. El prototipo tiene una interfaz en la cual los usuarios pueden consultar información de suplementos dietarios a través del uso de preguntas preestablecidas y ver en pantalla toda la información relacionada de acuerdo a la pregunta.

En [43] aporta una interfaz que permite al usuario consultar rápidamente en un grafo de conocimiento sin la necesidad de comandos o código de computadora, permitiéndole a los usuarios crear consultas en un formato fácil de usar, almacenar los resultados, clasificar la relevancia de las consultas. y explorar gráficamente los resultados.

En [44] muestra cómo el análisis visual puede ayudar a los usuarios a buscar atributos al permitir que los usuarios exploren visualmente el conjunto de datos disponibles y al servir como una interfaz para la construcción de consultas. Proporciona visualizaciones del propio gráfico de conocimiento para ayudar a los usuarios a comprender uniones complejas.

En [45] se crea una interfaz de usuario para la búsqueda de información relacionada con suplementos dietéticos de un grafo de conocimiento de iDsisk que contiene información sobre suplementos dietéticos, la base de conocimiento está almacenada en Neo4j. Es la tercera iteración del trabajo con mejoras de usabilidad, en la cual se han incorporado nuevas funcionalidades con respecto a otras iteraciones y como resultado se ha obtenido un mejor desempeño y uso de la aplicación web. [46] se hace la creación de un sistema que permite a las personas una mejor manipulación de tablas con datos obtenidos desde los grafos de conocimientos y su debida visualización de los

resultados.

En [47] los grafos de conocimiento son utilizados para la recolección de información de Medicina Tradicional China (TCM). Mediante la aplicación de un análisis profundo a la extracción de texto, reconocimiento de entidades, extracción de atributos, extracción de eventos, etc. Con la ayuda de una ontología se logra el reconocimiento de entidades de los datos no estructurados alineando este conocimiento al gráfico de conocimiento permitiendo proporcionar una referencia auxiliar a diagnósticos y tratamientos de médicos en medicina china a través de una plataforma.

En [48] se menciona la gran cantidad de grafos de conocimientos que se pueden encontrar en la web, pero que son realmente difíciles de consultar por el motivo de que se debe descargar al ordenador, por ello se trabaja un marco para analizar las interfaces de publicaciones de datos vinculados (grafos de conocimiento) que se encuentran en la red. [49] se usan los gráficos de conocimientos para respaldar consultas mediante una interfaz.

Tabla 3. Cadena de búsqueda para pregunta 2.

Base de datos	Búsqueda	Nº Resultados
Scopus	TITLE-ABS-KEY (visualization OR representation OR render OR "web application") AND TITLE-ABS-KEY ("medical health record" OR "electronic health record" OR "public health record" OR "personal health record" OR hl7) AND PUBYEAR > 2016 AND PUBYEAR < 2022 AND (LIMIT-TO (OA , "all")) AND (LIMIT-TO (DOCTYPE , "cp") OR LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (SUBJAREA , "COMP") OR LIMIT-TO (SUBJAREA , "ENGI"))	183

Para la **P2**, se encontraron 183 artículos, de los cuales por medio de criterios de inclusión se seleccionaron 8 artículos. En el contexto de visualización de información electrónica en salud, los diferentes autores muestran la necesidad de llevar la información recolectada de diferentes registros médicos con distintos formatos, a una forma que permita la visualización, análisis, usos y otros métodos para la fácil interpretación de dicha información a sus usuarios. El artículo [50] es un claro ejemplo de la necesidad de interpretar los diferentes datos que son encontrados en medicina, de diferentes pacientes o conocimientos médicos para su acceso y consulta rápida. Para esto se lleva a desarrollar una metodología y herramienta que permita su interpretación y análisis. Por otra parte, también está [51], donde para muchos investigadores la usabilidad de los datos EHR son obstaculizados por la falta de experiencia en programación o por familiaridad con la estructura de una base de datos EHR. Por lo cual se lleva a desarrollar una aplicación que permita manipular y consultar los datos de una forma más eficaz.

En [52] se presenta una metodología y la implementación de un marco de gestión de datos con el objetivo de apoyar los servicios de atención del paciente aprovechando una ontología basada en recurso HL7 FHIR. En [53] se propone un espacio de diseño de ocho dimensiones para guiar a los futuros diseñadores de aplicaciones a acceder a la historia clínica de un paciente. Las dimensiones de nuestro espacio de diseño provienen del método utilizado para recopilar información. Luego, ubicamos algunos sistemas existentes en comparación con las dimensiones de nuestro espacio de diseño para identificar las carencias en la visualización de registros médicos electrónicos como potencial para nuestro trabajo futuro. En [54] se presenta una visualización

técnica que segmenta los historiales de los pacientes en lugar de tratarlos como secuencias de eventos sin procesar, agrega los segmentos utilizando criterios como el historial completo o combinaciones de tratamientos, y luego visualiza los segmentos agregados como cuadros de mando estáticos que se organizan en una red de cuadros de mando para mostrar cambios longitudinales. La investigación representa un cambio radical en el detalle de los datos a gran escala que pueden visualizarse con éxito utilizando paneles de control y proporciona una guía sobre cómo se puede generalizar el enfoque. En [55] se menciona la adopción del software de soporte de decisiones clínicas que produce un gran volumen de datos, lo que brinda oportunidades para la investigación y para visualizar y analizar mejor las tendencias en estos datos, se creó un tablero utilizando una plataforma de visualización de datos disponible comercialmente (Tableau, Seattle, WA). Esto permitió la visualización en tiempo real de los datos generados por el software. La integración de herramientas de visualización de datos en herramientas de apoyo a la toma de decisiones clínicas permite un análisis de datos más sencillo y puede agilizar la investigación y los esfuerzos de mejora de la calidad. En [56] se propone una novedosa plataforma sanitaria móvil para la visualización y gestión de los informes médicos de los pacientes, denominada MyHealthCloud. La investigación ofrece un nuevo enfoque impulsado por el paciente para almacenar, recuperar y compartir informes médicos para pacientes y médicos. Esta nueva plataforma maximiza los beneficios de la tecnología de salud móvil al proporcionar una mejor manera para que los profesionales de la salud compartan información con sus pacientes de manera eficiente y efectiva. En [57] el objetivo de su estudio es diseñar, desarrollar y

evaluar una interfaz de visualización de código abierto y fácil de usar para explorar la especificación del estándar HL7 FHIR, donde se crea un prototipo de una herramienta de visualización interactiva para navegar y manipular los recursos, perfiles y extensiones principales de FHIR, demostrando que las técnicas de visualización son útiles para navegar por la especificación mencionadas y ayudar a su creación de perfiles.

Tabla 4. Cadena de búsqueda para pregunta 3.

Base de datos	Búsqueda	Nº Resultados
Scopus	TITLE ("knowledge graph" OR "knowledge graph application") AND TITLE-ABS-KEY ("medical health record" OR "electronic health record" OR "public health record" OR "personal health record" OR "linked data" OR "ontology" OR "Clinical knowledge graph") AND PUBYEAR > 2016 AND PUBYEAR < 2022 AND (LIMIT-TO (OA , "all")) AND (LIMIT-TO (DOCTYPE , "cp") OR LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (SUBJAREA , "COMP") OR LIMIT-TO (SUBJAREA , "ENGI"))	79

En la **P3**, se encontraron 79 artículos, de los cuales por medio de criterios de inclusión se seleccionaron 5 artículos donde los grafos de conocimientos (KG) se han usado como apoyo a la representación de información en salud. Debido a su fácil representación de la información, permite a las personas utilizar un sistema de apoyo para resultados más precisos. en [58] se usan los grafos de conocimientos a partir de datos recolectados de registros médicos en salud HCE, para proporcionar información para un análisis más útil en la similitud de pacientes. De igual importancia en [59] se realiza una recolección de información de una base de datos médica para construir un grafo de

conocimiento para recomendación en recetas médicas adecuadamente o advertencias de sus efectos secundarios mediante una interfaz interpretable por los profesionales de la salud.

En [60], capturan las relaciones mediante la construcción de un grafo heterogéneo grande y de alta calidad que vincula a pacientes, enfermedades y fármacos (PDD) en EMR. Específicamente, proponen un marco novedoso para extraer entidades médicas importantes de MIMIC-III (Centro de Información Médica para Cuidados Intensivos III) y vincularlas automáticamente con los grafos de conocimiento biomédico existentes, incluyendo la ontología ICD-9 y DrugBank. El grafo PDD presentado en este documento es accesible en la Web a través del punto final SPARQL, aplicaciones, como recomendaciones de tratamiento eficaces.

En [61] se trabaja, diseña e implementa una plataforma de preguntas y respuestas basada en grafos de conocimiento (KnowHealth) para administrar esos datos de salud bajo el dominio de la enfermedad envejecida. Basándose en ontologías, rastrean y extraen información relacionada con la salud de diferentes fuentes para construir un gráfico de conocimiento, donde finalmente completan una plataforma de respuesta a preguntas impulsada por el comportamiento histórico para atender consultas para personas mayores y dar respuestas con mayor precisión.

[41] tiene como objetivo permitir a los investigadores biomédicos acceder, consultar y dar sentido a la literatura relacionada con COVID-19. Para ello, adapta, combina y amplía herramientas para procesar, analizar y enriquecer el “COVID-19 Open Research Dataset” (CORD-19) que reúne más de 50.000

artículos científicos de texto completo relacionados con los coronavirus. Informamos sobre el conjunto de datos RDF y los recursos de software producidos en este proyecto aprovechando las habilidades en representación del conocimiento, extracción de texto, datos y argumentos, así como visualización y exploración de datos.

2.3.6. Resultados.

P1. ¿Como los grafos de conocimiento se han utilizado para la visualización de información?

Para la pregunta P1, de los artículos 393 resultados de la búsqueda por criterios de exclusión e inclusión, de los artículos seleccionados de la búsqueda que en total fueron 11 el cual representa el 2.8%, aportan a la pregunta en que los grafos de conocimiento apoyan la visualización con el uso de interfaces, lo cual apoya a la extracción de información, esto a su vez apoyado de elementos como preguntas que ayudan a los usuarios a gestionar la información que desean visualizar en la interfaz.

P2. ¿Cómo visualizar información electrónica en salud de forma más usable?

De los 183 artículos resultantes de la cadena de búsqueda elaborada para responder la pregunta 2 y mediante los criterios de inclusión y exclusión solo el 4.4% de los artículos fueron seleccionados, donde cada uno de los autores dan a conocer la necesidad de visualizar sus diferentes registros médicos almacenados con distintos formatos permitiendo una mayor análisis de la información, cada uno de los artículos propone una metodología y/o una

herramienta existente o desarrollada por ellos mismo, para cumplir con su objetivo de investigación.

P3. ¿Cómo se han usado los grafos de conocimiento para la representación de información en salud?

Como resultados a nuestra tercera cadena de búsqueda obtuvimos 79 artículos, de los cuales por medio de criterios de inclusión y exclusión se seleccionaron solo el 6.3% de esos artículos, donde sus respectivas investigación y desarrollo proponen que los grafos de conocimientos (KG) son usados como apoyo a la representación de información en salud. La utilización de los grafos de conocimiento, es realizado mediante la recolección y administración de datos de diferentes fuentes de salud en ontologías que permiten que la información sea de mayor relevancia para generar una fuente de datos en el grafo de conocimiento que brinde resultados más útiles y precisos para todo tipo de usuario.

Durante el mapeo sistemático se evidencian numerosos artículos que se tomaron como antecedentes relacionados con el tema de investigación del presente proyecto, también se pudo evidenciar que el campo de los grafos de conocimiento ha estado en desarrollo durante los últimos años y sus diferentes contribuciones en los contextos han sido de gran ayuda para lograr un análisis más detallado de la información. En la siguiente imagen se muestran los artículos obtenidos por año, en un intervalo de 2017-2021 (5 años), en el último año se evidencia pocos resultados, pero esto es debido a que los temas de investigación se encuentran en una curva de crecimiento.

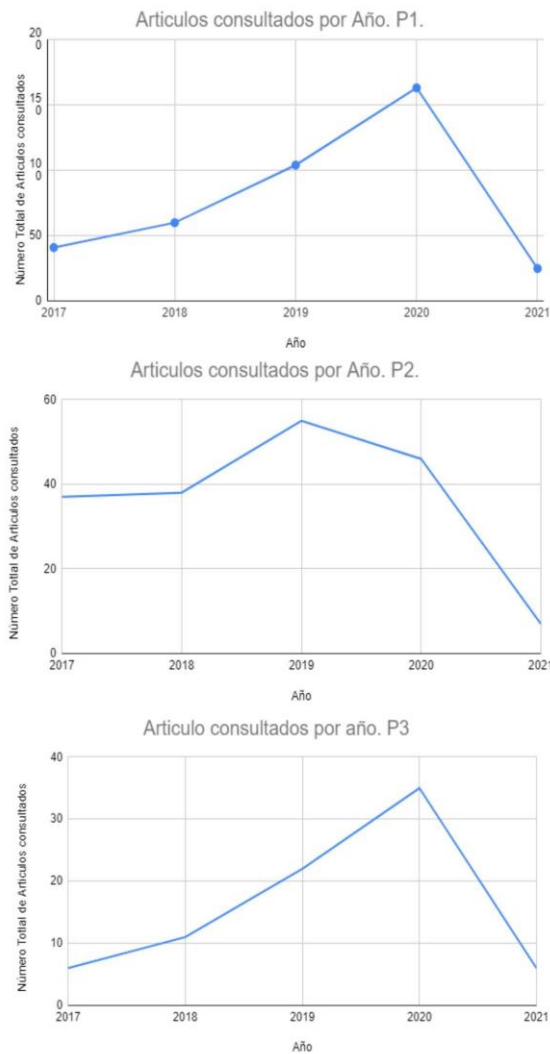


Imagen 1. Tendencia de Investigación de los últimos 5 años. Creación propia

2.3.7. Discusión

A partir de los aportes y brechas definidas en las categorías, mecanismos y estrategias analizados en el presente estudio, se genera un escenario base a nivel teórico que responde de la siguiente manera a las temáticas relacionadas con las preguntas de investigación planteadas.

P1, grafos de conocimiento y visualización de información: La visualización de información está apoyada con la creación de prototipos o interfaces aplicados con diferentes estándares, permitiendo la invención de

nuevas plataformas o aplicaciones más rápidas y fáciles de entender. Además, con el apoyo de los grafos de conocimiento se tiene un valor agregado en el análisis de la información. No obstante, se recomienda tener en cuenta evaluaciones de usabilidad para revelar posibles problemas que se encuentren relacionados con interfaces o plataformas. Otro de los inconvenientes se evidencia en la creación de grafos de conocimiento, debido a que algunos autores la definen como costosa y su elaboración toma tiempo. Pero con procesos definidos y diferentes herramientas se puede llegar a obtener resultados que benefician su uso de muchas maneras. En este sentido se debe considerar a futuro, investigaciones relacionadas con la búsqueda de información en escenarios de uso atendidos por sistemas o aplicaciones de análisis visual.

P2, visualización de información electrónica en salud: La visualización de la información electrónica en salud beneficia la utilización de estos datos en un tiempo más real, permitiendo a médicos optimizar su tiempo mientras gestionan grandes cantidades de datos generados por pacientes. En cuanto al uso de información, se deben tener en cuenta las herramientas necesarias para que exista una interoperabilidad completa entre los diferentes formatos manejados en las diferentes dimensiones médicas. También se hace importante la incorporación de mecanismos para manejar los flujos de datos de forma más optimizada, ya que entre más datos se estén procesando, más cantidad de recursos se utilizan en procesar dicha información. Además, invertir en mecanismos para consultas fáciles y flexibles de los diferentes datos basados en la semántica y la calidad de los datos vinculados, es algo que se debe tener en cuenta para facilitar el análisis y visualización de los datos.

P3, grafos de conocimiento y representación de información en salud: Se han encontrado diferentes trabajos con implementación de los de grafos de conocimiento en la representación de información relacionada en campos de salud o medicina, donde se puede observar los beneficios para los médicos, investigadores y usuarios en la implementación de los grafos de conocimiento para visualizar los diferentes formatos recogidos de los diferentes registros médicos en salud. La visualización permite una fácil consulta de los datos obtenidos de los grafos de conocimiento sin tener experiencia en los lenguajes de consulta, haciendo así más fácil la interpretación de la información manejada en salud obtenida por la consulta.

2.3.8. Conclusión mapeo sistemático

Se puede evidenciar que los grafos de conocimiento tienen un gran aporte dentro de los distintos sistemas que se pueden encontrar hoy en día, como es el caso de la visualización de información en distintos contextos, ya que los grafos de conocimiento brindan una mejor forma de consultar y visualizar información permitiendo que se tomen mejores decisiones. Dentro de este campo, en el uso de grafos de conocimiento se han desarrollado varias plataformas, sistemas y herramientas que han potencializado la usabilidad en cuanto a permitir un mejor análisis y uso de la información brindada hacia los diferentes usuarios o empresas.

En el campo de la salud, la gran cantidad de datos que se manejan dificultan mucho la consulta y análisis de la información generada de cada uno de los pacientes, causando pérdida de tiempo en investigaciones, atención a pacientes y demás procesos de la salud, por esto la visualización y consulta

de información con el apoyo de herramientas, se ha estado implementando cada vez más tratando de eliminar este problema en el sector salud.

Por lo tanto, se espera que, con el uso de grafos de conocimiento, la representación de información electrónica en salud se beneficie en la velocidad y manera de consultas la información de forma más eficiente, mejorando la capacidad de realizar análisis rápidos y por ende tomar mejores decisiones frente a las diferentes situaciones necesarias, incluso en entornos en los que se requiera interoperabilidad entre sistemas.

2.4. Identificación y priorización de áreas y dominios del sector salud en Colombia

En Colombia, de acuerdo con la Ley 1122 de 2007 [62], la salud pública está constituida por un conjunto de políticas que busca garantizar de manera integrada, la salud de la población por medio de acciones dirigidas tanto de manera individual como colectiva. De acuerdo con el sector de salud pública dividido en: Salud ambiental, salud sexual y reproductiva, estilos saludables, salud mental, entre otras. Además, se cuenta con el plan decenal de salud pública 2012-2021 dividido en 8 dimensiones como vida saludable y enfermedades transmisibles, vida saludable y enfermedades No transmisibles, salud sexual y reproductiva entre otras, orientadas a mejorar las condiciones de vida de los ciudadanos [63]. Con esta información encontrada se analizó y se llevó a representar la salud pública en Colombia de la siguiente forma.

Ver Anexo 1. División de Dimensiones Salud Pública en Colombia.

Sin embargo, para fines del proyecto es conveniente priorizar un solo dominio

de información debido a la cantidad de información que estos manejan. Para el proyecto se tiene en cuenta el dominio de vacunación en el área de COVID19. A continuación se nombran las dimensiones priorizadas en Colombia según estadísticas del DANE donde se menciona que, en el año 2020, el COVID-19 confirmado se convirtió en la primera causa de defunción en Colombia con un total de 50.071 casos, que corresponden al 16,9% del total de los fallecimientos ocurridos en el país, siendo más afectados los hombres con el 64,3% de los casos. En segundo lugar y con 10.597 defunciones de diferencia le siguen los infartos agudos del miocardio con 39.474 casos que representan el 13,3%, que al igual que con el COVID-19 confirmado, también afecta en mayor proporción a los hombres con el 55,4% de los casos. El tercer lugar es ocupado por las defunciones por COVID-19 sospechoso que acumulan 13.291 defunciones que equivalen al 4,5% del total de defunciones en 2020 afectando en mayor proporción a los hombres con un 56,1% del total de los casos [64]. En la siguiente imagen se hace mención de las estadísticas mencionadas anteriormente brindadas por el DANE.

Código CIE-10	Causa de defunción a 4 dígitos CIE-10 v2019	Total nacional			Hombres			Mujeres		
		Número	%	Tasa*	Número	%	Tasa*	Número	%	Tasa*
	Total nacional	296.800	100%		169.252	100,0%		127.446	100,0%	
U071	COVID-19 confirmado	50.071	16,9%	99,4	32.204	19,0%	130,9	17.867	14,0%	69,3
I219	Infarto agudo del miocardio, sin otra especificación	39.474	13,3%	78,4	21.882	12,9%	89,0	17.590	13,8%	68,2
U072	COVID-19 sospechoso	13.291	4,5%	26,4	7.452	4,4%	30,3	5.839	4,6%	22,7
J449	Enfermedad pulmonar obstructiva crónica (EPOC), no especificada	5.830	2,0%	11,6	3.158	1,9%	12,8	2.672	2,1%	10,4
C169	Tumor maligno del estómago, parte no especificada	4.705	1,6%	9,3	2.842	1,7%	11,6	1.863	1,5%	7,2
X954	Agresión con disparo de otras armas de fuego	4.298	1,4%	8,5	4.058	2,4%	16,5	240	0,2%	0,9
I10	Hipertensión esencial (primaria)	4.221	1,4%	8,4	1.934	1,1%	7,9	2.287	1,8%	8,9
C349	Tumor maligno de los bronquios o del pulmón, parte no especificada	4.196	1,4%	8,3	2.470	1,5%	10,0	1.726	1,4%	6,7
J189	Neumonía, no especificada	4.008	1,4%	8,0	2.163	1,3%	8,8	1.845	1,4%	7,2
J440	EPOC con infección aguda de las vías respiratorias inferiores	3.799	1,3%	7,5	2.069	1,2%	8,4	1.730	1,4%	6,7
C509	Tumor maligno de la mama, parte no especificada	3.456	1,2%	6,9	27	0,0%	0,1	3.429	2,7%	13,3
C61	Tumor maligno de la próstata	3.298	1,1%	6,5	3.298	1,9%	13,4	0	0,0%	0,0
I110	Enfermedad cardíaca hipertensiva con insuficiencia cardíaca (congestiva)	2.999	1,0%	6,0	1.353	0,8%	5,5	1.646	1,3%	6,4
E146	Diabetes mellitus, no especificada con otras complicaciones especificadas	2.853	1,0%	5,7	1.303	0,8%	5,3	1.550	1,2%	6,0
C189	Tumor maligno del colon, parte no especificada	2.664	0,9%	5,3	1.233	0,7%	5,0	1.431	1,1%	5,6
I64	Accidente vascular encefálico agudo, no especificado	2.664	0,9%	5,3	1.290	0,8%	5,2	1.374	1,1%	5,3
N390	Infección de vías urinarias, sitio no especificado	2.642	0,9%	5,2	1.104	0,7%	4,5	1.538	1,2%	6,0
I619	Hemorragia intraencefálica, no especificada	2.479	0,8%	4,9	1.257	0,7%	5,1	1.222	1,0%	4,7
I639	Infarto cerebral, no especificado	2.142	0,7%	4,3	1.052	0,6%	4,3	1.090	0,9%	4,2
R98	Muerte sin asistencia	1.885	0,6%	3,7	1.205	0,7%	4,9	665	0,5%	2,6
	Otras causas de defunción	135.825	45,8%		75.898	44,8%		59.842	47,0%	

Imagen 2 Estadísticas Vitales DANE. Defunciones Acumuladas 2020.

