

GRAFOS VIRTUALES DE CONOCIMIENTO PARA LA INTEGRACIÓN DE DATOS EMPRESARIALES EN UNA EMPRESA CUBANA

VIRTUAL KNOWLEDGE GRAPHS FOR INTEGRATION OF BUSINESS DATA BASED IN ENTERPRISE CUBAN

Mavis Lis Stuart Cárdenas ^{I*}  <https://orcid.org/0000-0002-0461-7118>

Tatiana Delgado Fernández ^I  <https://orcid.org/0000-0002-4323-9674>

Mercedes Delgado Fernández ^{II}  <https://orcid.org/0000-0003-2556-1712>

Manuel de la Iglesia Campos ^I  <https://orcid.org/0000-0002-2866-8454>

^I Universidad Tecnológica de La Habana “José Antonio Echeverría”, La Habana, Cuba

^{II} Escuela Superior de Cuadros del Estado y del Gobierno, La Habana, Cuba

*Autor para dirigir correspondencia: mavis@ind.cujae.edu.cu

Clasificación JEL: C02, C63, C65

DOI: <https://doi.org/10.5281/zenodo.6472957>

Recibido: 12/12/2021

Aceptado: 27/03/2022

Resumen

La integración de datos empresariales, en entornos de alta heterogeneidad sintáctica y semántica, utilizando grafos virtuales de conocimiento y alineado a las necesidades de la organización contribuye a mejorar el desempeño de los procesos operacionales. El artículo tiene como objetivo mostrar la aplicación de los grafos virtuales de conocimientos en la Empresa de Mensajería y Cambio Internacional (EMCI) de Correos de Cuba, en particular al proceso de planificación de la extracción de la carga del aeropuerto, donde se requirió la integración de datos heterogéneos de las prealertas o información adelantada de múltiples agencias, con la incidencia en un mejor desempeño del proceso operacional de las prealertas debido a la reducción del indicador del tiempo de estancia en el aeropuerto.

Palabras clave: datos heterogéneos, gestión de datos, datos integrados empresariales, grafos virtuales de conocimiento, proceso operacional

Abstract

The integration of business data, in environments with high syntactic and semantic heterogeneity, using virtual knowledge graphs and aligned to the needs of the organization contributes to improving the performance of operational processes. The article aims to show the application of virtual graphs of knowledge in the International Exchange and Messaging Company (EMCI) of Correos de Cuba, in particular to the planning process for the extraction of cargo from the airport, where the integration of heterogeneous data from pre-alerts or advanced information from multiple agencies, with the impact on a better performance of the operational process of pre-alerts due to the reduction of the indicator of time spent at the airport.

Keywords: heterogeneous data, data management, business integrated data, virtual knowledge graphs, operational process

Introducción

En las empresas existe una creciente heterogeneidad de fuentes, formatos y estructuras de información,^{1,2} así como, de aplicaciones informáticas no integradas entre sí y grandes volúmenes de datos distribuidos en varios sistemas que complejizan el tratamiento adecuado de este importante recurso.^{3,4} Esto provoca inconsistencia de los datos e información poco fiable, todo lo cual repercute negativamente en el uso de los datos para la toma de decisiones, lo que ha podido ser comprobado en más de 15 empresas cubanas.^{5,6,7}

Con la creciente necesidad de integrar los múltiples tipos de datos de diversas fuentes, los sistemas de gestión de datos a menudo se enfrentan con diferentes tipos de heterogeneidad,⁸ como la sintáctica, la de nivel de esquemas o de estructuras y la heterogeneidad semántica.^{9,10} Los grafos empresariales de conocimientos constituyen un enfoque emergente para la integración de datos empresariales. Debido a su versátil capacidad de representación, los grafos pueden ser usados para integrar diferentes fuentes de datos heterogéneas, tanto dentro, como entre organizaciones.^{11,12} Se difunden por Google, y ya en el año 2019, su grafo de conocimientos incluía mil millones de entidades y alrededor de 70 mil millones de significados.¹³ Se usan en muchos sistemas de información que requieren acceso a conocimiento estructurado y pueden ser dependientes o independientes del dominio.¹⁴ Constituyen un paradigma flexible de representación del conocimiento y son considerados como un habilitador clave para varios casos de uso.¹⁵

También conocido como grafo de datos, un KG (por sus siglas en inglés, Knowledge Graph) 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.¹⁶ Los grafos de conocimiento pueden generarse a partir de la extracción de bases de conocimiento de la Web, como el conjunto de datos DBpedia, que está publicado en el proyecto de Datos Abiertos Enlazados¹⁷; 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,¹⁸ por lo que se entiende que el grafo está construido sobre otras bases de datos, que al integrarse se enriquecen e incrementan su valor.¹⁹ Los grafos de conocimiento son fácilmente integrables, lo cual está determinando su uso preferencial respecto a otros enfoques similares y contribuyen a organizar los datos al interior de la empresa.²⁰

GRAFOS VIRTUALES DE CONOCIMIENTO PARA LA INTEGRACIÓN DE DATOS EMPRESARIALES EN UNA EMPRESA CUBANA

El artículo tiene como objetivo mostrar las dimensiones utilizadas para la integración de datos empresariales basada en grafos virtuales de conocimiento en una empresa cubana. Como resultado se aplican las dos dimensiones definidas en el proceso de planificación de la extracción de la carga del aeropuerto de la Empresa de Mensajería y Cambio Internacional (EMCI). La solución requirió la integración de datos heterogéneos de las prealertas o información adelantada de múltiples agencias, con la incidencia en un mejor desempeño del proceso operacional debido a la reducción del indicador del tiempo de estancia en el aeropuerto.

Materiales y métodos

La integración de datos empresariales utilizando grafos virtuales de conocimiento y alineado a las necesidades de la organización se lleva a cabo según las dos dimensiones mostradas en la **Figura 1**. Estas dimensiones se refieren a la Alineación de los Objetivos Empresariales (AOE) y el Desarrollo de la Solución de Integración (DSI).



Figura 1. Dimensiones para la integración de los datos empresariales

Fuente: elaboración propia

La aplicación de la integración de datos empresariales con grafos virtuales de conocimientos se desarrolla en la Empresa de Mensajería y Cambio Internacional (EMCI),²¹ que es una de las 20 empresas que integran la Organización Superior de Dirección Empresarial (OSDE) Correos de Cuba. La EMCI brinda servicios de importación y exportación de correspondencia y encomiendas (bultos) postales, de mensajería y paquetería expresa, aduanales y transitarlos, y que asegura el correo oficial a los organismos y organizaciones del Estado cubano, además, de la relación con los 192 países pertenecientes a la Unión Postal Universal.

Resultados y discusiones

La aplicación de las dimensiones para la integración de datos empresariales en la EMCI se presenta, con sus fases y actividades.

