



# Técnicas semánticas en la localización geoespacial de los suelos en Cuba utilizando medidas de similitud espacial

## *Semantic techniques in the spatial location of soils in Cuba using spatial similarity measures*

Dr.C. Neili Machado-García<sup>1</sup>, M.Sc. Minelkis Machado-Molina<sup>1</sup>, M.Sc. C. Izarys Rodríguez-Lohuiz<sup>1</sup>,  
Dr.C. Carlos Balmaseda-Espinosa<sup>II</sup>

<sup>1</sup> Universidad Agraria de la Habana, Facultad de Ciencias Técnicas, Departamento de Informática, San José de las Lajas, Mayabeque, Cuba.

<sup>II</sup> Universidad Estatal Península de Santa Elena, Facultad de Ciencias Agrarias, Ecuador.

**RESUMEN.** Durante los últimos años, múltiples organizaciones se han dado a la tarea de recopilar información geográfica, la cual en diferentes contextos puede servir para diversos propósitos. Consecuentemente, localizar esta información se convierte en una prioridad para los usuarios que necesitan acceder a ella. En este trabajo se presenta un sistema de recuperación que manipula relaciones espaciales entre objetos geográficos y mejora la gestión de esta información. Esta propuesta introduce de forma novedosa las relaciones espaciales en la determinación de la similitud semántica entre conceptos geoespaciales para resolver consultas que combinan aspectos textuales y espaciales. Otra cualidad importante de este sistema es presentar de manera amigable los resultados, permitiendo navegar por la cartografía que proporciona el sistema. La información indexada puede ser consultada fácilmente por los usuarios y se visualiza de manera clara y organizada para el apoyo a la toma de decisiones. Esta herramienta puede ser utilizada tanto para la caracterización y localización geoespacial de los suelos, como para el manejo de los suelos agrícolas en Cuba.

**Palabras clave:** Relaciones espaciales, Similitud semántica, Ontología, Recuperación de Información Geográfica.

**ABSTRACT.** In last years, many organizations have been given the task of collecting geographic information, which can be used in different contexts for different purposes. Consequently, locating this information becomes a priority for users who need to access it. In this paper a recovery system that manipulates spatial relations between geographic objects is presented. This proposal introduces of novel way the spatial relations to determinate semantic similarity between concepts to solve geospatial queries combining textual and spatial aspects. Another important feature of this system is to present the results in a friendly way, allowing navigate the mapping provided by the system. The indexed information can be easily accessed by users and displayed in a clear and organized to support decision-making way. This tool can be used for both characterization and geospatial location of the soil, as for the management of agricultural soils in Cuba.

**Keywords:** Spatial relation, Semantic similarity, Ontology, Geographic Information Retrieval.

## INTRODUCCIÓN

En el dominio de la información geográfica el desarrollo de tecnologías para capturar información sobre la Tierra y el incremento en la distribución de los Sistemas de Información Geográfica (SIG) en la Web ha generado el almacenamiento de gran cantidad de información geoespacial. Debido a este gran cúmulo de información las técnicas de organización y búsqueda de la información geográfica han cobrado gran im-

## INTRODUCTION

In the domain of geographic information, technology development to capture information about Earth and the increase in the distribution of Geographic Information Systems (GIS) in the Web has generated storing large amount of geospatial information. Because of this wealth of information organizing and finding geographic information techniques have become very important to extract from data all possible useful infor-

portancia para poder extraer de los datos toda la información útil que sea posible. Sin embargo, los datos geográficos poseen características específicas que dificultan su manipulación, la ubicación espacial, o sea el estar en una localización sobre la superficie de la tierra referida a un sistema de coordenadas, la temporalidad y las relaciones espaciales con otros objetos o datos, además, presentan gran heterogeneidad y volumen de almacenamiento.

Las ontologías han sido analizadas en la Geociencia como un procedimiento de estandarización que facilita la traducción entre diferentes fuentes de información (Batet *et al.*, 2013; Gan *et al.*, 2013). La similitud semántica es fundamental para este tipo de procesamiento de datos geoespaciales, establece el grado de interoperabilidad entre ellos o los diferentes SIG y constituyen las bases para la recuperación y la integración de información semántica (Janowicz *et al.*, 2011; Li *et al.*, 2012; Ballatore *et al.*, 2013a, b).

La investigación dentro de la recuperación semántica de información geográfica enfrenta diversos retos como la extensión del enfoque sintáctico al semántico y espacial y la determinación de intención del usuario expresada en las consultas. Algunos autores sugieren un modelo de recuperación<sup>1</sup>, que integra un criterio semántico con criterios geoespaciales y en el que se propone como trabajos futuros complementar el modelo de recuperación con un módulo de Procesamiento de Lenguaje Natural, para procesar consultas y proponer componentes de análisis topológico a través de la implementación de medidas de similitud semántica, propuesta que coincide con la realizada por Harispe *et al.* (2015).

