



## CIENCIAS TÉCNICAS

### Premio Anual de la Academia de Ciencias de Cuba, 2020

# Enfoques semánticos basados en ontologías para la recuperación de información en Sistemas de Información Geográfica

Alfredo Javier Simón Cuevas <sup>1\*</sup> <https://orcid.org/0000-0002-6776-9434>  
María Matilde García Lorenzo <sup>2</sup> <https://orcid.org/0000-0002-1663-5794>  
Manuel Enrique Puebla Martínez <sup>3</sup> <https://orcid.org/0000-0001-6008-0370>  
Nayi Sánchez Fleitas <sup>2</sup> <https://orcid.org/0000-0001-5305-537X>  
José Manuel Perea Ortega <sup>4</sup> <https://orcid.org/0000-0002-7929-3963>  
Raúl Comas Rodríguez <sup>5</sup> <https://orcid.org/0000-0003-1353-2279>

<sup>1</sup> Universidad Tecnológica de La Habana José Antonio Echeverría. La Habana, Cuba

<sup>2</sup> Universidad Central Marta Abreu de Las Villas. Santa Clara, Cuba

<sup>3</sup> Universidad de las Ciencias Informáticas. La Habana, Cuba

<sup>4</sup> Universidad de Extremadura. Extremadura, España

<sup>5</sup> Universidad Regional Autónoma de los Andes. Ambato, Tungurahua, Ecuador

\*Autor para la Correspondencia: [asimon@ceis.cujae.edu.cu](mailto:asimon@ceis.cujae.edu.cu), [asimon77@gmail.com](mailto:asimon77@gmail.com)

## RESUMEN

### Palabras clave

*ontologías geográficas; recuperación información geográfica; sistemas de información geográfica; razonamiento basado en casos*

**Introducción.** Los Sistemas de Información Geográfica (SIG) han adquirido gran relevancia, al aumentar el valor de la información geográfica para tomar decisiones. Mejorar la eficacia del procesamiento y la recuperación de información en SIG, teniendo en cuenta los grandes volúmenes de datos geográficos disponibles, su heterogeneidad, y la ausencia de significados semánticos formalizados, constituye un desafío y objetivo general de esta investigación.

**Métodos.** Los enfoques basados en ontologías (ej. ontologías geográficas) constituye una alternativa muy prometedora para dotar de semántica a los datos y su procesamiento. Los paradigmas de acceso e integración de datos basados en ontologías permiten la administración de grandes ontologías y la interoperabilidad de datos heterogéneos, problemáticas también tratadas en la investigación. En el trabajo se presenta un nuevo método semiautomático para la generación de ontologías geográficas a partir de fuentes de datos espaciales heterogéneas y un modelo de gestión de datos basado en ontologías y Razonamiento Basado en Casos (RBC). **Resultados.** El método propuesto permite dotar a los SIG de recursos de conocimiento más robustos, incrementando sus capacidades de recuperar información y su eficacia. Además, el paradigma de ontologías basadas en bases de datos aplicado permite un mejor manejo del volumen de conocimiento generado. Este paradigma, aplicado también en el modelo de gestión de datos basado en ontologías propuesto, permitió resolver los problemas de interoperabilidad existentes en sistemas de la Unión Eléctrica entre datos geográficos y de dominio eléctrico, al mismo tiempo que se extendieron las capacidades de formulación de consultas y la recuperación de información con el RBC.



# Ontology based semantic approaches for information retrieval in Geographic Information Systems

## ABSTRACT

### Keywords

*geographic ontologies; geographic information retrieval; geographic information systems; case-based reasoning*

**Introduction.** Geographic Information Systems (GIS) have become very relevant by increasing the value of geographic information for decision making. Improving the efficacy of information processing and retrieval in GIS -taking into account the large geographic data volumes available, their heterogeneity, and the absence of formalized semantic meanings- is a challenge and constitutes the general objective of this research. **Methods.** Ontology-based approaches (e.g. geographic ontologies) are a very promising alternative to provide semantics to data and its processing. Ontology-based data access and integration paradigms enable the management of large ontologies and the interoperability of heterogeneous data, issues also addressed in the research. The paper presents a new semi-automatic method for the generation of geographic ontologies from heterogeneous spatial data sources and a data management model based on ontologies and Case-Based Reasoning (CBR). **Results.** The proposed method provides more robust knowledge resources to the GIS, increasing their information retrieval capabilities and efficacy. In addition, the paradigm of database-supported ontologies applied allows for better management of the knowledge volume generated. This paradigm, also applied to the proposed ontology-based data management model, permitted solving the interoperability problems existing in Electric Union systems between geographic and electric domain data, while extending query formulation capabilities and information retrieval with the RBC.