Dimensión Alineación a los Objetivos Empresariales (AOE) en la EMCI

La Dimensión Alineación a los Objetivos Empresariales garantiza la coherencia de la solución de TI para la integración con los objetivos y procesos de la EMCI, mediante la visión y desarrollo de una arquitectura mínima viable.

Fase AOE.01 - Visión de la arquitectura empresarial para la integración

Actividad AOE.01.01 Identificar problemas de desempeño en procesos operacionales

El análisis de la documentación de la empresa permitió constatar que, a pesar de los resultados alcanzados en el desempeño operacional de la EMCI, en el transcurso de los últimos cuatro (4) años persisten problemáticas. Algunas de ellas son: 1) períodos de sobre saturación de productos en la planta, que incluye la llegada de contenedores por encima de la capacidad que se puede procesar y 2) demoras en la extracción de los envíos del aeropuerto. Estas problemáticas ocasionan demoras excesivas para la entrega de los productos (envíos) y crean insatisfacciones en los clientes con el consecuente aumento de solicitudes de información a las dependencias correspondientes de la EMCI y la OSDE. El indicador tiempo de entrega de los pedidos con frecuencia tiene valores por encima de lo establecido en las normas de calidad.

Actividad AOE.01.02 Visionar la arquitectura empresarial

Al analizar la cadena logística de los envíos se evidencia el desarrollo de diferentes momentos: primero en el aeropuerto, segundo en la planta de procesamiento y el último en las oficinas de correo de los destinos, antes de llegar al cliente final. El tiempo asociado a cada uno de los momentos influye a favor o en detrimento de la entrega en tiempo de los envíos a los clientes. Una de las problemáticas identificadas, a partir del análisis del desempeño de la empresa, se relaciona directamente con la demora en la extracción de los envíos del aeropuerto, con una permanencia de los envíos en el aeropuerto por encima de los cinco días normados.

Conjuntamente con los controles aduanales para la extracción de la carga del aeropuerto se debe realizar la planificación para la extracción de dicha carga. Esa operación la realiza la Dirección de Operaciones para lo cual debe considerar los datos relacionados con el contenido de los envíos incluidos en la carga, tales como: volumen total, peso total, totales de tipo de carga, entre otros datos de operación. La obtención de estos datos es posible a partir del análisis de la información que se brinda de forma adelantada, por los operadores Courier en las Prealertas. Sin embargo, se desaprovecha el potencial de esta información para la planificación de medios y recursos que garanticen en tiempo y forma el desarrollo exitoso de las operaciones, al no disponerse de una herramienta que brinde una visión consolidada de todas las prealertas, ni se dispone de los datos globales de operación.

La solución de esta problemática a la que se enfrenta la Dirección General de la EMCI y la Dirección de Operaciones requiere de la integración de datos con enfoque de arquitectura empresarial. En línea con esta problemática, el objetivo de la arquitectura empresarial para la integración de datos en la EMCI es introducir una solución de integración con orientación a los procesos operacionales de la EMCI donde las insuficiencias informacionales y la alta heterogeneidad estén incidiendo negativamente en el desempeño de la empresa (alineación TI - empresa). Tal solución debe coexistir con el resto de las

GRAFOS VIRTUALES DE CONOCIMIENTO PARA LA INTEGRACIÓN DE DATOS EMPRESARIALES EN UNA EMPRESA CUBANA

aplicaciones que existen en la EMCI, y consumir datos de dichas bases de datos, lo que proporciona la información integrada. Debe ser fácil y amigable, y también flexible y escalable.

Para definir este objetivo, el equipo de arquitectura empresarial (profesores y especialistas de la Universidad Tecnológica de La Habana) sostuvo varias entrevistas con los directivos principales de la empresa, del área de operaciones y del departamento de informática. La **Tabla 1** despliega los roles en la EMCI para la Arquitectura Mínima Viable (AMV) de la solución de integración de datos.

Tabla 1. Roles en la EMCI de la AMV

Roles AMV	Equipo de trabajo
Arquitecto empresarial	Consultor externo (CUJAE)
Arquitecto de negocio	
Ingeniero ontológico	Especialista Informático (Empresa Software contratada)
Arquitecto de datos	
Arquitecto de aplicación/ tecnología	
Especialista de operaciones	Experto en procesos operacionales de la EMCI
Especialista Soporte técnico	Especialista EMCI

Fuente: elaboración propia

En esta investigación se aplica la integración de datos con grafos virtuales de conocimiento al proceso planificación de la extracción de la carga del aeropuerto que se lleva a cabo en el Servicio de Paquetería y Carga no comercial. Los diagnósticos de la gestión de la información evidenciaron, que existe un entono de alta heterogeneidad sintáctica y semántica de los datos, que requiere de una integración alineada a los objetivos empresariales.²¹ Se aplica el procedimiento de diagnóstico²² que detectó las insuficiencias informacionales:

- No existe un proceso formal para procesar las prealertas, que brinda la información adelantada de la carga para los procesos posteriores.
- Cada área considera la prealerta de forma independiente con el riesgo de introducir errores.
- El contenido de la prealerta (Excel, pdf, XML según la agencia Courier) se incorpora en un sistema automatizado, pero no se procesa la información para la planificación del despacho de la carga; sólo para su uso aduanero.
- La prealerta llega a las áreas: comercial, operaciones y la aduana y no se procesa ni se aportan datos generales sobre las mismas, que contribuyan a organizar y planificar el servicio.
- Carencia de datos consolidados con el contenido de los envíos incluidos en la carga

Los objetivos y principios generales de la AMV para la solución de integración en la EMCI se muestran en la **Tabla 2**.

Tabla 2. Artefacto: principios referenciales de la AMV en la EMCI

Capa de	Principio
Negocio	La AMV se circunscribe al proceso de planificación de la extracción de la carga del aeropuerto, que incide de forma directa en el desempeño operacional identificado.
Información	Integración de datos con grafos virtuales de conocimiento de los datos de prealertas enviadas por las agencias Courier y DHL.
Aplicación	Consumo de datos provenientes de ficheros de prealertas, sin afectar los sistemas informáticos de EMCI.
Tecnología	Al ecosistema tecnológico de la EMCI se incorpora el grafo virtual de conocimiento que integra información de ficheros de prealertas.

Fuente: elaboración propia

GRAFOS VIRTUALES DE CONOCIMIENTO PARA LA INTEGRACIÓN DE DATOS EMPRESARIALES EN UNA EMPRESA CUBANA

Fase AOE.02 - Arquitectura empresarial de referencia para la integración de datos

En esta segunda fase se define la AMV de referencia para la integración de información basada en grafos de conocimientos en la EMCI.

Actividad AOE.02.01 – Gestionar la capa de negocio de la arquitectura empresarial

La carga, al llegar al aeropuerto se almacena en un depósito temporal para luego coordinar la fecha de distribución en función de las capacidades que posea el depósito de la EMCI y en función de la fecha de entrada a este depósito se coordina entonces la fecha de salida para su futuro traslado hacia el depósito de la EMCI. Desde aquí, la unidad organizativa de la Aduana General de la República otorga la fecha de salida de estos paquetes del depósito según el orden de llegada al mismo, y pasan entonces a la planta de tratamiento postal de la Oficina de Cambio Internacional (OCI).