Larín (2013), desarrolla un método para la representación semántica multidimensional de objetos geoespaciales integrados en un entorno de Sistemas de Información Geográfica Gobernados por Ontologías (SIGGO). No obstante, no considera las relaciones topológicas.

En este trabajo se presenta un sistema de recuperación semántica de información geoespacial, el cual utiliza una ontología que conceptualiza la Nueva Clasificación Genética de los Suelos de Cuba (Hernández *et al.*, 1999), y aplica la medida de similitud semántica DIS-C (Vizcarra *et al.*, 2013), entre los conceptos representados para identificar y recuperar los elementos que comparten propiedades similares.

## MÉTODOS

El principal objetivo de un sistema de recuperación de información geográfica es resolver consultas que combinen una componente textual y una componente espacial. En ambos casos es importante la interpretación de su semántica. La búsqueda en el sistema propuesto comienza con una solicitud del usuario (Figura 1). Para mejorar el criterio de búsqueda se puede especificar el tipo de relación.

However, geographic data have specific characteristics that hinder their manipulation, spatial location, which is, being in a location on the Earth surface referred to a coordinate system, the temporality and spatial relationships with other objects or data. Besides, they also present great heterogeneity and storage volume.

Ontologies have been analyzed in Geoscience as a standardization procedure that facilitates the translation between different sources of information (Batet *et al.*, 2013; Gan *et al.*, 2013). The semantic similarity is essential for this type of processing geospatial data, sets the degree of interoperability between them or different GIS and forms the basis for the recovery and integration of semantic information (Ballatore *et al.*, 2013a, 2013b; Janowicz *et al.*, 2011; Li *et al.*, 2012).

Research into semantic retrieval of geographic information faces several challenges such as the extension of syntactic approach to the semantic and spatial one and the determination<sup>1</sup> of user's intention expressed in consultations. Rentería Agualimpia (2009) suggests a recovery model that integrates a semantic approach to geospatial criteria and in which it is proposed as future work to complement the recovery model with a module of Natural Language Processing to process queries and propose components of topological analysis through implementation of semantic similarity measures, proposal that matches with the one made by Harispe *et al.* (2015).

Larín (2013), develops a method for the multidimensional semantic representation of integrated geospatial objects in an environment of Geographic Information Systems Governed by Ontology (SIGGO). However, it does not consider the topological relationships.

In this paper a system of semantic retrieval of geospatial information, which uses an ontology that conceptualizes the New Genetic Classification of Soils of Cuba (Hernández *et al.*, 1999), and applies the measure of semantic similarity DIS-C (Vizcarra *et al.*, 2013), among the concepts represented to identify and retrieve items that share similar properties.

## METHOD

The main objective of a system of geographic information retrieval is to resolve queries that combine a textual component and a spatial component. In both cases the interpretation of their semantics is important. The search in the proposed system begins with a user request (see Figure 1). To improve the search criterion you can specify the type of relationship.

The first step in the search string entered by the user is to identify the entities and subsequently, simplify and standardize the words of the query. In this step, the morphological analyzer Freeling is used (Atserias *et al.*, 2006). Here there is an exception with the words of stops because prepositions are essential to identify spatial relationships.

<sup>1</sup> RENTERÍA, A.W.: Recuperación controlada de información cualitativa desde repositorios de datos, Instituto Politécnico Nacional. Centro de Investigación en Computación, Tesis de Maestría, México, D. F., 164pp., [en línea] 2009, [en línea] 2009, Disponible en: <http://www.repositoriodigital.ipn.mx/handle/123456789/5693>, [Consulta: 28 de agosto de 2016].

FIGURA 1. Interface de usuario para la búsqueda de información geográfica.  
 FIGURE 1. User interface for geographic information search.

El primer paso en la cadena de búsqueda introducida por el usuario consiste en identificar las entidades y posteriormente simplificar y normalizar las palabras de la consulta. Para este paso se utiliza el analizador morfológico Freeling (Atserias *et al.*, 2006). Aquí hay una excepción con las palabras de paradas porque las preposiciones son fundamentales para identificar las relaciones espaciales.

### Recuperación semántica de objetos geográficos

Para llevar a cabo la recuperación de la información geográfica se propone la API (en inglés Application Program Interface) de Jena que incluye un motor de inferencia basado en reglas. Con el puente que se genera entre Jena y Protégé (software editor de ontologías) se lleva a cabo la recuperación de las instancias de los conceptos geográficos.

Posteriormente se activan los términos similares a los términos de búsqueda. Para identificar los conceptos similares en la ontología se utiliza el algoritmo de DIS-C. La idea principal de este algoritmo es determinar el valor de la distancia conceptual entre cada uno de los tipos de relaciones en la ontología. Esto se logra al convertir esta última en un grafo dirigido y ponderado en el que cada concepto se convierte en un nodo y cada relación se convierte en un par de aristas. Una vez que se tiene el grafo, se calculan las distancias mínimas, con lo que se puede hallar la distancia entre conceptos que no se encuentran directamente relacionados.