De acuerdo a las estadísticas reflejadas para el año 2020, se hace necesario el plan de vacunación población para la inmunización del COVID-19 para lo cual el gobierno nacional confiere una serie de leyes, normas y decretos como los siguientes: **Ley 2064 del 9 de diciembre del 2020** donde el **ARTÍCULO 1. Objeto.** Menciona que *“la presente ley tiene por objeto declarar de interés general la estrategia para la inmunización de la población colombiana contra la Covid-19 y establecer medidas administrativas y tributarias para la financiación y la gestión de los asuntos relacionados con la inmunización contra la Covid-19 y otras pandemias”* [65]. Igualmente, se encuentra el **decreto 109 del 29 de enero del 2021 “Por el cual se adopta el Plan Nacional de Vacunación contra el COVID -19 y se dictan otras disposiciones”** donde se menciona en el **CAPÍTULO 1 ARTÍCULO 1** que *“El presente decreto tiene por objeto adoptar el Plan Nacional de Vacunación contra el COVID-19 y establecer la población objeto, los criterios de priorización, las fases y la ruta para la aplicación de la vacuna”* [34]. En la imagen 3 se muestra la población objetivo durante el plan nacional para la inmunización contra el COVID-19, con el objetivo de conseguir inmunidad de rebaño en Colombia.

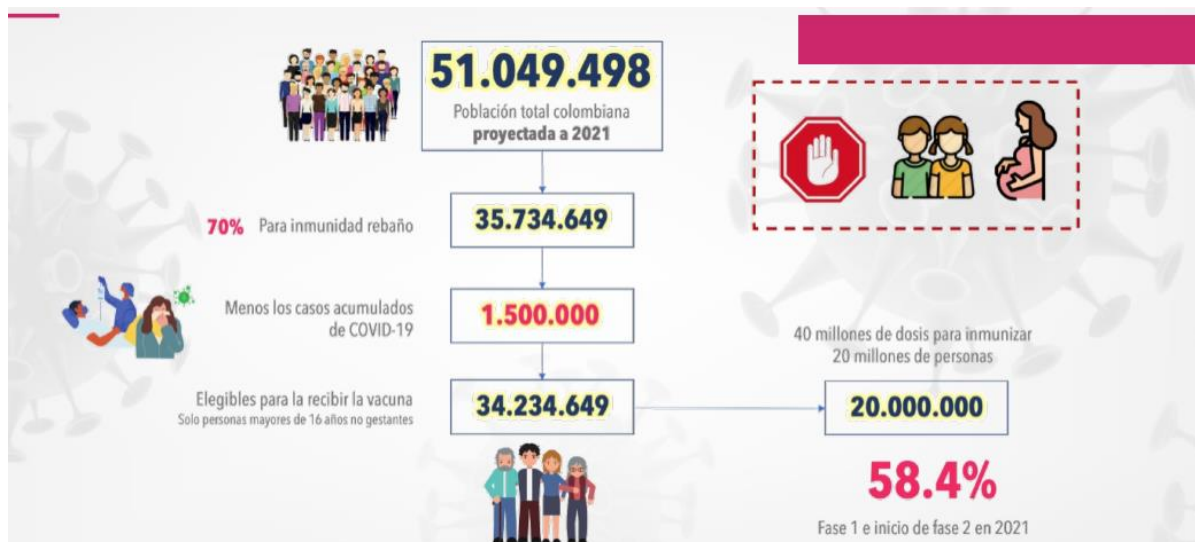


Imagen 3. Cobertura plan de vacunación de población para la inmunización del COVID-19

Así mismo, también se encuentra la Resolución número 537 de 2020 por la cual se modifica la resolución 3495 de 2019 que establece la Clasificación Única de Procedimientos en Salud (CUPS) [66], donde se modifica la lista de códigos de reporte para los distintos procedimientos en referencia a COVID-19. Por esto se hizo prioritario el dominio de información enfocado a la vacunación. Frente a la situación actual del país, se hace proceso de elección del área específica COVID-19 que pertenece a la dimensión de **Vida saludable y enfermedades transmisibles**. En la siguiente imagen se hace referencia a la dimensión en la cual se escoge el área específica enfocada a vacunación contra COVID-19.

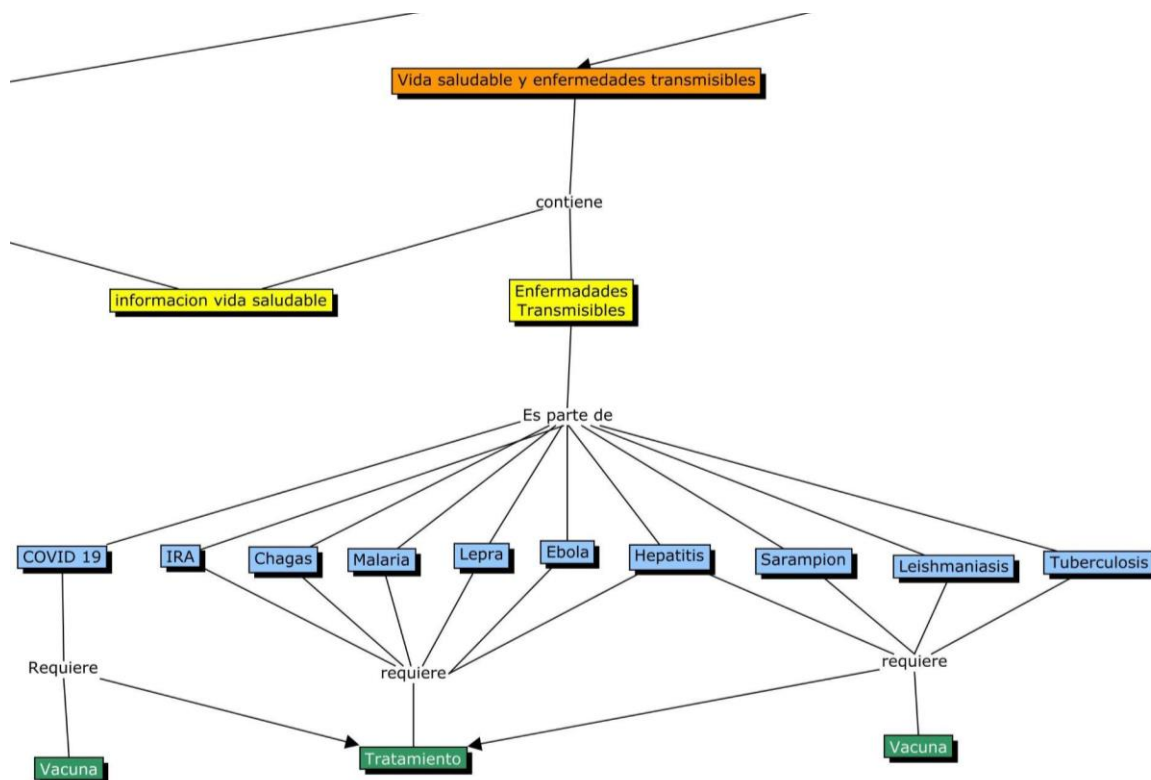


Imagen 4. Dimensión vida saludable y enfermedades transmisibles. Creación propia.

2.5. Caracterización de conceptos y estándares para la vacunación covid-19

De acuerdo a los diferentes lineamientos y esquemas que hay para la vacunación, desde fuentes como: El catálogo de información donde se hace referencia con el indicador 242 al Programa Ampliado de Inmunizaciones Web (PAIWEB), y el catálogo de datos de la plantilla de registro diario de vacunación digital 2016, donde se encuentran determinadas variables de vacunación indicadas con la referencia de PAIWEB para niños y adultos [67]. Así mismo, se encuentra el documento de “*los lineamientos técnicos y operativos para la vacunación contra COVID-19*”, donde se evidencia el modelo de carnet con diferentes variables de vacunación más el registro de vacunación contra COVID-19 realizado [68]. En la siguiente imagen se puede evidenciar el modelo

de carnet de vacunación adoptado para el registro de vacunación en Colombia.



MiVacuna
Covid-19
www.minsalud.gov.co



La salud es de todos

Minsalud



Certificado de vacunación

Nombres:

Apellidos:

Documento de identidad: C.C. ☐ T.I. ☐ Pasaporte ☐ PEP ☐ otro ☐ cuál:

No.

Fecha de nacimiento: Día Mes Año

Biología	Dosis	Fecha	Fabricante	Lote	IPS vacunadora	Nombre vacunador	Código del vacunador
COVID-19	1						
	2						

Imagen 5. Registro de la vacuna contra el COVID-19 en el carné de vacunación de adulto.

También, con el uso de plantilla reporte del ministerio de salud [69], se identificaron las variables utilizadas por el ministerio de salud para cada una de las fases del plan nacional de vacunación contra COVID-19.

De las distintas fuentes de datos como lo son: El registro de COVID-19, los datos requeridos por el carnet de vacunación nacional y las variables manejadas por el reporte departamental de salud del Cauca, se obtienen las variables del catálogo de datos. Con la identificación de las variables de las distintas fuentes de datos, se realizó la comparación de variables comunes entre los esquemas hallados, con esta información se crea una tabla comparativa como la siguiente:

Ver Anexo 2. Variables de Vacunación Covid-19.

2.6. Revisión de estándares hl7

2.6.1. Identificación de conceptos y estándares de hl7

Los estándares o protocolos que brinda HL7 indican cómo se organiza y comunica la información entre dos partes interesadas. Definen el idioma, la estructura y los tipos de datos para que se integren de forma fluida entre sistemas de salud. Muchas organizaciones de salud en el mundo utilizan estos estándares. La organización agrupa sus estándares en primarios, los cuales son los más populares para integraciones de varios sistemas. Por orden cronológico los cuatro estándares más importantes de la organización son: **HL7 v2, HL7 v.3, HL7 CDA y HL7 FHIR.**

HL7 v2 se conoce como “mensajería HL7” ya que está basado en el intercambio de mensajes con una estructura definida, existen muchos tipos de mensajes en HL7 v2 dedicados a la gestión del paciente, órdenes y resultados. Estos mensajes se ajustan a segmentos y campos, asociados a un evento que lo desencadena. Lo que se estandariza son la estructura de los mensajes, la codificación de los mensajes y los eventos que generan la información y son disparadores de estos mensajes. Para la codificación desde su versión cinco se utiliza el formato XML.

El estándar permite conectar muchos sistemas informáticos dentro de un mismo entorno, incluso sistemas de varios entornos. En lugar de crear conexiones a medida entre sistemas, el estándar permite que todos se adapten, permitiendo la interoperabilidad entre ellos.

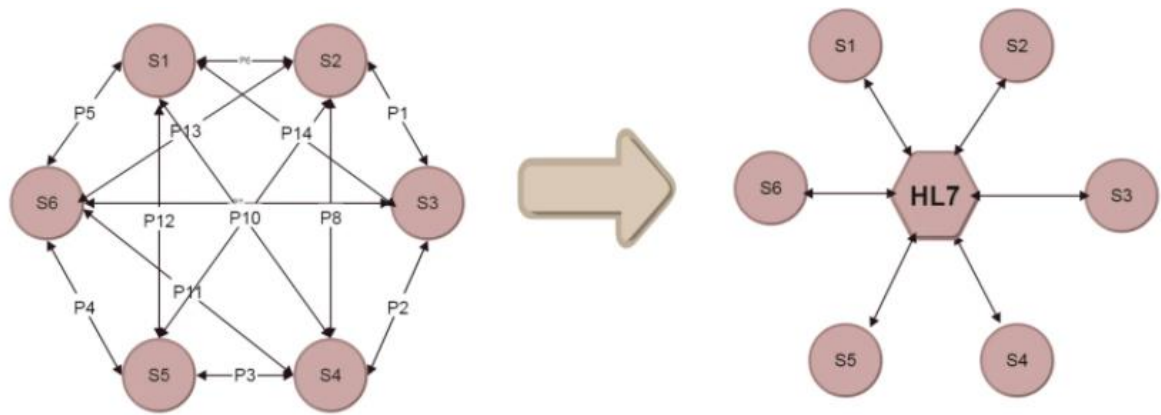


Imagen 6. El estándar HL7 facilita la interoperabilidad.

Aún sigue siendo el estándar más utilizado en mensajería en el sector salud, pero tiene algunos inconvenientes. Tiene mucha flexibilidad y es adaptable para la mayoría de los casos, pero conseguir la flexibilidad no es tan fácil, requiere de un profundo análisis y negociación para conseguir la integración apropiada entre los sistemas.

Esta situación impulsó al desarrollo del estándar HL7 V3. Este estándar pretende ser más ambicioso y amplio que la versión anterior.

HL7 v3 RIM intenta solventar algunos problemas reflejados en la versión 2, ya que tenía ciertos problemas de interoperabilidad en determinados contextos. La versión 3 al incorporar evoluciones tecnológicas como la programación orientada a objetos y la especificación en UML.

Una de las principales características del estándar HL7 V3, el cual está basado sobre **RIM (Reference Information Model)** un amplio modelo de objetos de referencia de los datos clínicos. Este especifica un modelo de la información de los servicios clínicos, identifica el ciclo de vida de la mensajería dentro de todas las actividades clínicas.

Tiene un concepto central que es rol, el tipo de rol (médico enfermero, paciente, familiar de paciente, contador), la persona, donde se almacena los datos

personales de la persona sin importar el rol y el concepto de relación entre roles, no necesariamente es par, puede ser de 2 a muchos como un equipo de cirugía.

HL7 CDA favorece el intercambio de documentos clínicos, mejora la legibilidad humana al formato de transferencia, por lo que se eligió el marcado XML.

HL7 CDA contiene seis características necesarias para el documento clínico:

- Persistencia: el documento debe permanecer inalterado y en uso durante un largo periodo de tiempo.
- Custodia: el documento debe ser mantenido por una organización confiable, como los hospitales.
- Autenticación: certificación legal de la información de la clínica.
- Contexto: El documento debe definir un contexto predeterminado para el registro y participantes, como la identidad del paciente, quién creó el documento, acto clínico, etc.
- Completitud: el documento debe ser una unidad de información, es decir, que esté completo y no por partes de él.
- Legibilidad Humana: Aunque está diseñado para ser procesado por sistemas automáticos, el documento debe ser legible por una persona en un navegador.

El documento está compuesto por un header que contiene la información del documento, tales como datos de los participantes (paciente y médico) o sistemas que han creado el documento. El body del documento contiene la información que constituye el documento y debe tener una parte legible por el ser humano.

HL7 FHIR se empezó a trabajar en el año 2011, con la intención de resolver definitivamente los problemas que contenían la versión 2 y 3 de HL7. Sin embargo, hasta diciembre de 2018 no se ha publicado la primera versión normativa de FHIR, y ha sido en su Release 4. Uno de los problemas de la V2, fue que estaba tecnológicamente anticuada de acuerdo a los avances, y se hacía necesario una renovación para solventar estos problemas incluyendo nuevas tendencias. La V3 fue más allá de la V2, pero su implementación resulta ser compleja y por ende no tuvo éxito.

El objetivo principal de FHIR es facilitar el intercambio de información clínica, e intenta combinar lo mejor de sus dos predecesores. Para ello busca facilitar la implementación y actualizar tecnológicamente la obsolescencia de la versión v2. Además, usa los estándares web más frecuentes, como XML, JSON y HTTP.

FHIR se basa en **recursos**, los cuales son bloques que contienen información básica que va a ser intercambiada. Cada resource o recurso representa un concepto de la realidad de la atención sanitaria, como pacientes, citas, organizaciones o resultados de pruebas, que funcionan como bloques de construcción que permiten componer estructuras de mensajes y/o documentos. Para hacer esto se apoya de una arquitectura REST con el uso de APIs para facilitar el intercambio, la manipulación y búsqueda de recursos, con una filosofía basada en cliente-servidor. El servidor ofrece acceso a los recursos a un cliente REST, permitiendo consultar o modificar dichos recursos a través de peticiones HTTP. Estas peticiones permiten crear, actualizar, leer y borrar (CRUD). Los datos se intercambian tanto en formato JSON como en

XML. HL7 FHIR cuenta con 143 recursos, de los cuales, solo 11 son normativos (estables), los demás tienen un cierto grado de madurez.

Un recurso consta de las siguientes partes:

- Una URL que identifica el recurso.
- Metadatos que incluyen la definición del recurso.
- Un resumen legible para humanos.
- Extensiones, que permiten incluir información adicional que hace parte de la definición básica de un recurso.

Los tipos de datos que maneja el estándar son 2: primitivos y complejos; los datos primitivos son aquellos que son representables en XML o JSON, tales como números, cadenas de texto, fechas, etc. Los datos complejos son estructuras de mayor complejidad, estos se definen como recursos FHIR.

La especificación de FHIR se divide en 3 partes:

- **Documentación general:** mecanismos en los que se define los recursos, el material de referencia que incluye el tipo de dato, vocabularios en formato XML o JSON.
- **Implementación:** la forma en que se utilizan los recursos definidos utilizando tecnología REST, mensajes, documentos clínicos o empleando arquitecturas basadas en servicios.
- **Lista de recursos:** cuenta con listas específicas de recursos clínicos (medicaciones, diagnósticos, interacción con dispositivos, etc.), administrativos (paciente, organización, ordenes, etc.) y de infraestructura (listas, manifiesto de documentos, encabezado de mensajes, consultas, etc.).

En resumen, como estándar HL7 V3 no obtuvo tanto éxito en comparación con la V2, la única parte del estándar que aún se utiliza es su especificación de documentos clínicos CDA. Uno de los mayores problemas según los expertos es:

- El proceso de diseño de V3 impone consistencia en nuestros estándares (es uno de los objetivos deseados) aunque la inconsistencia no se debía a HL7, sino a la inconsistencia de los procesos de negocio clínicos.
- HL7 V3 es independiente de tecnología y plataforma, es un pro que se vuelve en contra cuando hace que la implementación sea más costosa.
- Con HL7 v3 se pretende anticipar todos los modelos posibles, y esto se traduce en que el coste del cambio es aterrador. Terminó produciendo modelos que no eran implementables, demasiado grandes.

Según Garnet Group, en su informe de 2009 declara que, HL7 V3 es considerado como obsoleto antes de llegar a la meseta.

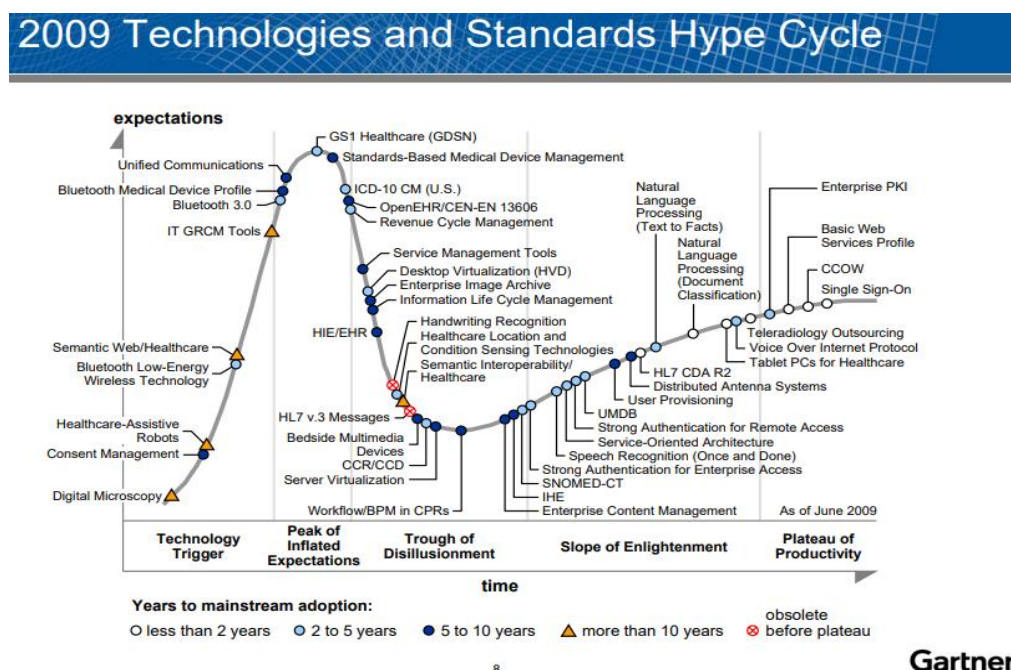


Imagen 7. Tecnologías emergentes año 2009 [70].

Para resolver la obsolescencia tecnológica que es uno de los más grandes problemas de la V2, surge FHIR para también cubrir la complejidad y formalidad de HL7 V3, obtiene el foco de cambio en cuatro puntos:

- Facilitar la implementación, y superar así el problema de HL7 v3.
- Flexibilidad y adaptabilidad, y superar así lo inflexible de HL7 v2.
- Enfocarse en la comunicación de información clínica entre sistemas, que es la materialización real de la necesidad clínica.
- Actualización tecnológica de la que carecía HL7 v2: Basarse en los estándares web: XML, JSON, OAuth y REST [71].

El hecho de que HL7 FHIR soporta cinco arquitecturas para el intercambio de información:

- **RESTful API:** Las siglas REST se refieren a Representational State Transfer. Esto quiere decir que cuando solicitamos información a una aplicación web RESTful (que cumple con REST), esta nos transferirá una representación del estado del recurso solicitado [72].
- **Mensajes:** similar a la comunicación de HL7 V2, mensajes que incluyen recursos. Un request mensaje es enviado desde la aplicación hacia un destino. El mensaje contiene un recurso, el destino replica uno o más mensajes.
- **Documentos:** similar a CDA, enfocado en documentos clínicos que pueden ser firmados e incluir autenticación.
- **Servicios/SOA:** invocación de interfaces
- **Bases de datos / persistencia:** se puede tener recurso de forma nativa, es decir sin almacenarlos.

Dentro de los estándares HL7 FHIR se destaca por ser el componente innovador y disruptivo ya que en 2019 se destaca como tendencia según la siguiente gráfica.

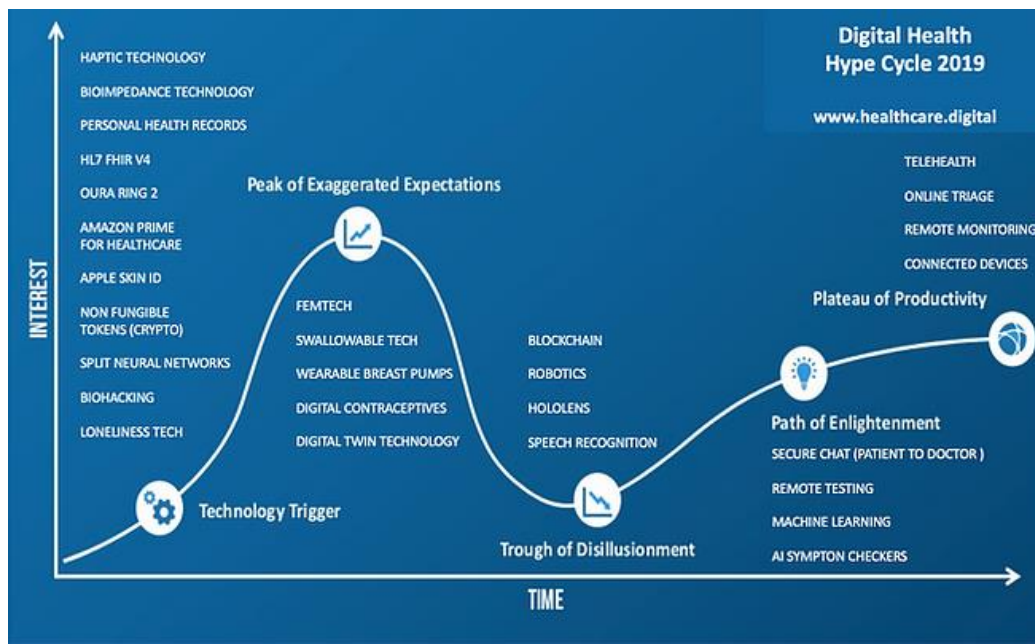


Imagen 8. Gráfica de las tendencias en salud digital 2019 [73].