## INTRODUCCIÓN

Actualmente, resulta extraordinario el acelerado crecimiento de la información digital disponible, siendo la mayor parte información georreferenciada. <sup>(1)</sup> En este escenario, la información geográfica adquiere mayor relevancia para la toma de decisiones en ámbitos gubernamentales, empresariales, turísticos, entre otros, generando un incremento del interés por los Sistemas de Información Geográfica (SIG). Mejorar la eficacia del procesamiento y la recuperación de información geográfica en estos sistemas constituye un tema de investigación actual y relevante, siendo las fuentes de información geográfica que se emplean un factor clave y el tratamiento de la semántica de la información geográfica uno de los principales desafíos. <sup>(2)</sup>

Los enfoques semánticos basados en ontologías se manifiestan como tendencia en contextos donde se requiere que el conocimiento sea semánticamente formalizado para su procesamiento computacional. Por tanto, adoptar este tipo de enfoques mediante ontologías geográficas resulta una alternativa muy prometedora para mejorar la eficacia de la recuperación de información en SIG. <sup>(3)</sup> Estos enfoques ofrecen los métodos, lenguajes y herramientas para dotar de significados semánticos a los datos, información y el conocimiento. Aunque hay avances científicos en este sentido, aún hay problemáticas abiertas a nuevas contribuciones, entre ellas:

generación automática de ontologías geográficas a partir de fuentes heterogéneas, incrementando el aprovechamiento de la gran diversidad de fuentes de información geográficas disponibles, por ejemplo, fuentes de información geográficas voluntarias, <sup>(4, 5)</sup> Bases de Datos Espaciales (BDE), entre otras; administración grandes ontologías geográficas y la inferencia sobre ellas; interoperabilidad (acceso e integración) de datos heterogéneos en SIG; la flexibilización de los mecanismos de consulta para la recuperación de información sobre fuentes de datos complejas, entre otras.

En este trabajo se presentan novedosas soluciones a problemáticas expuestas en el ámbito del procesamiento semántico y la recuperación de información geográfica en SIG. Específicamente, se presenta un método semiautomático para la generación de ontologías geográficas sobre zonas geográficas de interés con alcance global (todo el planeta), a partir de fuentes de datos espaciales heterogéneas y con alto nivel de expresividad semántica, <sup>(6, 7, 8, 9, 10, 11)</sup> y un modelo de gestión de datos basado en ontologías y el Razonamiento Basado en Casos (RBC), <sup>(12, 13, 14, 15, 16)</sup> para el análisis y recuperación de información en un SIG.

El método de generación de ontologías permite integrar, en único recurso semántico, datos e información geográfica proveniente de fuentes geográficas heterogéneas y aprovechar la semántica capturada para generar conocimiento nuevo, y está articulado con una solución para procesamiento e

inferencia sobre grandes ontologías; como las resultantes del método. Por otro lado, el modelo de gestión de datos basado en ontologías ofrece solución a las problemáticas de interoperabilidad (acceso e integración) de datos heterogéneos y el incremento de la flexibilidad en la formulación de consultas para la recuperación información geográfica en SIG.

Las propuestas presentadas fueron validadas mediante métodos de evaluación funcional basados en tareas, análisis estadístico, criterio de expertos, la aplicación de las propuestas en escenarios reales y estratégicos para nuestro país (electro-energético), entre otros. El objetivo general de la investigación fue mejorar la eficacia del procesamiento y la recuperación de información, teniendo en cuenta los grandes volúmenes de datos geográficos disponibles, su heterogeneidad, y la ausencia de significados semánticos formalizados.

## MÉTODOS

A continuación, se describen los aspectos generales de las contribuciones incluidas en la propuesta.

### Nuevo método semiautomático para la generación de ontologías geográficas a partir de fuentes de datos espaciales heterogéneas

El método tiene el propósito de generar ontologías geográficas sobre una zona de interés (región, municipio, ciudad o país) en el que se represente e integre información geográfica capturada de fuentes de información geográfica heterogéneas. Este método forma parte del proyecto de desarrollo de un Sistema de Recuperación de Información Geográfica basado en ontología, en el que se integran las técnicas de recuperación de información geográfica basada en consulta e interactiva (o browsing).<sup>(6, 7)</sup> El método se concibió en cua-

tro fases (figura 1) e incluye un mecanismo de generación de relaciones espaciales entre objetos geográficos (como conocimiento nuevo).<sup>(8)</sup> La integración de conceptos y objetos provenientes de fuentes diferentes está soportada en análisis de la similitud semántica entre ellos, y en las relaciones espaciales que se generan automáticamente.

El método toma como base una ontología preliminar (resultante también de esta investigación) en la que se representan conceptos, propiedades de datos (características geográficas) y propiedades de objetos (relaciones espaciales) comunes dentro del dominio geográfico. Mediante la ejecución de cada fase se va enriqueciendo automáticamente esta ontología, a partir de la información capturada de las fuentes de información geográficas utilizadas y que está relacionada con la zona de interés.

El proceso de enriquecimiento de la ontología preliminar inicia con *Geonames*, de donde se extraen automáticamente los conceptos y descripciones con el soporte de técnicas de procesamiento de lenguaje natural. En esta 1<sup>ra</sup> Fase, *Geonames* se utiliza principalmente para ampliar la taxonomía de conceptos de la ontología inicial. En la 2<sup>da</sup> Fase se extiende la ontología utilizando una BDE<sup>(10)</sup> como fuente de información geográfica. Se extrae de la BDE o cartografía vectorial toda la información relacionada con la zona de interés, la cual luego es integrada al conocimiento ya representado en la ontología geográfica resultante de la fase anterior.<sup>(10)</sup> La BDE a utilizar debe tener una tabla o relación por cada capa de información a modelar, cada tabla debe tener al menos un campo que permita ubicar espacialmente a las tuplas y debe estar al menos en primera forma normal.<sup>(10)</sup> En la 3<sup>ra</sup> Fase se incluye mecanismo para enriquecer la ontología desde cualquier otro tipo de fuente de datos espaciales que permita la identificación

