

Grafos de conocimiento para gestionar información epidemiológica sobre COVID-19

Knowledge graphs to manage epidemiological information about COVID-19

Tatiana Delgado Fernández^{1*} <https://orcid.org/0000-0002-4323-9674>

Mavis Lis Stuart Cárdenas¹ <https://orcid.org/0000-0002-0461-7118>

Mercedes Delgado Fernández¹ <https://orcid.org/0000-0003-2556-1712>

¹Universidad Tecnológica de La Habana "José Antonio Echeverría". La Habana, Cuba.

*Autor para la correspondencia: tatiana.delgado@uic.cu

RESUMEN

El control de la propagación de las enfermedades infecciosas requiere investigaciones epidemiológicas exhaustivas, lo que ha quedado validado con el desempeño del Ministerio de Salud Pública a lo largo de varias décadas en el combate a numerosas enfermedades como el dengue, el cólera y varios tipos de influenza, entre otras. Sin embargo, la pandemia COVID-19 está poniendo a prueba los más rigurosos protocolos epidemiológicos de Cuba y del mundo por su elevada capacidad de contagio y propagación. Ante este contexto, el presente artículo se propuso emplear los grafos de conocimiento para el apoyo a los estudios epidemiológicos de la COVID-19, haciendo mayor énfasis en los factores de exposición y rastreo de los contactos. Para alcanzar este objetivo se realizó un estudio relacionado con el estado del arte sobre grafos de conocimiento y su empleo en el sector de la salud, particularmente en la lucha contra el nuevo coronavirus SARS-CoV-2. La investigación tuvo como soporte un enfoque

metodológico de creación y uso de grafos de conocimiento adaptado al campo de estudio. Los resultados se simulan en el escenario del brote producido a mediados del mes de julio del año 2020 en el municipio de Bauta de la provincia de Artemisa, empleando para esto datos de la realidad, extraídos de la Web, combinados con otros datos simulados.

Palabras clave: Epidemiología; infecciones por coronavirus; grafo de conocimiento.

ABSTRACT

Control of the spread of infectious diseases requires exhaustive epidemiological research, as has been validated by the performance of the Ministry of Public Health during several decades of combat against numerous diseases, such as dengue, cholera and various types of influenza, among others. However, the COVID-19 pandemic is testing the limits of the most rigorous epidemiological protocols in Cuba and worldwide, due to its high transmissibility and fast spread. In this context, the present study had the purpose of using knowledge graphs to support epidemiological research about COVID-19, with greater emphasis on exposure factors and contact tracing. To achieve this end, a study was conducted about the state of the art of knowledge graphs and their use in the health care sector, particularly in the combat against the novel coronavirus SARS-CoV-2. The research applied a methodological approach based on the development and use of knowledge graphs adjusted to the study field. Results are simulated in the context of the outbreak occurring in mid July 2020 in the municipality of Bauta, Artemisa province, using real data obtained from the Internet and combined with other simulated data.

Key words: Epidemiology; coronavirus infections; knowledge graph.

Recibido: 12/08/2020

Aceptado: 01/02/2021

Introducción

La especial situación que vive el planeta por la pandemia de la COVID-19, causada por el nuevo coronavirus SARS-CoV-2, que apenas un año después de reportado el primer caso en China ha generado contagios a 100,2 millones de personas, con una letalidad de 2,1 millones^a, está disparando investigaciones e innovaciones en comunidades científicas y de práctica a nivel mundial para unir esfuerzos mancomunados en pos de mitigar su impacto. La hiperconectividad y la colaboración son principios clave de la transformación digital⁽¹⁾ y facilitan las interconexiones en las condiciones de aislamiento que impone el nuevo coronavirus, lo que convierte a la Web hoy en un instrumento habilitador de la gestión del conocimiento alrededor de la pandemia.

La experiencia del usuario en la búsqueda en la Web ha ido transitando desde el uso exclusivo de palabras clave e hipervínculos, a una experiencia más precisa y eficaz, que se basa en la Web semántica. Un ejemplo que ilustra esta afirmación es el panel de conocimiento de Google, que desde hace algunos años aparece a la derecha de algunas de las búsquedas que utilizan este motor, el cual es capaz de “entender” la semántica de la intención de búsqueda, y para esto se soporta en el grafo de conocimiento de Google^b. Bajo el mismo paradigma popularizado por Google “cosas, no cadenas”, descansa un amplio movimiento derivado de la Web Semántica, en el cual se defiende la estructuración de la información como nodos que representan entidades o hechos, y aristas que expresan las interrelaciones entre los nodos. Aunque el término grafo de conocimiento data de varias décadas atrás, fue precisamente Google el que disparó su popularidad en el año 2012, cuando difundió su propio grafo de conocimiento, presentando el cambio paradigmático que tendría lugar en las búsquedas de su motor a partir de este. En el año 2019, el grafo de conocimiento de Google incluía ya mil millones de entidades y alrededor de 70 mil millones de significados.⁽²⁾

Los grafos de conocimiento sobre la Web constituyen una columna vertebral de muchos sistemas de información que requieren acceso al conocimiento estructurado; pueden ser dependientes o independientes del dominio.⁽³⁾ Constituyen un paradigma flexible de representación del conocimiento destinado a facilitar el procesamiento del conocimiento tanto para humanos como para máquinas. Los grafos de conocimiento son considerados como un habilitador clave para varios casos de uso, cada vez más populares, incluida la búsqueda en la Web, la respuesta a preguntas y los asistentes personales, así como la habilitación de otras aplicaciones basadas en la inteligencia artificial para su uso en la mayoría de los sectores.⁽⁴⁾

Debido a su versátil capacidad de representación, los grafos pueden ser usados para integrar diferentes fuentes de datos heterogéneas, tanto dentro como entre organizaciones.⁽⁵⁾ Gracias a esta función, se están convirtiendo en una poderosa herramienta organizacional de gestión de datos.