### Análisis espacial

Se realiza un análisis de los objetos espaciales recuperados. Para esto a cada objeto le corresponde una cadena de texto estándar, la cual es propuesta por la OGC (Open Geospatial Consortium - por sus siglas en inglés) para el intercambio de información geométrica denominada WKT (Well Known Text- por sus siglas en inglés). Esta cadena se encarga de enumerar todos los nodos y vértices que conforman un objeto espacial.

Las cadenas de texto WKT contienen un tipo de dato “primitivo” en las bases de datos extendidas espacialmente, como es

### Semantic retrieval of geographic objects

To carry out the recovering of geographic information, Jena API (Application Program Interface) which includes an inference engine based on rules is proposed. With the bridge generated between Jena and Protégé (ontology editor software) the recovery of geographic concepts instances is carried out.

Later on similar terms to the search terms are activated. To identify similar concepts in the ontology, algorithm DIS-C is used. The main idea of this algorithm is to determine the value of the conceptual distance between each of the types of relationships in the ontology. This is accomplished by converting the latter in a weighted directed graph and where each concept becomes a node and each relationship becomes a pair of edges. Once the graph is established minimum distances are calculated, thus it is possible to find the distances between concepts that are not directly related are calculated.

### Spatial analysis

An analysis of the retrieved space objects is performed. For that, a standard text string corresponds to each object, and that string is proposed by the OGC (Open Geospatial Consortium) for the exchange of geometric information called WKT (Well-Known Text). This string is responsible for listing all nodes and vertices that make a space object.

Text strings WKT contain a data type “primitive” in the databases extended spatially, as in the case of PostGIS (Postgress + Extensions GIS), which can be stored in a field in any database table data (BD). Once the WKT string is obtained, all methods of topological and geometric analysis that supports JTS can be invoked, in order to calculate the spatial similarity.

The TDD model is implemented as an extension of Plane Parallel Coordinates (PCP-by its acronym in English) (Insel-

el caso de PostGIS (Postgress + Extensiones GIS), por lo que se pueden almacenar en un campo dentro de cualquier tabla de la base de datos (BD). Una vez que se tiene la cadena WKT, se puede invocar a todos los métodos de análisis topológico y geométrico que soporta JTS, con el fin de calcular la similitud espacial.

El modelo TDD se implementa como una extensión del Plano de Coordenadas Paralelas (PCP-por sus siglas en inglés) (Inselberg, 1985). En esta implementación, PCP se trabaja con tres ejes que representan las relaciones topológicas, de dirección y de distancia.

Para visualizar las relaciones espaciales cualitativas, el eje vertical del PCP tradicional se reemplaza por una red del vecindario conceptual. En el eje de las relaciones topológicas, se representan seis nodos para los diferentes tipos de relaciones topológicas: igual, separado, adyacentes, superpuesto, contiene y cubre a (Figura 2a).

berg, 1985). In this implementation, PCP is worked with three axes representing topological relations of direction and distance.

To display the qualitative spatial relationships, the vertical axis of traditional PCP is replaced by a conceptual neighborhood network. In the axis of topological relations six nodes are represented for the different types of them like, separate, adjacent, overlapping, contains and covers. See Figure 2 (a).

The lengths of the arcs on the axis between the nodes reflect the weight defined in the topological network. The relation between weight and length of each arc metric on the axis is described as:

$$\frac{\text{weight 1}}{\text{weight 2}} = \frac{\text{arc length 1}}{\text{arc length 2}} \quad (1)$$