La **Figura 2** muestra el diagrama del nuevo servicio de integración de datos a partir de las prealertas.

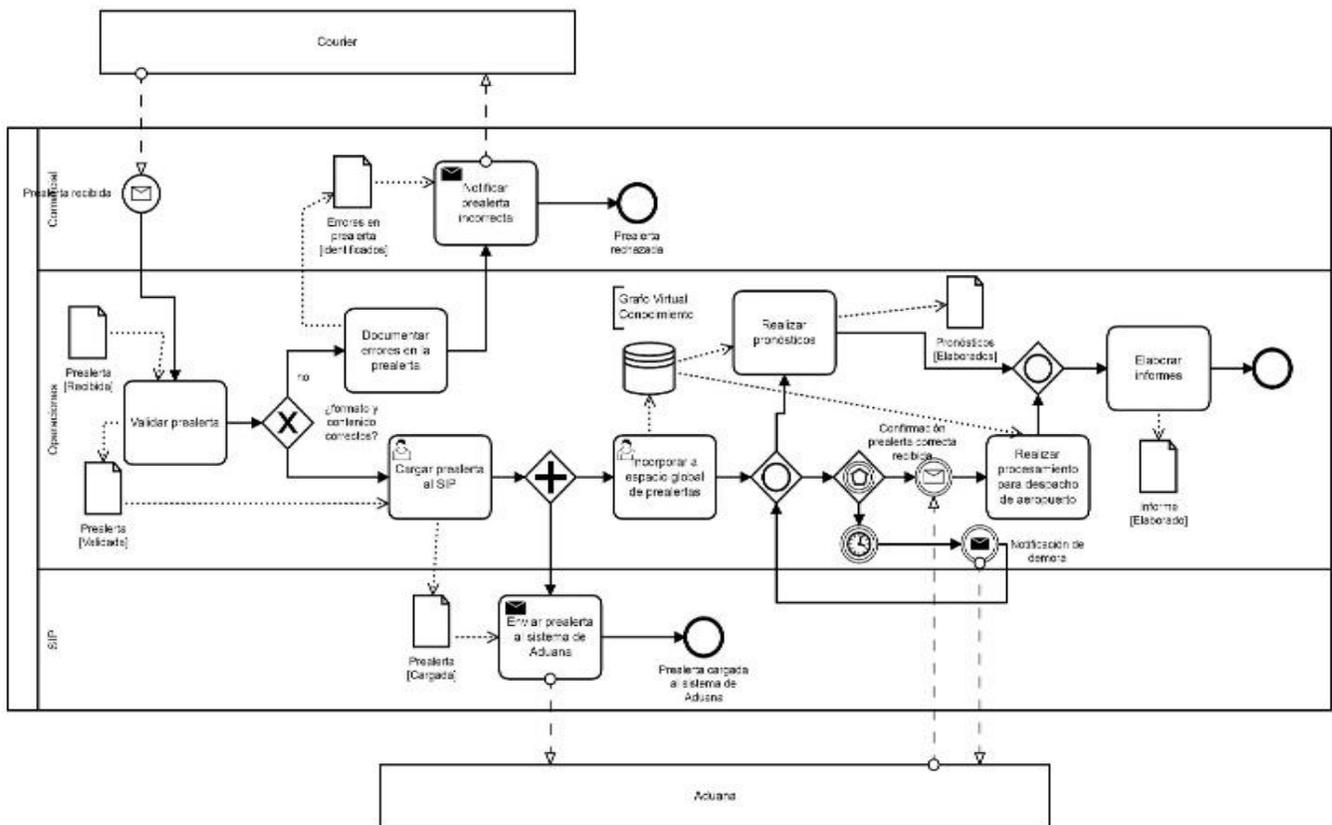


Figura 2. Diagrama del Mapa del nuevo proceso basado en prealertas

Fuente: elaboración propia

La propuesta del nuevo proceso de prealertas soporta el uso del Grafo Virtual de Conocimiento que genera la solución de integración, y se incluyen nuevos procesamientos en la Dirección de Operaciones de la EMCI. Dicha propuesta dota a esta dirección de la capacidad de consultar integradamente los datos

GRAFOS VIRTUALES DE CONOCIMIENTO PARA LA INTEGRACIÓN DE DATOS EMPRESARIALES EN UNA EMPRESA CUBANA

de los respectivos envíos, a partir de los datos recibidos en las prealertas, sin afectar el despliegue actual de dichos ficheros.

Actividad AOE.02.02 – Gestionar la capa de información de la arquitectura empresarial

Las prealertas constituyen las fuentes de datos sujetas a la integración para resolver la problemática de la EMCI relacionada con los tiempos de entrega del envío. Específicamente, las prealertas contienen la individualización de cada una de las Guías Courier que transporta el mensajero internacional y contiene información sobre la vía (aérea, terrestre o marítima) por la que se traslada la cantidad de bultos que contiene el envío y su peso en kilogramo, la agencia que lo envía, la línea área que lo transporta, el número de vuelo y el país origen. De cada envío en particular contiene un número identificativo, su peso individual, fecha, datos del remitente como nombre y número de identidad; descripción del contenido y datos del destinatario como nombre, carnet de identidad, dirección, provincia y municipio. Esta información se recibe y se procesa de forma separada por distintos especialistas, y según sea la agencia que envíe las prealertas, varía el formato. esta información se recibe y se procesa de forma separada por distintos especialistas, y según sea la agencia que envíe las prealertas, varía el formato. La **Tabla 3** se corresponde con la identificación y caracterización de las fuentes de datos de las prealertas, para la creación de la Ontología de Integración.

Tabla 3. Diversidad de formatos de prealertas

Fuentes	Ubicación	Formato	Uso
Agencias COURIER	Estación de trabajo Especialista Comercial de la empresa	xls	Validación Contenido
			Diseño del despacho de aeropuerto
			Control de Seguridad
Agencia DHL	Sistema DHL operado en la UEB de Mensajería Express	xml	Validación Contenido
			Planificación del despacho de aeropuerto
			Control de Seguridad
Agencia Copa	Estación de trabajo Especialista Comercial de la paquetería Copa	pdf	Validación Contenido
			Diseño del despacho de aeropuerto
			Control de Seguridad

Fuente: elaboración propia

Se analizan, los esquemas de las fuentes de datos de prealertas que serán integradas y se precisan las tablas y las columnas. La **Figura 3** representa el esquema de la tabla de prealerta de un Courier. Las prealertas, semánticamente, describen información sobre los envíos, destinatarios y remitente. La **Tabla 4** refleja tres conceptos importantes en el dominio de aplicación.