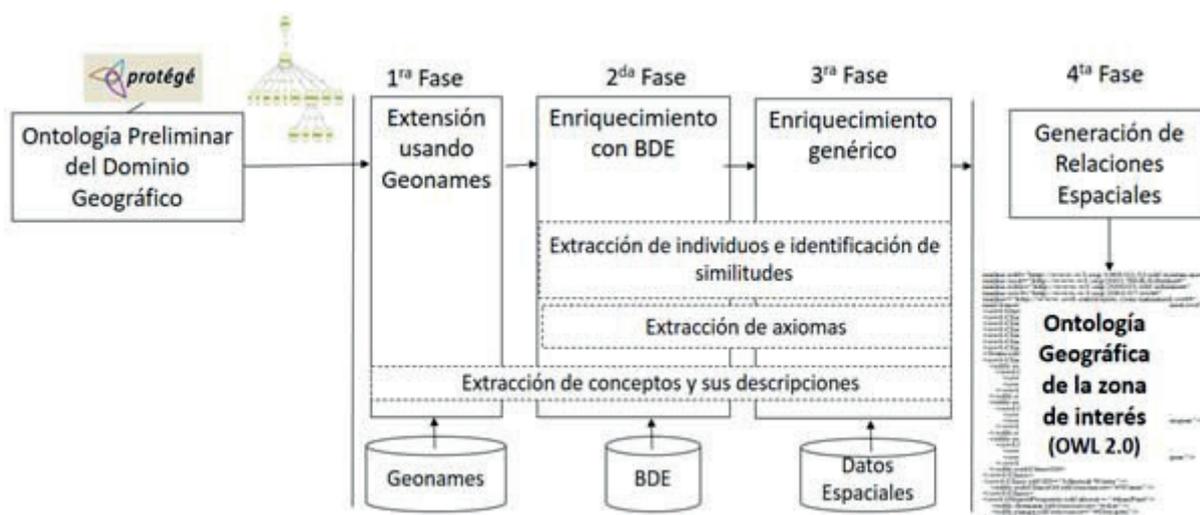


Fig. 1. Método de generación de ontologías geográficas a partir de fuentes heterogéneas

de conceptos y objetos geográficos. Esta última fase tiene el objetivo de aportar mayor generalidad al método, dada la diversidad de fuentes de datos espaciales disponibles.

En cada fase se llevan a cabo procesos de: extracción de información, integración y formalización de significados semánticos. Estos procesos están soportados en un mecanismo de análisis de la similitud semántica entre objetos espaciales o topónimos, el cual se basa en tres elementos fundamentales: a) distancia espacial; b) concepto geográfico o tipo al que pertenecen; y c) similitud entre los nombres.

Una vez concluido el procesamiento de las fuentes de información geográfica, se lleva a cabo una última fase, en la cual se aprovecha la semántica descrita en la ontología resultante para generar automáticamente nuevas relaciones espaciales entre los objetos geográficos representados; proceso no identificado en otras propuestas reportadas. Estas relaciones no se encuentran de forma explícita en las fuentes originales y propicia la generación de conocimiento nuevo. Las nuevas relaciones generadas, además de aportar conocimiento nuevo y enriquecer la semántica de los objetos espaciales, permite dotar de mayor capacidad de respuesta a consultas complejas por parte de los SIG.

La ontología resultante se codifica en OWL (Ontology Web Language) 2.0, lográndose un nivel de descripción semántica superior a otros recursos disponibles, y con ello disponer de más información semántica para mejorar la eficacia en la recuperación de información geográfica. La integración de la información geográfica proveniente de diversas fuentes (algunas con grandes volúmenes de datos) en una sola ontología, y la generación de nuevas relaciones espaciales, produce un crecimiento significativo del volumen conocimiento a ser representado y procesado. Sin embargo, las principales herramientas disponibles para la construcción y administración de ontologías muestran limitaciones para el manejo e inferencia sobre grandes ontologías, y más aún en OWL 2.0. Esta problemática es resuelta en esta investigación a través de una arquitectura de representación de la ontología basada en el paradigma de ontologías basadas en bases de datos<sup>(17)</sup> y la definición de operaciones de inferencia que dan soporte a las necesidades fundamentales de recuperación de información dentro del dominio geográfico, todo ello soportado en el *framework* OWLAPI.<sup>(11)</sup> La arquitectura de representación consiste en compartir la conceptualización de la ontología y los datos entre el fichero OWL (donde se almacena la ontología) y una base de datos relacional. Específicamente, los datos pertenecientes al ABox son extraídos de la ontología y almacenados en una base de datos relacional, manteniendo íntegramente el TBox (conceptualización) en el fichero OWL, el cual sería gestionado por el *framework* OWLAPI.

## Modelo de gestión de datos basado en ontologías y razonamiento basado en casos

El Sistema Integral de Gestión Empresarial (SIGE) de la Unión Eléctrica incluye un SIG denominado SIGOBE. Estudios realizados sobre este entorno arrojaron que el sistema SIGOBE mostraba limitaciones que dificultaban la toma de decisiones en el ámbito de los procesos de distribución y transmisión, entre ellas: a) restricción de la recuperación de información a partir de consultas predefinidas con muy poca flexibilidad; b) falta de integración entre los modelos de datos de los sistemas y la información geográfica en la cartografía; c) ausencia de correspondencia semántica entre los elementos de ambos dominios (eléctrico y geográfico).