Inspirado en el amplio y exitoso uso de los grafos de conocimiento en las más diversas esferas, este artículo se propuso emplear los grafos de conocimiento para el apoyo a los estudios epidemiológicos de la COVID-19, haciendo mayor énfasis en los factores de exposición y rastreo de los contactos.

Bases conceptuales sobre grafos de conocimientos

Para alcanzar el objetivo propuesto, el artículo se estructuró en tres secciones. La primera está dedicada a ofrecer las principales bases conceptuales sobre las que descansan los grafos de conocimiento, que concluye con su uso en el campo de la salud y, en particular, en el escenario concreto del enfrentamiento a la COVID-19. Le sigue el acápite donde se describen los métodos que se seguirán para alcanzar la principal contribución del artículo.

Para el propósito definido en este artículo, se tomó como definición de grafo de conocimiento (knowledge graph - KG), la de un sistema que “adquiere e integra

información dentro de una ontología y aplica un razonador para derivar un nuevo conocimiento”.⁽⁶⁾ En esta definición, se consideró la integración como la capacidad de incorporar información de diferentes fuentes de datos -estructurados y semi-estructurados-, por lo que se entiende que el grafo está construido sobre otras bases de datos, que al integrarse se enriquecen e incrementan su valor.⁽⁷⁾

Aunque para apoyar la contribución principal de este artículo no se requiere adentrarse en una comparación de los grafos con otros sistemas de organización del conocimiento (como las propias ontologías, las taxonomías o las bases de conocimiento), para aquellos lectores interesados en profundizar en esto, se recomienda la comparación fundamentada que ofrecen *Ehrlinger & Wöß*.⁽⁶⁾ Para una mayor comprensión del resto del artículo, bastaría resaltar que el foco de un KG es el nivel de instancias/individuos (A-box) por encima del enfoque a nivel de taxonomías de conceptos (T-box).⁽⁶⁾

También conocido como grafo de datos, un KG está destinado a acumular y transmitir conocimiento del mundo real, donde los nodos del grafo representan entidades de interés y sus aristas representan relaciones entre estas entidades.⁽⁸⁾ Los grafos de conocimiento pueden generarse a partir de la extracción de bases de conocimiento de la Web, como es el caso del conjunto de datos DBpedia, que está publicado en el proyecto de Datos Abiertos Enlazados (*Linked Open Data*);⁽⁹⁾ pueden ser editados por la colaboración de usuarios de la Web o a partir de métodos de extracción de información de fuentes semi-estructuradas o no estructuradas.⁽¹⁰⁾

Aquellos grafos de conocimiento generados desde fuentes no estructuradas o con estructuras lingüísticas (como documentos de texto), requieren incorporar procesamiento en lenguaje natural para minar y descubrir las relaciones semánticas entre las clases. En estos casos, las relaciones semánticas son generalmente “aprendidas” directamente del contexto y mapeadas en la ontología que soporta el grafo.⁽¹¹⁾ En cambio, cuando el grafo se construye desde fuentes

estructuradas, como las bases de datos, o semiestructuradas, como otros grafos de datos (ej, Linked Open Data - LOD), las relaciones semánticas se derivan de las relaciones entre clases de sus fuentes de datos originales, o son previstas en la ontología que soportará al grafo de acuerdo con el propósito de este. La creación de un grafo de conocimiento argumentativo, a partir de minar la base de datos abierta para la COVID-19 (CORD-19) y con el establecimiento de un enlace al Web de Datos, ilustra claramente este proceso de completamiento de las clases y las relaciones semánticas entre ellas partiendo de fuentes estructuradas y semi-estructuradas.⁽¹²⁾ Existen en la actualidad plataformas que guían este proceso como Ontotext^c y Ontop^d, las cuales pueden utilizarse conjuntamente para ofrecer un poderoso motor de código abierto de virtualización de datos que soporta la construcción de grafos de conocimiento.

Tras una investigación de varias definiciones se han determinado 4 categorías de grafos de conocimiento,⁽⁸⁾ las cuales se resumen en el cuadro 1, que incorpora una columna adicional dedicada a esgrimir un breve análisis crítico de cada una.⁽⁸⁾

Cuadro 1 - Categorías de grafos de conocimientos

Categoría	Definición	Análisis crítico
Categoría I	Define el grafo de conocimiento como un gráfico donde los nodos representan entidades y las aristas representan relaciones entre esas entidades. ⁽⁸⁾	Es la definición más simple, y no parece tener en cuenta el rol que juega el conocimiento.
Categoría II	Gráfico estructurado de una base de conocimiento. ⁽⁸⁾	Aunque le ofrecen un rol al conocimiento, a diferencia de la primera categoría, es una definición reduccionista al hacer equivalente los grafos a las bases de conocimiento.
Categoría III	Se define mediante cuatro características: ⁽³⁾ 1) describe principalmente entidades del mundo real y sus interrelaciones, organizadas en un gráfico; 2) define posibles clases y relaciones de entidades en un esquema; 3) permite	En la definición de <i>Paulheim</i> , ⁽³⁾ las dos primeras características distinguen como foco de los KG el nivel de instancias/individuos (A-box en la terminología de lógica