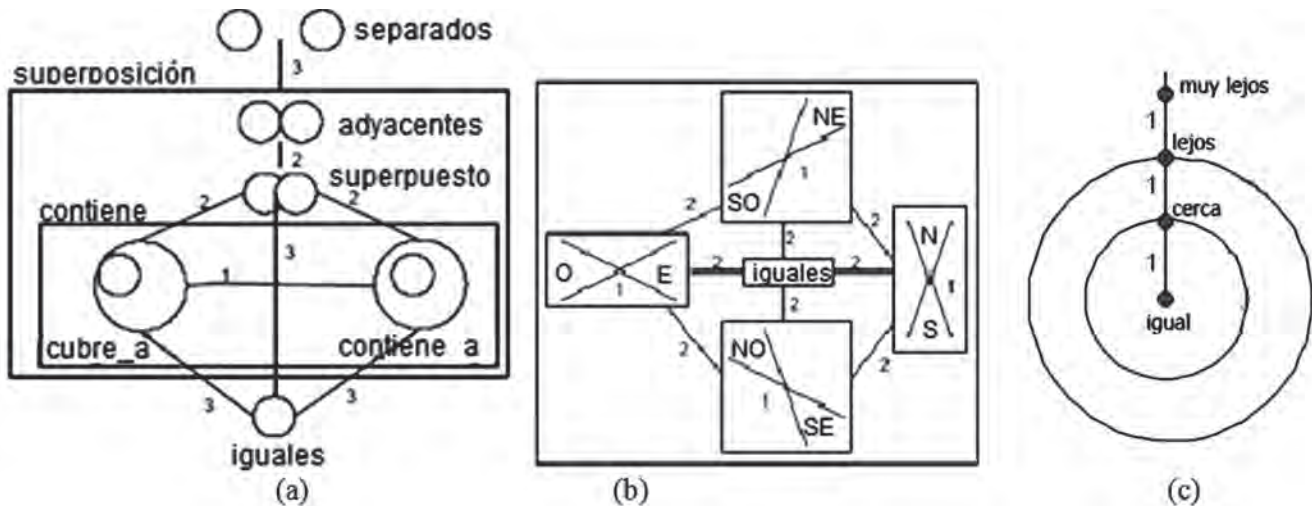


FIGURA 2. (a) relaciones topológicas, (b) relaciones de dirección, (c) relaciones de distancia.  
 FIGURE 2. (a) topological relationships, (b) management relations, (c) distance relationships.

Las longitudes de los arcos en el eje entre los nodos reflejan el peso definido en la red topológica. La relación entre el peso y la longitud métrica de cada arco en el eje se describe como:

$$\frac{\text{weight 1}}{\text{weight 2}} = \frac{\text{arc length 1}}{\text{arc length 2}} \quad (1)$$

Una excepción es el arco de contiene a “cubre a” en este caso el peso es 1, de “superposición” a los otros nodos es 2 y de “igual a” a otros nodos es 3. El análisis de las relaciones topológicas utiliza los métodos que soporta la API JTS a partir de la especificación de la OGC del DE-9IN (Dimensionally-Extended 9 IntersectionMatrix –por sus siglas en inglés).

En el eje de la relación de dirección (Figura 2(b)) las cinco relaciones son iguales, Este, Oeste, Noreste, Sureste, Norte, Sur, Noroeste y/o Suroeste. El peso puesto para cada arco es 2. Es posible establecer dichas relaciones con JTS porque se pueden considerar las coordenadas (x,y) del centroide de un objeto espacial directamente o de su polígono convexo. Después de comparar dichas coordenadas, y asumiendo que la región espacial en estudio está al Norte orientado hacia “arriba”, es fácil comparar la coordenada X de dos objetos geográficos y determinar la que tenga un valor mayor (más a la derecha) estará “al Este” de dicho objeto (que entonces está al Oeste).

An exception is the arch of contains “covers to” in this case the weight is 1, “overlay” to the other nodes is 2 and the “equal” to other nodes is 3. The analysis of the topological relationships uses methods supported by API JTS from the specification of DE-9IN OGC (Dimensionally-Extended 9 IntersectionMatrix).

In the axis of the steering ratio (see Figure 2 (b)) the five relationships are equal, East, West, Northeast, Southeast, North, South, West and / or West. The weight set for each arc is 2. It is possible to establish such relations with JTS because the coordinates (x, y) of the centroid of a space object directly or of its convex polygon can be considered. After comparing these coordinates and assuming that the spatial region under study is in the North, facing “up”, it is easy to compare the X coordinate of two geographic objects and determine which has a higher value (rightmost), it will be “East “of the object (which is then West).

In the same way, it is possible to compare the two geographical coordinates and objects and set the coordinate that has higher value (up) then it will be north of the other object (located south).

The axis of the distance ratio (see Figure 2 (c)) represents geometrically the metric distance with a line and four points

De la misma forma, se pueden comparar las coordenadas Y de dos objetos geográficos y establecer el que tenga la coordenada con mayor valor (hacia arriba) entonces estará al Norte del otro objeto (que se encuentra al Sur).

El eje de la relación de distancia (ver Figura 2 (c)) representa geoméricamente la distancia métrica con una línea y cuatro puntos que representan las cuatro relaciones de distancia: igual, cerca, medio y lejos.

### Integración de los criterios de recuperación

Debido a que se tienen definidos dos criterios en la valoración de la similitud, conceptual y espacial, es necesario establecer cómo se integrarán.

En la Figura 3 se visualiza que los valores más cercanos a cero satisfacen mejor la consulta del usuario. Para un objeto que satisface una consulta, será representado como una tupla de valores en el espacio. Con las siguientes definiciones:

- El valor de la similitud conceptual se define como el valor de similitud obtenido al aplicar el algoritmo DIS-C: si entonces el objeto es muy similar al buscado, si entonces el objeto satisface exactamente a la consulta.
- El valor de la similitud espacial se define como el valor de similitud obtenido al aplicar el algoritmo TDD: si entonces el objeto es muy similar, en cuanto a sus relaciones espaciales, al buscado, si entonces el objeto satisface exactamente a la consulta.
- El valor de relevancia, de cada resultado, es decir, cada punto en el espacio para la integración de criterios, es calculado como señala la ecuación 2:

$$R_w(c_i, w_i) = 1 - \left( \frac{\sum_{i=1}^n (c_i)^{w_i}}{n} \right) \quad (2)$$

Donde  $c_i$  son los criterios de recuperación,  $w_i$  los pesos asignados a cada criterio y  $n$  el número de criterios a utilizar, en este caso.