La solución diseñada para resolver estas limitaciones consistió en el desarrollo de un modelo de gestión de datos basado en ontologías<sup>(12, 13, 14, 15, 16)</sup>, donde también se aplica el paradigma de gestión de datos basada en ontología<sup>(17)</sup> utilizado en la solución descrita en la sección anterior, pero enfocado a la solución del acceso e integración de datos heterogéneos. Integrado a este modelo se concibió un proceso inteligente de recuperación de información a partir de consultas, cuyos resultados se construyen sobre la base del procesamiento semántico de los datos, soportado en una ontología y RBC. El modelo se concibió para dar respuesta a cualquier pregunta que hiciera el usuario sobre el sistema, mediante la interpretación de la consulta y la localización en un mapa de los resultados. Una vista general del modelo se muestra en la figura 2.

Como parte del modelo se construyó la ontología de dominio OntoSIGOBE que describe y formaliza la semántica de los conceptos y datos en el ámbito de los procesos de distribución y transmisión, y su relación con la geografía sobre la que se despliegan los elementos eléctricos. La integración de conocimiento eléctrico y geográfico en la ontología, permite realizar consultas donde se integren conceptos de ambos dominios, logrando mayor integralidad en el análisis de los datos. A través de *OntoSIGOBE* se logró resolver los problemas de interoperabilidad semántica entre los datos del dominio eléctrico y geográfico, así como su entendimiento a la hora de construir nuevas consultas, ofreciendo mayor flexibilidad a los usuarios en su formulación. Esta flexibilidad se logró aplicando técnicas de procesamiento de lenguaje natural sobre el texto de la consulta y *OntoSIGOBE* como base de conocimiento.

En *OntoSIGOBE* los datos primarios (instancias o Abox) se mantienen en las bases de datos relacionales de los diferentes sistemas contemplados en SIGE, mientras que la conceptualización se almacena en el fichero OWL, de forma similar a la arquitectura empleada en la solución anterior (como

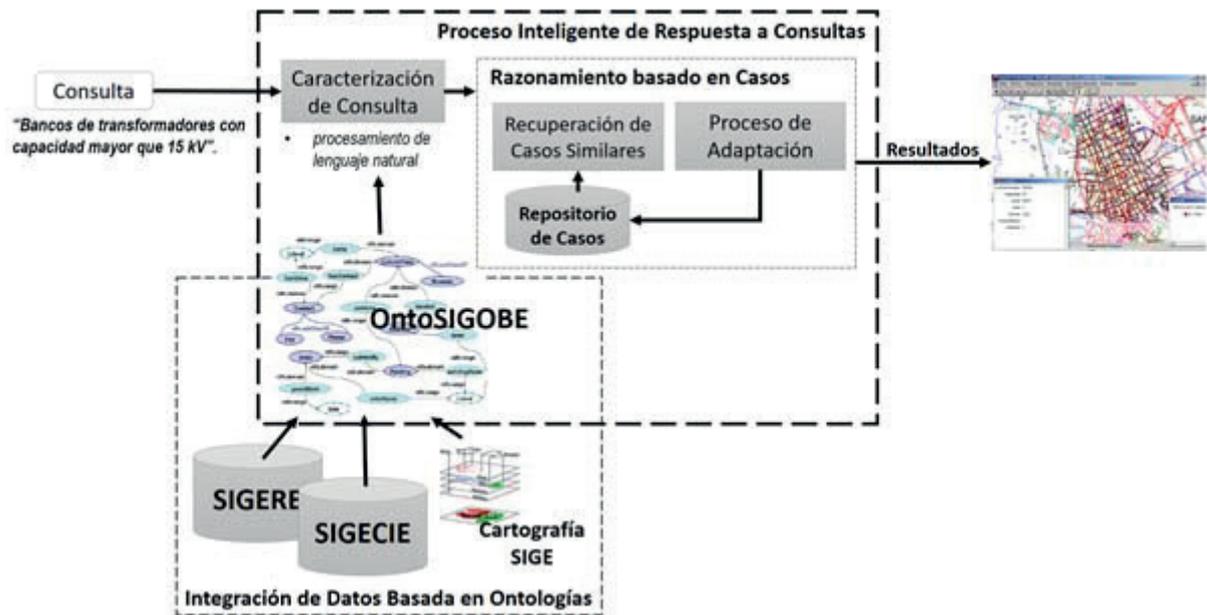


Fig. 2. Modelo de gestión de datos basado en ontología de SIGOBE. <sup>(12)</sup>

se muestra en la figura 3). El acceso a los datos primarios se realiza mediante sentencias codificadas en el lenguaje SQL. La aplicación de la técnica de RBC permite establecer un puente entre el resultado de la interpretación de la consulta del usuario en lenguaje natural y la codificación SQL necesaria para acceder y recuperar los datos. <sup>(13)</sup> La integración de estas técnicas permite diseñar consultas complejas (interengan conceptos del dominio eléctrico y geográfico) de manera flexible y lograr un análisis más integral desde el proceso de recuperación de información geográfica para la toma de decisiones.