Por otra parte, se destaca a HL7 FHIR con respecto a los demás estándares en que tiene una curva de aprendizaje muy rápida, por ende, tiende a prevalecer y crecer más rápido que los otros estándares.

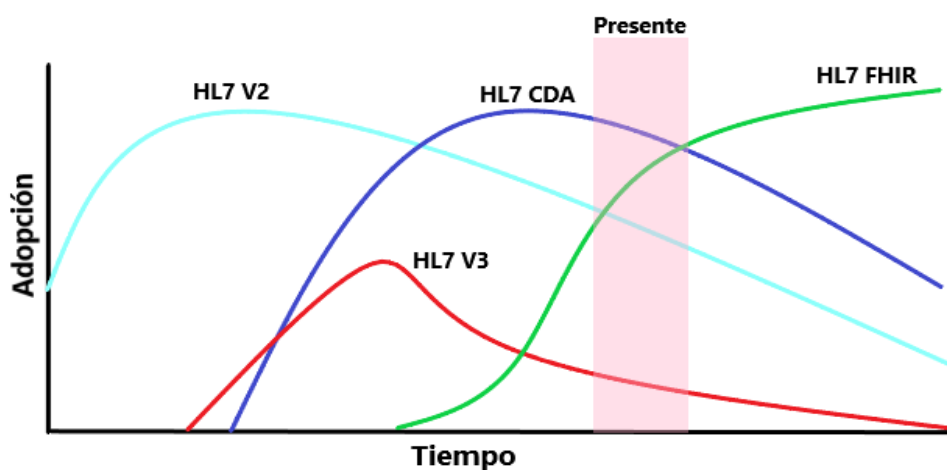


Imagen 9. Comparación de los estándares base de HL7 durante el tiempo. Creación Propia.

2.6.2. Adopción del estándar FHIR

Con respecto al análisis de cada uno de los estándares base de HL7, se adoptó el estándar HL7 FHIR, debido a que facilita la especificación de variables de vacunación en el respectivo recurso manejado por el propio estándar, además de ser orientado a servicios web facilitando la interoperabilidad entre sistemas y de estar a la vanguardia en el manejo de tecnologías y estándares web.

Uno de los recursos clave dentro del estándar relacionado directamente al campo de la vacunación, es el llamado ***Immunization***, donde el estándar lo denomina de la siguiente forma: *“El caso de que a un paciente se le administre una vacuna o un registro de una inmunización según lo informado por un paciente, un médico u otra parte”* [74]. Este recurso especifica las variables necesarias para el campo de la vacunación en cualquier contexto.

El recurso de *Patient* se describe como: *“Datos demográficos y otra información administrativa sobre una persona o animal que recibe atención u otros servicios relacionados con la salud”* [75]. Este es un recurso fundamental dentro del dominio ya que es la base de las variables básicas del paciente, que en este caso dentro del dominio de vacunación se establece los datos del paciente. A continuación, se especifican algunas de las variables que demuestra el recurso:

Tabla 5. Recursos de HL7 con algunos atributos.

Recurso	Variable
Person	Name
	Gender
	Address
	BirthDate
	Identifier
Immunization	Status
	VaccineCode
	TargetDisease
	Date
	Locate

2.6.3. Conclusiones

De acuerdo al segundo capítulo donde se hizo revisión del estándar HL7 y de las áreas y dominios del sector público de Colombia, se orientó desarrollo del proyecto en el dominio de vacunación en el área de covid-19 teniendo en cuenta la priorización para Colombia y la crisis vivida por la pandemia que se está transcurriendo tanto de forma nacional e internacional y a la necesidad de combatirla con una vacuna, es necesario el manejo de información de los vacunados sea de una forma adecuada. Por esto, fue identificado que el estándar FHIR de HL7 suple las necesidades para el manejo de la información de vacunación, ya que está dedicado al manejo de información del sector

salud. También, se logró la identificación de las variables necesarias para la representación de información en el dominio de vacunación, a continuación, se resumen las conclusiones del capítulo.

- Se analizaron diferentes dominios de salud en Colombia, conociendo el estado actual de cada uno de ellos, priorizando en base a estadísticas y datos, identificó el dominio priorizado en salud para Colombia, enfocado a la vacunación contra Covid-19.
- Tras el análisis y revisión del estándar FHIR se identificó los recursos necesarios para el uso de sus variables en el dominio de vacunación.
- En la revisión se identificó que los estándares HL7 CDA y V3 son obsoletos, puesto que no se adaptan a las nuevas tecnologías, y se quedan con paradigmas antiguos al momento en que fueron lanzados.
- La adopción del estándar pretende que las variables establecidas durante el análisis, sean utilizadas dentro de la definición de la información en capítulos posteriores del proyecto, así como aplicar el modelo arquitectónico RESTFull como lo utiliza el estándar FHIR de HL7.
- Fue posible hacer la integración de algunas variables del plan nacional de vacunación contra el Covid-19 junto con algunas del estándar HL7 FHIR.

3. CAPÍTULO 3 CONSTRUCCIÓN DE GRAFO DE CONOCIMIENTO

En este capítulo se aborda la construcción de grafos de conocimiento, donde se hace necesario el uso de ontologías que se utilizan para aprovechar los mecanismos de inferencia, mapear modelos de datos relacionales en el mundo gráfico y también son importantes para detectar posibles inconsistencias en los datos. Las ontologías pueden clasificarse y convertirse en una instancia de una o más clases para expresar restricciones y desarrollar un modelo de conocimiento más consistente esencial para las tareas de integración de datos [76].

A continuación, se describe la jerarquía de las metodologías a aplicar dentro del documento, en la cual se desarrolla la estructura de cada una como se muestra en la siguiente imagen.

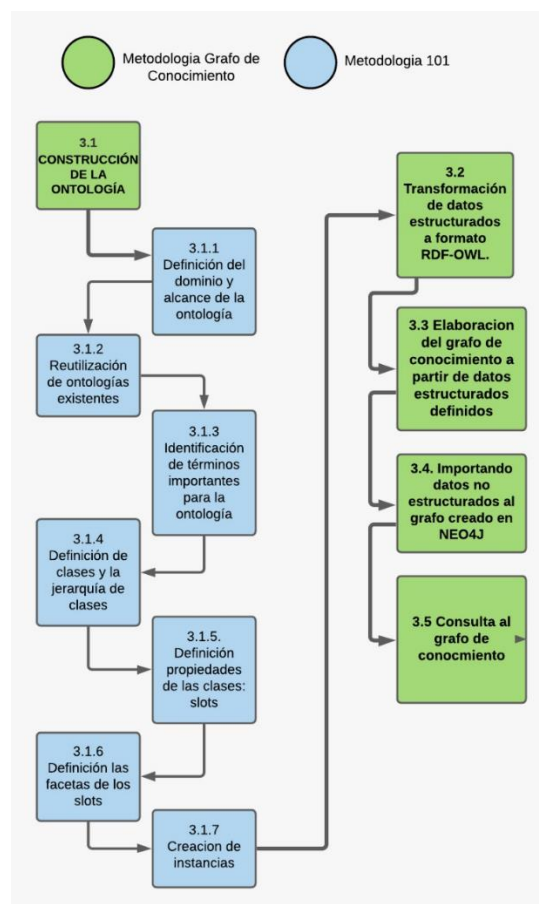


Imagen 10. Metodologías para la creación Grafos de Conocimiento. Creación Propia

3.1. Construcción de la ontología

Para la construcción de la ontología, se tuvo en cuenta *Ontology Development 101: A Guide to Creating Your First Ontology* [77], donde se define la metodología en las siguientes secciones que se describen a continuación.

3.1.1. Definición del dominio y alcance de la ontología

Como se identificó en la sección 2.4 del documento, se identificó el dominio de vacunación del sector salud en Colombia para el desarrollo de la ontología, donde se hace énfasis en el área específica de vacunación contra COVID -19 la cual pertenece a la dimensión de vida saludable y enfermedades transmisibles. El propósito de la ontología es servir de insumo para la posterior construcción de un grafo de conocimiento específico en el área de aplicación, además de definir los términos más importantes relacionados con el área, la identificación de individuos partícipes en la vacunación, sus clases, roles, restricciones y relaciones son partes de la ontología.

La ontología desarrollada debe proveer respuestas a preguntas relacionadas con los pacientes, ubicación, vacuna y vacunación, donde es dirigida para diferentes entidades de salud o personas que requieran conocer información relacionada con el dominio identificado en este proyecto.

3.1.2. Reutilización de ontologías existentes

En esta sección, se mencionan ontologías identificadas en el dominio de vacunación con el fin de la reutilización de conceptos.

3.1.2.1. Vo vaccine ontology

Ontología de vacunas basada en la comunidad que se encuentra en la siguiente página web (VO [78]), donde se ha desarrollado la ontología a través de la colaboración con investigadores en vacunación, centros y programas nacionales e internacionales para abordar desafíos de: Estandarizar la anotación de vacunas, integrar datos sobre diversos tipos y recursos de vacunas, también respaldar el análisis y la inferencia avanzada de datos de vacunación.

Tabla 6. Conceptos importados de VO Vaccine Ontology.

vaccine	manufacturer
dose	company
age range at vaccination	vaccinated
vaccination	Date

3.1.2.2. Schema.org

Schema.org [79] es una actividad comunitaria colaborativa, con la misión de crear, mantener y promover esquemas para datos estructurados en Internet, en páginas web, en mensajes de correo electrónico y más. Se puede utilizar con muchas codificaciones diferentes, incluidas RDFa, Microdata y JSON-LD. Estos vocabularios cubren entidades, relaciones entre entidades y acciones, y pueden extenderse fácilmente a través de un modelo de extensión bien documentado.

Tabla 7. Conceptos importados de Schema.org.

person	organization
name	manufacturer
address	State
affiliation	city
email	dose value - dose unit
gender	occupation
Family Name	patient
insurance Agency	birth date
adverse Outcome	Typical age range
employee	location

3.1.3. Identificación de términos importantes para la ontología

En la siguiente tabla se nombran los términos relacionados con el dominio de vacunación, importantes para la elaboración de la ontología.

Tabla 8. Términos usados para la construcción de Ontología.

Términos	Tipo	Términos	Tipo
Vacuna	Clase	Fecha_de_nacimiento	Atributo
Vacunación	Clase	Edad	Atributo
Persona	Clase	Sexo	Atributo
Tipo De Identificación	Atributo	Nombre	Atributo
Número De Identificación	Atributo	Apellido	Atributo
Teléfono	Atributo	Rango De Edad	<i>Clase definida</i>

Grupo_etnico	Atributo	Dept_de_residen cia	Atributos
Correo_electronic o	Atributo	Municipio	Atributos
Condición De Desplazamiento	Atributo	Área_de_residen cia	Atributos
Condición De discapacidad	Atributo	Barrio	Atributos
Organización	Clase	Dirección	Atributos
Grupo Empleo	clase definida	Fecha Vacunación	Atributo
Ubicación	Clase	Lote Dosis 1	Atributo
Vacunado	<i>Subclase</i>	Dosis	Clase
Vacunador	<i>Subclase</i>	Fabricante	Subclase
Régimen De Afiliación	Atributo	Confianza	Atributo
Aseguradora	Clase	Profesión	Atributo
Lote Dosis 2	Atributo	Acompañante	Subclase
Parentesco	Atributo	Comorbilidad	Atributo
Efecto Secundario	Clase	N Dosis	Atributo
Nombre Vacuna	Atributo	Nombre Efecto Secundario	Atributo
Nombre Aseguradora	Atributo	Nombre Organización	Atributo
Creencia	Atributo	Confianza	Atributo

3.1.4. Definición de clases y la jerarquía de clases

Se definen las principales clases y su respectiva jerarquía, relacionadas con el dominio principal de Vacunación:

Tabla 9. Clases y Jerarquía de Clases.

Clase	Tipo	Descripción
Persona	primitiva	información referente a la persona (paciente, vacunador o acompañante)
Vacuna	primitiva	Información referente por vacunación
Ubicación	primitiva	Dirección de la persona, lugar de vacunación
Vacunación	definida	Información referente a el proceso de vacunación
Dosis	primitiva	Dosis a administrar según el tipo de vacuna
Aseguradora	primitiva	Aseguradora a la que pertenece la persona
Organización	primitiva	Información referente a la organización que realiza vacunación
Grupo Empleo	definida	Información referente al empleo de la persona vacunada
Rango de Edad	definida	Rangos de edades de las personas a vacunar
Vacunado	subclase	Información referente a la persona vacunada
Vacunador	subclase	Información referente a la persona que realiza la vacunación
Acompañante	subclase	Información referente del acompañante de la persona vacunada
Efecto Secundario	Clase	Efectos secundarios producidos por la vacuna

En la siguiente imagen se especifica la jerarquía de clases de la ontología.

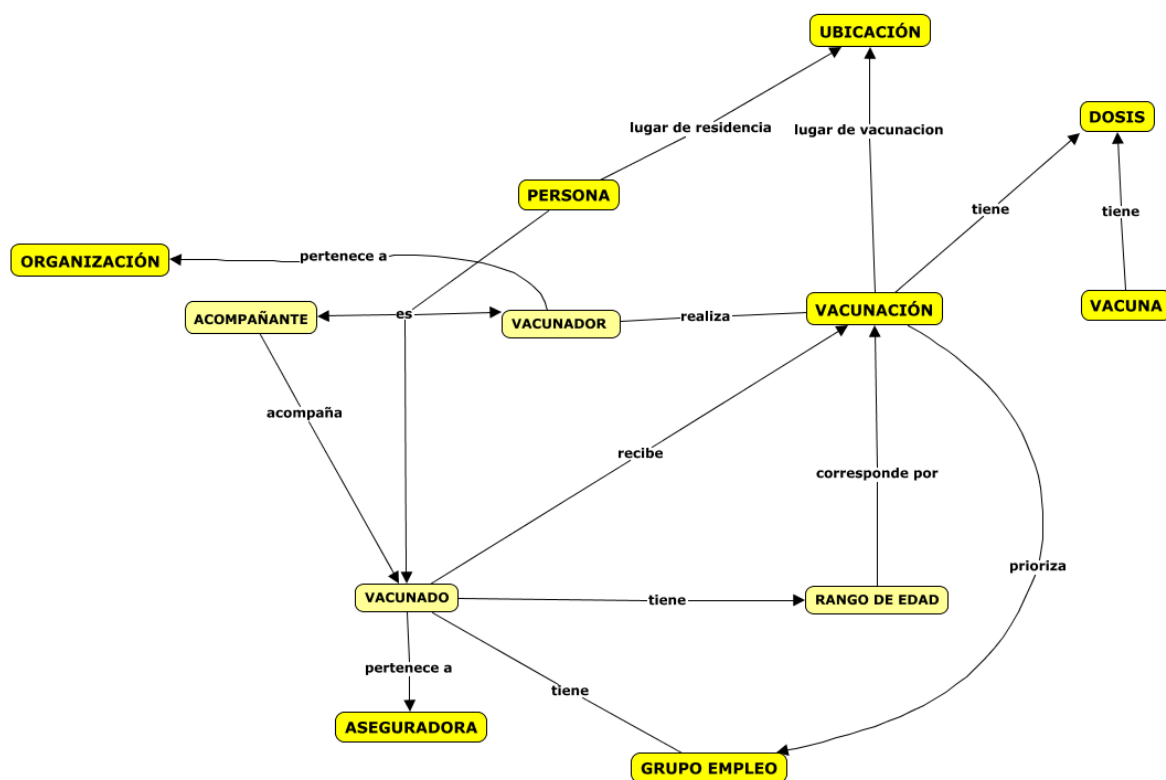


Imagen 11. Jerarquía de clases. Creación Propia.

Ver Anexo 3. Plantilla de Jerarquía de Clases.

3.1.5. Definición propiedades de las clases: slots

En la siguiente tabla se definen las propiedades de las clases y subclases.

Tabla 10. Propiedades de clases de ontología.

Clase	Atributos de Clase
Vacuna	Nombre Vacuna
Vacuna	fabricante
Vacuna	organización
Vacuna	Lote dosis 1

Clase	Atributos de Clase
Persona	Edad
Persona	Fecha Nacimiento
Persona	Tipo Identificación
Persona	Número Identificación

Vacuna	Lote Dosis 2
Ubicación	Departamento
Ubicación	Municipio
Ubicación	Área
Ubicación	Dirección
Persona	Nombre
Persona	Apellido
Persona	Sexo

Persona	Teléfono
Persona	Correo Electrónico
Aseguradora	Nombre Aseguradora
Organización	Nombre Organización
Organización	Tipo
Organización	Nivel
Dosis	Fecha Vacunación
Efecto	Nombre Efecto
Secundario	Secundario

3.1.6. Definir las facetas de los slots

En la siguiente tabla se definen cada una de las restricciones, demostrando cada uno de los atributos que corresponden a las clases:

Tabla 11. Facetas de los Slots.

Nombre de restricción	Nombre Atributo (data property)	Tipo de Valor	Cardinalidad	Valores Permitidos
máximo de dosis	Dosis	integer	3	1,2,3
grupo étnico	Grupo_Etnico	String	1	Indígena, Raizal, Gitano, Palenquero,

				Negro, Mulato o Afrocolombiano, sin pertenecía étnica
tipo de documento	Tipo de Documento	String	1	Cédula de ciudadanía, Pasaporte, cédula de extranjería, tarjeta de identidad, registro civil, adulto sin identificación
Municipio	Municipio	String	1	listados de municipio de Colombia
Vacuna	Vacuna	String	1	Colocar nombres de vacunas
Departamento	Departamento	String	1	listado de departamentos de Colombia

condición de desplazamiento	Condición De Desplazamiento	String	1	sí, no
condición de discapacidad	Condición De Discapacidad	String	1	sí o no
regimen de afiliación	Regimen de Afiliación	String	1	Contributivo, Subsidiado, Pobre no asegurado, Especial y de Excepción
parentesco	Parentesco	String	1	padre, madre, hermano(a), primo(a), tío(a), abuelo(a), amigo(a), suegro(a), cuñado(a), esposo(a). Ver en el DANE
profesión	Profesión	String	1	Profesionales de la salud
régimen (aseguradoras)	Régimen	String	1	

grupo empleo	Grupo Empleo	String	1	Se encuentran descritos en el diagrama
sexo	Sexo	String	1	femenino, masculino ver en el DANE
área	Área	String	1	rural, centro poblado, cabecera ver en el DANE
tipo (organización)	Tipo	String	1	publica, privada, pública-privada
nivel	Nivel	Integer	1	1,2,3,4
país (fabricante)	País	String	1	China, Reino Unido, Estados Unidos, Rusia
rango de edad	Rango Edad	String	1	+16, 18-22, 23-27, 28-32, 33-37, 38-42, 43-47, 48-52, 53-57, 58-59, +60, +70, +80

3.1.7. Crear instancias

Finalmente, las instancias se crearán en el momento de importar los datos en el grafo a partir del modelo definido previamente.

3.2. Transformación de datos estructurados a formato rdf-owl

Con la identificación de las clases, subclasses y atributos en el dominio de vacunación, se hace importante el uso de una herramienta que permita la creación de la ontología propuesta en la sección 3.1. Para ello, se hace uso de **Protégé** en su versión 5.5.0. Protégé Desktop es un entorno de edición de ontologías rico en funciones con soporte completo para OWL 2 Web Ontology Language en un solo espacio de trabajo, a través de una interfaz de usuario completamente personalizable como se muestra en la siguiente imagen. Las herramientas de visualización permiten la navegación interactiva de las relaciones ontológicas y su arquitectura de plug-in se puede adaptar para construir aplicaciones basadas en ontologías simples y complejas [80].

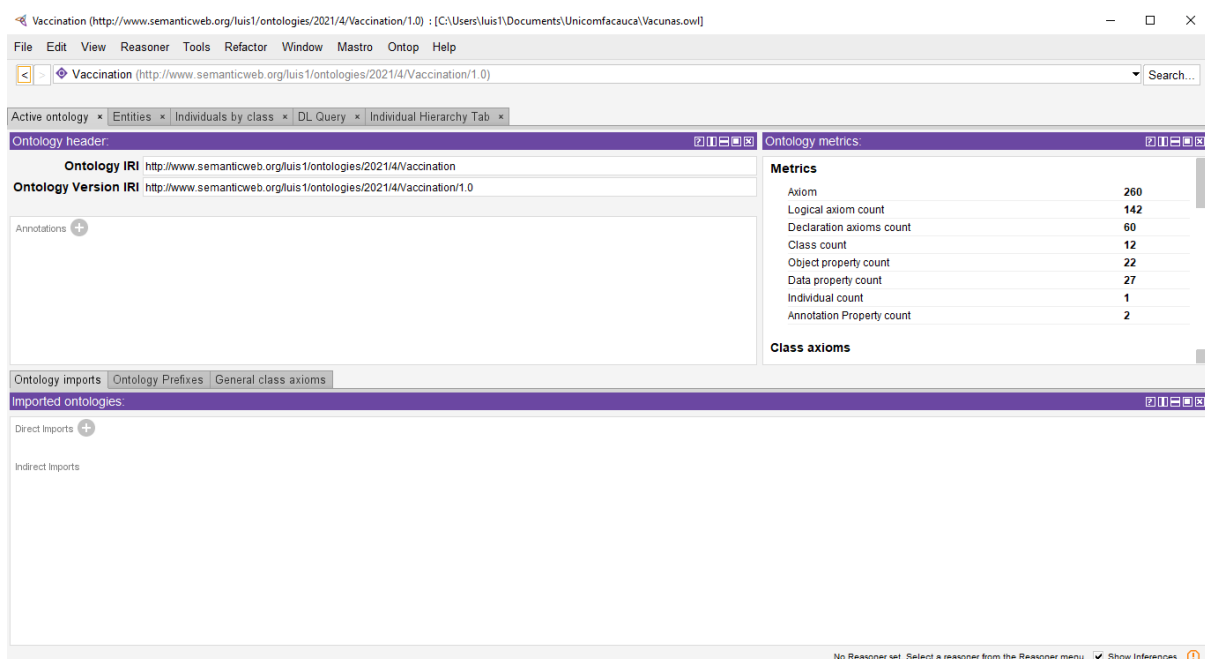


Imagen 12. Interfaz principal de PROTÉGÉ.

En la siguiente imagen se muestran los prefijos utilizados para hacer referencia a las ontologías mencionadas anteriormente, en donde solo se utiliza la URI de la ontología a referenciar.