La ecuación 2 permite determinar los resultados que más se aproximan o satisfacen la consulta. De manera tal que, si entonces el resultado satisface totalmente la consulta. Según la Figura 3 el objeto es más satisfactorio por estar más próximo al origen. En este ejemplo, y,

representing the four distance relationships: same, near, middle and far.

### Integration of recovery criteria

Due to the existence of two defined criteria in the assessment of the similarity: conceptual and spatial, it is necessary to establish how to integrate them.

Figure 3 shows that values closest to zero better meet the user's query. For  $O_i$  object satisfying a query, it will be represented as a tuple of  $O_i$  values  $(\epsilon, \sigma)$  in space. With the following definitions:

- The value of the conceptual similarity  $\epsilon$  is defined as the similarity value obtained by applying the algorithm DIS-C: If  $\epsilon \rightarrow 0$ , then the  $O_i$  object is very similar to that sought; if  $\epsilon = 0$ , then the object satisfies the query exactly.
- The value of the spatial similarity  $\sigma$  is defined as the similarity value obtained by applying the algorithm TDD: if  $\sigma \rightarrow 0$ , then the  $O_i$  object is very similar, in their spatial relationships to that searched; if  $\sigma = 0$ , then the object exactly satisfies the query.
- The relevance value  $R$  of each result, that is, each point in space for integrating criteria, is calculated as Equation 3.5 points:

$$R_w(c_i, w_i) = 1 - \left( \frac{\sum_{i=1}^n (c_i)^{w_i}}{n} \right) \quad (2)$$

Where  $C_i$  are the recovery criteria,  $w_i$  are weights assigned to each criterion and  $n$  is the number of criteria to be used in this case  $n = 2$ .

Equation 2 determines the results that better approximate or satisfy the query. So that, if  $R_w \rightarrow 1$  then the result fully satisfies the query. According to Figure 3 the  $O_1$  object  $(\epsilon, \sigma)$  is more satisfying to be closest to the origin. In this example  $\sigma_1 < \sigma_2$  and  $\epsilon_1 < \epsilon_2$ .

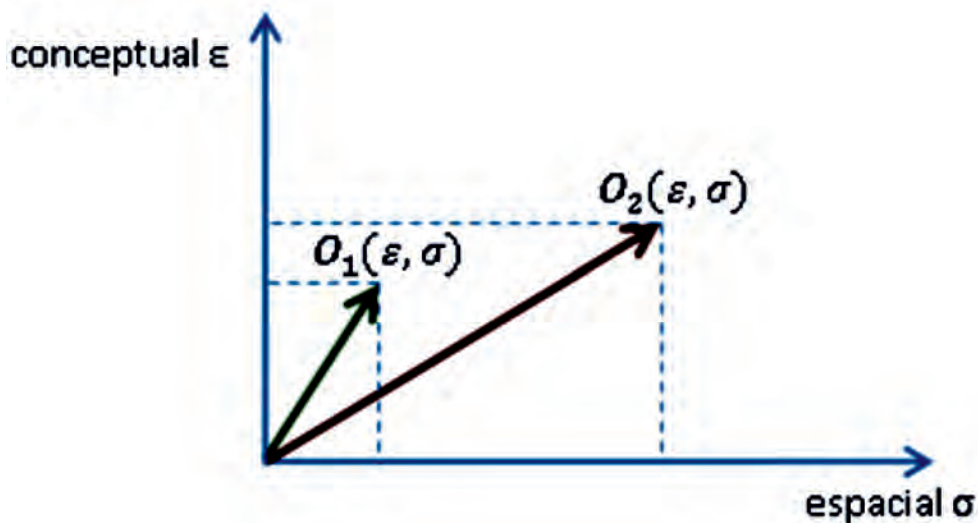


FIGURA 3. Espacio para la comparación de criterios.  
FIGURE 3. Space for comparison criteria.

## RESULTADOS Y DISCUSIÓN

Los datos con los que se ha trabajado proceden de los mapas de suelos de la región de San José de Las Lajas en la provincia de Mayabeque, el mapa de la propia provincia de Mayabeque y el de la provincia de Pinar del Río, región occidental de Cuba.

Se tiene una superficie del municipio San José de Las Lajas que posee diversos tipos de temáticas. La región está compuesta por polígonos de tipos de suelos (Ps), por perfiles (Pf) y por elementos hidrológicos (Hd) dentro de los que se encuentran 21 objetos de tipo polígono, 1 objeto de tipo línea y 19 de tipo punto.

Además, se cuenta con 26 imágenes de los diferentes tipos de suelos que están representados en esos polígonos. Estas imágenes son utilizadas en la carrera de Ingeniería Agronómica, en la cual un alto porcentaje de sus asignaturas emplea este tipo de objeto como apoyo a los diferentes procesos de identificación y diagnóstico que se estudian en esta carrera.