GRAFOS VIRTUALES DE CONOCIMIENTO PARA LA INTEGRACIÓN DE DATOS EMPRESARIALES EN UNA EMPRESA CUBANA



Figura 3. Esquema de datos Courier

Fuente: elaboración propia

Tabla 4. Descripción de los conceptos representados en las fuentes de datos

Fuente	Concepto	Descripción	Tabla/Campo
Courier	Agencia	Nombre de la Agencia de envío	Agencia
DHL			Agency
Courier	País	Nombre del país del que proviene el envío	PaísOrigen
DHL			MovementOriginCountry
Courier	Peso	Peso en Kg del envío que se recibe	PesoKg
DHL			ShipmentWeight

Fuente: elaboración propia

La **Tabla 4** evidencia la heterogeneidad semántica presente en el proceso de prealertas, con varios términos que se corresponden al mismo significado o conflicto de sinónimos. Se identifican los conceptos para el tratamiento de la prealerta, referidos a: manifiesto o prealerta, Courier, paquete y personas. Se

GRAFOS VIRTUALES DE CONOCIMIENTO PARA LA INTEGRACIÓN DE DATOS EMPRESARIALES EN UNA EMPRESA CUBANA

definen las propiedades: código postal, descripción, dirección, peso, nombre, apellidos, teléfono, provincia, país, número de guía y de agencia, fecha de nacimiento y de imposición. La **Figura 4** muestra el modelo de dominio de la prealerta.

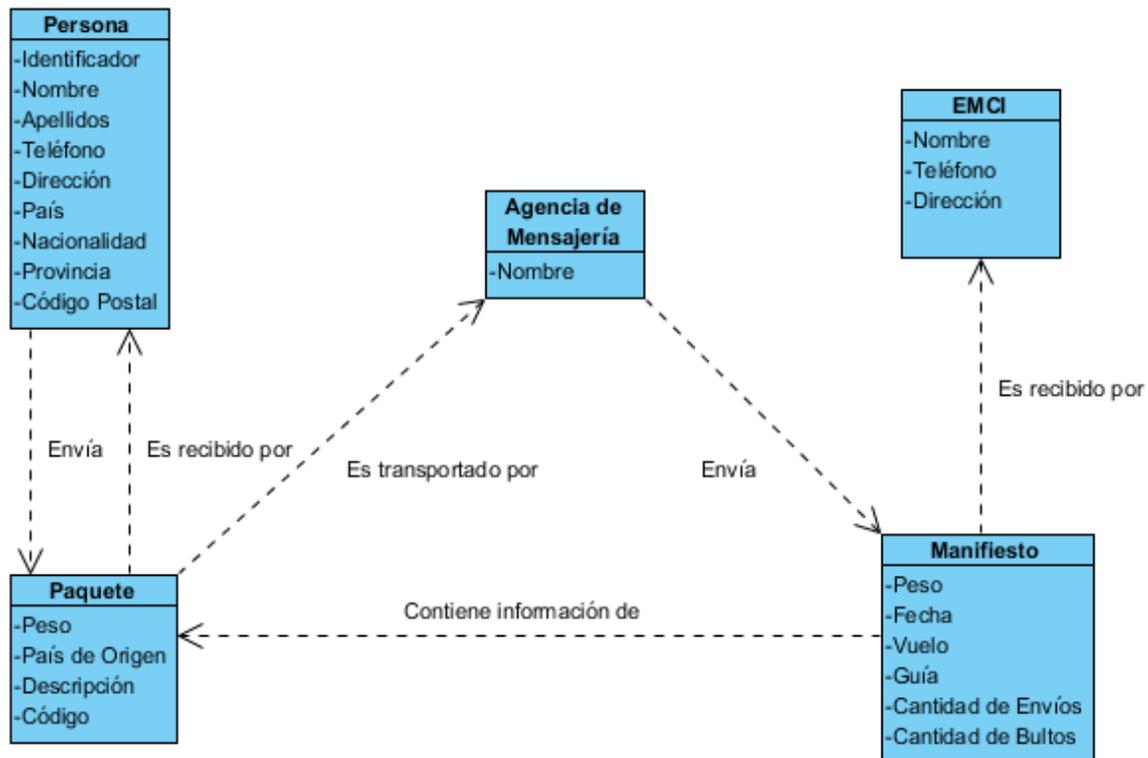


Figura 4. Modelo de dominio de la prealerta

Fuente: elaboración propia

Para el diseño de la ontología de integración, participó, como líder de la etapa, un especialista informático con el rol de ingeniero ontológico, subcontratado por la empresa para llevar a cabo el desarrollo del grafo virtual de conocimiento sobre Ontop. Tomando en consideración el modelo de dominio, se crea una red de ontologías, a partir de la reutilización de clases y propiedades de los vocabularios foaf y vcard y de la creación de una ontología propia para estos efectos, la ontología de integración Operaciones. Del vocabulario FOAF (<https://es.wikipedia.org/wiki/FOAF>), se utilizan clases y propiedades en la modelación de personas que representan a los remitentes y destinatarios. FOAF es una ontología legible para las máquinas que describe a las personas, sus actividades y sus relaciones con otras personas y objetos. Del vocabulario VCARD (<https://es.wikipedia.org/wiki/VCard>), se utilizan las propiedades para completar la modelación de los datos de las direcciones de remitentes y destinatarios. VCARD es un formato estándar para el intercambio de información personal. Pueden contener nombre, dirección, números telefónicos, entre otras informaciones. Para el diseño de la ontología se desplegó sobre Protégé una guía para la construcción del grafo basado en Ontop en la EMCI, resultado del contrato entre la EMCI y la CUJAE. La ontología de integración Operaciones se muestra en la **Figura 5**.

GRAFOS VIRTUALES DE CONOCIMIENTO PARA LA INTEGRACIÓN DE DATOS EMPRESARIALES EN UNA EMPRESA CUBANA

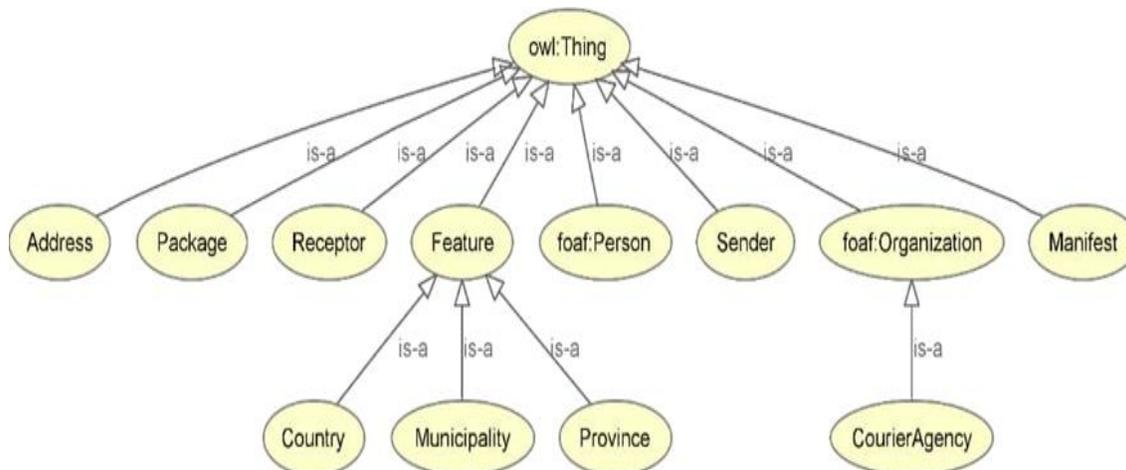


Figura 5. Modelo ontológico de caso de integración de datos de prealertas

Fuente: elaboración propia

Actividad AOE.02.03 – Gestionar la capa de aplicación de la arquitectura empresarial

El patrón de la arquitectura de solución del modelo INTEGRAL es reutilizado en esta actividad. La vista gráfica de la arquitectura se muestra en la **Figura 6**.