	<p>potencialmente interrelacionar entidades arbitrarias entre sí; 4) cubre varios dominios de actualidad.</p> <p>Otra definición que encaja en esta categoría expresa que “un grafo de conocimiento adquiere e integra información en una ontología y aplica un razonador para derivar nuevos conocimientos”.⁽⁷⁾</p>	<p>descriptiva), lo cual constituye una distinción clave respecto a otros términos de la Web semántica. Resulta también particularmente interesante el acotamiento sobre la inclusión de “razonamiento” en la definición proporcionada por <i>Ehrlinger & Wöβ</i>⁽⁷⁾</p>
<p>Categoría IV</p>	<p>Esta categoría evita ofrecer una definición específica; más bien adopta un enfoque extensivo, mencionando las características comunes de los grafos. Los autores más genuinos de esta categoría son <i>Noy N, Gao Y, Jain A, Narayanan A, Patterson A y Taylor J</i>,⁽²⁾ quienes ofrecen una comparación de tales características en varios KGs (Google, Microsoft, Facebook, eBay, IBM). Sobre esa base plantea que “un grafo de conocimiento describe objetos de interés y conexiones entre ellos; proporcionan un sustrato compartido de conocimiento dentro de una organización, lo que permite que diferentes productos y aplicaciones usen vocabulario similar y reutilicen definiciones y descripciones que otros crean. Además, generalmente proporcionan una representación formal compacta que los desarrolladores pueden usar para inferir nuevos hechos y desarrollar el conocimiento”.</p>	<p>No aporta una definición concretamente, sino un grupo de características que presentan los KG. El mérito clave de esta última categoría es que reúne el consenso de cinco principales exponentes de aplicaciones basadas en grafos: Google, Microsoft, Facebook, eBay e IBM Watson.</p>

Adaptado de: Hogan A, Blomqvist E, Cochez M, d'Amato C, de Melo G, Gutiérrez C, Gayo JE, Kirrane S, Neumaier S, Polleres A, Navigli R. Knowledge graphs. arXiv Preprint; 2020 [12/08/2020]. Disponible en:

<https://arxiv.org/abs/2003.02320>

Manejo de bases de datos para construir grafos de conocimiento

Para crear grafos de conocimiento, una de la herramientas que gestionan bases de datos más utilizadas es la base de datos gráfica RDF (*Resources Data Framework*), estandarizada por el consorcio de la Web W3C^e y que ha sido ampliamente utilizada

para publicar grafos de conocimiento por su poderosa facilidad de abstracción para representar grafos heterogéneos, incompletos, y potencialmente ruidosos.⁽¹³⁾ RDF se basa en ontologías en forma de tripletas sujeto-predicado-objeto. Está soportada por un rico ecosistema de sistemas de gestión de datos y herramientas que han evolucionado en las últimas décadas.

Este ecosistema incluye formatos y procesamientos estándares de serialización y el propio sistema de gestión de base de datos o almacén de tripletas que soporta SPARQL; el lenguaje de consulta estándar del consorcio W3C. En SPARQL, lo más común es gestionar las consultas directamente desde el lenguaje nativo, mediante lo que se conoce como un *EndPoint SPARQL*, aunque existen algunas implementaciones que encapsulan sus operaciones en interfaces de programación de aplicaciones (*Application Programming Interface - API*) y marcos de procesamiento de consultas más flexibles o personalizadas a un entorno de aplicación específico. Un ejemplo lo constituye el marco *RDFFrame*,⁽¹³⁾ que ofrece un proceso de generación de consultas SPARQL combinando patrones para optimizar las interacciones con el motor RDF, a la vez que asegura simplicidad de cara al usuario.

Existen otras populares bases de datos para el almacenamiento persistente de grafos de conocimiento, como Neo4J. Esta herramienta tiene un enfoque basado en propiedades, no en ontologías, lo que permite importar sus nodos y ejes desde ficheros CVS. Puede ser usado como una herramienta visual y soporta su propio lenguaje de consulta, Cypher.⁽¹⁴⁾

Aplicaciones de los grafos de conocimiento

Los grafos de conocimiento cubren muchos dominios de aplicación. Se destacan los siguientes grupos de aplicaciones que emplean grafos:⁽⁸⁾

- Sistemas de integración de información empresarial (EIS)
- Sistemas de recomendaciones

- Agentes conversacionales o chatbots
- Vigilancia tecnológica e inteligencia empresarial
- Evaluación y mitigación del riesgo
- Rastreo de eventos de noticias
- Comunidades enlazadas de usuarios

Sin embargo, su uso no está limitado a estos campos; es cada vez más frecuente la aparición de nuevos ámbitos de aplicación de los grafos de conocimiento. El uso de KG para acceder y enriquecer información relativa a ecosistemas de innovación expande notablemente sus beneficios de cara al tomador de decisión.⁽¹⁵⁾

Paralelamente, está emergiendo una familia de grafos de conocimientos especializados en problemáticas empresariales e industriales. Los grafos de conocimiento empresariales contribuyen a organizar los datos al interior de la empresa. Se están convirtiendo en una poderosa tecnología, principalmente en aquellas organizaciones que gestionan una variada y voluminosa cantidad de datos.