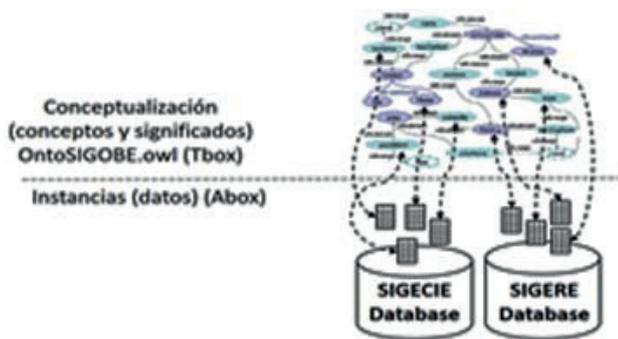


Fig. 3. Arquitectura de la representación de conocimiento de OntoSIGOBE <sup>(12)</sup>

Cada consulta de un usuario es tratada como un nuevo caso (necesidad de información) a procesar y mediante análisis de similitud se recuperan automáticamente otros casos similares (necesidades de información), en los cuales se en-

cuentra codificado el mecanismo de acceso a los datos primarios <sup>(13,14,15,16)</sup>. Los casos similares recuperados pueden ser adaptados si no satisfacen todos los elementos incluidos en la petición del usuario, la cual es descrita a través de los rasgos predictores. Los casos adaptados son almacenados en la base de casos para enriquecer la base de conocimiento. Los casos se representan mediante rasgos predictores (capturan elementos del dominio y las necesidades de información) y objetivos (garantizan acceso a la base de datos). <sup>(13)</sup>

## RESULTADOS Y DISCUSION

Las propuestas presentadas se validaron utilizando métodos reconocidos en el ámbito de la recuperación de información y la ingeniería de ontologías, a saber: métodos de evaluación funcional midiendo resultados con métricas y comparando resultados con otras soluciones, análisis estadístico, criterio de expertos, y la aplicación en escenarios reales, como el electro-energético.

El método de generación de ontologías geográficas se aplicó para la obtención de una ontología de Marianao (municipio de La Habana), obteniendo como resultado OntoMarianao, cuyo volumen de conocimiento se caracteriza por: 681 clases o conceptos geográficos; 836 relaciones taxonómicas; 3,777 instancias u objetos geográficos (por ejemplo, "Plaza de Marianao"); 29,236,214 relaciones entre objetos geográficos, fundamentalmente relaciones espaciales (por ejemplo, "cerca\_de", "al\_norte\_de"); 35,662 relaciones de propiedad-valor de las clases; 6,328 pares de objetos espaciales similares.

La tabla 1 muestra el proceso de construcción (fase por fase) de OntoMarianao, donde se aprecia el crecimiento del conocimiento representado al transitar por cada fase, destacándose la cuarta fase con la generación automática de las relaciones espaciales. Aunque se reportan algunas soluciones dirigidas a obtener recursos ontológicos con información geográfica, en ningún caso se logra integrar la diversidad de fuentes de información geográficas como lo permite el método propuesto para este propósito.

Las ontologías que se generen mediante este método, constituyen nuevos recursos de conocimientos a ser utilizados por los SIG, como fuente de información primaria (fundamental en recuperación de información geográfica interactiva) o complementaria. La aplicabilidad de OntoMarianao se ejemplifica en la tabla 2, donde se recogen un conjunto de consultas (y sus resultados recuperados) sobre elementos geográficos del territorio, las cuales pueden ser de gran utilidad para la toma de decisiones en sectores del transporte, económicos, sociales, entre otros. En la tabla 2 se evidencia una mayor eficacia de la recuperación de información geográfica sobre la ontología obtenida con el método propuesto, respecto a otros sistemas muy reconocidos con los que se compararon los resultados. También es importante resaltar el incremento de la capacidad de respuesta a consultas complejas (requiere búsqueda sobre varios conceptos geográficos) que se logra cuando se integra en la ontología geográfica mayor variedad de fuentes heterogéneas, y se incorporan me-

canismos de generación de relaciones espaciales entre los objetos geográficos representados.

Aunque la aplicación del método se enfocó en el territorio de Marianao, es aplicable para obtener ontologías geográficas de cualquier área geográfica del país, mientras se tenga información en las fuentes de información geográficas utilizadas. OntoMarianao o cualquier otra ontología que se obtenga puede ser utilizada directamente como fuente de búsqueda principal de un SIG (por ejemplo, en servicio de guía virtual a turistas dentro la zona de la Habana Vieja), o como fuente de información complementaria (por ejemplo, para enriquecer la información geográfica en SIGOBE).

La ontología OntoSIGOBE constituye uno de los principales resultados del modelo de gestión de datos basado en ontologías concebido para SIGOBE. Aunque se ha identificado en la literatura el desarrollo de otras ontologías del dominio eléctrico, estas no abordan en profundidad el dominio de conocimiento de los procesos de transmisión y distribución, y no integran información geográfica sobre los datos (fundamental para lograr más integralidad en la toma de decisiones). En OntoSIGOBE se representa e integran conocimientos asociados a: los conceptos eléctricos definidos como las nomenclaturas en las bases de datos (estructuras administrativas y políticas, las capacidades de carga, tipos de voltajes, fabricantes); elementos eléctricos geográficamente representados en la cartografía con información espacial y alfanumérica (circuitos, banco de transformadores, lámparas, desconec-