### Ontología espacial como estructura de indexación

La ontología propuesta modela este espacio conceptual, por lo que permite instanciar todos los objetos que se encuentran dentro de los mapas utilizados, así como las relaciones que existen entre ellos. Además, tiene instancias de los documentos y archivos con extensión *.jpg* que se encuentran almacenados en una base de datos relacional.

Una vez que se lanza la recuperación, es necesario seleccionar las instancias que representan objetos geográficos, con el objetivo de verificar si existe relación entre las mismas; entonces si se cumple esta condición se genera una parte de la descripción para este par de instancias. Por ejemplo, de acuerdo con el escenario presentado, el cual está compuesto de varios objetos geográficos, que reflejan una relación de existencia “”: “”. Además, el concepto relación “” está vinculado con y varios de los polígonos de suelos; donde es un objeto lineal e instancia del concepto “”. A su vez, la relación “” esta generalizada como una relación topológica, que dentro de la ontología hereda de la clase “”. De esta forma se instancian tanto las imágenes provenientes de la base de datos relacional como los objetos geográficos de la BD geográfica.

### Recuperación semántica de información geoespacial

Para llevar a cabo la recuperación dentro del sistema propuesto por Machado *et al.* (2014), se implementaron las funcionalidades básicas existentes en los SIG convencionales accesibles a través de la barra de herramientas, estas son: carga y visualización de capas de datos, interacción con el mapa (Paneo, Zoom, Restablecer) y consulta a los datos.

La aplicación también posee un visor de mapas en el cual se pueden observar los datos cargados a través de su componente espacial permitiendo realizar búsquedas de información de los diferentes tipos de suelos. Al seleccionar la opción buscar, por ejemplo, si se lanza una búsqueda de determinado tipo de suelo “ferralítico rojo”, se recupera un mapa de la región en el cual se visualiza la localización de este tipo de suelo. Además, se mostrarán todas las imágenes relacionadas con la solicitud.

Con la implementación de esta aplicación se logra un ahorro de tiempo en la realización de los análisis espaciales. Esto es debido a que se eliminan una serie de pasos que conllevan a un gasto de tiempo. Además, se le simplifica el trabajo al usuario

## RESULTS AND DISCUSSION

The data that have been worked come from the soil maps of the region of San Jose de Las Lajas in the province of Mayabeque, the map of Mayabeque province and the one from Pinar del Rio province, western region of Cuba.

It has an area of San José de las Lajas municipality possessing various types of themes. The region is composed of polygons of soil types (Ps), profiles (Pf) and hydrological elements (Hd), among them there are 21 objects of type polygon, 1 object line type and 19 point type.

In addition, it has 26 images of different soil types that are represented in these polygons. These images are used in Agricultural Engineering major, where a high percentage of its courses use this type of objects to support the different processes of identification and diagnosis studied.

### Ontology as spatial indexing structure

The proposed ontology models this conceptual space, allowing instantiate all objects that are within the maps used, and the relationships among them. It also has instances of documents and files with *.jpg* extension are stored in a relational database.

Once the recovery is launched, it is necessary to select the instances that represent geographic objects, in order to verify whether a relationship exists between them; then if this condition is met, part of the description for this pair of instances is generated. For example, according to the scenario presented, which is composed of several geographic objects that reflect a relationship of existence “it is”: “R1 is Linear Object”. In addition, the relationship “Intersect” concept is linked to R1 and several soil polygons; where R1 is a linear object and instance of the concept “river”. In turn, the “Intersect” relationship is generalized as a topological relationship that within the ontology inherits from the “geographic\_relation” class. Thus, both the images from the relational database as geographical objects of the geographical BD are instantiated.

### Semantic retrieval of geospatial information

To carry out recovery within the proposed system Machado *et al.* (2014), the basic features existing in the available conventional GIS through the toolbar were implemented, these are: load and display of data layers, interaction with the map (Pan, Zoom, Reset) and data consultation.

The application also has a map viewer where it is possible to see the data loaded through its spatial component allowing searching for information on the different types of soils. By selecting the option searching, for example, if a search for a particular “red ferrallitic” soil type is launched, a map of the region in which the location of this soil type is displayed, is recovered. In addition, all images related to the application are going to be displayed.

With the implementation of this application a time saving is achieved in performing spatial analysis. This is because a series of time wasting steps are eliminated. In addition, it simplifies the work to the user and prevents mistakes in recovery because

y se evita que cometa errores en la recuperación ya que no tiene que generar una consulta en lenguaje SQL, porque puede plantear su búsqueda en lenguaje natural.

En la Tabla 1 se presentan los resultados comparativos en cuanto a eficiencia de los métodos propuestos. Estos resultados ponen de manifiesto como las relaciones existentes entre los objetos, en este caso las relaciones topológicas, resultan útiles en tareas como la visualización automática de la clasificación de los suelos.