Uno de los propósitos más interesantes en el marco de las empresas para el cual se están empleando con éxito los grafos de conocimiento son los sistemas de integración de información empresarial. Los grafos de conocimiento absorben la filosofía de los datos enlazados, y agrega nuevos enfoques particularmente apropiados desde la perspectiva organizacional.^(16,17) También, a raíz del advenimiento de la 4ta. Revolución Industrial, modelos de Industria 4.0 están siendo representados mediante grafos de conocimientos.⁽¹⁸⁾

Grafos de conocimiento aplicados a la gestión de información en salud y a la COVID-19

En la esfera de la salud, los grafos de conocimiento han sido ampliamente empleados. En *Rotmensch* y otros⁽¹⁹⁾ se presenta una metodología para derivar automáticamente un grafo que relaciona las enfermedades con los síntomas que pueden causar a partir de los datos de la historia clínica o registros médicos

electrónicos. Los autores evalúan la metodología con un grafo adaptado a un entorno de atención a partir de los registros del departamento de emergencias de más de 270,000 visitas de pacientes. Otro interesante estudio en este campo presenta un grafo de conocimiento personalizado para la salud, que agrega conocimientos de diversas fuentes heterogéneas, como los dispositivos de Internet de las cosas (IoT), las notas clínicas y los registros médicos electrónicos.⁽²⁰⁾

En la medicina tradicional, también existen trabajos relacionados. En un estudio en China,⁽²¹⁾ por ejemplo, se construye un grafo de conocimiento de gran escala, para facilitar varios servicios, como visualización, recuperación y recomendaciones, con el fin ayudar a compartir, interpretar y utilizar el conocimiento de la atención médica de la medicina tradicional china.

A partir de la crisis sanitaria acarreada con la pandemia generada por el nuevo coronavirus SARS-CoV-2, han surgido varios estudios que se soportan en grafos de conocimientos. El cuadro 2 resume algunos de estos trabajos de investigación.^(22,23,24,25,26,27)

Cuadro 2 - Principales trabajos de grafos de conocimiento en apoyo a la lucha contra la COVID-19

Tipo de solución	Descripción
Extracción de información de literatura científica para crear grafos de conocimiento y explotarlos en función de la COVID-19	Presentan un grafo de conocimiento de COVID-19, construido a partir de la literatura científica sobre el nuevo coronavirus con el objetivo de proporcionar una visión integral de su fisiopatología. Implementan una aplicación web y lanzan el KG en múltiples formatos estándar. ⁽²²⁾ https://bikmi.covid19-knowledgespace.de
	Desarrollan un grafo de conocimiento para extraer entidades, relaciones y eventos de la literatura científica sobre la COVID-19, y lo emplean para responder preguntas en los informes de reutilización de medicamentos. ⁽²³⁾
	Construyen un grafo de conocimiento, basado en Amazon Neptune, para la COVID-19, combinando el esquema inherente a la información de artículos científicos de COVID-19 (del conjunto de datos CORD-19); así como, extrayendo relaciones de entidades biomédicas e información sobre temas similares. Además de aplicar metodologías normalizadas de curación de las fuentes, demuestran las capacidades de recuperación, visualización y descubrimiento. Brinda un sistema de recomendaciones basado en similitud. ⁽²⁴⁾
	Sugieren posibles técnicas para predecir enfermedades relacionadas, candidatos a fármacos, genes, mutaciones genéticas y compuestos relacionados como parte de un esfuerzo sistemático para aplicar métodos de descubrimiento de conocimientos para ayudar a los investigadores biomédicos a abordar la pandemia, utilizando grafos de conocimiento creados a partir de conjuntos de datos PubMed y CORD-19. ⁽²⁵⁾
Herramienta genérica de gestión de generación y gestión de grafos de conocimiento	Presentan una herramienta para la manipulación y análisis de grandes grafos de conocimiento que aprovecha las bibliotecas populares desarrolladas para aplicaciones de ciencia de datos, lo que permite una amplia audiencia de desarrolladores para construir fácilmente suministro de datos para grafos de conocimiento en aplicaciones. Se ilustra con escenarios del mundo real para integrar y manipular grandes KG, como Wikidata, DBpedia y ConceptNet. Usan como caso de estudio el conjunto de datos CORD-19 de COVID-19. ⁽²⁶⁾
Seguimiento epidemiológico	Aborda el rastreo de los contactos de los casos confirmados es crucial para el control exitoso de Taiwán en la propagación temprana de COVID-19.

Desarrollan una nueva herramienta que pueda mejorar la práctica actual de investigación epidemiológica, incorporando grafos de conocimiento.⁽²⁷⁾

Se aprecia una mayor representatividad de grafos de conocimiento generados para integrar literatura científica en soporte a investigadores de varias disciplinas y con fines principalmente de diagnóstico, mientras están mucho menos representadas aquellas publicaciones que empleando grafos de conocimiento se orientan a apoyar la gestión del comportamiento epidemiológico de la COVID-19.

Considerando los nuevos brotes de la enfermedad y la limitada cantidad de publicaciones halladas que abordan concretamente el control de la propagación a partir de contactos usando grafos de conocimiento, en este artículo se propone una metodología para crear grafos de conocimiento en apoyo al control epidemiológico de la pandemia.

Todas las experiencias levantadas en el escenario de la COVID-19 podrían aportar elementos interesantes a la propuesta, debido a que las operaciones de un grafo de conocimiento siguen ciertos patrones que facilitan la generación de estructuras flexibles y escalables, capaces de ofrecer mayor eficacia a la búsqueda, realizar minería y otras técnicas de aprendizaje automático para extraer información relevante y generar nuevo conocimiento a partir de la incorporación de “razonamiento” basado en reglas.

Métodos

El método está alineado a las categorías III^(3,7) y IV⁽²⁾ de las definiciones ofrecidas en la tabla 1 de las bases conceptuales. Esta decisión está fundamentada en que ambas categorías representan en su conjunto las mejores prácticas de grandes empresas exitosas en el uso intensivo de datos de gran escala, además que ofrecen

una definición inclusiva que parte de hechos (supone instancias de datos) y no solo taxonomías, y se habilita el razonamiento para generar nuevo conocimiento, lo cual resulta compatible con el interés de estudiar el comportamiento epidemiológico complejo de la COVID-19.