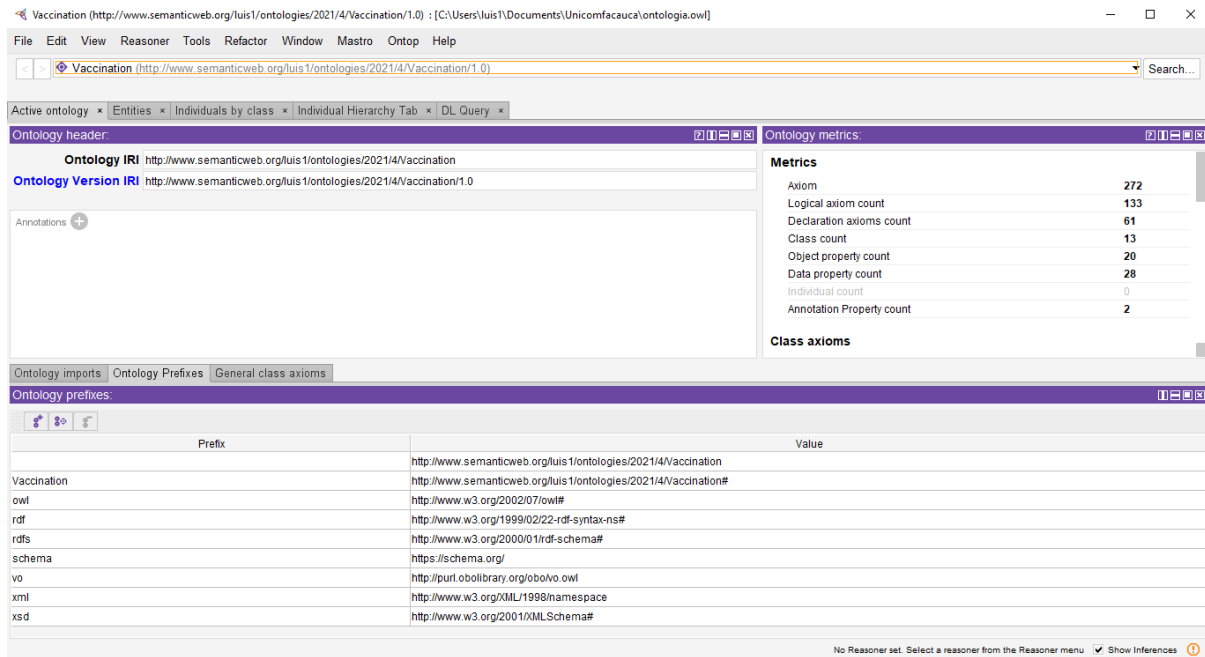
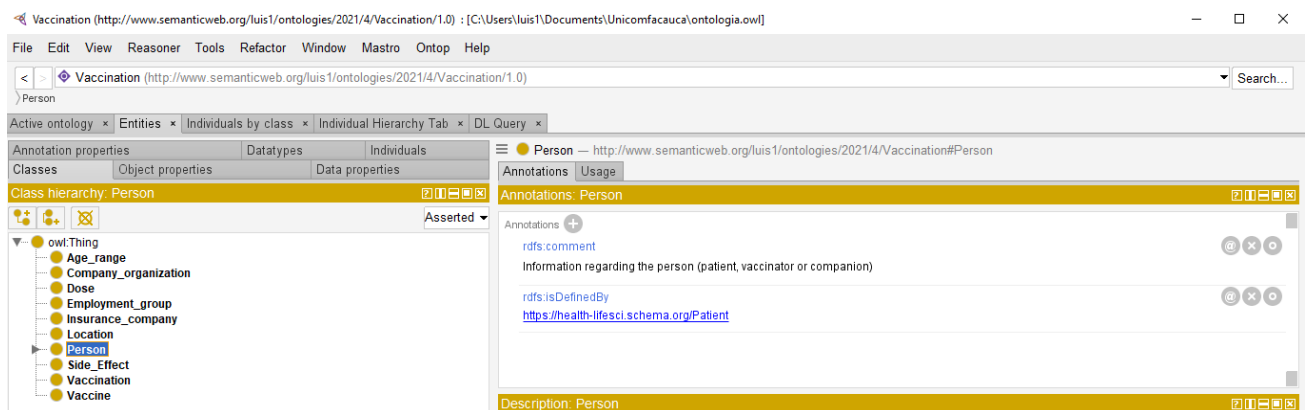


Imagen 13. Interfaz principal de PROTÉGÉ.

Los conceptos importados se pueden evidenciar en la siguiente imagen en donde se utilizó la dirección que propone las ontologías definidas en la **sección 3.1.2** para definir el concepto, en este caso se define a Person basado en la ontología de Schemas.org.



3.3.Elaboración del grafo de conocimiento a partir de datos estructurados definidos en la ontología

Para almacenar la información generada a partir de la ontología propuesta, se hace necesario el uso de una base de datos orientada a grafos, para ayudar a encontrar relaciones, dar sentido a la búsqueda de información y sobre todo la visualización y análisis de datos. Una de las herramientas conocidas en este campo es **Neo4j**, la cual es esencial para representar los datos y las relaciones entre ellos. A continuación, se describen las principales ventajas al usar esta herramienta:

- **Rendimiento:** Las bases de datos orientadas a grafos como Neo4j tienen mejor rendimiento que las relacionales (SQL) y las no relacionales (NoSQL). La clave es que, aunque las consultas de datos aumenten exponencialmente, el rendimiento de Neo4j no desciende, frente a lo que sí sucede con las BD relacionales como MySQL.
- **Agilidad:** Neo4J tiene muchas ventajas, pero una es su agilidad en la gestión de datos. Si nosotros quisiéramos llevar al límite sus capacidades, tendríamos que superar un volumen total de 34.000 millones de nodos (datos), 34.000 millones de relaciones entre esos datos, 68.000 millones de propiedades y 32.000 tipos de relaciones.
- **Flexibilidad y escalabilidad:** Cuando los desarrolladores de una empresa trabajan con grandes datos, buscan flexibilidad y escalabilidad. Las bases de datos orientadas a grafos aportan mucho en este sentido porque cuando aumentan las necesidades, **las posibilidades de añadir más nodos y relaciones a un grafo ya existente son enormes [81].**

Para desarrollo de este proyecto se hizo la creación de la ontología haciendo uso de la herramienta Protégé y su respectiva importación a Neo4j de la siguiente forma:

- La ontología creada en Protégé debe ser guardada en formato RDF/XML, para no tener inconvenientes es su posterior importación en Neo4j. RDF es un modelo estándar de W3C para el intercambio de datos.
- Ahora desde Neo4j, creamos un nuevo proyecto, en el cual se establece una configuración base en la que se crea un Graph DBMS, en donde se aloja toda la información referente al grafo.
- Dentro del proyecto es necesario instalar desde la interfaz una serie de plugins (APOC, NeoSemantics(n10s)) los cuales son necesarios para la importación de archivos en formato RDF y OWL, el más importantes es Neosemantics (n10s) que es un complemento que permite el **uso de RDF en Neo4j**.
- Para usar el complemento dentro del proyecto hay que realizar una serie de configuraciones globales, en las cuales se definen cómo manejar los datos en el gráfico. Para establecer la configuración se necesita ejecutar el siguiente código: **“CALL n10s.graphconfig.init();”**, el siguiente comando es importante para ejecutar procedimientos de importación de RDF:

```
“CREATE CONSTRAINT n10s_unique_uri ON (r:Resource)  
ASSERT r.uri IS UNIQUE;”.
```

El propósito de esta restricción es garantizar la unicidad de los recursos por URI y también acelerar el proceso de ingestión al agregarlos a un índice.

- Por último, se usa el siguiente comando para la importación de la ontología dentro de Neo4j:

CALL

```
n10s.onto.preview.fetch("file:///C:\\Users\\luis1\\Documents\\Unicomfacau  
ca\\Vacunas1.0.owl","RDF/XML");
```

En la anterior consulta se llama al método con el cual se ubica el archivo dentro del sistema, en este caso la dirección donde se encuentra guardada la ontología a importar en Neo4j, más el formato en el cual es compatible con Neo4j, para este caso es el formato RDF/XML.

Como resultado final después de hacer la importación de la ontología creada a Neo4j se puede ver el grafo resultante en la siguiente imagen.

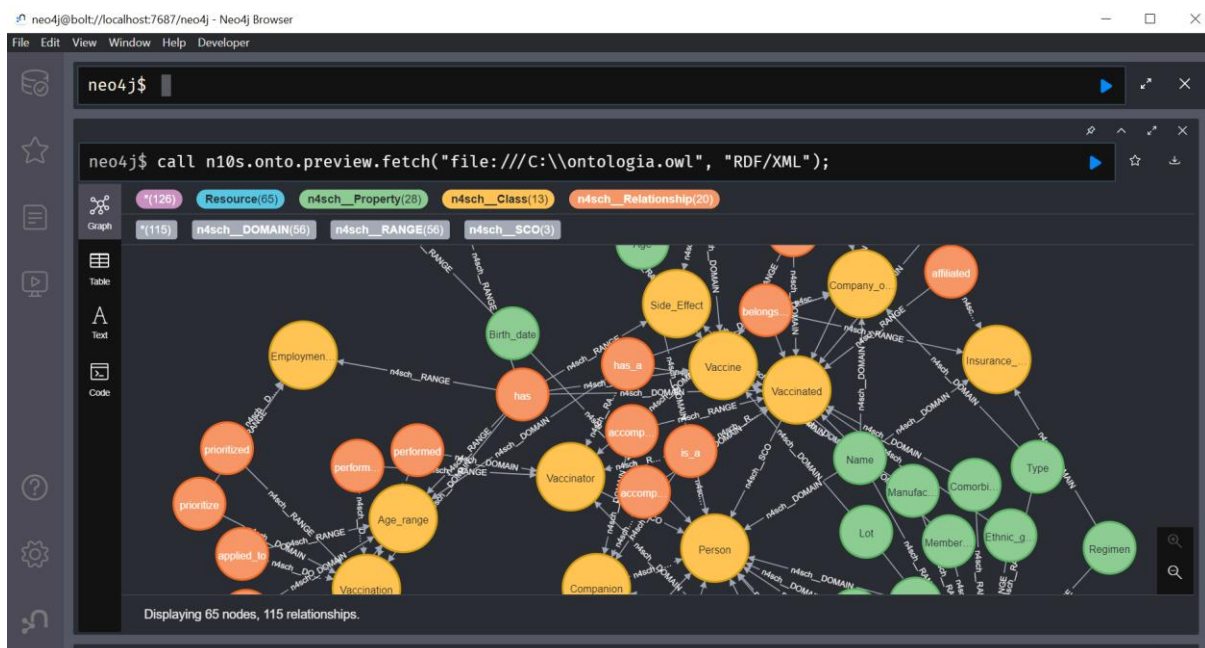


Imagen 15. Grafo resultante de la importación de la ontología en Neo4j. Creación Propia

3.4. Importando datos no estructurados al grafo creado en neo4j

3.4.1. Recolección de datos (extracción de entidad)

Para la recolección de datos, se realizó una encuesta en Google Forms, a personas que han recibido una o más dosis de las diferentes vacunas disponibles en el país, las preguntas realizadas extraen información del vacunado de una forma general donde no se implican datos personales como teléfono, número de identificación, nombre, apellido y demás datos debido a que se generaron de manera propia para evitar incluir información privada de los encuestados por lo cual se generan datos aleatorios. La información está relacionada a los efectos secundarios generados por la vacuna, el lugar de vacunación, EPS que realizó el proceso de vacunación, la confianza que tienen por dicha vacuna, entre otras. Todo lo anterior en relación a los campos especificados de ontología en la sección 3.1 del documento.

Una vez realizada la encuesta se descargó la información recolectada en un archivo con formato CSV, donde los encabezados se cambian a una palabra clave omitiendo signos de puntuación, tildes debido a que neo4j solo reconoce los encabezados en forma base del idioma inglés, como se muestra en la siguiente imagen.

A1	Marca temporal	Vacunado	Nombre	Apellido	Sexo	parentesco	anio	DIA	mes	nacimiento	Edad	Rango_edad	N_iden
1	2021/06/13 3:51	Vacunado1	Maria	Garcia	Mujer	Padre	1988	23	12	1988-12-23	33	33-37	4627590
2	2021/06/13 4:00	Vacunado2	Julietta	Fernández	Mujer	Hermano	1987	04	12	1987-12-04	34	33-37	1166463300
3	2021/06/13 4:10	Vacunado3	Gerardo	González	Hombre	Hermana	1964	18	06	1964-06-18	57	53-57	10724085
4	2021/06/13 4:44	Vacunado4	Luis	Rodríguez	Hombre	Hijo	1958	25	09	1958-09-25	63	60-69	1061019080
5	2021/06/13 4:50	Vacunado5	Miguel	López	Hombre	Hija	1981	08	07	1981-07-08	40	38-42	1061726120
6	2021/06/13 6:49	Vacunado6	Sara	Martínez	Mujer	Amigo	1982	22	04	1982-04-22	39	38-42	4666235
7	2021/06/13 7:57	Vacunado7	Saray	Sánchez	Mujer	Amigo	1958	11	09	1958-09-11	63	60-69	1059907840
8	2021/06/14 11:5	Vacunado8	Maicol	Pérez	Hombre	Esposa	1962	03	06	1962-06-03	59	58-59	10519512
9	2021/06/14 2:38	Vacunado9	Luz Dary	Martín	Mujer	Esposo	1949	17	11	1949-11-17	72	70-79	1029602560
10	2021/06/14 6:42	Vacunado10	Mario	Gómez	Hombre	Madre	1968	18	09	1968-09-18	53	53-57	10292245
11	2021/06/15 9:10	Vacunado11	Antonia	Ruiz	Mujer	Tio	1954	20	05	1954-05-20	67	60-69	10320382
12	2021/06/15 9:13	Vacunado12	Joaquín	Viveros	Hombre	Tia	1952	21	06	1952-06-21	69	60-69	1061782120
13	2021/06/16 8:49	Vacunado13	Yolanda	Aranda	Mujer	Primo	1933	28	06	1933-06-28	88	Mayor de 80	25649627
14	2021/06/21 9:48	Vacunado14	Maximo	Espinosa	Hombre	Prima	1971	13	12	1971-12-13	50	48-52	1061810650
15	2021/06/21 9:53	Vacunado15	Nidia	González	Mujer	Madre	1969	07	02	1969-02-07	53	48-52	1063813440
16	2021/06/21 10:0	Vacunado16	William	Sanchez	Hombre	Hijo	1960	30	11	1960-11-30	61	60-69	25398093
17	2021/06/21 10:1	Vacunado17	Johana	Rodríguez	Mujer	Hijo	1960	13	03	1960-03-13	61	60-69	25290312
18	2021/06/21 11:2	Vacunado18	Mario	Herrera	Hombre	Amiga	1968	09	03	1968-03-09	53	53-57	1061826760
19	2021/06/21 2:46	Vacunado19	Antonia	Franco	Mujer	Hijo	1956	16	02	1956-02-16	65	60-69	48627382

Imagen 16. Datos recolectados de encuesta desde EXCEL como archivo de formato CSV. Creación Propia

3.4.2. Importación de datos en neo4j

Inicialmente se importó en la versión Neo4j Desktop, para importar el archivo CSV en neo4j es necesario guardarlo dentro de una carpeta con el nombre Import dentro del proyecto creado, desde la interfaz de Neo4j se usa el siguiente comando para obtener la información de archivo CSV. El siguiente comando, carga el archivo de formato CSV de la carpeta Import con cada uno de los encabezados para permitir especificar qué datos importar en el grafo.

LOAD CSV WITH HEADERS FROM "file:///Encuesta.csv" as Encuesta

Para crear los nodos y relaciones en Neo4j se hace uso del lenguaje Cypher propio de Neo4j para obtener y vincular la información importada del CSV con el grafo de la ontología importada en la sección 3.3.

A continuación, se muestra un ejemplo de la importación del archivo CSV haciendo uso del lenguaje Cypher donde se explica el uso de cláusulas como

create, match y merge [82] para crear los nodos referentes a los nombres de los encuestados encontrados en el archivo CSV.

```
LOAD CSV WITH HEADERS FROM "file:///Encuesta.csv" as Encuesta  
MERGE (p:Name {Name:Encuesta.Nombre})  
WITH p MATCH(n:n4sch__Property)  
WHERE n.n4sch__name = 'Name'  
MERGE(n)-[r:IS]->(p) RETURN p,n
```

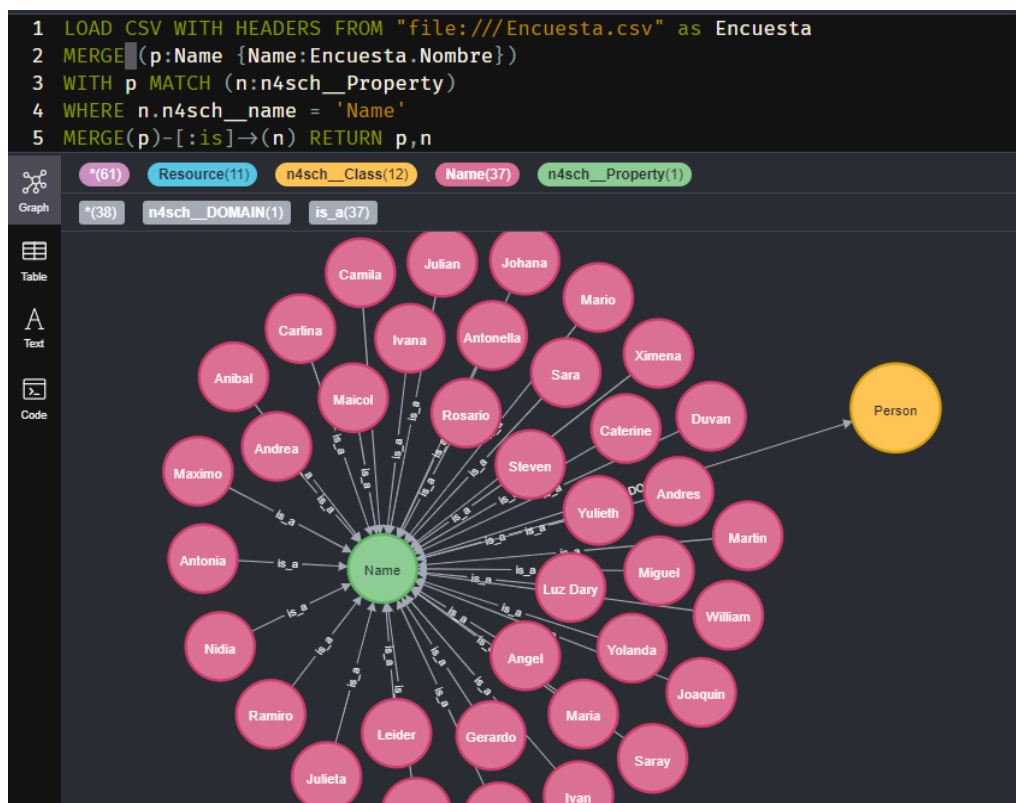


Imagen 17. Aplicación y uso de importación de datos y uso de lenguaje Cypher en Neo4j. Creación Propia

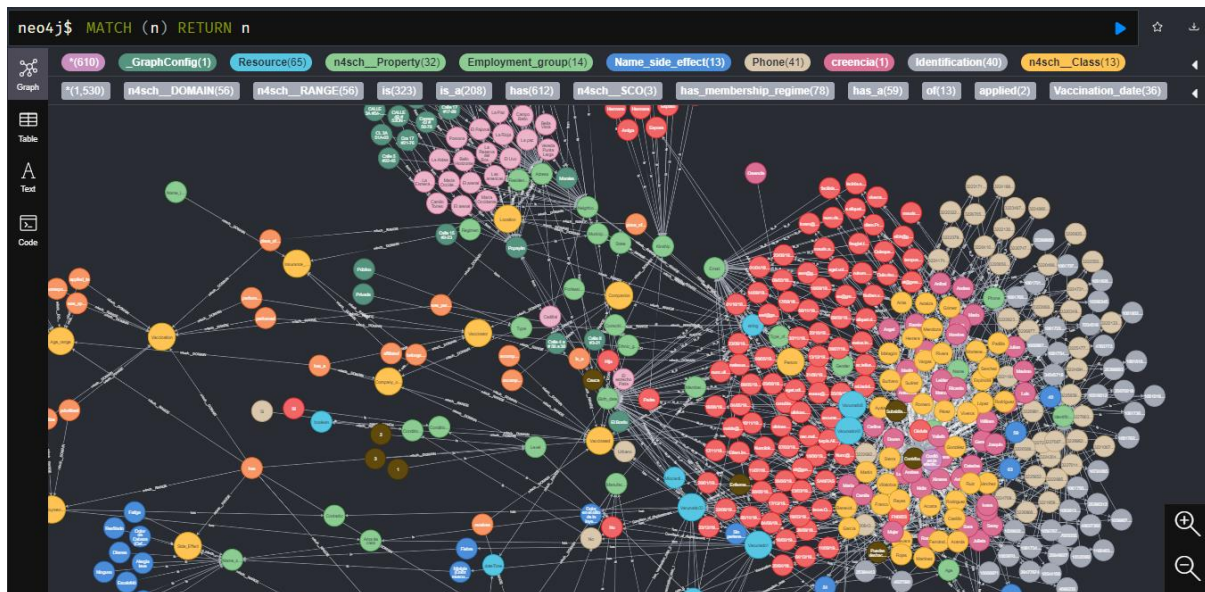


Imagen 18. Grafo resultante de toda la información importada. Creación Propia

3.5. Consultas al grafo de conocimiento

Con el grafo de conocimiento creado se puede proceder a realizar algunas consultas con el lenguaje utilizado en neo4j: Cypher (Lenguaje de consultas), para su posterior análisis y uso sobre el proyecto. Estas consultas son en base a la información recolectada e importada en el grafo apoyado de la ontología, en la cual cada nodo hace referencia a datos individuales de la persona, esto con el fin de que las consultas sean más específicas y lograr hacer un análisis en profundidad con el uso de un algoritmo específico.

Ver Anexo 4. Queries Neo4j

Una de las consultas necesarias para el proyecto es conocer los efectos secundarios relacionados a cada usuario. En la siguiente imagen, en la parte superior de la interfaz de neo4j se puede evidenciar la query realizada para conocer los vacunados relacionados a sus efectos secundarios seleccionados en la encuesta realizada y en la parte inferior de la imagen se puede observar los resultados de la consulta. Estos resultados se pueden ver en forma de grafo donde

se visualiza que un vacunado puede estar relacionado a los mismos efectos secundarios de otro vacunado.

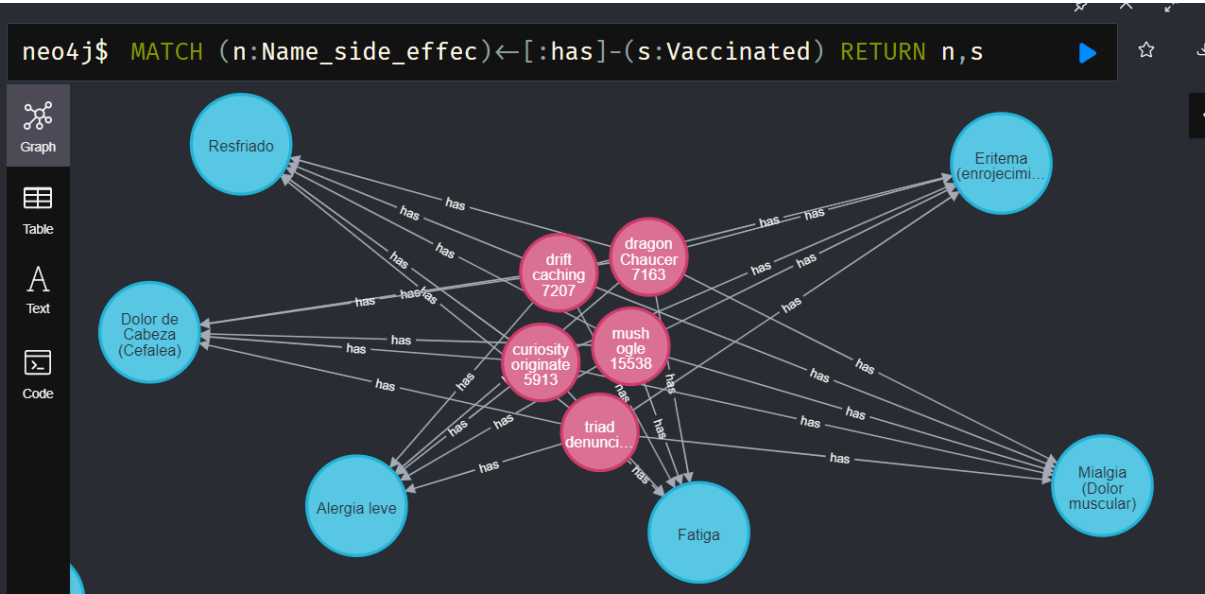


Imagen 19. Consulta sobre efectos secundarios. Creación Propia

Neo4j también nos permite visualizar los datos resultados en forma de tabla como se muestra en la siguiente imagen.