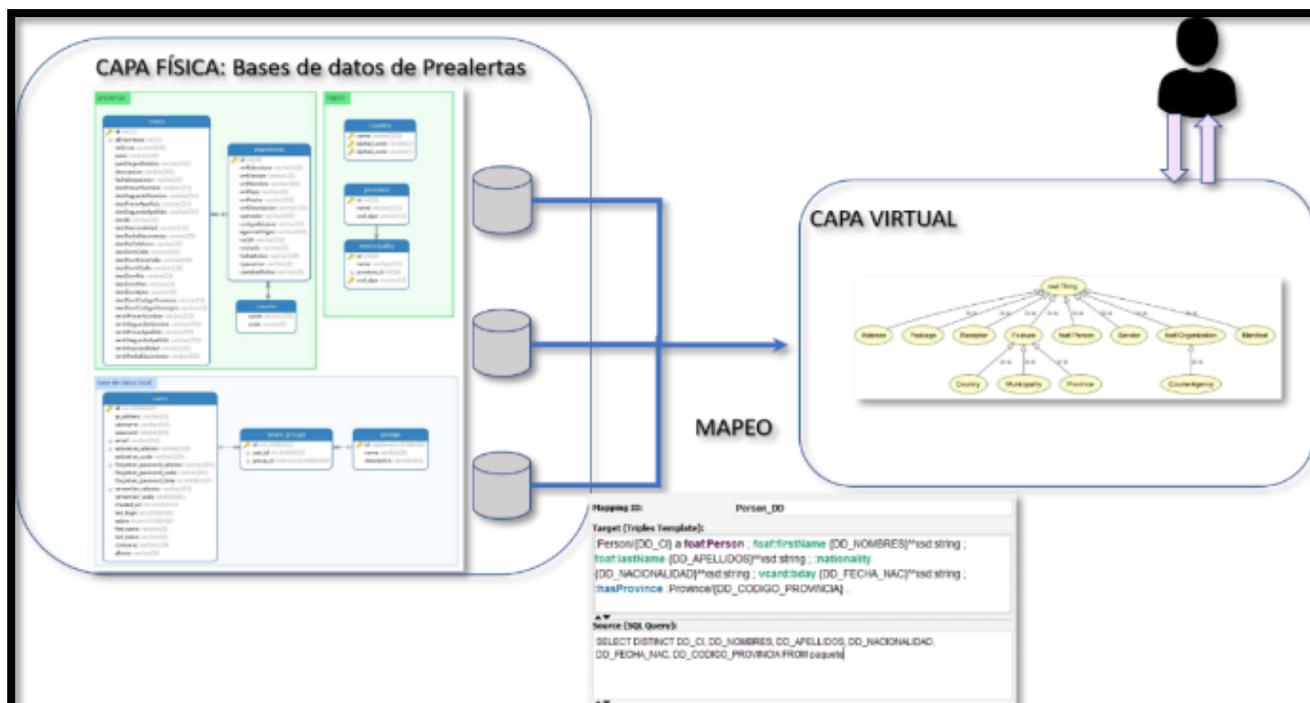


Figura 6. Arquitectura de solución de las prealertas

Fuente: elaboración propia

GRAFOS VIRTUALES DE CONOCIMIENTO PARA LA INTEGRACIÓN DE DATOS EMPRESARIALES EN UNA EMPRESA CUBANA

Los componentes de la arquitectura de solución de las prealertas se describen a continuación:

- Conjunto de datos a integrar: Este componente está en la capa física de la arquitectura y corresponde a las bases de datos de las prealertas.

Las informaciones a obtener a partir de la consulta del grafo se definen de los criterios de decisión generada con los directivos, siendo algunos:

- Los envíos próximos a arribar según la fecha, especificando para cada envío, de qué agencia proceden, la cantidad de bultos y la cantidad específica de envíos.
- Las provincias destinatarias de los próximos envíos, con su peso total y la cantidad de paquetes.

Actividad AOE.02.04 – Gestionar la capa de tecnología de la arquitectura empresarial

En esta capa se definen las prestaciones y características de los sistemas que permiten la materialización de las capas anteriores de la Arquitectura Empresarial (AE), según las variables que se necesitan para el grafo virtual de conocimiento (GVC). La **Tabla 5** muestra estas variables.

Tabla 5. Artefacto: Requerimientos y tecnologías de referencia para las variables de GVC

Variable	Tecnología
Ontología	Sistema Protégé (https://protege.stanford.edu/) 
Conjuntos de Datos	SQL Server
Mapeos	Lenguaje R2RML 
Consultas Típicas	SPARQL
Sistema OBDA	Ontop ⁸

Fuente: elaboración propia

OnTop, implementa Grafos Virtuales de Conocimiento y permite consultar fuentes de datos relacionales a través de una representación conceptual del dominio de interés, proporcionados por los términos de una ontología, a la cual se mapean las fuentes de datos. Las características principales de Ontop son: sólidos fundamentos teóricos, acercamiento virtual a la OBDA, que evita materializar las tripletas y se implementa a través de la técnica de reescritura de las consultas, optimizaciones extensivas, explotando todos los elementos de la arquitectura OBDA, el cumplimiento de estándares como las consultas SPARQL, los mapeos R2RML, y las ontologías OWL 2 QL y RDFS, y finalmente su soporte a los principales gestores de bases de datos relacionales.²³

Cuando es necesario integrar varias fuentes de datos, los sistemas de respuesta a consultas de GVC a menudo se usan juntos con herramientas de federación de datos.²⁴ Los federadores SQL proporcionan una capa relacional unificada sobre múltiples fuentes de datos y evalúan Consultas SQL sobre la capa unificada. Estos sistemas a menudo también admiten fuentes de datos no relacionales, por ejemplo, XML archivos, archivos JSON, MongoDB o API web, proporcionando una vista relacional sobre su contenido. Con la ayuda de un federador SQL, los sistemas GVC pueden acceder al contenido de múltiples fuentes de datos sin tener que realizar un posprocesamiento complejo, como unir los datos procedentes de diferentes fuentes de datos. En esta investigación se utiliza Dremio (<https://www.dremio.com>) como

GRAFOS VIRTUALES DE CONOCIMIENTO PARA LA INTEGRACIÓN DE DATOS EMPRESARIALES EN UNA EMPRESA CUBANA

servidor de bases de datos federadas, que modifica la Arquitectura de la aplicación, incorporándole una capa de la Base de Datos virtualizada, quedando el gráfico de la arquitectura como se representa en la **Figura 7**.