Los resultados demuestran que la recuperación de la información mediante el método propuesto aumenta la calidad de los análisis pues reduce los posibles errores en las operaciones que deben realizar, por ejemplo, al crear la consulta.

it does not have to generate a query in SQL language, because it can raise its natural language search.

In Table 1 the comparative results in efficiency of the proposed methods are presented. These results show how that the relationships between objects, in this case the topological relationships, are useful in tasks such as automatic display of soil classification.

The results show that the retrieval of information by the proposed method increases the quality of the analysis as it reduces possible errors in the operations to be carried out, for example, when creating the query.

**TABLA 1. Resultados de las comparaciones entre tiempos de la realización de las operaciones de búsqueda y análisis de la información espacial utilizando el método propuesto y métodos de recuperación basada en el lenguaje SQL**  
**TABLE 1. Results of comparisons between times conducting search operations and analysis of spatial information using the proposed method and recovery methods based on language SQL**

Methodological steps	Selecting soil type regions	Time (min)
GIS Operations	Manual selection of the layer. Manual Conformation of the SQL statement. Automatic selection of objects.	3,0
Operation in the proposed system	Manual search criteria Conformation of natural language. Automatic selection of objects.	0,5

Estos resultados muestran que el uso de la ontología como una capa intermedia entre el sistema de recuperación y las fuentes de datos permite el manejo de los objetos a través de su representación conceptual (abstracción). Dando la posibilidad de que el sistema disponga de herramientas para el procesamiento semántico de los datos geoespaciales. Esto permite que los diferentes tipos de análisis realizados sobre esos tipos de datos puedan ser realizados tanto por especialistas como por usuarios no expertos en el uso de SIG.

These results show that the use of ontology as an intermediate layer between the recovery system and data sources enables the handling of objects through its conceptual representation (abstraction) by giving the possibility that the system has tools for semantic processing of geospatial data. This allows different types of analyses of these data types can be made by both specialists and non-expert users in the use of GIS.

**CONCLUSIONES**

- La aplicación desarrollada utiliza las relaciones espaciales en la determinación de la similitud semántica entre conceptos geoespaciales para resolver consultas que combinan aspectos textuales y espaciales.
- Los resultados de la búsqueda se presentan de manera amigable, permitiendo navegar por la cartografía que proporciona el sistema.
- Con base en los elementos característicos identificados de los objetos geoespaciales se definió de forma inédita una ontología que conceptualiza la Nueva Versión de Clasificación Genética de los Suelos de Cuba con la anotación semántica de los mapas de suelos de las regiones de Pinar del Río, Mayabeque y San José de Las Lajas. Esta ontología permite simular el rol de un experto al que se le puede preguntar acerca de un tópico.
- Se implementó un módulo con técnicas de Procesamiento del Lenguaje Natural para que pueda ser gestionado por personas NO expertas y una aplicación de experimentación en la cual se evaluaron los métodos propuestos.

**CONCLUSIONS**

- The developed application uses spatial relationships in determining semantic similarity between concepts to solve geospatial queries combining textual and spatial aspects.
- The search results are presented in a friendly manner, allowing browsing the mapping provided by the system.
- Based on the characteristic elements identified of geospatial objects, it was defined, in an unprecedented way, an ontology that conceptualizes the New Version of Genetic Classification of Soils of Cuba with the semantic annotation of soil maps of the regions of Pinar del Rio, Mayabeque and San Jose de Las Lajas. This ontology allows simulating the role of an expert who can be asked about a topic.
- A module with techniques for Natural Language Processing that can be managed by non-experts and an application of experimentation in which the proposed methods were evaluated was implemented.