El grafo debe ser capaz de extraer estructuradamente la información epidemiológica (fuente de exposición e historial de cada contacto) y combinarlo con datos de lugares extraídos de otras fuentes. A partir del grafo, se habilitarían análisis de múltiple naturaleza, además de ubicar los potenciales sospechosos, debe ser posible proponer otras medidas de búsqueda activa de casos por áreas, minar el grafo y razonar para descubrir nuevo conocimiento basándose en reglas cuidadosamente introducidas, e incluso, debe ser capaz de producir recomendaciones a los ciudadanos que pudieran tener riesgo de exposición.

El enfoque metodológico para construir el grafo de conocimiento de apoyo al estudio epidemiológico de la COVID-19 se desglosa en los cuatro pasos que muestra la figura 1.

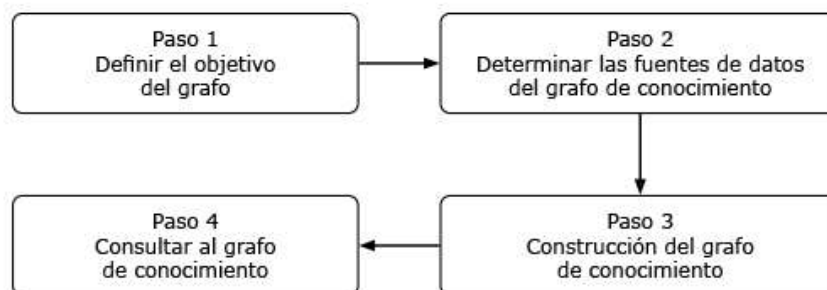


Fig. 1 - Enfoque metodológico para construir el grafo de conocimiento de apoyo al estudio epidemiológico de la COVID-19.

Resultados

Se presenta una simulación de creación de un grafo de conocimiento a partir del brote de COVID-19 del municipio de Bauta, de la provincia de Artemisa, Cuba, que desde las últimas semanas de julio del año 2020 ha generado varias decenas de casos en la provincia y con impacto en municipios habaneros colindantes.

Simulación del método en el escenario COVID-19 desencadenado en la “Fiesta de Bauta”

Se sigue el enfoque metodológico expuesto para la generación de un grafo de conocimiento en apoyo a los estudios epidemiológicos de la COVID-19 en un caso simulado.

Paso 1 - Definir el objetivo del grafo:

Desde mediados de julio del año 2020, La Habana y Artemisa se comportaron como focos de nuevos brotes de la COVID-19. Uno de ellos fue el que se produjo en el municipio Bauta, de la provincia Artemisa, que a los efectos de esta simulación será etiquetado como “Fiesta de Bauta”. Según refiere la Redacción Digital del periódico Granma,⁽²⁸⁾ el primer caso identificado de este brote fue una mujer de 53 años con antecedentes de diabetes e hipertensión, quien acudió al Cuerpo de Guardia el 18 de julio con un cuadro de astenia o fatiga marcada, tos seca y pérdida del olfato y del gusto. La paciente se reportó el 20 de julio como positiva, según la base de datos bajada del sitio Cuba Data COVID-19^g.

Partiendo de este caso real, el trabajo realizado se propuso construir un grafo de conocimiento experimental para apoyar un estudio epidemiológico, el cual estaría dirigido a obtener información sobre los lugares de exposición y el rastreo de contactos. Para esto se empleó una ventana de tiempo desde el 20 al 30 de julio de 2020, período en el cual se identificaron 24 personas sospechosas a partir de este brote, de las cuales resultaron positivas alrededor de 13, que dieron lugar a otras cadenas de contagio. Algunos de los casos del escenario “Fiesta de Bauta” se involucraron en una nueva cadena porque también asistieron al “Bar de Bauta”, en lo que generó una propagación en el municipio Baracoa, de la propia provincia

Artemisa, y en el municipio La Lisa, en La Habana, y siguió su recorrido en días posteriores a otros municipios.

Paso 2 - Determinar las fuentes de datos del grafo de conocimiento:

Atendiendo a las investigaciones previas,⁽²⁷⁾ existen dos tipos de información que son fundamentales para el control epidemiológico de la COVID-19: fuente de exposición e historial de contactos. Por eso, el método empleado presupone un levantamiento flexible de esta información que incluya los datos de cada caso, fecha, lugar y personas de contacto.

- *Fuente de exposición:* Como ha informado el Ministerio de Salud Pública y otras autoridades de la Organización Mundial de la Salud, los últimos 14 días de actividades de cada caso son determinantes por el período de incubación del virus, por lo que se investigan todas las actividades que haya realizado en ese período y los lugares que hubiera visitado. Eso puede identificar la fuente de infección y la posible fecha en que el paciente contrajo la enfermedad.
- *Historial del contacto:* Todas las personas que tuvieron contacto con el caso de análisis deben entonces ser levantadas y declaradas como “sospechosas” para ser sujetas a las medidas de los protocolos establecidos. Algunas veces el caso infectado puede olvidar o no saber exactamente el nombre de un contacto cercano (por ejemplo, un taxista), en cuyo caso, hay que buscar información externa basada en la información de fecha-ubicación obtenida de la investigación epidemiológica.

Para la creación de la data de entrada se utilizaron datos reales publicados del Dashboard: Cuba COVID-19,⁽⁷⁾ en la ventana de tiempo entre el 20 y el 30 de julio del año 2020, para simular un escenario de estudio. Los atributos de cada caso de esta fuente son: caso, sexo, edad, país, municipio, provincia, fecha_confirmación, fecha_ingreso, tipo_contagio.

Se pueden incluir otras fuentes de datos de las redes sociales usando las API asociadas a sus los grafos de conocimientos, en dependencia del tipo de consulta o análisis que se desee.