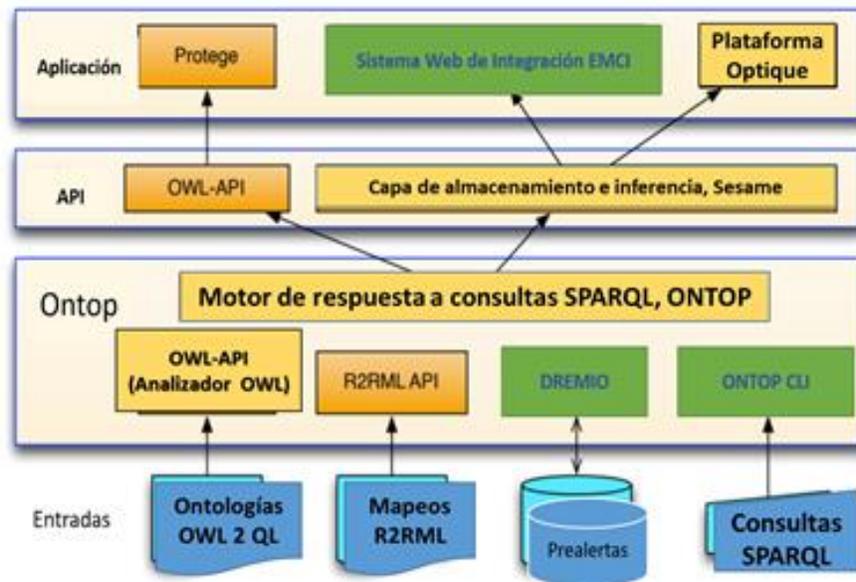


Figura 7. Arquitectura de Tecnología de la solución

Fuente: adaptado de²⁴

Dimensión Desarrollo de la Solución de Integración (DSI) para las prealertas

La Dimensión DSI en la EMCI se encarga de crear la solución de integración de información de prealertas basada en grafos virtuales de conocimiento para facilitar la realización de consultas a dicho grafo y proporcionar la información que se necesita.

Fase DSI.01 Implementación de la solución de integración

Actividad DSI.01.01 – Generar la Información Integrada

En las reglas de mapeo intervienen las bases de datos de las prealertas con las clases y propiedades de la ontología siguientes:

- Clase Person: Person_DD, Person_DR (mapeos de destinatarios y remitentes de paquetes)
- Clase Manifest: Manifest
- Clase Package: Package
- Clases geográficas: Country, Province
- Propiedades: receivesPackage, sendsPackage

Las consultas fueron predefinidas en la etapa de modelación de la arquitectura, de acuerdo a los criterios de decisión aportados por los directivos del área de operaciones de la EMCI. Un ejemplo de corrida de una consulta que genera información integrada, se ilustra a través de las clases involucradas:

- persons: para recuperar la información concerniente tanto a remitentes como a destinatarios

GRAFOS VIRTUALES DE CONOCIMIENTO PARA LA INTEGRACIÓN DE DATOS EMPRESARIALES EN UNA EMPRESA CUBANA

- provincias: para recuperar información correspondiente a las provincias que recibirán paquetes: cantidad de paquetes y su peso total.
- manifests: para recuperar información correspondiente a los manifiestos

Para facilitar el acceso a la información integrada resultante se desarrolló un sistema web visualizador de reportes: Sistema Integrado de Prealertas, que consume el grafo virtual de conocimientos y visualiza las consultas sobre el grafo en una manera amigable y accesible para los directivos de la empresa. Las funcionalidades de una primera versión del sistema, de forma general son:

- Envíos prealertados en un período próximo de tiempo.
- Envíos, relacionados con las Agencias u Operadores Internacionales.
- Para un período dado, consultas relativas a:
 - o Volumen de operación Total (cantidad de bultos; cantidad de envíos; peso en kg)
 - o Provincias que deben recibir envíos en el período
 - o Envíos de una Provincia.
 - o Volúmenes de operación previstos por Provincia.

El uso del sistema, incorporado en el proceso de planificación de la extracción de la carga del aeropuerto y considerando los datos de los volúmenes de las cargas permite determinar las variables relacionadas con: cantidad de medios de transporte, cantidad de recursos humanos y el tipo de los medios de transporte necesarios para la realización de dicho proceso.

Fase DSI.02 Gestión de la capacidad y disponibilidad de la solución de integración

Actividad DSI.02.01 – Gestionar la capacidad y disponibilidad de la solución de integración

El desempeño de la solución de integración, depende principalmente del desempeño propio del sistema Ontop. Es de destacar que el sistema Ontop, se reconoce en la literatura como el sistema GVC de código abierto, actualmente disponible, más maduro y con un desempeño muy robusto.²⁴ La evaluación del desempeño de Ontop reflejó que en general las consultas requerían entre uno y pocos segundos para su ejecución teniendo unas pocas prealertas insertadas, y estando desplegado en un entorno de trabajo de escasos recursos de cómputo.

Adicionalmente, existen limitaciones de la versión actual de Ontop con el empleo de esta herramienta:

- El reporte de errores de la versión más reciente de Ontop no es intuitivo, lo que afecta la gestión y corrección de los mapeos. Se limita a mostrar la traza de errores de Java. Por ejemplo, un error de “uso ilegal de null” en tiempo de ejecución de una consulta, sin más especificaciones, puede deberse a cualquier error leve de sintaxis, por ejemplo: al nombrar una instancia sin el ‘/’ que divide el nombre de la clase, del campo de identificador entre llaves, o al incluir un dato literal sin tipo de datos especificado, fallando el mecanismo de inferencia. También puede deberse al uso de consultas SQL algo más complejas en los mapeos, que requieran el empleo de alias, por ejemplo.
- El lenguaje R2RML genera fallas con el uso de campos con guiones (‘-’), estos deben sustituirse por alias desde el SQL. También genera errores la inserción de valores por defecto, lo que debe hacerse en cambio desde el propio SQL.

Finamente, aunque los errores se hayan rectificado, en muchos casos sólo el reinicio de Ontop actualiza estos cambios.

GRAFOS VIRTUALES DE CONOCIMIENTO PARA LA INTEGRACIÓN DE DATOS EMPRESARIALES EN UNA EMPRESA CUBANA

Fase DSI.03 – Gestión del cambio

Actividad DSI.03.01 – Realizar pruebas y proponer cambios

Para el análisis del desempeño de la solución, se realizaron pruebas que se muestra en la **Tabla 6**.

