Cienc. Tierra y Esp., II Época, 2011 / Julio - Diciembre/ No. 12 / pp. 47-57
ISSN 1729-3790

Relaciones topológicas para la generación automática de ontologías de aplicación en sistemas de información geográfica

Rainer Larin-Fonseca y Eduardo Garea-Llano

Centro de Aplicaciones de Tecnologías de Avanzada, 7a# 21812, Siboney, Playa, La Habana-12200, Cuba. Email: rlarin@cenatav.co.cu; egarea@cenatav.co.cu

Recibido: agosto 6, 2010; Aceptado: junio 14, 2011; Versión final recibida: junio 28, 2011

RESUMEN

Proveer a los Sistemas de Información Geográfica (SIG) con los mecanismos para el procesamiento de los datos geográficos basados en su abstracción semántica es una tarea que en el presente es enfrentada por un gran número de investigaciones dada su amplia posibilidad de aplicaciones en la vida real las cuales muchas veces implican la toma de decisiones. Pasos de avances en este sentido pueden ayudar a resolver muchos problemas de los datos geográficos como su heterogeneidad, ya que los SIG podrían contar con herramientas para procesar los datos geográficos sobre la base de su significado y no en su sintaxis y/o estructura. Un aspecto importante para alcanzar estos objetivos es el establecimiento de forma automática de la correspondencia entre los datos geográficos y su conceptualización. En este trabajo se proponen reglas para la representación de las relaciones topológicas en Ontologías de Aplicación. El uso de estas relaciones enriquece este tipo de Ontologías y sirve de mucha ayuda en tareas de consultas o recuperación de datos. El método es aplicado con buenos resultados en un caso de estudio para la evaluación de riesgos.

Palabras clave: Ontologías, Relaciones Topológicas, Representación Semántica, Sistemas de Información Geográfica

Topological relationships for automatic generation of application ontologies in geographical information systems

ABSTRACT

Providing Geographical Information Systems (GIS) with mechanisms for processing geographical data based on their semantic abstraction is very common in current researches. It is due to the wide scope that this type of processing has in real applications. Improvements in this sense may help to solve many problems of geographical data e.g. heterogeneity, since GIS could process geographical data focusing on their meaning and not on their syntax and/or structure. An important aspect for achieving these objectives is the correspondence between geographical data and their conceptualization in an automatic way. In this paper, rules for the representation of topological relationships in Application Ontologies are stated. The representation of these relationships between geographical data enriches these kinds of Ontologies. It may help in data consult or data retrieval tasks. The proposed approach was applied in a case of study for risk evaluation and acceptable results were obtained.

Key words: Ontologies, Topologic Relationships, Semantic Representation, Geographical Information Systems

INTRODUCCIÓN

Debido al amplio uso y la actual relevancia del manejo de la información espacial, resulta de gran importancia encontrar y recuperar información útil en los ambientes abiertos y distribuidos de las actuales Infraestructuras de Datos Espaciales (IDE) (Klien *et al.*, 2004). Esto conlleva buscar nuevas formas de manipular este tipo de información de una manera mucho más eficiente apoyándonos en la semántica subyacente en estos datos. Esto además trae consigo un conjunto de problemas derivados de la interpretación del significado, producto de las diferencias en cuanto a significado y conceptualización en los procesos de descubrimiento y recuperación de información geográfica (Lutz y Klien, 2008). Estos problemas han sido denominados ya por varios autores como "*heterogeneidad semántica*".