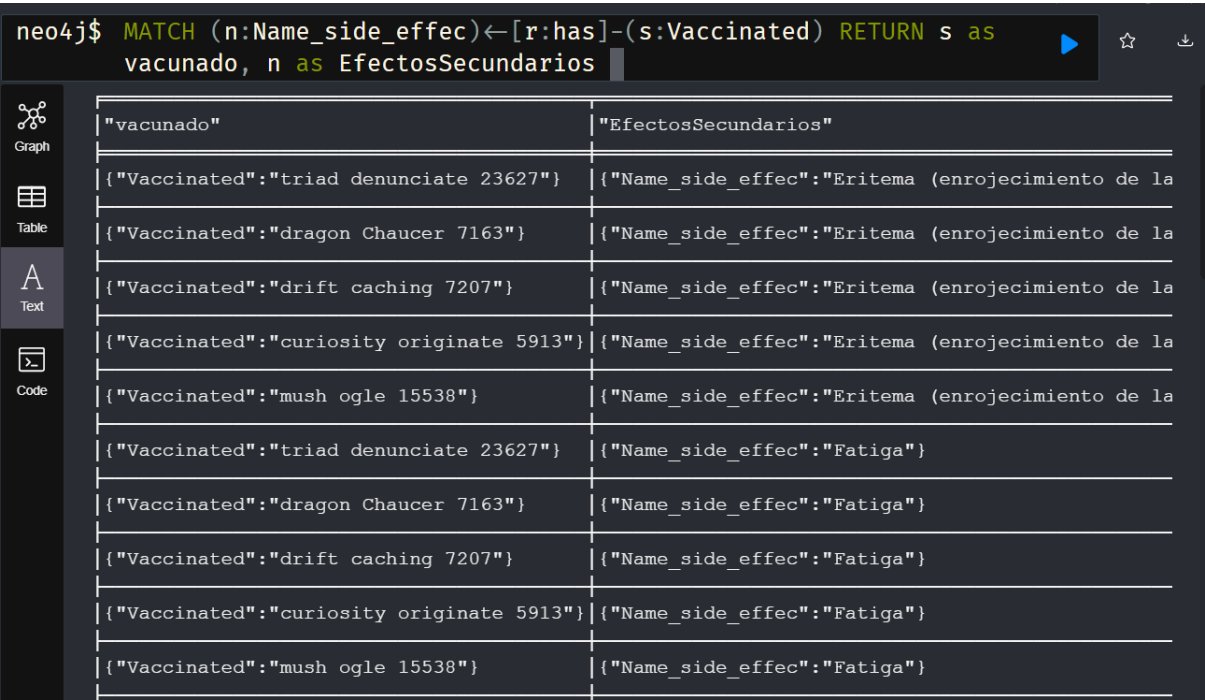


Imagen 20. Tabla de resultados de la consulta en Cypher. Creación Propia

Otro ejemplo de consulta, es conocer las creencias relacionadas a cada Vacunado según lo que seleccionaron en la encuesta.

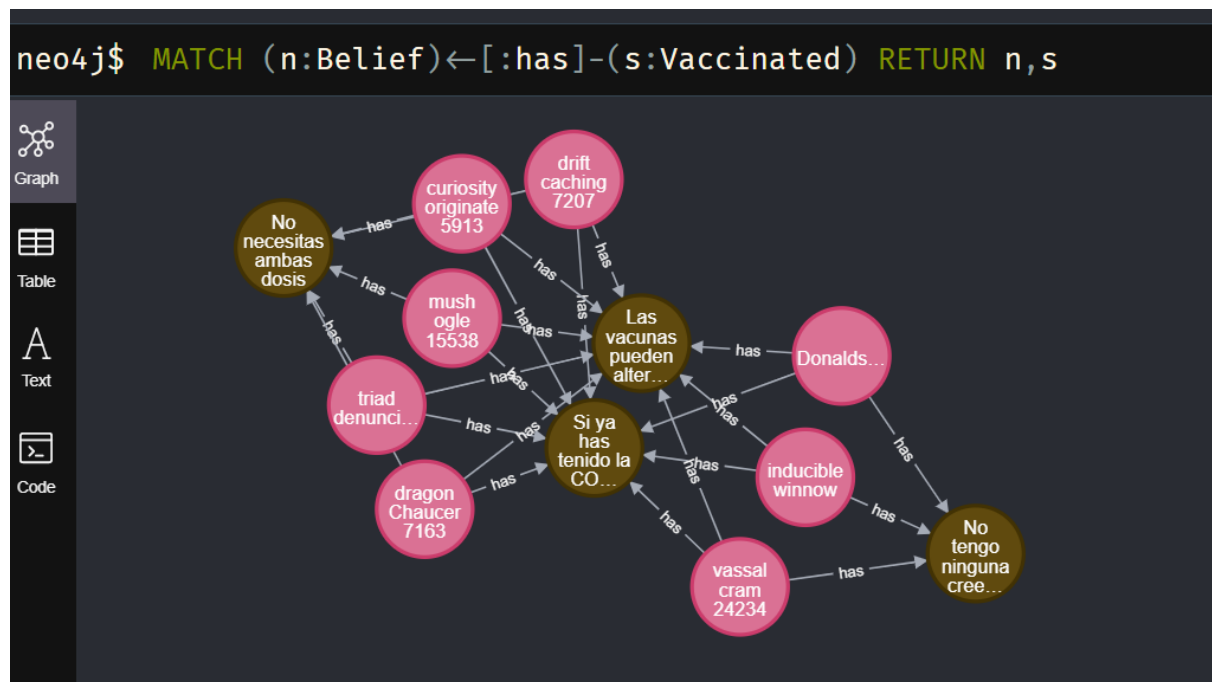


Imagen 21. Consulta sobre creencias relacionadas a la vacuna de los vacunados. Creación Propia

3.5.1. Razonamiento del grafo

Para el análisis de datos en el grafo, es necesario contar con el uso de algoritmos, en este caso neo4j cuenta con una serie de algoritmos de grafos. Los algoritmos de grafos se utilizan para calcular métricas de grafos, nodos o relaciones. Estos algoritmos pueden proporcionar información sobre las entidades relevantes (centralidades, clasificación) en el gráfico o estructuras inherentes como comunidades (detección de comunidades, partición de grafos, agrupamiento).

Las características del grafo anteriormente creado son importantes para el uso de algoritmos, a continuación, se describen las características que se identificaron del grafo realizado:

- **Grafo Dirigido:** Las relaciones entre los nodos cuenta con dirección como por ejemplo (Vacunados)----tiene--->(síntoma).
- **Grafo no Ponderado:** Las relaciones que se especifican en el grafo no tienen ninguna propiedades o pesos que representan la unión de los nodos, simplemente no existe una fuerza en la relación.
- **Grafo Multipartito:** El grafo cuenta con muchas etiquetas para diferenciar los nodos, por ejemplo, podemos diferenciar los nodos de Vacunados, con el grupo de los efectos secundarios ya que cada uno cuenta con una etiqueta diferente que los agrupa. En la siguiente imagen podemos ver las distintas etiquetas para cada uno de los nodos en el grafo.



Imagen 22. Etiquetas por cada grupo de nodos. Creación Propia

Dentro de los grupos de algoritmos de grafos manejados por Neo4j [83], encontramos varios grupos de algoritmos que se utilizan dependiendo de las características únicas de cada grafo.

Los algoritmos encontrados en Neo4j son:

- Búsqueda de Ruta.
- Centralidad.
- Detección de Comunidades.
- Similitud.

- Predicción de Enlaces.
- Incrustación de nodos.
- Clasificación de nodos.

Por eso, teniendo en cuenta las características del grafo mencionadas antes, se trabajó con solo un algoritmo brindado por neo4j; **el algoritmo de similitud**, que es usado para encontrar nodos similares en una red según sea la estructura del grafo. Estos algoritmos se apoyan de la librería Neo4j GDS (Graph Data Science), la cual facilita el uso de los algoritmos en el lenguaje de consulta cypher, si se usa la versión de Neo4j Desktop es necesario instalar el plug-in de Graph Data Science Library dentro del proyecto a crear para usar los diferentes algoritmos.

Jaccard Similarity (Alpha) es un algoritmo diseñado para aplicar los conceptos de similitud desarrollados por Paul Jaccard que ayuda a medir las similitudes entre conjuntos. Este se define como el tamaño de la intersección dividida por el tamaño de la unión de dos conjuntos [64]. Este se rige bajo la siguiente fórmula:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

Imagen 23. Fórmula para el cálculo de la similitud de Jaccard.

El algoritmo se usó en el proyecto para calcular la similitud entre las relaciones de un vacunado específico con respecto a sus creencias de la vacuna o los síntomas que presentó en común, con el conjunto en general de los demás vacunados.

El algoritmo realiza la consulta de las creencias o síntomas de la persona a consultar reuniendo los Id de cada uno de los nodos que tiene y los compara con las creencias o síntomas de cada uno de los demás vacunados, obteniendo dos conjuntos (los nodos de síntomas/ creencias de la persona) y (los nodos de todas las creencias/ síntomas de los demás vacunados), los cuales son de insumo para calcular la similitud de los 2 conjuntos y obtener el promedio de la similitud de la persona con respecto a las demás, hay que tener en cuenta que hay una similitud del 100% cuando el resultado es cercano al 1.0, de lo contrario varía la similitud según sea el caso, como se ve en las siguientes imágenes.

```
neo4j$ match (n:Vaccinated{Vaccinated:'inducible winnow 11993'}) -
[:has]→(s:Name_side_effec) WITH n, collect(id(s)) AS
n1Efectos MATCH (v2:Vaccinated)-[:has] → (Name_side_effec)
WHERE n < v2 WITH n, n1Efectos, v2,
collect(id(Name_side_effec)) AS n2Efectos RETURN
AVG(gds.alpha.similarity.jaccard(n1Efectos, n2Efectos)) AS
similarity
```

similarity	
1	0.046365914786967416

Imagen 24. Ejemplo de cálculo de similitud de efectos secundarios de un vacunado específico con respecto a todos los vacunados. Creación Propia

```
neo4j$ MATCH (n:Vaccinated{Vaccinated:'inducible winnow 11993'}) -
[:has]→(s:Belief)WITH n, collect(id(s)) AS n1Creencias MATCH
(v2:Vaccinated)-[:has] → (Belief) WHERE n < v2 WITH n,
n1Creencias, v2, collect(id(Belief)) AS n2Creencias RETURN
AVG(gds.alpha.similarity.jaccard(n1Creencias, n2Creencias))
AS similarity
```

similarity	
1	0.10697197539302801

Imagen 25. Ejemplo de cálculo de similitud de creencias de la vacuna de un vacunado con respecto a todos los vacunados. Creación Propia

Realizando el algoritmo en general para obtener la similaridades de cada uno de los vacunados con respecto a sus creencias sobre la vacunación contra el

Covid-19 como se muestra en la siguiente imagen, se puede observar a dos vacunados que su similaridad en creencias es muy cercana a 1.0, por lo que se puede decir que sus respuestas son muy similares a los demás. Por ejemplo, tomando al vacunado “enzyme breastwork 7883” su similaridad en creencia con respecto a los demás es de 0.07, el algoritmo realizó el análisis donde identifica que respuestas de las creencias fueron seleccionados por el vacunado “enzyme breastwork 7883” y cuales tiene el vacunado 2, para así sacar un promedio de similaridad de estos dos, luego el algoritmo procede a hacer el análisis del vacunado “enzyme breastwork 7883” con el siguiente vacunado y así sucesivamente con cada vacunado, para al final realizar un promedio general de la similaridad de creencia de los vacunados.

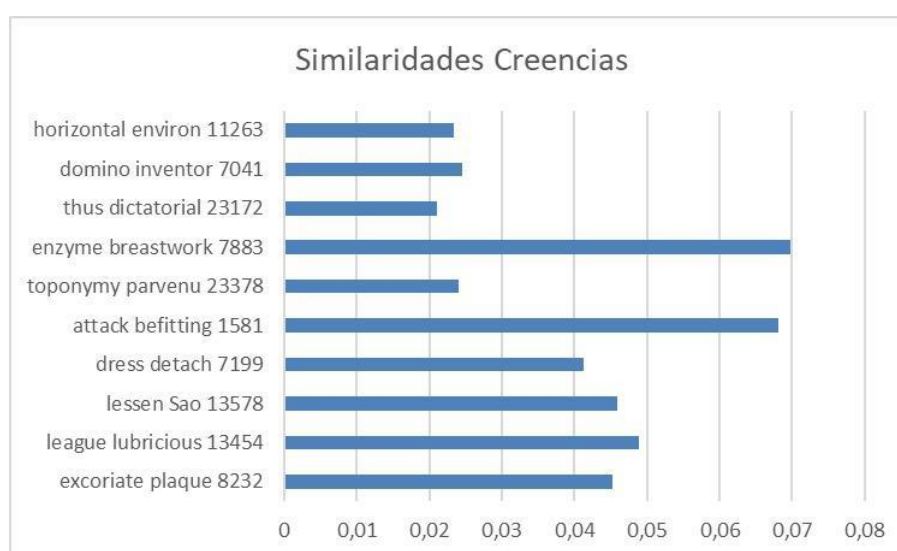


Imagen 26 Comparación similaridades de Vacunados con respecto a Creencias. Creación Propia.

3.6. Conclusiones

- Para la construcción ontología se tuvo en cuenta la metodología 101, la cual se realiza con base a los términos importantes del dominio elegido y también se hace uso de términos de otras ontologías ya existentes, realizar

este método del uso de términos facilitó y aceleró la construcción de la ontología.

- El uso de herramientas existentes como Protégé y Neo4j facilitó el desarrollo del proyecto. Protégé fue importante, debido a que facilitó la creación de la ontología con una estructura ya definida (Clases, subclases y atributos) y neo4j permitió la elaboración del grafo a partir de la ontología.
- El uso de algoritmos de grafos, en este caso de similitud; según el tipo de nuestro grafo. Fue importante para obtener datos relevantes de similitud de variables como: Efectos secundarios y creencias, relacionadas a las vacunas contra el COVID-19 de cada vacunado con la población en total de los vacunados.

4. CAPÍTULO 4: VALIDACIÓN DEL GRAFO DE CONOCIMIENTO

Para la validación del grafo presentado en capítulo 3, se realizó el desarrollo del prototipo en la sección 4.1. Seguido en la sección 4.2, se definen los aspectos para la evaluación de la calidad y efectividad del grafo de conocimiento. Y para terminar este capítulo, en sección 4.3 se definen las preguntas necesarias para medir la efectividad y calidad del grafo de conocimiento, mediante una comparativa entre la página web Mi Vacuna y la solución propuesta durante el proyecto.

4.1. Desarrollo del prototipo de visualización basado en el grafo

Para realizar el prototipo se tomó como referencia el grafo de conocimiento creado. En este apartado se muestra una serie de diagramas como: Diagrama de secuencias, diagrama de componentes y demás. Donde se evidencia la lógica del prototipo al momento de usarlo.

- **Diseño de Mockups**

En el diseño de los mockups se detalló la estructura de las interfaces necesarias para el prototipo. En la siguiente imagen se muestra el diseño para la interfaz de presentación, la cual contiene dos botones que redirigen al usuario a otras interfaces necesarias en el prototipo:

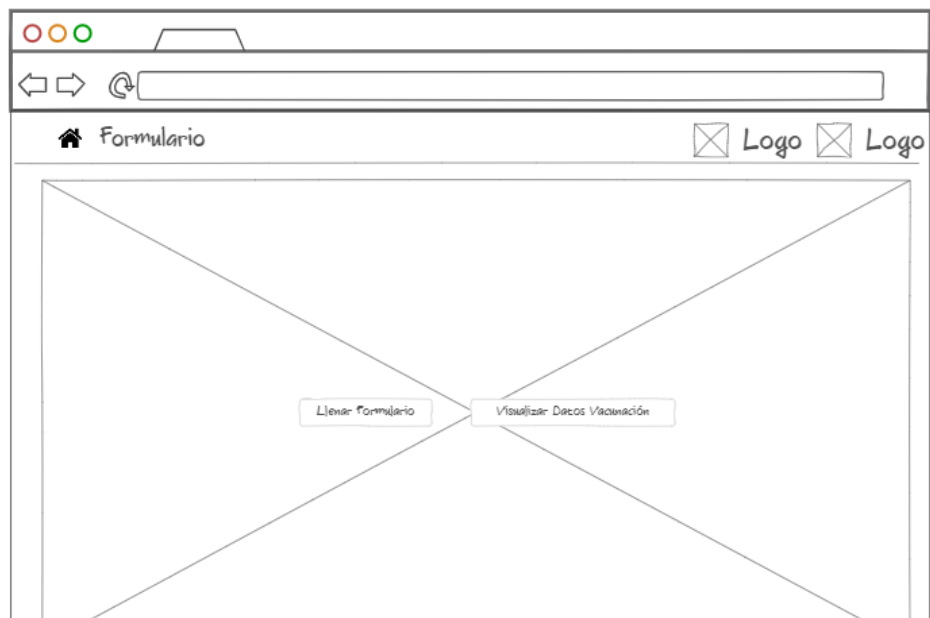


Imagen 27. Mockup de Interfaz Principal. Creación Propia

En el siguiente mockup, se muestra la interfaz necesaria para el diseño del formulario, la cual contiene cada una de las preguntas necesarias para el formulario, donde al final de la interfaz se guarda la información el grafo de conocimiento.

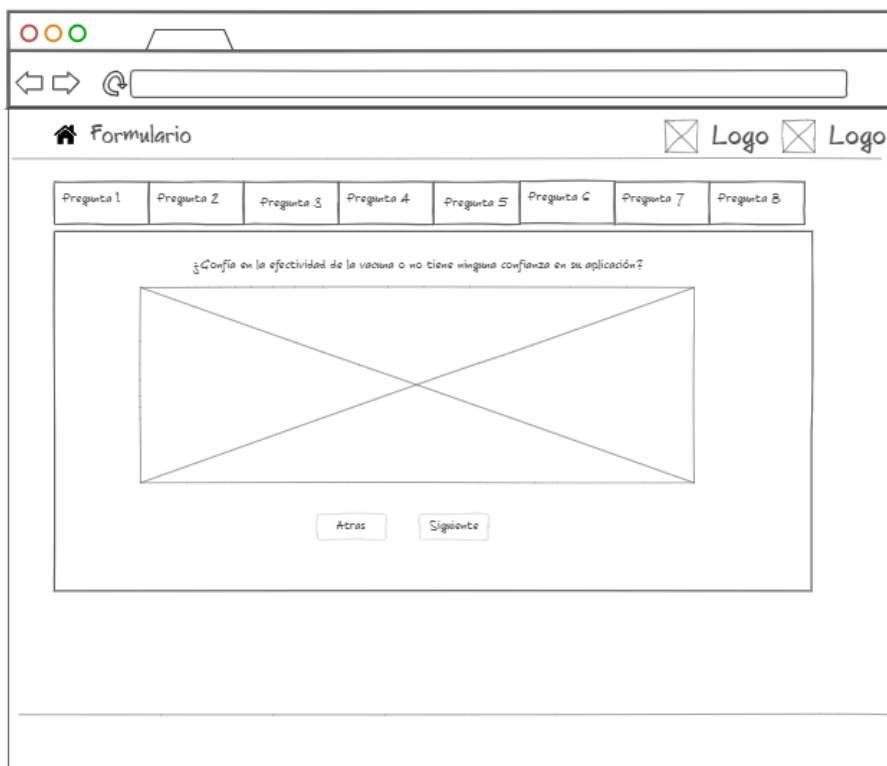


Imagen 28. Mockup prototipo de visualización. Creación Propia

El siguiente mockup va a enfocado al diseño de la interfaz de visualización de los datos de vacunación del usuario, la cual contiene tres secciones: la primera un grafo dinámico en el cual se muestran los nodos de los usuarios, la segunda una sección con las queries de consulta y por último una sección de resultados en los cuales el usuario podrá observar los datos de vacunación en específico.

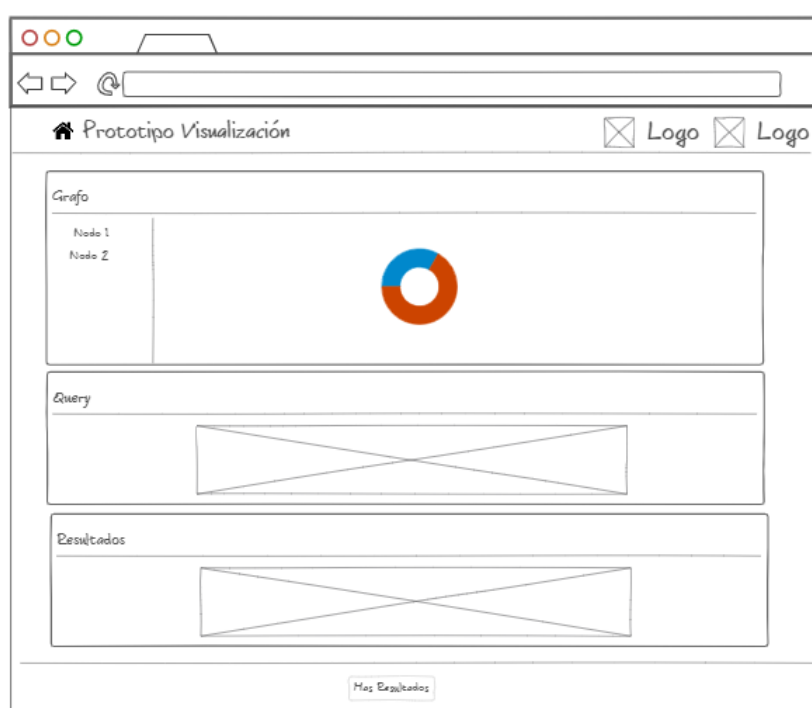


Imagen 29. Mockup prototipo de visualización. Creación Propia

- **Casos de Uso**

Los casos de uso del prototipo fueron elaborados para expresar los requisitos funcionales del prototipo, en la siguiente imagen se especifica cada una de las funcionalidades que pueden hacer los usuarios en el prototipo. Los casos de uso se dividen en cinco principales funcionalidades:

- El usuario puede ingresar información, para ello primero el usuario debe ingresar a la interfaz que contiene el formulario de

registro en el prototipo y por último diligenciar cada uno de los campos requeridos.

- El usuario podrá guardar la información diligenciada, previamente tendrá que haber diligenciado cada uno de los campos, donde al final del formulario, se guardará la información en el grafo de conocimiento.
- Sobre el grafo de conocimiento, se usó un algoritmo de grafo, en este caso el algoritmo de Similitud, en el cual se calcula la similitud de los usuarios con respecto a variables específicas, como las creencias y efectos secundarios frente a la vacuna.
- Por último, el usuario visualiza los resultados de la información, a través de una interfaz, la cual obtiene los datos previamente guardados en el grafo de conocimiento.