Tabla 6. Definición del plan de pruebas

	Objetivos	Resultados esperados	Resultados obtenidos	Grado
Alcance y escenario	Integrar ficheros de prealertas en dos formatos diferentes			
	Validar posible propuesta para la integración de la información de fuentes de datos heterogéneas.	Obtener la información integrada.	5 ficheros de prealertas (xls y xml) con 2403 envíos. información integrada de prealertas verificando capacidad de integración de la solución, para fuentes de datos heterogéneas.	Satisfactorio.
	Integrar ficheros de prealertas en dos formatos diferentes			
	Verificar el tiempo de respuesta de la solución de integración.	Obtener la respuesta en un tiempo inferior a un minuto.	5 ficheros de prealertas (xls y xml) con 2403 envíos. Tiempo de respuesta fue de pocos segundos.	Satisfactorio.
	Comprobación del sistema con datos reales			
Verificar el funcionamiento estable y adecuado, de solución de integración.	Obtener respuesta y mantener el funcionamiento estable.	ficheros de prealertas recibidos en dos días de trabajo con 40 ficheros. El resultado fue la saturación del sistema.	Insatisfactorio.	

Fuente: elaboración propia

El plan de acciones de los resultados insatisfactorios se refiere a los cambios que deben ejecutarse para optimizar y mejorar el rendimiento del sistema, a partir de:

- Reducir la ambigüedad en la ontología.
- Reducir la longitud de caracteres de la ontología y las consultas.
- Simplificar las consultas SPARQL.

Actividad DSI.03.02 – Crear habilidades en la operación de la solución

Considerando los roles de la AMV definidos, la arquitectura de tecnología definida en la Actividad AOE.02.04 y la aplicación de una encuesta de diagnóstico de conocimientos, se definieron las necesidades de formación y el plan del curso.

GRAFOS VIRTUALES DE CONOCIMIENTO PARA LA INTEGRACIÓN DE DATOS EMPRESARIALES EN UNA EMPRESA CUBANA

A partir de las necesidades formativas, el sistema de conocimiento del curso se enfocó en las temáticas de ontologías, servidores de bases de datos federadas, Ontop como sistema de grafos virtuales de conocimiento, lenguaje para la especificación de las reglas de mapeo (R2RML) y el lenguaje para la especificación de las consultas (SPARQL). El curso se desarrolló durante el mes de junio de 2021, con actividades presenciales y semipresenciales. Al culminar el curso se aplicó la encuesta sobre la cual se realizó el diagnóstico inicial, cuyos resultados evidenciaron un avance significativo por la comprensión y asimilación de los contenidos. La valoración de los contenidos, se muestran en la **Tabla 7**.

Tabla 7. Valoración general del curso

Contenidos	Valor Medio
ONTOP	4.47
Dremio	4.47
SPARQL	4.00
Ontologías	3.72
R2RML	3.46

Fuente: elaboración propia

Un experimento se desarrolló para valorar la incidencia de la integración de los datos con grafos virtuales de conocimiento en la reducción del tiempo de extracción del envío del aeropuerto. Para ello se tomó una muestra de 89 guías de los meses de junio a septiembre 2021, equivalentes a 6 días de trabajo (último viernes y últimos días de agosto y septiembre). Se realizó una prueba de hipótesis de comparación de medias en relación a la variable tiempo de extracción del envío (TEA) tal como ocurrió en relación al tiempo que hubiese transcurrido si se usaran los datos integrados con las prealertas. Se obtuvo una diferencia significativa de 2,764 días como promedio con la solución integrada de las prealertas, con un Intervalo de confianza del 95 % (1,952; 3.576). La planificación de la extracción del envío con la solución integrada de las prealertas garantiza un tiempo menor de extracción del envío y un mejor desempeño del proceso operacional. De esta forma, se confirma con la aplicación práctica de los grafos virtuales de conocimiento la tendencia a su uso en función de la creación de valor derivado de un uso más efectivo e integrado de los datos de la empresa.²⁵

Conclusiones

La aplicación de las dimensiones Alineación a los Objetivos Empresariales (AOE) y Desarrollo de la Solución de Integración (DSI) para la integración de datos empresariales basado en grafos virtuales de conocimiento en la empresa de Mensajería y Cambio Internacional de Correos de Cuba (EMCI) evidencia la viabilidad y utilidad de la propuesta.

La aplicación de la Dimensión Alineación a los Objetivos Empresariales (AOE) en la EMCI garantiza la alineación de la solución que se propone con la satisfacción del requerimiento del tiempo de extracción del envío del aeropuerto.

La aplicación de la Dimensión Desarrollo de la Solución de Integración (DSI) permite, a través de la generación de un grafo virtual de conocimiento que contiene información integrada de envíos, provincias y personas, da respuesta a los criterios de decisión de los directivos del área de operaciones.

GRAFOS VIRTUALES DE CONOCIMIENTO PARA LA INTEGRACIÓN DE DATOS EMPRESARIALES EN UNA EMPRESA CUBANA

La mejora en el desempeño operacional de los envíos con la integración de datos empresariales basado en grafos virtuales de conocimiento para las prealertas se evidencia con la disminución aproximada de tres días por cada envío.