Paso 3 - Construir el grafo de conocimiento:

Se simula la creación del grafo a partir de los casos, partiendo del primer caso detectado asociado al escenario de estudio “Fiesta de Bauta”, diagnosticado el 20 de julio del 2020. La figura 2 muestra un fragmento del grafo de conocimiento generado combinando datos reales y simulados para el estudio.

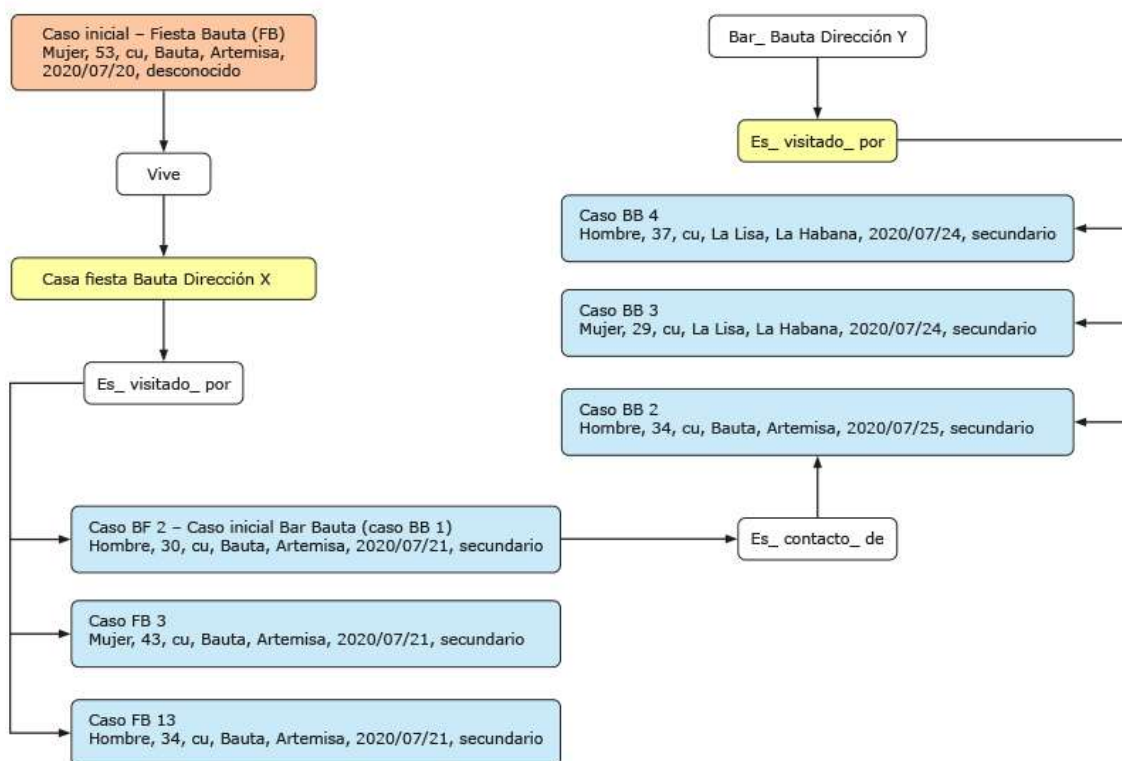


Fig. 2 - Fragmento del grafo de conocimiento del escenario “Fiesta de Bauta”.

Paso 4 - Consultar el grafo de conocimiento:

Para el fragmento del grafo generado, se puede fundamentar la pertinencia de, al menos, dos consultas Bar para el rastreo de los contactos y la obtención de los lugares de exposición contra casos de COVID-19, basándose en patrones de grafos⁽⁸⁾ estandarizados. Se emplea un patrón de consulta denominado *semántica basada*

en *homomorfismo*, que permite mapear múltiples variables asociadas al mismo término, de modo que todas las asignaciones de una variable se considerarían resultados. En el cuadro 3 se muestran los resultados de dos consultas con este patrón; una va encaminada a buscar qué casos visitaron un lugar, y la segunda a obtener los lugares que fueron visitados por un caso.

Cuadro 3 - Consultas al grafo de conocimiento del brote “Fiesta de Bauta” empleando el patrón de semántica basado en homomorfismo

Consulta	Patrón de consulta/Resultados
¿Cuáles casos visitaron la casa de la “Fiesta de Bauta”?	Casa Fiesta Bauta (dirección X) — Es_visitado_ por → (caso)
	Caso FB1, hombre,30,cu,Bauta,Artemisa,2020/07/21, secundario
	Caso FB13, hombre,34,cu,Bauta,Artemisa,2020/07/21,secundario
¿Qué lugares fueron visitados por el caso FB 2”?	Lugar- Es_visitado_ por → Caso FB2 - Caso inicial Bar Bauta (BB)
	hombre,30,cu,Bauta,Artemisa,2020/07/21, secundario
	Casa de la Fiesta de Bauta Bar de Bauta

Es preciso notar que, aunque la relación formal es la misma en ambos patrones de consulta, en el primero se busca el caso a partir de un lugar, mientras que en el segundo se busca el lugar a partir de un caso conocido.

La relación semántica empleada en ambos patrones es del tipo “par sustantivo-verbo”, que corresponde a un conjunto de funciones léxicas, usadas para expresar realización y representan el proceso mediante el cual algo ocurre.⁽²⁹⁾ En el ejemplo del cuadro 3, el infinitivo del *verbo* “visitar” es empleado en su forma verbal de pasado de participio, porque se quiso hacer mayor énfasis durante la modelación al *sustantivo* “lugar” donde ocurrieron los contagios (casos).

El fragmento del grafo solo muestra una vista simplificada de lo que pudiera ser el grafo real para el control epidemiológico de la enfermedad. Hay que tener en cuenta que el potencial analítico de este es mayor en la medida en que se disponga

de más información sobre las entidades (nodos) y sus atributos, así como que se pueda contar con la mayor cantidad de relaciones posibles (aristas) detectadas entre tales nodos.