La necesidad de acceder y manipular la información geográfica, tomando en cuenta las diferencias de significado en los términos y los objetos representados se ha incrementado en la actualidad (Harvey *et al.*, 1999). Sin embargo los catálogos actuales proveen herramientas de búsqueda basadas en mecanismos ineficientes de recuperación y descubrimiento de la información (Klien *et al.*, 2004), por lo que el uso de mecanismos para el manejo de la información espacial desde un punto de vista semántico resulta hoy en día una necesidad. De esta manera, además, se darían pasos de avance en la resolución de problemas derivados de la compleja naturaleza de este tipo de información como por ejemplo, la heterogeneidad sintáctico-estructural(Larin-Fonseca y Garea-Llano, 2010)

A fin de afrontar los problemas generados por la interoperabilidad sintáctica de los servicios geoespaciales el *Open GeospatialConsortium*[1] (OGC) proporciona un rango completo de especificaciones y establece protocolos para el intercambio de la información espacial. De la misma manera la creación de los estándares de metadatos como los ISO 19115 constituye un paso de avance para alcanzar la interoperabilidad sintáctica y facilitar el procesamiento automático de las descripciones de los datos geográficos.

La heterogeneidad estructural está determinada por los diferentes métodos de representación en el cual están basadas las estructuras de almacenamiento de los datos espaciales. Un elemento importante para evitar este tipo de heterogeneidad es el uso de los metadatos ya que describen la estructura del esquema de representación en el que los datos han sido almacenados.

Por otro lado, la heterogeneidad semántica se refiere la integración de estos datos geoespaciales basados no en su estructura o formato de almacenamiento sino en lo que estos significan realmente, o sea, se refiere a la integración de las diferencias en las convenciones de nomenclatura y conceptualizaciones en los diferentes grupos en los que estos pueden ser clasificados (Bishr, 1998), garantizando un entendimiento mutuo en un contexto determinado entre los diferentes sistemas, incluyendo además a aquellos humanos que puedan interactuar con ellos. Raubal en el 2005 plantea que "*… la interoperabilidad semántica de la información y servicios geográficos ha sido la principal área de investigación dentro de las Ciencias de la Información Geográfica*" (Raubal, 2005) y esta puede ser alcanzada mediante el uso de ontologías de dominio y anotaciones semánticas de los datos espaciales que utilicen estas ontologías (Klien y Lutz, 2005).

El uso de ontologías, traducidas como un componente activo dentro de los sistemas de información, conduce a los Sistemas de Información Guiados por Ontologías y, en el caso específico de los SIG, conduce hacia los Sistemas de Información Geográfica Guiados o Gobernados por Ontologías (ODGIS por sus siglas en inglés) (Fonseca y Egenhofer, 1999). Los ODGIS son construidos utilizando componentes de software derivados de varias ontologías. Estos componentes de software son clases contienen el conocimiento extraído de las ontologías.

La estructura del ODGIS tal y como ha definido Fonseca (Fonseca y Egenhofer, 1999) comprende dos aspectos fundamentales: la generación del conocimiento y el uso del conocimiento. La generación del conocimiento involucra la especificación de las ontologías, la generación de nuevas ontologías a partir de las ya existentes utilizando un editor de ontologías así como la traducción de estas ontologías a componentes de software.

En este proceso de generación del conocimiento uno de los aspectos que constituyen un tema abierto se centra en la necesidad de crear nuevas ontologías, especialmente las llamadas ontologías de aplicación. Estas ontologías tienen que ser construidas manualmente a partir de los datos integrados a la aplicación SIG mediante el análisis de la estructura de la base de datos así como de las relaciones topológicas y espaciales entre los objetos espaciales representados. Este es un proceso complejo, caro computacionalmente y está sujeto a errores y pérdida de información.

Recientemente en la literatura se ha venido notando un interés creciente en el desarrollo de métodos para la extracción automática de la semántica subyacente en bases de datos. La mayoría de las metodologías propuestas

[1]<http://www.opengeospatial.org>

utilizan ontologías como forma de representación del conocimiento. En el caso de la información geográfica esta tarea es mucho más compleja debido a las complejidades de este tipo de información (Leung, 2010).