**Tabla 1.** Resultados por fases de la generación de la ontología geográfica OntoMarianao

Componentes Ontológicos	Ontología Preliminar	1ra Fase	2da Fase	3ra Fase		4ta Fase
		Geonames	BDE*	OSM**	Wikipedia***	Generación de Relaciones Espaciales
Conceptos	15	635	679	664	681	681
Conceptos Equivalentes	0	14	24	17	24	24
Relaciones Taxonómicas	14	650	832	707	836	836
Individuos	0	48	3,773	2,368	3,777	3,777
Propiedades de Objeto	0	270	278	270	278	29,236,214
Propiedades de Dato	0	1,344	35,654	23,248	35,662	35,662
Pares de individuos similares	0	2	6,321	6,134	6,328	6,328
Anotaciones	80	1,384	5,271	3,758	5,279	5,279

\* BDE construida a partir de información sobre topónimos de Cuba, suministrados por GEOCUBA. \*\* OpenStreetMap: mapa del mundo editable de forma voluntaria (<https://taginfo.openstreetmap.org>). \*\*\* WikiProject Geographical Coordinates ([https://en.wikipedia.org/wiki/Wikipedia:WikiProject\\_Geographical\\_coordinates](https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Geographical_coordinates))

**Tabla 2.** Ejemplificación de la recuperación de información sobre OntoMarianao

Requisito Informativo (consulta)	Sistemas de Información Geográfica						OntoMarianao
	A	B	C	D	E	F	
“Recursos hidrográficos en Marianao”	6 resultados (ej Arroyo Marinero; Arroyo Paila; Presa Teresita; Presa Naranjito)	SR	NP	1 resultado: Marianao (no relevante)	6 resultados, ninguno relevante. (ej. Café Enlace; Casa Doña Eneida; Cine Encuentro)	SR	47 Resultados (ej.: Arroyo Marinero; Presa-Teresita; Presa-Naranjito; Río-Quibú; Río-Orengo)
“Escuelas cerca (1km) de hospitales y de puntos de transporte público”	NP	NP	NP	17 resultados	Varios resultados, pero ninguno relevante	SR	10 resultados (ej.: Instituto Tecnológico-Militar José Martí (ITM); Escuela Secundaria Aristides Viera Gonzáles; Ciudad Escolar Libertad; ESBU Mártires del 13 de Marzo)
“Rutas que pasan cerca (1km) de 100 y 51 y cerca de un hospital”	NP	NP	NP	1 resultado, no relevante (Universidad Tecnológica de La Habana)	Varios resultados, pero ninguno relevante	SR	2 resultados (ej.: Metrobus P14 Parque Fraternidad => Cerro =>San Agustín; Metrobus P14 Parque Fraternidad => Cerro => San Agustín)
“Rutas que se intersectan con la Avenida 51”	NP	NP	NP	1 resultado, no relevante (Avenida 51)	Varios resultados, pero ninguno relevante	SR	95 resultados (ej.: Habana-Güira-Artemisa; Habana-Guanajay-Artemisa; 118; 120; Calle 100)
“Construcciones de interés cerca (1km) del Hospital La Ceguera”	1 resultado ( Hotel “San Alejandro”)	NP	NP	6 resultados, ninguno está a menos de 1 km.	Varios resultados, pero ninguno relevante	SR	222 resultados (ej. San-Alejandro; Conservatorio; Calle-41; Funeraria; Avenida-31)

**Leyenda:** A : Geonames Semantic Web (<https://www.geonames.org/>); B: SearchOSM (<https://searchosm.com> ); C: GeoData (<https://www.mediawiki.org/wiki/Extension:GeoData#API> ); D: Google Maps; E: FOURSQUARE (<https://es.foursquare.com> ); F: Wikimapia (<https://www.geonames.org/export/wikipedia-webservice.html#wikipediaSearch>); SR (Sin resultado): indica que el sistema no devolvió ningún resultado y no se logra satisfacer la petición del usuario; NP (No posible): indica que en el sistema correspondiente no se ofrecen las funcionalidades necesarias para poder realizar ese tipo de consulta.

tivos); y conceptos geográficos (áreas geográficas, regiones, accidentes geográficos naturales o contruidos, entre otros). El volumen de conocimiento en OntoSIGOBE se caracteriza en: 53 clases o conceptos, 63 subclases; 6 clases equivalentes; 10 propiedades de objetos; 155 axiomas; y 929 relaciones semánticas. <sup>(12)</sup>

Otro importante resultado es la construcción de una base de casos formado por 201 casos, con los cuales se extiende significativamente la diversidad de tipos de consultas que se

pueden realizar en SIGOBE. El funcionamiento del modelo y la respuesta de OntoSIGOBE fue evaluado en el ámbito del trabajo de operaciones diarias de 3 departamentos de la Unión Eléctrica: Oficina, Departamento de Ingeniería, y Servicio al Cliente, y se midió la precisión de las respuestas a las consultas realizadas por los especialistas de esos departamentos. En la tabla 3 se muestran los resultados obtenidos, los cuales reflejan que el modelo propuesto no solo extiende las capacidades para dar respuesta a una mayor diversidad de consul-

tas, sino que ofrece una calidad superior al 90 % de fiabilidad de las respuestas.

Otro aspecto muy importante es que el modelo logra enriquecer de forma significativa la base de datos, como otra alternativa para incrementar la eficacia en la recuperación de información.