Referencias bibliográficas

1. DAMA I. The DAMA Guide to the Data Management Body of Knowledge (DAMA-DMBOK Guide) (Second Edition ed.). ISBN-13: 978-1634622349; 2014.
2. Kim Sk, Wang W. An Integrated View of Data: Application of Knowledge Modeling to Data Management, Journal of International Technology and Information Management. 2020;29(2, Article 4). [Consultado 1 julio 2021] Disponible en: <https://scholarworks.lib.csusb.edu/cgi/viewcontent.cgi?article=1448&context=jitim>
3. Akter S, Wamba SF. Big data analytics in E-commerce: a systematic review and agenda for future research, Electronic Markets. 2016;26(2):173-194. [Consultado 5 julio 2021] Disponible en: <https://ro.uow.edu.au/buspapers/886/>.
4. Mikalef P, Pappas IO, Krogstie J, Giannakos M. Big data analytics capabilities: a systematic literature review and research agenda Information Systems and e-Business Management, 2018;16:1-32. [Consultado 3 julio 2021] Disponible en: https://ideas.repec.org/a/spr/infsem/v16y2018i3d10.1007_s10257-017-0362-y.ht
5. Pérez SJ, Novoa B. Diagnóstico de Gestión de Información del Servicio Postal Universal de la Empresa de Mensajería y Cambio Internacional con Enfoque de Arquitectura de Información Empresarial (AIE). [Trabajo de Diploma para optar por el título de Ingeniería Industrial], Tesis tutorada por Stuart, M.L., La Habana: Facultad de Ingeniería Industrial CUJAE, 2018.
6. Stuart ML, Menéndez DD, Martínez Y, Cordero A, Delgado T. Experiencia en el diagnóstico de la Gestión de Información con Enfoque de Arquitectura de Información Empresarial, GECONTEC: Revista Internacional de Gestión del Conocimiento y la Tecnología. 2017;5(1). [Consultado 13 julio 2021] Disponible en: <https://www.upo.es/revistas/index.php/gecontec/article/view/1897>.
7. Stuart ML, Delgado T, Delgado M, Prieto D.R, Quial J. Enfoque de datos empresariales enlazados aplicado en una empresa cubana. Ingeniería Industrial. 2020;41(2): 235-247. [Consultado 15 julio 2021] Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1815-59362020000200007
8. Stuart ML, Prieto DR, Delgado T, Delgado M. Enfoque de Integración Basado en Datos Enlazados Empresariales. Revista Cubana de Administración Pública y Empresarial. 2018;2(3): 268-279. [Consultado 15 julio 2021] Disponible en: <https://apye.esceg.cu/index.php/apye/article/view/56>
9. Alamir E, Urgessa T, GopiKrishna T, Ellappan V. Application of machine learning with Big data analytics in the insurance industry. IAEME Publications; 2020.
10. Giunchiglia F, Zamboni A, Bagchi M, Bocca S. Stratified data integration. 2021 [Consultado 13 julio 2021] Disponible en: <https://arxiv.org/abs/2105.09432>
11. Calvanese D, Liuzzo P, Mosca A, Remesal J, Rezk M, Rull G. Ontology-based data integration in EPNet: Production and distribution of food during the Roman Empire. Engineering Applications of Artificial Intelligence 2016;(51):212-229. [Consultado 21 julio 2021] Disponible en: <https://doi.org/10.1016/j.engappai.2016.01.005>
12. Heist N, Hertling S, Ringler D, Paulheim H. Knowledge Graphs on the Web – an Overview. Mar Preprint; 2020. [Consultado 1 julio 2021] Disponible en: <https://doi.org/10.48550/arXiv.2003.00719>
13. Noy N, Gao Y, Jain A, Narayanan A, Patterson A, Taylor J. Industry-scale Knowledge Graphs:

- Lessons and Challenges. Communications of the ACM, August, 2019;62(8):36-43. [Consultado 1 julio 2021] Disponible en: <https://doi.org/10.1145/3331166>.
14. Paulheim H. Knowledge graph refinement: A survey of approaches and evaluation methods”, Semantic web, 2017;8(3):489-508. [Consultado 7 julio 2021] Disponible en: <http://semantic-web-journal.net/system/files/swj1167.pdf>.
 15. Dirschl C, Kent J, Schram J, Reul Q. Enabling Digital Business Transformation Through an Enterprise Knowledge Graph. In: Harth A. et al. (eds) The Semantic Web: ESWC 2020 Satellite Events. ESWC 2020. Lecture Notes in Computer Science. 2020;12124. Springer, Cham. [Consultado 7 julio 2021] Disponible en: https://doi.org/10.1007/978-3-030-62327-2_45.
 16. Hogan A, Blomqvist E., Cochez M, d'Amato C, de Melo G, Gutierrez C, Gayo JE, Kirrane S, Neumaier S, Polleres A, Navigli R. Knowledge graphs, preprint Mar 4; 2020. [Consultado 13 septiembre 2021] Disponible en: <https://arxiv.org/abs/2003.02320>
 17. Lehmann J, Isele R, Jakob M, Jentzsch A, Kontokostas D, Mendes PN, Hellmann S, Morsey M, Van KP, Auer S, Bizer C. DBpedia – a large-scale, multilingual knowledge base extracted from Wikipedia. Semantic Web Journal. 2015;6(2):167-195. [Consultado 15 julio 2021] Disponible en: http://svn.aksw.org/papers/2013/SWJ_DBpedia/public.pdf.
 18. Kondreddi SK, Triantafyllou P, Weikum G. Combining information extraction and human computing for crowdsourced knowledge acquisition. IEEE 30th International Conference on Data Engineering. Mar 31; 2014:988-999. [Consultado 1 octubre 2021] Disponible en: <https://doi.org/10.1109/ICDE.2014.6816717>
 19. 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. Jan 2019;1(13). [Consultado 5 octubre 2021] Disponible en: <https://doi.org/10.3145/thinkepi.2019.e13f05>
 20. Delgado T, Stuart ML, Delgado M. Grafos de conocimiento para gestionar información epidemiológica de la COVID-19. Revista Cubana de Información en Ciencias de la Salud. octubre-diciembre. 2021;32(4). [Consultado 1 diciembre 2021] Disponible en: <http://www.acimed.sld.cu/index.php/acimed/article/view/1686>.
 21. Stuart ML, Prieto D, Quial J, Delgado T, Delgado M. Mejora a la gestión de información en el proceso de Servicio Postal Universal. Revista Cubana de Transformación Digital. abril-junio 2021;2 (2):24-40. [Consultado 1 octubre 2021] Disponible en: <https://rctd.uic.cu/rctd/article/view/118>.
 22. Stuart ML, Cuyar L, Prieto DR, Delgado T. Procedimiento de Diagnóstico de Gestión de Información con Enfoque de Arquitectura de Información Empresarial Congreso Internacional de Información. INFO 2018. V Foro sobre Gestión de Información y el Conocimiento, La Habana, Cuba; 2018 http://www.congreso-info.cu/public/conferences/1/INFO2018-Foro%20FIKM_ES.pdf
 23. Calvanese D, Cogrel B, Komla-Ebri S, Kontchakov R, Lanti D, Rezk M, Rodriguez-Muro M, Xiao G. Ontop: Answering SPARQL queries over relational databases. Semantic Web J. 2017;8(3):471-87. [Consultado 12 octubre 2021] Disponible en: <https://doi.org/10.3233/SW-160217>.
 24. Xiao G, Ding L, Cogrel B, Calvanese D. Virtual knowledge graphs: An overview of systems and use cases. Data Intelligence. 2019;1(3):201-223. [Consultado 12 octubre 2021] Disponible en: https://doi.org/10.1162/dint_a_00011
 25. Stuart ML, Delgado T, Delgado M, Piedra Y. Datos empresariales enlazados: Revisión sistemática desde una perspectiva organizacional. Alcance, 2020;9(23):153-176. [Consultado 1 julio 2021] Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2411-99702020000200153.
-

Conflicto de intereses

Los autores declaran no tener conflicto de intereses.

Contribución de los autores

- Mavis Lis Stuart Cárdenas: Conceptualización, metodología, escritura, revisión, edición, validación y visualización.
- Tatiana Delgado Fernández: Conceptualización, metodología, escritura, revisión, edición, validación y visualización.
- Mercedes Delgado Fernández: Conceptualización, análisis formal, escritura, revisión, edición, validación.
- Manuel de la Iglesia Campos: Software, validación, visualización.