Un primer paso en la definición de una metodología para la construcción automática de una ontología geoespacial como vista semántica del esquema de una base de datos espacial para el soporte de un sistema de búsquedas orientado al usuario ha sido presentado por Baglioni y colaboradores (Baglioni *et al.*, 2007). Ellos se basan en una serie de reglas para la generación de ontologías a partir de bases de datos que se encuentren en al menos tercera forma normal propuestas por Man Li (Li *et al.*, 2005) e incluyen elementos para el almacenamiento de posición espacial y la geometría del objeto representado pero no toman en cuenta las relaciones espaciales y topológicas que existen entre los objetos espaciales representados, por lo que estas Ontologías que se generan pueden ser enriquecidas semánticamente aún más tomando en cuenta estas relaciones.

En este trabajo se presenta un método que considera las relaciones topológicas como una regla más para la generación de una ontología geoespacial a partir de los datos almacenados en bases de datos espaciales. Introducimos el uso de las relaciones topológicas para el enriquecimiento semántico de la ontología geoespacial generada, la cual puede servir como base para un sistema de búsqueda inteligente que pueda además aprovechar las relaciones topológicas existentes entre los objetos, incrementando así la calidad de los análisis espaciales en los diferentes procesos como la recuperación de información o la de toma de decisiones. Se propone además una nueva estructura para la generación de la ontología de aplicación en la que estén representadas explícitamente tanto la abstracción semántica de los datos como las relaciones existentes entre ellos.

El uso de los SIG como herramienta para la toma de decisiones en aplicaciones relacionadas con la evaluación de peligros, riesgos y vulnerabilidades ante fenómenos meteorológicos de gran envergadura ha tenido un amplio uso por su importancia en la preservación de vidas humanas y la minimización de pérdidas materiales, así como en la evaluación de posibles afectaciones a futuras instalaciones industriales o asentamientos humanos. En este trabajo presentamos un caso de uso del método propuesto para la evaluación de riesgos meteorológicos para la localización de instalaciones de energía renovable (solar y eólica) que demuestran la efectividad del método propuesto.

El trabajo se estructura de la siguiente forma: En el epígrafe *Base teórica* se presentan los principales trabajos que sirven de base teórica al problema. En el epígrafe *Reglas de extracción de los rasgos espaciales* se explica el método propuesto para la extracción semántica mediante la generación automática de ontologías. En el epígrafe *Resultados de la aplicación del método propuesto a un caso de estudio en la evaluación de riesgos* mostramos los resultados de la aplicación del método propuesto a un caso de estudio en la evaluación de riesgos. Finalmente ofrecemos las conclusiones e ideas para trabajos futuros.

BASE TEÓRICA

En años recientes, han sido propuestas varias estrategias para el proceso de integración semántica de datos geoespaciales, como el propuesto por Torres (Torres *et al.*, 2005). Uno de los principales problemas en este tipo de integración radica en dos aspectos principales. La generación de ontologías de aplicación a partir de los datos almacenados y la interrelación de estas ontologías con ontologías de mayor nivel de abstracción como las ontologías de dominio.

Por interrelación de ontologías se entiende, al proceso en el cual términos en una ontología son relacionados con sus semejantes en una segunda ontología. Esta interrelación va a estar determinada a partir de la semejanza entre los términos de ambas ontologías sobre la base de sus propiedades.

Por otro lado, Por la generación de ontologías de aplicación es entendida como la conceptualización de los datos geográficos por medio de ontologías, la intención es generar un modelo conceptual a partir de modelo físico en el que se basan los datos. Este proceso está compuesto por dos fases: *Fase de extracción*: Acceso a los datos para recuperar las entidades almacenadas en el esquema físico dando como resultado un esquema lógico. *Fase de conceptualización*: Generación de un esquema conceptual mediante la aplicación de diferentes procesos sobre el esquema lógico obtenido en la fase de extracción. Los trabajos más relevantes en este sentido son los presentados por Li y An (An *et al.*, 2006; Li *et al.*, 2005)