## REFERENCES / REFERENCIAS BIBLIOGRÁFICAS

- ATSERIAS, J.; CASAS, B.; COMELLES, E.; GONZÁLEZ, M.; PADRÓ, L.; PADRÓ, M.: "FreeLing 1.3: Syntactic and semantic services in an open-source NLP library", [en línea], En: *V International Conference on Language Resources and Evaluation*, Ed. European Language Resources Association, Evaluation and Language resources Distribution Agency, Istituto di Linguistica Computazionale, Genoa, Italy, pp. 2281-2286, 2006, Disponible en: [http://www.lrec-conf.org/proceedings/lrec2006/pdf/198\\_pdf.pdf](http://www.lrec-conf.org/proceedings/lrec2006/pdf/198_pdf.pdf), [Consulta: 21 de septiembre de 2016].
- BALLATORE, A.; BERTOLOTTO, M.; WILSON, D.C.: "Geographic knowledge extraction and semantic similarity in OpenStreetMap", *Knowledge and Information Systems*, 37(1): 61-81, octubre de 2013a, ISSN: 0219-1377, 0219-3116, DOI: 10.1007/s10115-012-0571-0.
- BALLATORE, A.; WILSON, D.C.; BERTOLOTTO, M.: "Computing the semantic similarity of geographic terms using volunteered lexical definitions", *International Journal of Geographical Information Science*, 27(10): 2099-2118, octubre de 2013b, ISSN: 1365-8816, 1362-3087, DOI: 10.1080/13658816.2013.790548.
- BATET, M.; SÁNCHEZ, D.; VALLS, A.; GIBERT, K.: "Semantic similarity estimation from multiple ontologies", *Applied Intelligence*, 38(1): 29-44, enero de 2013, ISSN: 0924-669X, 1573-7497, DOI: 10.1007/s10489-012-0355-y.
- GAN, M.; DOU, X.; JIANG, R.: "From Ontology to Semantic Similarity: Calculation of Ontology-Based Semantic Similarity", *The Scientific World Journal*, 2013: 1-11, 2013, ISSN: 1537-744X, DOI: 10.1155/2013/793091.
- HARISPE, S.; RANWEZ, S.; JANAQI, S.; MONTMAIN, J.: "Semantic Similarity from Natural Language and Ontology Analysis", *Synthesis Lectures on Human Language Technologies*, 8(1): 1-254, 23 de mayo de 2015, ISSN: 1947-4040, 1947-4059, DOI: 10.2200/S00639ED-1V01Y201504HLT027.
- HERNÁNDEZ, J.A.; PÉREZ, J.M.; BOSCH, D.; RIVERO, L.; CAMACHO, E.; RUÍZ, J.; SALGADO, E.J.; MARSÁN, R.; OBREGÓN, A.; TORRES, J.M.; GONZÁLES, J.E.; ORELLANA, R.; PANEQUE, J.; RUIZ, J.M.; MESA, A.; FUENTES, E.; DURÁN, J.L.; PENA, J.; CID, G.; PONCE DE LEÓN, D.; HERNÁNDEZ, M.; FRÓMETA, E.; FERNÁNDEZ, L.; GARCÉS, N.; MORALES, M.; SUÁREZ, E.; MARTÍNEZ, E.: *Nueva versión de clasificación genética de los suelos de Cuba*, Ed. AGROINFOR, La Habana, Cuba, 64 p., 1999, ISBN: 959-246-022-1.
- INSELBERG, A.: "The plane with parallel coordinates", *The Visual Computer*, 1(2): 69-91, agosto de 1985, ISSN: 0178-2789, 1432-2315, DOI: 10.1007/BF01898350.
- JANOWICZ, K.; RAUBAL, M.; KUHN, W.: "The semantics of similarity in geographic information retrieval", *Journal of Spatial Information Science*, 2011(2): 29-57, 25 de mayo de 2011, ISSN: 1948-660X, DOI: 10.5311/JOSIS.2011.2.26.
- LARÍN, R.: *Nuevo tipo de ontología para la representación semántica de objetos espaciales*, Ed. CENATAV, La Habana, Cuba, 2013.
- LI, W.; RASKIN, R.; GOODCHILD, M.F.: "Semantic similarity measurement based on knowledge mining: an artificial neural net approach", *International Journal of Geographical Information Science*, 26(8): 1415-1435, agosto de 2012, ISSN: 1365-8816, 1362-3087, DOI: 10.1080/13658816.2011.635595.
- MACHADO, G.N.; GONZÁLEZ, R.L.; BALMASEDA, E.C.: "Recuperación de objetos geoespaciales utilizando medidas de similitud semántica", *Revista Cubana de Ciencias Informáticas*, 8(2): 132-144, junio de 2014, ISSN: 2227-1899.
- VIZCARRA, J.; TORRES, M.; QUINTERO, R.; MORENO-IBARRA, M.: "SemGsearch: An Approach to Semantically Retrieve Geospatial Objects from Different Geographic Servers", *Journal of Web Engineering*, 12(5): 403-421, octubre de 2013, ISSN: 1540-9589.

---

Received: 28/12/2015.

Approved: 08/07/2016.

*Neili Machado-García*, Prof. Titular, Universidad Agraria de la Habana, Facultad de Ciencias Técnicas, Departamento de Informática, San José de las Lajas, Mayabeque, Cuba. E-mail: [neili@unah.edu.cu](mailto:neili@unah.edu.cu)

*Minelkis Machado-Molina*, E-mail: [neili@unah.edu.cu](mailto:neili@unah.edu.cu)

*C. Izarys Rodríguez-Lohuiz*, E-mail: [neili@unah.edu.cu](mailto:neili@unah.edu.cu)

*Carlos Balmaseda-Espinosa*, E-mail: [cbalma59@gmail.com](mailto:cbalma59@gmail.com)

Nota: La mención de marcas comerciales de equipos, instrumentos o materiales específicos obedece a propósitos de identificación, no existiendo ningún compromiso promocional con relación a los mismos, ni por los autores ni por el editor.

Note: the mention of commercial equipment marks; instruments or specific materials obey identification purposes, not existing any promotional commitment with relationship to them, neither for the authors nor for the editor.