El razonamiento basado en reglas y otros algoritmos de aprendizaje automático que se pueden generar sobre grafos de conocimiento de gran escala, como podría ser el resultante de utilizar la data real de la pandemia en Cuba, ofrecerían un grupo significativo de análisis y habilitaría el descubrimiento de conocimiento en apoyo a la toma de decisión más efectiva y oportuna, como parte de los estudios epidemiológicos que se realizan en el combate contra la COVID-19.

Conclusiones

La pandemia COVID-19 ha generado un amplio movimiento de las comunidades científicas a nivel global en torno al diagnóstico, el tratamiento y la prevención de la enfermedad. Las medidas de aislamiento social y físico que se toman por el Gobierno encuentran brechas en el comportamiento de las personas, que desencadenan nuevos brotes, como los acaecidos en el municipio de Bauta de la provincia de Artemisa, en Cuba, desde mediados de julio del año 2020. De ahí la importancia de prestar una mayor atención por parte de la comunidad científica al apoyo a los rigurosos estudios epidemiológicos que se efectúan por el Ministerio de Salud Pública.

La construcción de grafos de conocimiento para apoyar los estudios epidemiológicos en el enfrentamiento a la pandemia es una solución viable y puede contribuir a alcanzar una mayor precisión y eficacia en el rastreo de los contactos y, por tanto, en la eficiencia de las medidas y los protocolos previstos.

Los grafos de conocimientos se han convertido en plataformas de análisis e inteligencia artificial para el manejo de grandes y/o complejos conjuntos de datos, como lo demuestran algunas de las grandes empresas de la Web: Facebook,

Google, eBay, Microsoft e IBM, además del cada vez mayor uso que se les está dando en el sector público y empresarial para gestionar datos complejos.

Una simulación del escenario del brote “Fiesta de Bauta” permitió demostrar la viabilidad en la creación de grafos de conocimiento para el seguimiento de casos de la pandemia, lo cual abre un amplio diapasón para nuevas investigaciones a partir de su desarrollo y explotación, con el fin de apoyar la mancomunada lucha que está desplegándose en toda la nación por combatirla.

Próximas investigaciones estarán encaminadas a incluir otras fuentes de datos (ej. de las redes sociales) y de otras fuentes semiestructuradas o no estructuradas, que pueden aportar más elementos del historial de contactos y los lugares de exposición al virus. Se fortalecerá también la línea de análisis sobre los grafos, con la experimentación de nuevos patrones de consulta y otros algoritmos de razonamiento y de inteligencia artificial.

Referencias bibliográficas

1. Delgado-Fernández T. Taxonomía de transformación digital. Rev Cubana Transform Dig. 2020 [12/08/2020];1(1):4-23. Disponible en: <https://rctd.uic.cu/rctd/article/view/62>
2. Noy N, Gao Y, Jain A, Narayanan A, Patterson A, Taylor J. Industry-scale knowledge graphs: lessons and challenges. Queue. 2019;17(2):48-75. DOI: <https://doi.org/10.1145/3331166>
3. Paulheim H. Knowledge graph refinement: A survey of approaches and evaluation methods. Sem Web. 2017 [12/08/2020];8(3):489-508. Disponible en: <http://www.semantic-web-journal.net/system/files/swj1167.pdf>
4. Dirschl C, Kent J, Schram J, Reul Q. Enabling Digital Business Transformation through an enterprise Knowledge Graph. ESWC - Industry_Track; 2020 [12/08/2020]. Disponible en: https://preprints.2020.eswc-conferences.org/industry_track/paper_277.pdf

5. Heist N, Hertling S, Ringler D, Paulheim H. Knowledge Graphs on the Web-an Overview. arXiv preprint arXiv:2003.00719; 2020 [12/08/2020]. Disponible en: <https://arxiv.org/abs/2003.00719>
6. Ehrlinger L, Wöß W. Towards a Definition of Knowledge Graphs. SEMANTiCS; 2016;48:1-4. Disponible en: <https://www.semanticscholar.org/paper>
7. Saorín T. Grafos de conocimiento y bases de datos en grafo: conceptos fundamentales a partir de una" obra maestra" del Museo del Prado. Anuario Think EPI; 2019.
8. Hogan A, Blomqvist E, Cochez M, d'Amato C, de Melo G, Gutiérrez C, Gayo JE, Kirrane S, Neumaier S, Polleres A, Navigli R. Knowledge graphs. arXiv Preprint; 2020 [12/08/2020]. Disponible en: <https://arxiv.org/abs/2003.02320>
9. Lehmann J, Isele R, Jakob M, Jentzsch A, Kontokostas D, Mendes PN, Hellmann S, Morsey M, Van Kleef P, Auer S, Bizer C. DBpedia-a large-scale, multilingual knowledge base extracted from Wikipedia. Sem Web; 2015 [12/08/2020];6(2):167-95. Disponible en: <https://content.iospress.com/articles/semantic-web/sw134>
10. Kondreddi SK, Triantafillou P, Weikum G. Combining information extraction and human computing for crowdsourced knowledge acquisition. IEEE 30th International Conference on Data Engineering; 2014 [12/08/2020]. pp. 988-99. Disponible en: <https://ieeexplore.ieee.org/abstract/document/6816717/>
11. Grainger T, AlJadda K, Korayem M, Smith A. The Semantic Knowledge Graph: A compact, auto-generated model for real-time traversal and ranking of any relationship within a domain. IEEE International Conference on Data Science and Advanced Analytics (DSAA); 2016.
12. Michel F, Gandon F, Ah-Kane V, Bobasheva A, Cabrio E, Corby O, Gazzotti R, et al. Covid-on-the-Web: Knowledge graph and services to advance COVID-19 research. International Semantic Web Conference; 2020.
13. Mohamed A, Abuoda G, Ghanem A, Kaoudi Z, Aboulnaga A. RDF Frames: Knowledge Graph Access for Machine Learning Tools. arXiv:2002.03614v1; 2020.