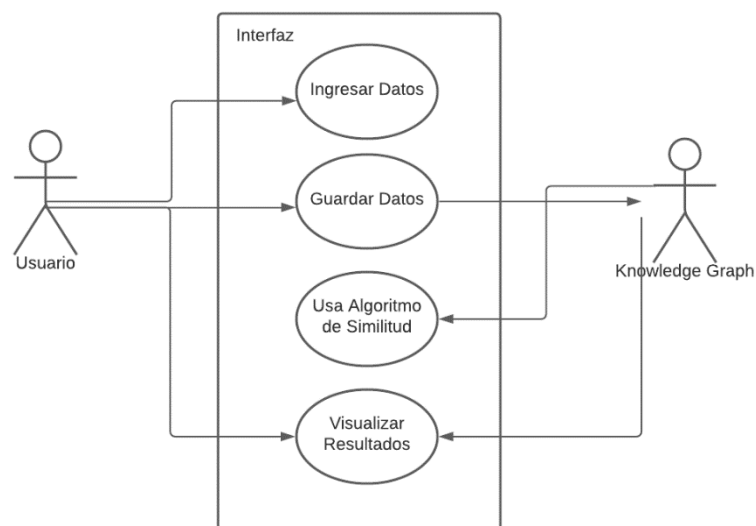


Imagen 30. Casos de uso. Creación Propia

• Diagrama de Componentes

En el siguiente diagrama se especifican cada uno de los componentes funcionales del prototipo. El diagrama se compone de dos endpoint

(para visualización de datos y registro de datos) por los cual el usuario puede acceder a las interfaces necesarias que ofrece en prototipo. Se compone de 3 paquetes de componentes:

- La interfaz de formulario, la cual contiene un componente encargado de guardar los datos que son registrados en el formulario. Los datos son enviados al componente de grafo de conocimiento.
- El componente de grafo de conocimiento se encarga de gestionar los datos que son enviados desde la interfaz de formulario, al igual que se encarga de enviar datos que son solicitados a la interfaz de prototipo de visualización.
- La interfaz de prototipo es la encargada de hacer las peticiones de información al grafo de conocimiento, en este caso la solicitud de nodos en específico, estos son entregados por el grafo y el componente de obtener el resultado es el encargado de visualizar los datos del grafo en la interfaz.

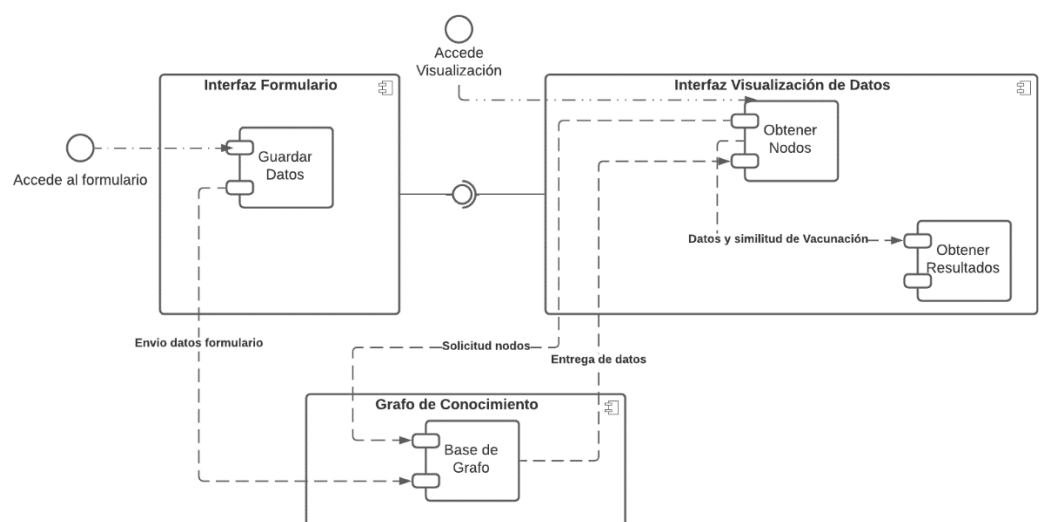


Imagen 31. Diagrama de Componentes. Creación Propia

- **Diagrama de Secuencia**

En el siguiente diagrama de secuencia se especifican los procesos que debe seguir un usuario con el uso del prototipo. Como primer paso el usuario accede a la interfaz de formulario, en la cual llena los datos requeridos por el formulario, al finalizar estos datos son enviados al grafo de conocimiento, el cual procesa y guarda los datos en la estructura definida en la ontología.

El usuario, accede a la interfaz de visualización, en la cual puede acceder a la información que brinda el prototipo seleccionando la información a visualizar. Se solicitan los datos seleccionados por el usuario al grafo de conocimiento, el cual retorna los datos solicitados y la interfaz se encarga de dar la visualización de los resultados.

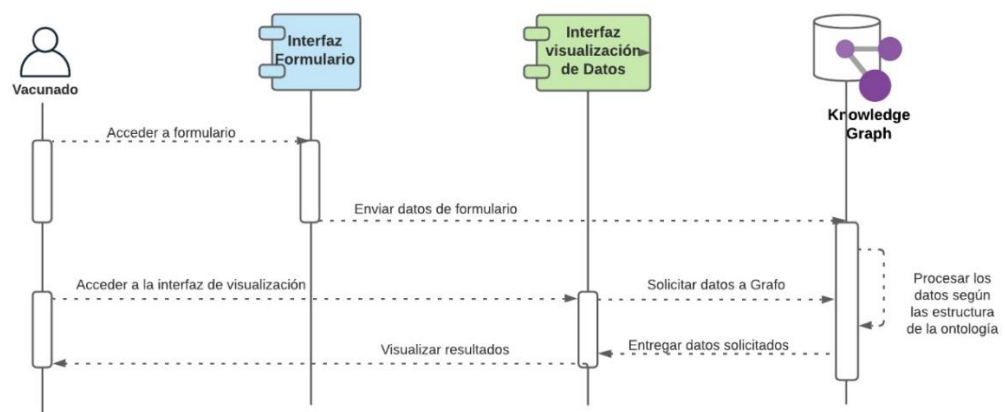


Imagen 32. Diagrama de Secuencia. Creación Propia

- **Formulario de Registro de Datos**

Para el desarrollo de la interfaz de registro de datos se usaron las siguientes herramientas: Material UI, React, Javascript, Npm (Node Js). La interfaz como se muestra a continuación, permite al usuario ingresar su información al gráfico de conocimientos a través de un formulario con

preguntas las cuales están relacionadas al proceso de vacunación contra el COVID-19. Esta interfaz hace uso del BackEnd para guardar los datos en el grafo de conocimiento.

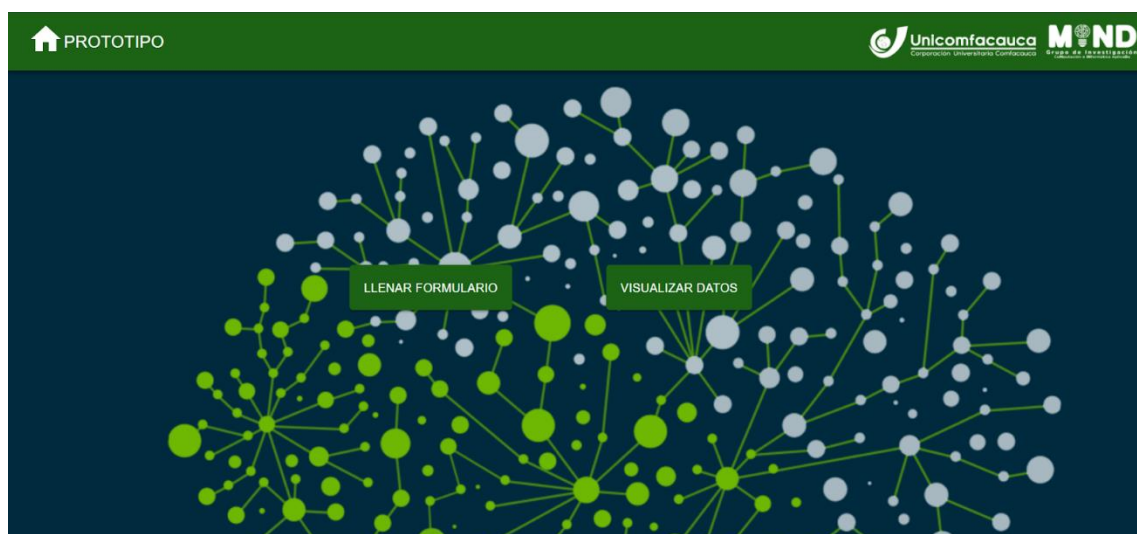


Imagen 33. Mockup prototipo de visualización. Creación Propia



Imagen 34. Interfaz de formulario. Creación Propia

- **Interfaz de Visualización de Datos**

Para la visualización de los datos guardados en el grafo de conocimiento, se creó la interfaz de visualización, con la ayuda de una librería llamada Popoto Js que trabaja con código HTML y Javascript. Popoto Js brinda una serie de herramientas, como un visualizador de

nodos provenientes del grafo, los cuales pueden seleccionar y obtener la información específica de cada usuario con sus respectivos datos de vacunación. Como valor agregado se muestra la similitud del usuario con respecto a las variables de creencias y efectos secundarios relacionados con la vacunación, las cuales están calculadas con el uso del algoritmo de Jaccard el cual compara los nodos del usuario con respecto a todos los usuarios del prototipo. A continuación, se muestra la interfaz de visualización de los datos del grafo de conocimiento:

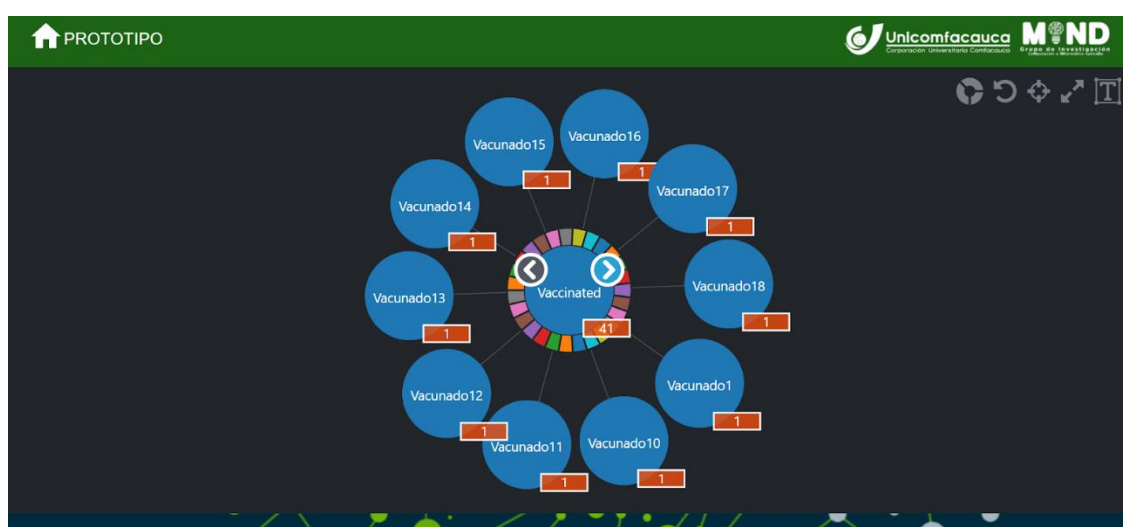


Imagen 35. Interfaz de Visualización. Creación Propia

También se añaden partes del código de BackEnd y FrontEnd esenciales para comprender su funcionamiento. El código completo se encuentra disponible en el siguiente repositorio [84].

Para hacer uso del grafo de conocimiento es necesario establecer una clase la cual maneja la conexión al grafo, donde requiere de una URL, la cual es brindada por Neo4j, un usuario y contraseña definidas por el administrador del grafo de conocimiento. Esto se hace con el uso del Driver de Python orientado al uso de Neo4j en este lenguaje de programación.

```
class Conexion(object):
    def __init__(self, uri, user, password):
        self._driver = GraphDatabase.driver(uri,auth=(user, password))
```

Imagen 36. Clase de Conexión al grafo de conocimiento. Creación Propia

Los parámetros que requiere la clase anterior se pasan de la siguiente forma:

```
from Neo4jService import Conexion

neo4j = Conexion("bolt://localhost:7687", 'neo4j', "12345")
```

Imagen 37. Parámetros de conexión a neo4j. Creación Propia

En la siguiente imagen se creó una función llamada nodo, que recibe los parámetros enviados desde el formulario. Aquí se definieron las queries necesarias para crear los nodos por cada parámetro recibido en el grafo de conocimiento.

```
def nodo(self, tx,TIden,Genero,Efec,EPS,Vacuna,LoteP,FechaDP,LoteS,FechaDS,TipVacunador,IPS,Regimen,
Empleo,Municipio,Barrio,Etnia,Discap,Desplaz,Name,Apell,Age):

    tx.run("MERGE (:Type_of_identification{Type_of_identification:$TIden})",TIden=TIden)
    tx.run("MERGE (:Name{Name:$Name})",Name=Name)
    tx.run("MERGE (:Surname{Surname:$Apell})",Apell=Apell)
    tx.run("MERGE (:Age{Age:$Age})",Age=Age)
    tx.run("MERGE (:Residence_Area{Residence_Area:'Rural'}) MERGE (:Residence_Area
{Residence_Area:'Urbano'})")
    tx.run("MERGE (:State{State:'Cauca'})")
    tx.run("MERGE (:Gender{Gender:$Genero})",Genero=Genero)
    tx.run("MERGE (:Confidence{Confidence:$Efec})",Efec=Efec)
    tx.run("MERGE (:Name_insurance_company{Name_insurance_company:$EPS})",EPS=EPS)
    tx.run("MERGE (:Name_Vaccine{Name_Vaccine:$Vacuna})",Vacuna=Vacuna)
    tx.run("MERGE (n:Dose_Number{Dose_Number:'+"1"+"}) MERGE(s:Dose_Number{Dose_Number:'+"2"+"})")
    tx.run("MERGE (:Lot{Lot:$LoteP})",LoteP=LoteP)
    tx.run("MERGE (:Vaccination_date{Vaccination_date:$FechaDP})",FechaDP=FechaDP)
    tx.run("MERGE (:Lot{Lot:$LoteS})",LoteS=LoteS)
    tx.run("MERGE (:Vaccination_date{Vaccination_date:$FechaDS})",FechaDS=FechaDS)
    tx.run("MERGE (:Profession{Profession:$TipVacunador})",TipVacunador=TipVacunador)
    tx.run("MERGE (:Name_company_org{Name_company_org:$IPS})",IPS=IPS)
    tx.run("MERGE (:Membership_regimen{Membership_regimen:$Regimen})",Regimen=Regimen)
    tx.run("MERGE (:Employment_group{Employment_group:$Empleo})",Empleo=Empleo)
    tx.run("MERGE (:Municipality{Municipality:$Municipio})",Municipio=Municipio)
    tx.run("MERGE (:Neighborhood{Neighborhood:$Barrio})",Barrio=Barrio)
    tx.run("MERGE (:Ethnic_group{Ethnic_group:$Etnia})",Etnia=Etnia)
    tx.run("MERGE (:Condition_of_disability{Condition_of_disability:$Discap})",Discap=Discap)
    tx.run("MERGE (:Condition_of_displacement{Condition_of_displacement:$Desplaz})",Desplaz=Desplaz)
```

Imagen 38. Función Nodo, crea nodos en el grafo de conocimiento. Creación Propia

En la siguiente imagen se realiza un `session.write_transaction()`, `write_transaction` ejecuta sentencias del lenguaje Cypher para sobrescribir el grafo, en este caso se ejecuta la función **nodo**, la cual requiere de todos los datos que son generados desde el formulario y `session` es el contenedor de conexión que permite ejecutar las transacciones necesarias.

```
session.write_transaction(neo4j.nodo,TIden,Genero,Efec,EPS,Vacuna,LoteP,FechaDP,LoteS,FechaDS,TipVacunador,IPS,Regimen,Empleo,Municipio,Barrio,Etnia,Discap,Desplaz,Name,Apell,Age)
```

Imagen 39. Transacción que pasa datos como parámetros a la función nodo. Creación Propia

Para calcular la similitud se creó una función que permita realizar la query de similitud del usuario, en la cual recibe como parámetro el nombre del usuario que se crea al momento de guardar la información del formulario, con ese parámetro se ejecuta la query la cual retorna el cálculo de la similitud. Al final ese resultado es retornado por la función en la variable de `similarity`.

```
def similiaridadBelief(self, tx, aleatVac):
    query = ("MATCH (n:Vaccinated{Vaccinated:$aleatVac}) -[:has]->(s:Belief)
    WITH n, collect(id(s)) AS n1Creencias MATCH (v2:Vaccinated)-[:has] ->
    (Belief) WHERE n <> v2 WITH n, n1Creencias, v2, collect(id(Belief)) AS
    n2Creencias RETURN AVG(gds.alpha.similarity.jaccard(n1Creencias,
    n2Creencias)) AS similarity")
    result = tx.run(query, aleatVac=aleatVac)
    return [record["similarity"] for record in result]
```

Imagen 40. Función de Similaridad. Creación Propia

Con la similaridad ya obtenida como se especificó anteriormente, este resultado es pasado a una nueva función para proceder a crear los nodos de similaridad y relaciones con el vacunado.

```
def similaridad (self,tx,resultado,aleatVac):
    tx.run("MERGE (:Similarity_B{Similarity_B:$resultado})",
    resultado=resultado)
    tx.run("MATCH (n:Similarity_B),(c:n4sch__Property),(s:Vaccinated) where s.
    Vaccinated='"+aleatVac+"'AND c.n4sch__name= 'Belief' AND n.Similarity_B='"+
    +resultado+"' MERGE (s)-[:has]->(n) MERGE (c)-[:is]->(s)",
    resultado=resultado,aleatVac=aleatVac)
```

Imagen 41. Función de similaridad en el grafo de conocimiento. Creación Propia

Por último, en la siguiente imagen se realiza un session.write_transaction con la función especificada en la imagen 36, el resultado se pasa a la variable w, este resultado lo convertimos a string y se guarda en la matriz w2 y se toma la primera posición de w2 en la variable resultadow para poderlo pasar al session.write_transaction de la función especificada en la imagen 37.

```
r1=[str(resultadoPantalla) for resultadoPantalla in resultadoPantalla] #r1 lista de nombres de
vacunados
print (r1)
for lista in r1:
    print(" ")
    print("-----SIMILARIDAD-----")

    print (lista) # lista es cada uno de los vacunados recorridos por el for
    w=session.write_transaction(neo4j.similiaridadBelief,lista)#realiza similaridad de creencias y
    guarda resultados en w
    z=session.write_transaction(neo4j.similaridadSintoma,lista)#realiza similaridad de efectos
    secundarios y guarda resultados en z
    w2=[str(w) for w in w]# w2 convierte resultados en w a String
    z2=[str(z) for z in z]
    resultadow=w2[0]# Resultadow toma la primera posición de w2
    resultadoz=z2[0]
    print("similaridad_creencia "+lista+": "+resultadow)
    print("similaridad_sintoma "+lista+": "+resultadoz)
    session.write_transaction(neo4j.similaridad,resultadow,lista)# funcion similaridad recibe los
    parametros de resultadow y el nombre recorrido por el
    session.write_transaction(neo4j.similaridad2,resultadoz,lista)
```

Imagen 42. Transacción de similaridad. Creación Propia

Con el uso de la herramienta Popoto se realiza una consulta al gráfico de conocimiento de los usuarios existentes en el parámetro llamado *LISTA*, la información traída por Popoto es guardada en un array llamado *vacunado*.

```


.append("h5").text(function (result) {
  lista = result.attributes.Vaccinated;
  Vacunado.push(lista)
  return lista;
});


```

Imagen 43. Mapeo de los vacunados existentes en la visualización. Creación Propia

Para el diseño de la interfaz de visualización de datos, se creó una función la cual, ejecuta una query en la que se obtienen los datos del usuario con sus nombres y propiedades, la query recibe como parámetro de búsqueda el nombre del usuario seleccionado desde la interfaz después de que se haya inicializado, la información que se obtiene de la query se guarda en un Array, en este caso denominada *Seleccionado*.

```

function Query (tx, ultimoObj){
  return tx.run('match (v1:Vaccinated{Vaccinated:$Vaccinated})-[r]->(a) with keys(a) as k1, a unwind k1 as k2 return k2 as Prop, a[k2] as value',
    {Vaccinated: ultimoObj}).then(result => {
      result.records.forEach(record => {
        var Propiedad = record.get('Prop')
        var informacion = record.get('value')
        Seleccionado.push(Propiedad + ":" + informacion)
      })).catch(error => {
        console.log(error)});
    }
    console.log(Seleccionado);
    },2000);
    setTimeout(()=>{Seleccionado.length=0;},000);

```

Imagen 44. Función de solicitud de información al grafo de conocimiento. Creación Propia

Con la información del usuario seleccionado almacenada en el Array, se procede a mostrar la información en la interfaz de prototipo de la siguiente forma.

```

setTimeout(()=>{
  Seleccionado.forEach(function(x){
    div.append("h5").text(function (result) {
      listaVacunado = result.attributes.Vaccinated;
      return x;
    });
  });
}, 6000)

```

Imagen 45. Mapeo de datos seleccionado por el usuario en pantalla. Creación Propia

4.2. Aspectos para la validación de la calidad de los datos en el grafo de conocimiento

Para el correcto funcionamiento del prototipo es necesario evaluar la calidad del grafo de conocimiento. Un grafo de conocimiento de baja calidad produce una aplicación de baja calidad basada en el mismo. Un grafo de conocimiento de alta calidad debe ser correcto, completo y actualizado. En [85] proponen un framework de evaluación de calidad para un grafo de conocimiento, diseñado para evaluar el "ajuste para el propósito" de un grafo de conocimiento para construir una aplicación basada en conocimiento.

A continuación, se enumeran los 10 requisitos sobre la calidad que debe tener cualquier grafo de conocimiento:

- **Los triples deben ser concisos.**

Las tripletas concisas permiten una mejor definición de cada uno de los objetos, relaciones y sujetos, dentro de la base de conocimiento para permitir un mejor análisis y organización de los datos.

- **Se debe capturar la información contextual de las entidades.**

La información o datos que se generan son concisos y precisos al momento de ser capturados en entidades o propiedades. Al momento de hacer consultas en el grafo es más sencillo y rápido, ya que la información está estructurada contextualmente en entidades.

- **El grafo de conocimiento no contiene triples redundantes.**

El grafo de conocimiento al tener una base ontológica está diseñado para no contener triples redundantes. Esto se debe a que en la ontología se especifican los triples básicos necesarios para evitar la redundancia en su definición.

- **El grafo de conocimiento se puede actualizar dinámicamente.**

El grafo de conocimiento permite su actualización dinámica, la información de los nodos se actualiza cada vez que se ingresa nueva información de los usuarios desde el formulario al grafo de conocimiento.

- **Las entidades deben estar densamente conectadas.**

Debido al uso de la ontología, las entidades definidas están conectadas a través de grupos, y existen conexiones entre cada uno de estos grupos. Por ejemplo, la información básica del usuario como nombre, apellido y edad que pertenece a un grupo denominado persona está conectada directamente con la información de un grupo llamado vacuna.