### Conclusiones

En este trabajo se han presentado novedosas soluciones basadas en ontologías enfocadas a mejorar la eficacia en la recuperación de información geográfica. El método de generación de ontologías permite dotar a los Sistemas de Información Geográfica de recursos de conocimiento más robustos, sobre la base de la integración eficaz del conocimiento proveniente de una mayor diversidad de fuentes de información geográfica y el incremento de la descripción semántica mediante lenguaje OWL 2.0. El mecanismo de generación de relaciones espaciales incluido en este método propicia la extensión del conocimiento representado en las fuentes primarias, incrementando las capacidades para recuperar información sobre consultas complejas. Las pruebas realizadas con la ontología generada con el método, arrojaron que se obtenían mejores resultados en la recuperación de información que los obtenidos por otros sistemas reconocidos. El paradigma de ontologías basadas en bases de datos permitió diseñar una solución soportada en el *framework* OWLAPI para el manejo de grandes ontologías, y usando un enfoque similar también se logró resolver los problemas de interoperabilidad existentes en entre los datos geográficos y del dominio eléctrico dentro del SIGOBE. La combinación de la ontología OntoSIGOBE y el RBC en el modelo de gestión de datos propuesto permitió extender las capacidades de formulación de consultas y recuperación del SIGOBE, alcanzando una fiabilidad en las respuestas superior al 90 %. En general, se ha comprobado que ambas contribuciones proporcionan mejoras en la eficacia y cobertura de la recuperación de información geográfica en los SIG, así como en la toma de decisiones soportada por este proceso.

### Agradecimientos

Los autores agradecen a la Universidad Tecnológica de La Habana José Antonio Echeverría, CUJAE, a la Universidad Central Marta Abreu de Las Villas (UCLV), a la Universidad de las Ciencias Informáticas (UCI), y a la Universidad Regional Autónoma de los Andes (UNIANDÉS) por su apoyo en el desarrollo de la investigación.

### REFERENCIAS BIBLIOGRÁFICAS

1. Vilches-Blázquez LM., Saavedra JA framework for connecting two interoperability universes: OGC Web Feature Services and Linked Data. *Transactions in GIS*. 2019, 23(1):22-47. DOI: <https://doi.org/0.1111/tgis.12496>
2. Purves RS., Clough P., Jones CB., Hall MH., Murdock V. Geographic Information Retrieval: Progress and Challenges in Spatial Search of Text. *Foundations and Trends® in Information Retrieval*. 2018; 12(2-3): 164-318. DOI: <https://doi.org/10.1561/15000000034>
3. Hu Y. Geospatial Semantics. In: Huang B., editor. *Comprehensive Geographic Information Systems*. Oxford: Elsevier; 2018, pp. 80–94. DOI: <https://doi.org/10.1016/B978-0-12-409548-9.09597-X>
4. Chen X., Vo H., Wang Y., Wang F. A framework for annotating OpenStreetMap objects using geo-tagged tweets. *Geoinformatica*. 2018; 22(3): 589–613. DOI: <https://doi.org/10.1007/s10707-018-0323-8>
5. See L., Fonte C. C., Antoniou V., Minghini M. Volunteered geographic information: looking towards the next 10 years. *Journal of Geographical Systems*. 2019; 21(1): 1-3. DOI: <https://doi.org/10.1007/s10109-018-00291-x>
6. Puebla Martínez ME., Perea Ortega JM., Simón Cuevas A. Acercamiento a un modelo de recuperación de información geográfica basado en técnicas de exploración y descripción semántica de datos geoespaciales. *CEUR-WS*. 2014, 1219: 72-86.
7. Puebla Martínez ME., Perea Ortega JM., Alfredo Simón Cuevas. Integración de técnicas browsing con ontologías en un modelo de recuperación de información geográfica. *Ciencias de la Información*. 2015; 46(1): 53-61.
8. Puebla Martínez M. E., Perea Ortega J. M., Simón-Cuevas A. Generación Semiautomática de una Ontología Geográfica a partir de Fuentes Heterogéneas. *CEUR-WS*. 2016;1797:67-79.
9. Puebla Martínez ME., Perea Ortega JM., Simón Cuevas A., Romero FP. Automatic Expansion of Spatial Ontologies for Geographic Information Retrieval. *Communications in Computer*

**Tabla 3.** Resultados de la evaluación del modelo de gestión de datos basado en OntoSIGOBE <sup>(12)</sup>

Áreas	Consultas	Respuestas correctas	Respuestas incorrectas	Nuevos casos aprendidos	Precisión (%)
Oficina	230	221	9	23	96,08
Departamento de Ingeniería	189	178	11	40	94,18
Servicio al Cliente	147	140	7	18	95,23
Total	566	539	27	81	
Precisión promedio					95,16

and Information Science. 2018; 855: 659-670. DOI: [https://doi.org/10.1007/978-3-319-91476-3\\_54](https://doi.org/10.1007/978-3-319-91476-3_54)