14. Zou Y, Liu Y. The Implementation Knowledge Graph of Air Crash Data based on Neo4j. IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC); 2020.
15. Tejero A, Rodríguez-Doncel V, Pau I. Knowledge Graphs for Innovation Ecosystems. arXiv preprint; 2020 [12/08/2020]. Disponible en: <https://arxiv.org/abs/2001.08615>
16. Stuart-Cárdenas ML, Delgado-Fernández T, Delgado-Fernández M, Piedra Y. Datos empresariales enlazados: Revisión sistemática desde una perspectiva organizacional. ALCANCE; 2020 [12/08/2020];9:23. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2411-99702020000200153
17. Galkin M, Auer S, Vidal ME, Scerri S. Enterprise Knowledge Graphs: A Semantic Approach for Knowledge Management in the Next Generation of Enterprise Information Systems. ICEIS; 2017 [18/04/2017]:88-98. Disponible en: <https://www.scitepress.org/Papers/2017/63252/63252.pdf>
18. Bader SR, Grangel-González I, Nanjappa P, Vidal ME, Maleshkova M. A Knowledge Graph for Industry 4.0. European Semantic Web Conference; 2020:465-80.
19. Rotmensch M, Halpern Y, Tlimat A, Horng S, Sontag D. Learning a health knowledge graph from electronic medical records. Scient Rep. 2017;7(1):5994. DOI: <https://doi.org/10.1038/s41598-017-05778-z>
20. Gyrard A, Gaur M, Shekarpour S, Thirunarayan K, Sheth A. Personalized health knowledge graph. Core Scholar Publications; 2018 [acceso: 28/07/2020]. Disponible en: <https://corescholar.libraries.wright.edu/>
21. Yu T, Li J, Yu Q, Tian Y, Shun X, Xu L, Zhu L, Gao H. Knowledge graph for TCM health preservation: design, construction, and applications. Artificial Intelligence in Medicine. 2017;77:48-52. DOI: <https://doi.org/10.1016/j.artmed.2017.04.001>
22. Domingo-Fernández D, Baksi S, Schultz B, Gadiya Y, Karki R, Raschka T, et al. COVID-19 Knowledge Graph: a computable, multi-modal, cause-and-effect knowledge model of COVID-19 pathophysiology. BioRxiv. 2020. DOI:

<https://doi.org/10.1101/2020.04.14.040667v1.full-text>

23. Wang Q, Li M, Wang X, Parulian N, Han G, Ma J, et al. COVID-19 Literature Knowledge Graph Construction and Drug Repurposing Report Generation. arXiv Preprint; 2007 [acceso: 01/07/2020]. Disponible en:

<https://arxiv.org/abs/2007.00576>

24. Wise C, Ioannidis VN, Calvo MR, Song X, Price G, Kulkarni N, et al. COVID-19 Knowledge Graph: Accelerating Information Retrieval and Discovery for Scientific Literature. arXiv Preprint; 2007 [acceso: 24/07/2020]. Disponible en:

<https://arxiv.org/abs/2007.12731>

25. Chen C, Ebeid IA, Bu Y, Ding Y. Coronavirus Knowledge Graph: A Case Study. arXiv Preprint; 2007 [acceso: 04/07/2020]. Disponible en:

<https://arxiv.org/abs/2007.10287>

26. Ilievski F, Garijo D, Chalupsky H, Divvala NT, Yao Y, Rogers C, et al. KGTK: A Toolkit for Large Knowledge Graph Manipulation and Analysis. arXiv Preprint; 2006 [acceso: 29/05/2020]. Disponible en: <https://arxiv.org/abs/2006.00088>

27. Chen WJ, Yang SY, Chang JC, Cheng WC, Lu TP, Wang YN, et al. Development of a semi-structured, multifaceted, computer-aided questionnaire for outbreak investigation: e-Outbreak Platform. Biomed J. 2020 [acceso: 20/06/2020].

Disponible en:

<https://www.sciencedirect.com/science/article/pii/S2319417020300949>

28. Arruda N, Venceslau AD, da Cruz MM, Vidal VM, Pequeno VM. Publishing and Consuming Semantic Views for Construction of Knowledge Graphs. InICEIS; 2020 [acceso: 20/06/2020];1:197-204. Disponible en:

<https://www.semanticscholar.org/paper/Publishing-and-Consuming-Semantic-Views-for-of-Arruda-Venceslau/3ddf6802b3eb40eac320ff34656ec23985166b40>

29. Claveau V, L'Homme MC. Discovering specific semantic relationships between nouns and verbs in a specialized French corpus. 3rd International Workshop on Computational Terminology; 2004.

Conflicto de intereses

Los autores declaran que no existe conflicto de intereses.

^aRecuperado el 26 de enero de 2021 del sitio del Centro de Recursos del Coronavirus de la Universidad Johns Hopkins:

<https://coronavirus.jhu.edu/map.html>

^bhttps://en.wikipedia.org/wiki/Google_Knowledge_Graph

^c<https://www.linkedin.com/company/ontotext-ad/?originalSubdomain=es>

^d<https://www.w3.org/2001/sw/wiki/Ontop>

^e<https://www.w3.org/RDF>

^f<https://www.w3.org/TR/sparql11-query>

^g<https://covid19cubadata.github.io/>