- **Deben incluirse las relaciones entre diferentes tipos de entidades.**

En el grafo de conocimiento se pueden encontrar relaciones entre las múltiples entidades, toda la información de los usuarios está relacionada entre sí, sin omitir ninguno de los datos enviados desde el formulario.

- **La fuente de datos debe ser de varios campos.**

La fuente de datos debe contener datos que son solicitados al usuario, para la definición de estos campos se tuvo en cuenta estándares como el HL7 FHIR para definir algunos de ellos.

- **La escalabilidad con respecto al tamaño de KG.**

El grafo de conocimiento permite aumentar su tamaño si se agregan más grupos de información relacionados al mismo campo, en este caso la salud, al igual que su estructura se puede modificar para contener más información si así se requiere.

- **Los atributos de las entidades no deben perderse.**

Los atributos de las entidades no se pueden modificar, debido a que en el BackEnd no se especifica ninguna modificación de los atributos, hasta que no haya una query que indique algún tipo de modificación a los atributos, esto no sucederá.

- **Para tareas específicas de dominio, el gráfico de conocimiento debe estar relacionado con ese campo.**

El dominio de vacunación está directamente relacionado con la temática del grafo, ya que las definiciones utilizadas son estrictamente de vacunación contra el COVID-19. Cada tripleta en el grafo contiene la información del dominio.

4.3. Validación de la calidad y efectividad de la visualización de información

Para la validación de la calidad y efectividad del prototipo creado, se realizó comparando otra página web con el mismo objetivo, para este caso se tomó algunos aspectos de la página Mi Vacuna, encargada de generar los certificados de vacunación en el país. Para medir la efectividad del prototipo es necesario obtener los mismos resultados de una página web enfocada el mismo dominio contando con menos recursos.

Para esta evaluación se realizó una encuesta comparativa para medir la efectividad y calidad entre el prototipo y la página web mi vacuna a personas con experiencias en software.

Para la encuesta, tomando como referencias las preguntas creadas con respecto a la calidad de grafos y preguntas relacionadas con la efectividad de aplicativos webs, se crearon las siguientes preguntas.

Efectividad: Basado en una encuesta para evaluar la experiencia de servicios de una página web [86], se realizaron las siguientes preguntas enfocadas a medir la efectividad de la visualización de datos. Las preguntas se realizaron de acuerdo a como el grafo entrega la información a ser visualizada por parte del usuario y la interacción que el usuario tiene con el aplicativo web.

- ¿Se siente satisfecho o cómodo con la visualización de la aplicación web?
- ¿La información que se visualiza en el aplicativo web es fácil de entender?
- ¿La forma en que se visualiza la información es completamente clara?
- ¿La forma en que se visualiza la información con respecto a cada uno de los componentes del aplicativo web (interfaces) es sencilla?

Calidad:

- ¿La forma en que se organizan los datos es concisa en la aplicación?
- ¿Los datos que entrega la aplicación son claros y precisos?
- ¿Considera que la información debe ser actualizable?
- ¿La información brindada tiene coherencia con el objetivo de la página?
- ¿La información que ingresó no ha sufrido algún tipo de modificación?
- ¿Todos los campos son totalmente necesarios para el objetivo de la página?
- ¿Considera que los datos manejados en la aplicación son suficientes?
- ¿Estás de acuerdo a que a futuro se puedan realizar cambios a la información del aplicativo web?
- ¿Considera que es importante que la información deba ser actualizable?

Las anteriores preguntas fueron aplicadas en formato encuesta a personas relacionadas con el tema.

4.4. Resultados

De las preguntas realizadas en la sección 4.3, se realizaron 2 encuestas enfocadas a evaluar el prototipo de visualización de datos de vacunación y así como la página del gobierno nacional llamada Mi Vacuna.

El formato de la mayoría de preguntas es de tipo Likert, con un total de 13 preguntas relacionadas a la efectividad y calidad del grafo de conocimiento.

La encuesta fue realizada por un total de 10 personas, donde evaluaron ambos aplicativos webs con las mismas preguntas, con el objetivo de comparar la efectividad para la visualización de los datos. Se encuestó a personas con conocimientos en aplicativos webs debido a que estas personas han tenido experiencia en el uso de los mismos, esto con el fin de evaluar la calidad y la efectividad (eficiencia y eficacia) del grafo de conocimiento desarrollado.

4.4.1. Análisis de resultados encuesta de evaluación

A continuación se describe cada una de las preguntas y sus respectivos resultados:

- **¿Se siente satisfecho o cómodo con la visualización de la aplicación web?**

La aplicación de las encuestas arrojó como resultado que los usuarios están entre totalmente de acuerdo y de acuerdo en un 70% en el uso del prototipo, mientras que un 50% se siente totalmente de acuerdo con el uso del aplicativo de Mi vacuna. En el prototipo se obtuvo un mayor grado de satisfacción con la visualización de los datos.

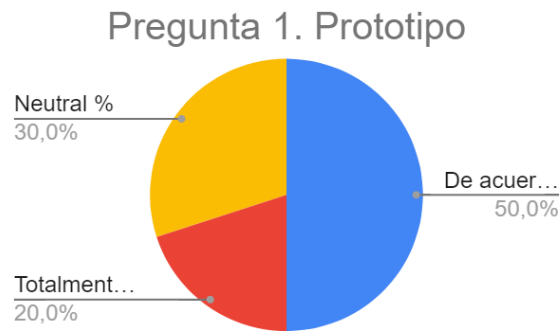


Imagen 46. Resultado Prototipo Pregunta 1. Creación Propia.

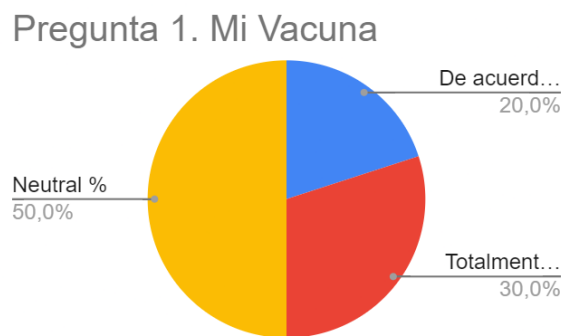


Imagen 47. Resultado Mi Vacuna Pregunta 1. Creación Propia

- **¿La información que se visualiza en el aplicativo web es fácil de entender?**

Los usuarios encuestados están de acuerdo un 70% en que usando el prototipo la información que se visualiza es fácil de entender, mientras que el 90% están de acuerdo que la información es fácil de entender con la página web de Mi Vacuna. En ambas aplicaciones se considera que es fácil de entender la información a pesar de que en la página web de mi Vacuna solo un 10% de los encuestados está en desacuerdo.

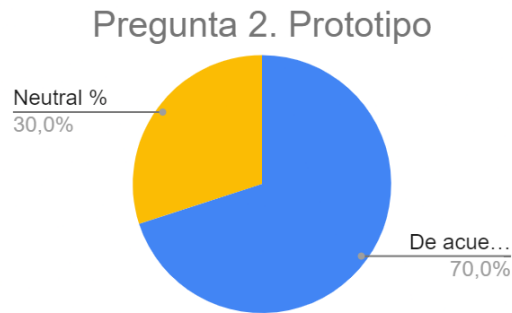


Imagen 48. Resultado Prototipo Pregunta 2. Creación Propia

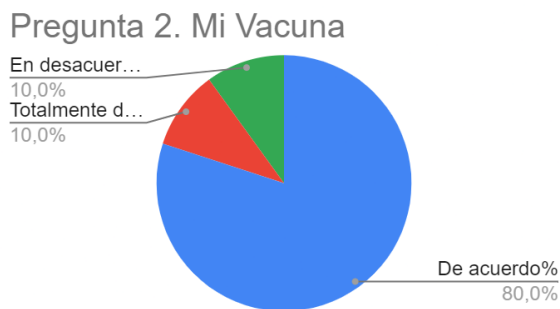


Imagen 49. Resultado Mi Vacuna Pregunta 2. Creación Propia

- **¿La forma en que se visualiza la información es completamente clara?**

En la claridad de la información que se visualiza los usuarios encuestados están de acuerdo un 70%. Un 90 % de los usuarios están de acuerdo con que la información que se visualiza en la página de web de mi vacuna es completamente clara. Los usuarios consideran que la información es completamente clara usando la página web de Mi vacuna, pero un 30% está totalmente de acuerdo con que mejor lo hace el prototipo.

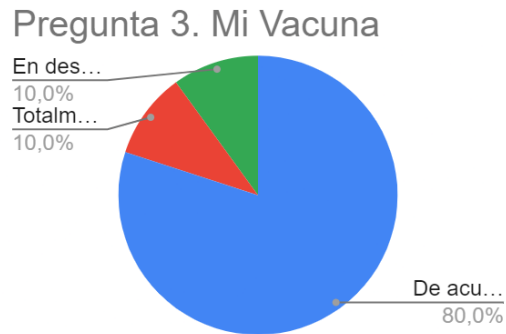


Imagen 50. Resultado Prototipo Pregunta 3. Creación Propia

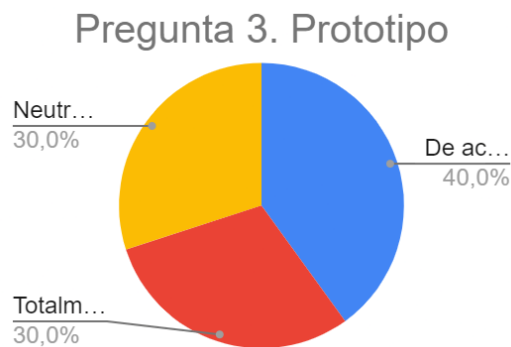


Imagen 51. Resultado Mi Vacuna Pregunta 3. Creación Propia

- **¿La forma en que se visualiza la información con respecto a cada uno de los componentes del aplicativo web (interfaces) es sencilla?**

El 90% de los usuarios están de acuerdo que las interfaces desarrolladas para el aplicativo brindan una visualización sencilla con respecto a la página web de Mi vacuna, ya que solo el 80% de los encuestados están de acuerdo con que la visualización a través de las interfaces es sencilla.

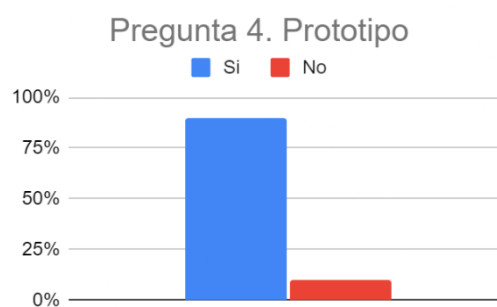


Imagen 52. Resultado Prototipo Pregunta 4. Creación Propia

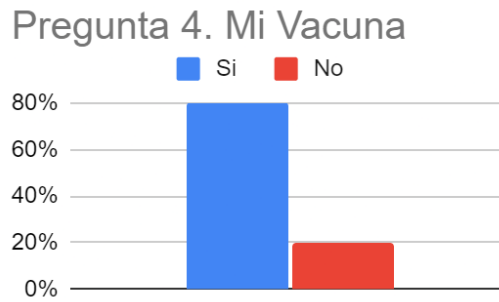


Imagen 53. Resultado Mi Vacuna Pregunta 4. Creación Propia

- **¿La forma en que se organizan los datos es concisa en la aplicación?**

La forma en como el prototipo organiza la información para el 90% de los encuestados están de acuerdo que es concisa, mientras que solo el 70% de los encuestados están de acuerdo con que la información es concisa en la página web de Mi Vacuna. Un 10% de los encuestados está en desacuerdo con la organización de los datos en Mi Vacuna. Por lo anterior el prototipo organiza mejor los datos.

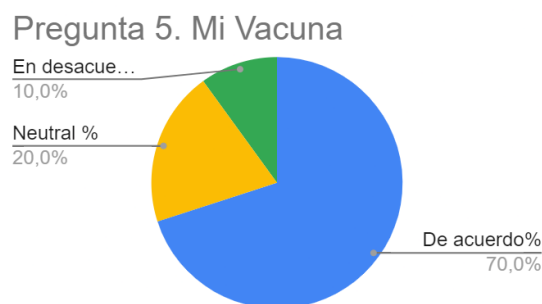


Imagen 54. Resultado Prototipo Pregunta 5. Creación Propia

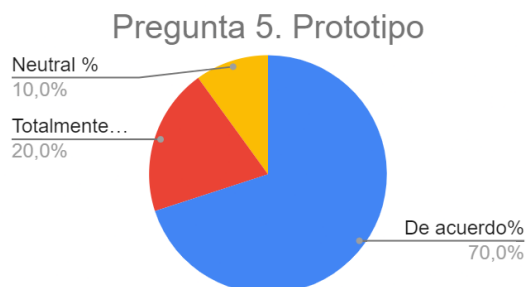


Imagen 55. Resultado Mi Vacuna Pregunta 5. Creación Propia

- **¿Los datos que entrega la aplicación son claros y precisos?**

Los encuestados consideran que los datos que entregan ambas soluciones son claros y precisos. En ambas encuestas el 70% de los encuestado está de acuerdo que los datos brindados por ambas soluciones son claros y precisos.

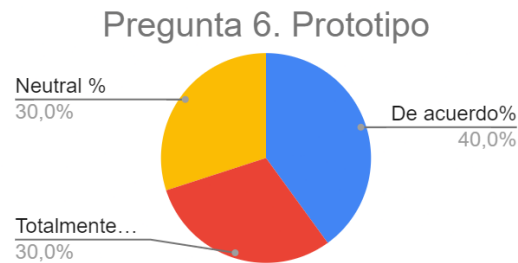


Imagen 56. Resultado Prototipo Pregunta 6. Creación Propia

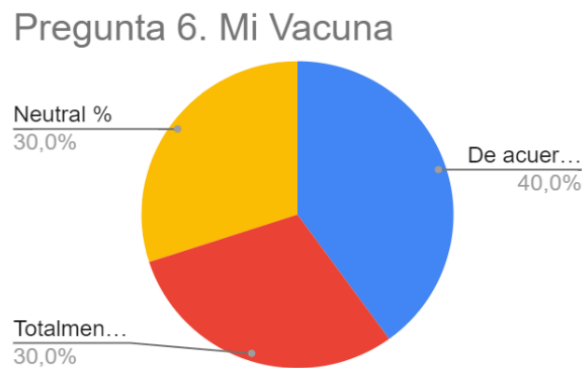


Imagen 57. Resultado Mi Vacuna Pregunta 6. Creación Propia

- **¿Considera que la información debe ser actualizable?**

Un 100% de los encuestados consideran que la información debe ser actualizable en el prototipo, mientras que el 90% de los encuestados considera que la información debe ser actualizable en la página web de Mi Vacuna. Es decir que los datos sean actualizables, en el caso del prototipo lo hace con el cálculo de la similitud de cada usuario, mientras que en la página web de mi vacuna los datos de registro son estáticos, ya que genera un certificado no modificable.

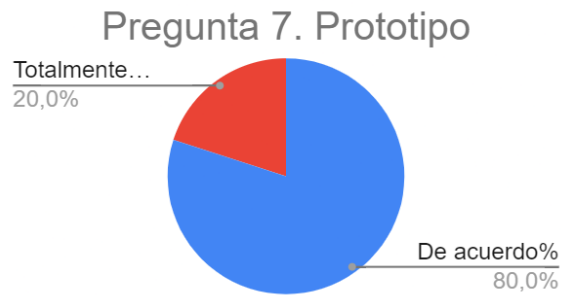


Imagen 58. Resultado Prototipo Pregunta 7. Creación Propia

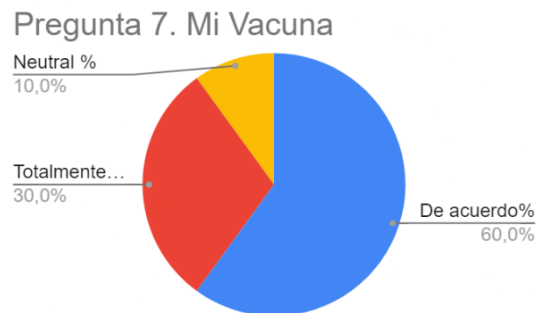


Imagen 59. Resultado Mi Vacuna Pregunta 7. Creación Propia

- **¿La información brindada tiene coherencia con el objetivo de la página?**

Se puede observar que el 90% de los usuarios están de acuerdo que el prototipo presenta información coherente al objetivo de la página, y el 100% piensa que la página web Mi Vacuna es totalmente coherente con su objetivo. Por lo tanto, el prototipo puede mejorar en esta parte.

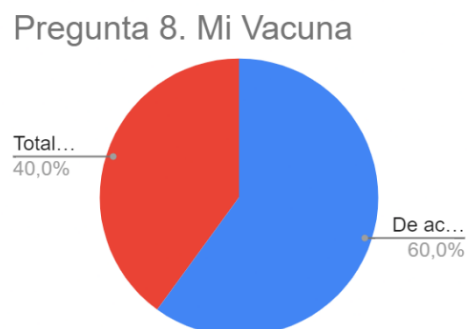


Imagen 60. Resultado Prototipo Pregunta 8. Creación Propia

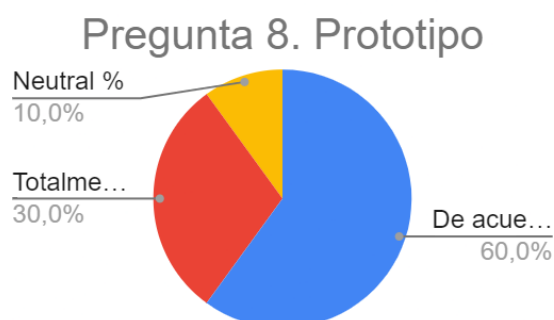


Imagen 61. Resultado Mi Vacuna Pregunta 8. Creación Propia

- **¿La información que ingresó no ha sufrido algún tipo de modificación?**

De los participantes de la encuesta un 20% ha dicho que ha sufrido modificaciones en la información en la página de Mi Vacuna, siendo esto algo totalmente desfavorable para esta página web, por lo tanto, el prototipo obtuvo un 90% de usuarios que están de acuerdo que su información ingresada en el prototipo no ha sufrido modificación.

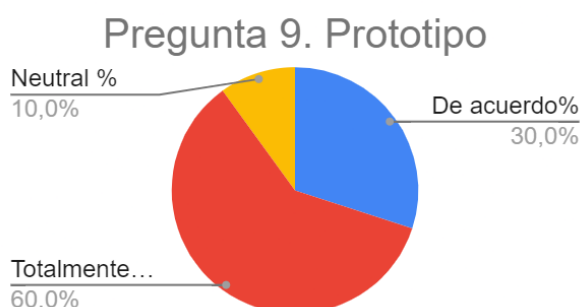


Imagen 62. Resultado Prototipo Pregunta 9. Creación Propia

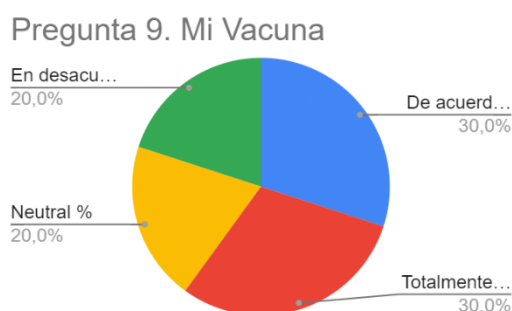


Imagen 63. Resultado Mi Vacuna Pregunta 9. Creación Propia

- **¿Todos los campos son totalmente necesarios para el objetivo de la página?**

De la encuesta realizada se puede observar que los usuarios están de acuerdo con un 80% que los campos presentados por el prototipo son necesarios para su objetivo, sin embargo, en la página de Mi Vacuna hay un 10% de personas que están totalmente de acuerdo, pero también se observa un 10% de personas que están totalmente en desacuerdo afectando negativamente el porcentaje anterior, por lo tanto, se puede decir que solo el 60% de los usuarios están de acuerdo.

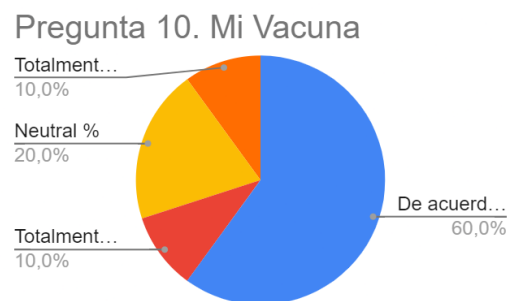


Imagen 64. Resultado Prototipo Pregunta 10. Creación Propia

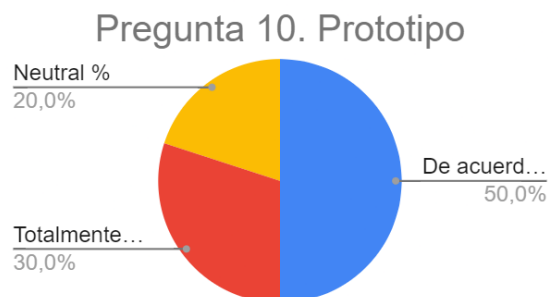


Imagen 65. Resultado Mi Vacuna Pregunta 10. Creación Propia

- **¿Considera que los datos manejados en la aplicación son suficientes?**

En la encuesta realizada, se puede observar que los datos manejados por el prototipo son suficientes, por una diferencia del 20% con la página web “Mi Vacuna”, aunque la diferencia de los usuarios que están de acuerdo es del 10% entre el prototipo y la página web, la página web tiene un 10% de usuarios que están en desacuerdo con que los datos presentados por la página sean suficientes. Esto afecta negativamente a la página web: Mi Vacuna, debido a que el prototipo puede estar presentando información más específica, fácil de entender e informativa para los usuarios.

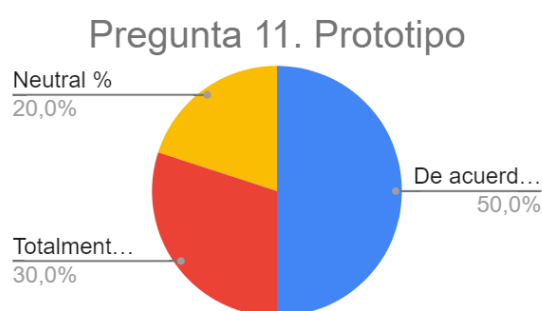


Imagen 66. Resultado Prototipo Pregunta 11. Creación Propia

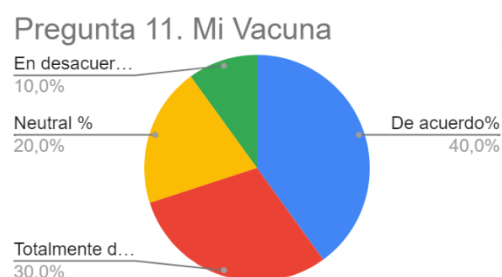


Imagen 67. Resultado Mi Vacuna Pregunta 11. Creación Propia

- **¿Estás de acuerdo a que a futuro se puedan realizar cambios a la información del aplicativo web?**

Los usuarios consideran que es más importante realizar cambios en el prototipo que en la página web mi vacuna, aunque en ambas encuestas los resultados muestran que el 90% de los usuarios están de acuerdo a

cambios futuros, la diferencia en porcentajes de usuarios que están “totalmente de acuerdo” es muy diferente.