10. Puebla Martínez M. E., Perea Ortega J. M., Simón-Cuevas A., Romero F. P., Olivás Varela J. A. Automated Large Geographic Ontologies Generation Method from Spatial Databases. *Communications in Computer and Information Science*. 2019;1029:118-33. DOI: [https://doi.org/10.1007/978-3-030-21395-4\\_9](https://doi.org/10.1007/978-3-030-21395-4_9)
11. Puebla Martínez M. E., Perea Ortega J. M., Simón-Cuevas A. Extensión del Framework OWLAPI para la Administración y Razonamiento sobre Grandes Ontologías. *CEUR-WS*. 2018;2096:124-34.
12. Comas Rodríguez R., Simón Cuevas A., Sánchez Fleitas N., García Lorenzo M. M. An Ontology-Based Data Management Model Applied to a Real Information System. *Advances in Intelligent Systems and Computing*. 2019; 1078: 36-50. DOI: [https://doi.org/10.1007/978-3-030-33614-1\\_3](https://doi.org/10.1007/978-3-030-33614-1_3)
13. Sánchez Fleitas N., Comas Rodríguez R., García Lorenzo MM., Riverol Quesada A. Intelligent Case-Based System for Decision-Making in the Electricity Sector. *Computación y Sistemas*. 2019;23(4): 1165-1180. DOI: <https://doi.org/10.13053/CyS-23-4-2720>
14. Sánchez Fleitas N., Comas Rodríguez R., García Lorenzo M. M. Sistema Inteligente de Información Geográfica para las empresas eléctricas cubanas. *Ingeniare. Revista chilena de ingeniería*. 2019; 27(2): 197-209. DOI: <https://doi.org/10.4067/S0718-33052019000200197>
15. Sánchez Fleitas N., Comas Rodríguez R., García Lorenzo M. M., Rivero Quesada A. Modelo de manejo de datos, con el uso de inteligencia artificial, para un sistema de información geográfica en el sector energético. *Enfoque UTE*. 2016; 7(3): 95-109.
16. Sánchez Fleitas N., Comas Rodríguez R., García Lorenzo M. M., Frankz Carrera. Geographical Information System Based on Artificial Intelligence Techniques. *Communications in Computer and Information Science*. 2018; 895: 446-461. DOI: [https://doi.org/10.1007/978-3-030-05532-5\\_33](https://doi.org/10.1007/978-3-030-05532-5_33)
17. Abiteboul S., Arenas M., Barceló P., Bienvenu M., Calvanese D., David C., Hull R., Hüllermeier E., Kimelfeld B., Libkin L., Martens W., Milo T., Murlak F., Neven F., Ortiz M., Schwentick T., Stoyanovich J., Su J., Suciu D., Vianu V., Yi K. Research Directions for Principles of Data Management. *ACM SIGMOD Record*. 2017;45(4):5-17. DOI: <https://doi.org/10.1145/3092931.3092933>

---

Recibido: 30/04/21  
Aprobado: 16/07/21

---

### Conflicto de intereses

Los autores declaran que no existen conflictos de intereses relacionados con el contenido del artículo, con las instituciones a las que pertenecen, ni con ninguna otra entidad o institución.

### Contribuciones de los autores

- Conceptualización: Alfredo Javier Simón Cuevas, María Matilde García Lorenzo, Manuel Enrique Puebla Martínez, Nayi Sánchez Fleitas, José Manuel Perea Ortega, Raúl Comas Rodríguez
- Curación de datos: María Matilde García Lorenzo, Manuel Enrique Puebla Martínez, Nayi Sánchez Fleitas
- Análisis formal: Alfredo Javier Simón Cuevas, María Matilde García Lorenzo, Nayi Sánchez Fleitas, José Manuel Perea Ortega, Raúl Comas Rodríguez
- Adquisición de fondos: Alfredo Javier Simón Cuevas
- Investigación: Alfredo Javier Simón Cuevas, María Matilde García Lorenzo, Manuel Enrique Puebla Martínez, Nayi Sánchez Fleitas, José Manuel Perea Ortega, Raúl Comas Rodríguez
- Metodologías: Alfredo Javier Simón Cuevas, María Matilde García Lorenzo, Manuel Enrique Puebla Martínez, Nayi Sánchez Fleitas, José Manuel Perea Ortega, Raúl Comas Rodríguez
- Administración de proyecto: Alfredo Javier Simón Cuevas, María Matilde García Lorenzo
- Recursos: Alfredo Javier Simón Cuevas, María Matilde García Lorenzo, Manuel Enrique Puebla Martínez, Nayi Sánchez Fleitas
- Software: Manuel Enrique Puebla Martínez, Nayi Sánchez Fleitas
- Supervisión: Alfredo Javier Simón Cuevas, María Matilde García Lorenzo, José Manuel Perea Ortega, Raúl Comas Rodríguez
- Validación: Manuel Enrique Puebla Martínez, Nayi Sánchez Fleitas
- Visualización: Alfredo Javier Simón Cuevas, María Matilde García Lorenzo, José Manuel Perea Ortega, Raúl Comas Rodríguez, Raúl Comas Rodríguez
- Redacción-borrador original: Alfredo Javier Simón Cuevas, María Matilde García Lorenzo, Manuel Enrique Puebla Martínez, Nayi Sánchez Fleitas, José Manuel Perea Ortega, Raúl Comas Rodríguez
- Redacción-revisión y edición: Alfredo Javier Simón Cuevas, María Matilde García Lorenzo

### Financiación

El desarrollo de esta investigación se llevó a cabo sin fondos de financiamiento externo.

### Cómo citar este artículo

Simón Cuevas AJ, García Lorenzo MM, Puebla Martínez ME, Sánchez Fleitas N *et al.*. Enfoques semánticos basados en ontologías para la recuperación de información en Sistemas de Información Geográfica. *Anales de la Academia de Ciencias de Cuba* [internet] 2022[citado en día, mes y año];12(1): e1124. Disponible en: <http://www.revistaccuba.cu/index.php/revacc/article/view/1124>