Pregunta 12. Mi Vacuna

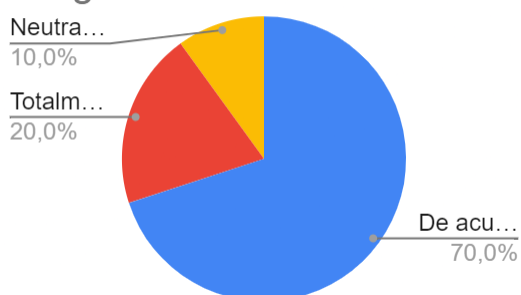


Imagen 68. Resultado Prototipo Pregunta 12. Creación Propia

Pregunta 12. Prototipo

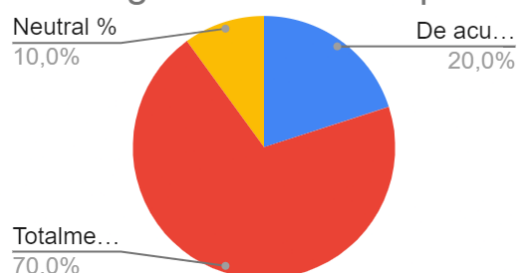


Imagen 69. Resultado Mi Vacuna Pregunta 12. Creación Propia

- **¿Considera que es importante que la información deba ser actualizable?**

Los usuarios encuestados consideran que es importante que la información sea actualizable. En ambas encuestas el 60% consideran que es muy importante la actualización de la información, y el resto aun en menor cantidad consideran que es importante la actualización de información.

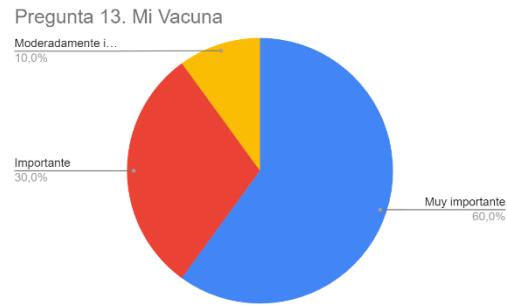


Imagen 70. Resultado Prototipo Pregunta 13. Creación Propia

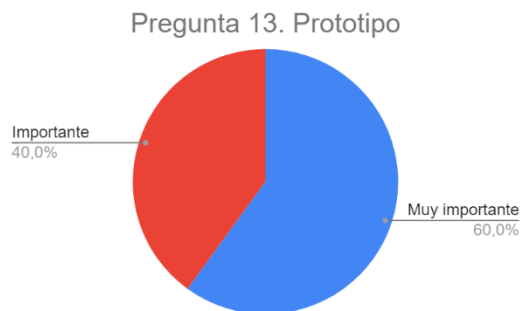


Imagen 71. Resultado Mi Vacuna Pregunta 13. Creación Propia

4.5. Discusión

En los resultados se pudo evidenciar que en un ámbito tecnológico la visualización de los datos apoyados por el uso de un grafo de conocimiento, aporta un valor agregado al análisis de la información. El prototipo recolecta la información de las personas vacunadas, así como también permita el cálculo de la similitud de ciertas variables como las creencias y efectos secundarios relacionados a las vacunas contra el COVID-19, hace que se obtenga el respectivo análisis gracias al uso de algoritmos de grafos, como es el caso del uso del algoritmo de similitud de Jaccard, es similar a lo encontrado en [70], se usan los grafos de conocimientos a partir de datos recolectados de registros médicos en salud HCE, para proporcionar información para un análisis más útil en la similitud de pacientes.

El uso de interfaces en el prototipo creado apoyó a la extracción de información referente a lo que los usuarios en general necesitan conocer sin tener

conocimientos en lenguajes de consultas o programación, como es el caso de la visualización de la información referente al proceso de vacunación, como se vio también en [88] donde se usó un grafo de conocimiento apoyado por interfaces, lo cual facilitó la extracción de información y por ende a mejorar la interpretación de los datos consultados.

El campo del uso de los grafos de conocimiento está en pleno desarrollo por lo cual como se evidencio en el mapeo sistemático no hay muchos artículos de investigación relacionados al tema y más en el campo de la salud.

4.6. Conclusiones y trabajo futuro

- La visualización permitió entender de manera más sencilla la gran cantidad de información que se almacena, para personas que no tienen ningún conocimiento en el manejo de esta. Esto generalmente no se logra con el desarrollo de aplicaciones que se generan dentro del contexto.
- Las diferentes fuentes de información fueron realmente importantes para la construcción de la ontología, siendo esta clave para la construcción del grafo de conocimiento.
- Con la construcción del grafo de conocimiento basado en el dominio de vacunación contra el COVID-19, permitió organizar los datos de forma más concisa evitando redundancia en la información, como se puede obtener de una consulta a una base de datos SQL, sin afectar la eficiencia de las consultas.
- Con la definición y el uso del algoritmo de similitud aplicado con la herramienta NEO4J se hizo comprensible el valor resultante de similitud

que tienen cada uno de los usuarios entre sí, respecto a las dos variables aplicadas: Los efectos secundarios generados por la vacuna y las creencias que tienen los usuarios frente a la vacuna

- Con los resultados de la implementación de la encuesta, se pudo evidenciar que el prototipo es lo suficientemente capaz de cumplir con el mismo objetivo de la página web Mi Vacuna, teniendo en cuenta que el prototipo solicita información en una cantidad menor a la página web.
- Haciendo uso del grafo de conocimiento como almacenamiento de datos y a través de un prototipo se permitió solucionar el problema identificado en cuanto a la falta de visualización de información médica bajo las definiciones de un mismo estándar, en este caso HL7 FHIR.
- El desarrollo del proyecto permitió dar como aporte una estandarización para el proceso de vacunación contra el COVID -19 gracias al uso de diferentes variables relacionadas al proceso de vacunación con el uso de recursos del estándar HL7 FHIR, también se hizo aporte de un grafo de conocimiento definido por una ontología dirigida al manejo de la información referente al dominio seleccionado.

Trabajo Futuro

Como trabajo futuro queda la implementación completa del prototipo utilizando servicios HL7 FHIR, así como la implementación de un login en el que los usuarios puedan registrarse y ver sus datos de forma privada. También el ampliar la información que recibe el grafo en distintos contextos relacionados con la salud o en distintos dominios de información.

REFERENCIAS

- [1] "Portabilidad - EcuRed." <https://www.ecured.cu/Portabilidad> (accessed Apr. 29, 2021).
- [2] "Portabilidad." <https://iso25000.com/index.php/normas-iso-25000/iso-25012/116-portabilidad> (accessed May 17, 2021).
- [3] Congreso de la Republica de Colombia, "Ley 2015 del 31 de enero de 2020," *Congreso de Colombia*, p. 7, 2020.
- [4] "Health Level Seven International - Homepage | HL7 International." <https://www.hl7.org/> (accessed Nov. 25, 2021).
- [5] "El estándar HL7 - Apuntes de Electromedicina Xavier Pardell." <https://www.pardell.es/el-estandar-hl7.html> (accessed Apr. 29, 2021).
- [6] "La verdadera historia de hospitales y clínicas al borde de la quiebra - Salud - EL TIEMPO.COM." <https://www.eltiempo.com/salud/la-verdadera-historia-de-hospitales-y-clinicas-al-borde-de-la-quiebra-247988> (accessed Apr. 29, 2021).
- [7] "Transformación social y grafos de conocimiento | CTIC." <https://www.fundacionctic.org/es/actualidad/transformacion-social-y-grafos-de-conocimiento> (accessed Apr. 29, 2021).
- [8] E. Berrocal De Luna, J. E. López, F. De, C. De, L. A. Educación, and U. de Granada, "UNIDAD 3. EL PROCESO DE INVESTIGACIÓN EDUCATIVA II: INVESTIGACIÓN-ACCIÓN".
- [9] "Definición de HCE – Historia Clínica Electrónica y sus aledaños." <https://jfnietopajares.wordpress.com/jimena/1-modelo/definicion/> (accessed May 17, 2021).
- [10] "Portabilidad." <https://iso25000.com/index.php/normas-iso-25000/iso-25012/116-portabilidad> (accessed Aug. 06, 2021).
- [11] "Widget - Qué es, definición y concepto | 2021 | Economipedia." <https://economipedia.com/definiciones/widget.html> (accessed May 17, 2021).
- [12] "¿Qué es la visualización de datos? Definición, ejemplos y recursos." <https://www.tableau.com/es-mx/learn/articles/data-visualization> (accessed May 17, 2021).
- [13] "estándar | Definición | Diccionario de la lengua española | RAE - ASALE." <https://dle.rae.es/estándar> (accessed May 17, 2021).
- [14] "HL7 un estándar de interoperabilidad en Salud - Disrupción Tecnológica." <https://www.disrupciontecnologica.com/hl7/> (accessed May 17, 2021).
- [15] "HL7 Standards Product Brief - HL7 Version 2 Product Suite | HL7 International." https://www.hl7.org/implement/standards/product_brief.cfm?product_id=185 (accessed May 17, 2021).

- [16] "HL7 Standards Product Brief - HL7 Version 3 Product Suite | HL7 International."
https://www.hl7.org/implement/standards/product_brief.cfm?product_id=186
 (accessed May 17, 2021).
- [17] "HL7 Standards Product Brief - CDA® (HL7 Clinical Document Architecture) | HL7 International."
https://www.hl7.org/implement/standards/product_brief.cfm?product_id=496
 (accessed May 17, 2021).
- [18] "HL7 Standards Product Brief - FHIR® R4 (HL7 Fast Healthcare Interoperability Resources, Release 4) | HL7 International."
https://www.hl7.org/implement/standards/product_brief.cfm?product_id=491
 (accessed May 17, 2021).
- [19] "Mapas de conocimiento: qué son y para qué sirven en una organización empresarial? | Magnetismo Aplicado."
<https://magnetismocnea.wordpress.com/2014/10/16/mapas-de-conocimiento-que-son-y-para-que-sirven-en-una-organizacion-empresarial/> (accessed May 17, 2021).
- [20] "Definición de entidad - Qué es, Significado y Concepto." <https://definicion.de/entidad/> (accessed May 17, 2021).
- [21] "Definición de relaciones - Qué es, Significado y Concepto."
<https://definicion.de/relaciones/> (accessed May 17, 2021).
- [22] "DATOS SIMPLES SEMIESTRUCTURADOS | Marketing Directo."
<https://www.marketingdirecto.com/marketing-general/marketing/datos-simples-semiestructurados> (accessed May 17, 2021).
- [23] "¿Qué son los datos no estructurados? - Máster Big Data & Analytics - MBDA Universidad Pablo de Olavide." <http://mbda.es/que-son-los-datos-no-estructurados/> (accessed May 17, 2021).
- [24] "RDF/XML." <https://www.w3.org/2001/sw/Europe/events/200406-esp/trabajo-final-extratesauros/node3.html> (accessed May 17, 2021).
- [25] "Arquitectura de software - Wikipedia, la enciclopedia libre."
https://es.wikipedia.org/wiki/Arquitectura_de_software (accessed May 17, 2021).
- [26] "Extracción de información: en busca de la semántica textual (I) - innovación en español." <https://www.innovaspain.com/extraccion-informacion-busca-la-semantica-textual-i/> (accessed May 17, 2021).
- [27] "Reconocimiento de entidades nombradas - Wikipedia, la enciclopedia libre."
https://es.wikipedia.org/wiki/Reconocimiento_de_entidades_nombradas (accessed May 17, 2021).
- [28] "¿Qué es Semántica? » Su Definición y Significado [2021]."
<https://conceptodefinicion.de/semantica/> (accessed May 17, 2021).

- [29] “Significado de Ontología (Qué es, Concepto y Definición) - Significados.” <https://www.significados.com/ontologia/> (accessed May 17, 2021).
- [30] “Tuplas — Programación.” <http://progra.usm.cl/apunte/materia/tuplas.html> (accessed May 17, 2021).
- [31] “Sistemas Expertos: Definición, Aplicaciones y Ejemplos.” <https://www.tecnologias-informacion.com/sistemas-expertos.html> (accessed May 17, 2021).
- [32] “Qué es una base de conocimientos y por qué es importante para mejorar tu servicio al cliente.” <https://blog.hubspot.es/service/que-es-una-base-de-conocimientos> (accessed Aug. 06, 2021).
- [33] L. A. L. Contra, C. Pandemia, and Y. S. E. Dictan, “D,c 202”.
- [34] J. Ui, “SOCl~!-Ap,” 2021.
- [35] Congreso de la República, “Ley Estatutaria 1751 de 2015,” *Ministerio de Salud y Protección Social*, p. 13, 2015.
- [36] C. de la Republica, “Ley 1581 de Octubre de 2012,” pp. 1–15, 2012.
- [37] K. Petersen, S. Vakkalanka, and L. Kuzniarz, “Guidelines for conducting systematic mapping studies in software engineering: An update,” *Information and Software Technology*, vol. 64, pp. 1–18, 2015, doi: 10.1016/j.infsof.2015.03.007.
- [38] S. E. Group, “Методи за автоматично управление на подедни устройства при Jack-up системите,” 2007, doi: 10.1145/1134285.1134500.
- [39] P. Haase, D. M. Herzig, A. Kozlov, A. Nikolov, and J. Trame, “Metaphactory: A platform for knowledge graph management,” *Semantic Web*, vol. 10, no. 6, pp. 1109–1125, 2019, doi: 10.3233/SW-190360.
- [40] S. M. Shamimul Hasan, D. Rivera, X. C. Wu, E. B. Durbin, J. B. Christian, and G. Tourassi, “Knowledge Graph-Enabled Cancer Data Analytics,” *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 7, pp. 1952–1967, 2020, doi: 10.1109/JBHI.2020.2990797.
- [41] J. Z. Pan, · Valentina, T. · Claudia D’amato, K. Janowicz, B. Fu, and A. Polleres, “The Semantic Web-ISWC 2020.” [Online]. Available: <http://www.springer.com/series/7409>
- [42] H. Zheng *et al.*, *Proceedings, 2018 IEEE International Conference on Bioinformatics and Biomedicine : 3-6 Dec. 2018, Madrid, Spain*.
- [43] “OP-CBIO190613 5382..5384 _ Enhanced Reader”.
- [44] D. Cashman *et al.*, “CAVA: A visual analytics system for exploratory columnar data augmentation using knowledge graphs,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 2, pp. 1731–1741, Feb. 2021, doi: 10.1109/TVCG.2020.3030443.
- [45] X. He *et al.*, “ALOHA: Developing an interactive graph-based visualization for dietary supplement knowledge graph through user-centered design,” *BMC Medical*

- Informatics and Decision Making*, vol. 19, Aug. 2019, doi: 10.1186/s12911-019-0857-1.
- [46] “QueDI_ From Knowledge Graph Querying to Data Visualization _ Enhanced Reader”.
 - [47] Z. Zheng, Y. Liu, Y. Zhang, and C. Wen, “TCMKG: A deep learning based traditional chinese medicine knowledge graph platform,” in *Proceedings - 11th IEEE International Conference on Knowledge Graph, ICKG 2020*, Aug. 2020, pp. 560–564. doi: 10.1109/ICBK50248.2020.00084.
 - [48] R. Verborgh *et al.*, “Triple Pattern Fragments: A low-cost knowledge graph interface for the Web,” *Journal of Web Semantics*, vol. 37–38, pp. 184–206, Mar. 2016, doi: 10.1016/j.websem.2016.03.003.
 - [49] S. Cox *et al.*, “Visualization environment for federated knowledge graphs: Development of an interactive biomedical query language and web application interface,” *JMIR Medical Informatics*, vol. 8, no. 11, Nov. 2020, doi: 10.2196/17964.
 - [50] S. Kropf, C. Chalopin, D. Lindner, and K. Denecke, “Domain Modeling and Application Development of an Archetype- and XML-based EHRS,” *Applied Clinical Informatics*, vol. 08, no. 02, pp. 660–679, 2017, doi: 10.4338/aci-2017-01-ra-0009.
 - [51] B. S. Glicksberg *et al.*, “PatientExploreR: An extensible application for dynamic visualization of patient clinical history from electronic health records in the OMOP common data model,” *Bioinformatics*, vol. 35, no. 21, pp. 4515–4518, 2019, doi: 10.1093/bioinformatics/btz409.
 - [52] V. Kilintzis, I. Chouvarda, N. Beredimas, P. Natsiavas, and N. Maglaveras, “Supporting integrated care with a flexible data management framework built upon Linked Data, HL7 FHIR and ontologies,” *Journal of Biomedical Informatics*, vol. 94, Jun. 2019, doi: 10.1016/j.jbi.2019.103179.
 - [53] I. Belkacem, I. Pecci, and B. Martin, “Exploring a design space for electronic health record visualization,” in *IHM 2017 - Actes de la 29ieme Conference Francophone sur Interaction Homme-Machine, Actes*, Aug. 2017, pp. 325–334. doi: 10.1145/3132129.3132165.
 - [54] J. Bernard, D. Sessler, J. Kohlhammer, and R. A. Ruddle, “Using dashboard networks to visualize multiple patient histories: A design study on post-operative prostate cancer,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 3, pp. 1615–1628, Mar. 2019, doi: 10.1109/TVCG.2018.2803829.
 - [55] T. C. Huber, A. Krishnaraj, D. Monaghan, and C. M. Gaskin, “Developing an Interactive Data Visualization Tool to Assess the Impact of Decision Support on Clinical Operations,” *Journal of Digital Imaging*, vol. 31, no. 5, pp. 640–645, Oct. 2018, doi: 10.1007/s10278-018-0065-z.

- [56] M. N. Khan, O. Ormandjieva, and K. Pitula, "Mobile viewing and self-management of patient's electronic health records (EHRs) with MyHealthCloud," Jul. 2017. doi: 10.1145/3084035.3084038.
- [57] N. Hong, K. Wang, S. Wu, F. Shen, L. Yao, and G. Jiang, "An Interactive Visualization Tool for HL7 FHIR Specification Browsing and Profiling," *Journal of Healthcare Informatics Research*, vol. 3, no. 3, pp. 329–344, Sep. 2019, doi: 10.1007/s41666-018-0043-8.
- [58] Z. Lin, D. Yang, and X. Yin, "Patient similarity via joint embeddings of medical knowledge graph and medical entity descriptions," *IEEE Access*, vol. 8, pp. 156663–156676, 2020, doi: 10.1109/ACCESS.2020.3019577.
- [59] F. Gong, M. Wang, H. Wang, S. Wang, and M. Liu, "SMR: Medical Knowledge Graph Embedding for Safe Medicine Recommendation," *Big Data Research*, vol. 23, p. 100174, 2021, doi: 10.1016/j.bdr.2020.100174.
- [60] "Enhanced Reader".
- [61] C. Li, S. Hang, X. Hu, D. Chu, and H. Zheng, "Knowhealth: A knowledge graph based question-answer platform for elderly people," in *Proceedings of the 12th EAI International Conference on Mobile Multimedia Communications, MOBIMEDIA 2019*, 2019, pp. 154–165. doi: 10.4108/eai.29-6-2019.2282866.
- [62] "Leyes desde 1992 - Vigencia expresa y control de constitucionalidad [LEY_1122_2007]."
http://www.secretariasenado.gov.co/senado/basedoc/ley_1122_2007.html (accessed Apr. 29, 2021).
- [63] "Salud pública." <https://www.minsalud.gov.co/salud/publica/Paginas/salud-publica.aspx> (accessed Apr. 29, 2021).
- [64] B. Dc, "Boletín Técnico Boletín Técnico," pp. 1–34, 2021.
- [65] Congreso de la República de Colombia, "Ley 2064 de 2020," pp. 1–6, 2020.
- [66] "resolucion-537-de-2020.pdf."
- [67] "Pages - Catálogo de información."
<https://www.sispro.gov.co/catalogos/Pages/catalogo-de-informacion.aspx> (accessed Apr. 30, 2021).
- [68] MinSalud, "LINEAMIENTOS TÉCNICOS Y OPERATIVOS PARA LA VACUNACIÓN CONTRA EL COVID-19 Dirección," *Minsalud*, p. 63, 2021.
- [69] "1 REGISTRO DIARIO COVID 19_Febrero2021."
- [70] B. Runyon, "The Hype Cycle for Healthcare Provider Technologies and Standards , 2009 Macro Level Workflows in the CDO," no. August, 2009.
- [71] "HL7 un estándar de interoperabilidad en Salud - Disrupción Tecnológica."
<https://www.disrupciontecnologica.com/hl7/> (accessed Apr. 30, 2021).

- [72] “Operaciones con HL7® FHIR® | El Blog de Nessy.” <https://blog.codeness.io/hl7-principios-y-fundamentos/modulo-3/> (accessed Nov. 25, 2021).
- [73] “The Digital Health Hype Cycle 2019.” <https://www.healthcare.digital/single-post/2019/01/12/the-digital-health-hype-cycle-2019> (accessed Apr. 30, 2021).
- [74] “Immunization - FHIR v4.0.1.” <https://www.hl7.org/fhir/immunization.html> (accessed Aug. 06, 2021).
- [75] “Patient - FHIR v4.0.1.” <https://www.hl7.org/fhir/patient.html> (accessed Aug. 06, 2021).
- [76] A. Blumauer and H. Nagy, *The Knowledge Graph Cookbook*, no. December. 2000.
- [77] L. Encalada, I. Boavida-Portugal, C. C. Ferreira, and J. Rocha, “Identifying tourist places of interest based on digital imprints: Towards a sustainable smart City,” *Sustainability (Switzerland)*, vol. 9, no. 12, pp. 1–25, 2017, doi: 10.3390/su9122317.
- [78] J. Hur, Z. Xiang, E. L. Feldman, and Y. He, “Ontology-based Brucella vaccine literature indexing and systematic analysis of gene-vaccine association network,” *BMC Immunology*, vol. 12, no. 1, p. 49, 2011, doi: 10.1186/1471-2172-12-49.
- [79] “Schema.org - Schema.org.” <https://schema.org/> (accessed Aug. 06, 2021).
- [80] “protégé.” <https://protege.stanford.edu/products.php> (accessed Aug. 07, 2021).
- [81] “¿Qué es y para qué sirve Neo4j? - BBVA.” <https://www.bbva.com/es/que-es-neo4j-y-para-que-sirve-una-base-de-datos-orientada-a-grafos/> (accessed Aug. 07, 2021).
- [82] “Introduction - Neo4j Cypher Manual.” <https://neo4j.com/docs/cypher-manual/current/introduction/> (accessed Aug. 07, 2021).
- [83] “Neo4j Graph Algorithms - Developer Guides.” <https://neo4j.com/developer/graph-data-science/graph-algorithms/> (accessed Nov. 13, 2021).
- [84] “Luis-xmart/TesisGraphNeo4j.” <https://github.com/Luis-xmart/TesisGraphNeo4j> (accessed Dec. 13, 2021).
- [85] Y. L. B, Q. Miao, T. Guo, J. Geng, and C. Hu, *Knowledge Graph and Semantic Computing*, vol. 1, no. January. 2018. doi: 10.1007/978-981-15-1956-7.
- [86] “ENCUESTA DE SATISFACCIÓN PARA USUARIOS DEL SERVICIO WEB DE LA CONTRALORÍA GENERAL DE LA REPÚBLICA DOMINICANA (CGR). Survey.” <https://es.surveymonkey.com/r/JGB88DV> (accessed Nov. 26, 2021).
- [87] B. S. Glicksberg *et al.*, “PatientExploreR: An extensible application for dynamic visualization of patient clinical history from electronic health records in the OMOP common data model,” *Bioinformatics*, vol. 35, no. 21, pp. 4515–4518, 2019, doi: 10.1093/bioinformatics/btz409.
- [88] S. E. Group, “Методи за автоматично управление на портативни устройства при Jack-up системите,” 2007, doi: 10.1145/1134285.1134500.