Existen métodos en la literatura que definen reglas para la extracción automática de ontologías a partir de bases de datos relacionales (Li *et al.*, 2005). Estas reglas producen conceptos y relaciones a partir del esquema y la estructura de la base de datos, la cual debe estar en al menos tercera forma normal. Baglioni y colaboradores (Baglioni *et al.*, 2007), presentaron una propuesta para la extracción de ontologías enriquecidas semánticamente a partir de bases de datos geográficas donde se definen dos nuevas reglas para la extracción de relaciones espaciales para el caso de las localizaciones directas e indirectas. Esta metodología genera una Ontología de aplicación a partir de una base de datos, esta ontología está compuesta por un grupo de términos y las relaciones existentes entre ellos los cuales representan la estructura interna de la base de datos en la cual las relaciones espaciales están

explícitamente representados, luego esta ontología es enriquecida por medio de una Ontología de Dominio definiendo así el dominio semántico en el cual se va a trabajar. Esta ontología enriquecida representa la visión semántica y la taxonomía de los datos, los que son mapeados para facilitar las búsquedas semánticas espaciales.

Tomando en cuenta que la ontología de aplicación es derivada de una base de datos geoespacial, esta depende estrictamente de la estructura y las relaciones de las tablas en la Base de Datos. El objetivo del Módulo de extracción de reglas presentado por Baglioni es la construcción automática de la Ontología de Representación partiendo del esquema de la base de datos.

Este método fue implementado por los autores del presente trabajo y probado en una aplicación SIG para la zonificación agroecológica de cultivos (Larin-Fonseca y Garea-Llano, 2009)

Dado que la ontología debe ser capaz de representar la abstracción semántica de los datos en la base de datos, es necesario definir explícitamente qué tipo de relaciones pueden estar presentes en ellos para poder expresar fielmente sus propiedades espaciales.

La ontología se define sobre la base de dos tipos de conceptos ("*terminales*" y "*no terminales*") y dos tipos de relaciones ("*tiene*" y "*es un*") (Torres *et al.*, 2005)

Conceptos:

- *Conceptos Terminales (CT):* No utilizan otros conceptos para definir su significado (son definidos por "*valores simples*").
- *Conceptos No Terminales (CN):* Conceptos que utilizan otros conceptos (terminales o no terminales) en sus definiciones.

Relaciones:

- *es_un*: Define que un objeto A puede ser definido, en un nivel mayor de abstracción como un objeto B, por ejemplo:
  - Un *Cambisol* es_un tipo de *Suelo.*
- *tiene*: Define que un objeto A tiene un objeto B, por ejemplo:
  - Un tipo de Suelo *tiene Profundidad Efectiva*.

Los conceptos y relaciones anteriores se definen en base a la semántica subyacente en el esquema lógico de la base de datos. Para el modelo relacional en la base de datos cada relación "R_i" corresponde a una tabla "T" y cada columna en "T" es llamada atributo. Basado en esto, en Li *et al.*(2005)se elaboran normas para la generación de ontologías, con la siguiente estructura (1)

$$O = \{C, R, H^c, rel, A\} \quad (1)$$

Donde:

- *C:* Conjunto finito de conceptos
- *R*: Conjunto finito de relaciones
- H^c: Jerarquía o taxonomía de conceptos
- *rel*: Relaciones no-taxonómicas entre conceptos; rel(R_1) = (C_1, C_2) especifica que C_1 y C_2 tienen la relación R_1
- *A*: Conjunto de axiomas.

Dependencias tales como la equivalencia o la inclusión se generan a partir de las relaciones entre los atributos. La generación de clases está enfocada en la integración de la información de una o más relaciones en una clase ontológica (*ontClass*), siempre y cuando estas relaciones se utilicen para describir una entidad. Las propiedades (*Object Property* y *Data Type Property*) son generadas a partir de las relaciones binarias existentes en la base de datos. Dado que el modelo relacional puede contener relaciones no-binarias estas tienen que ser descompuestas en binarias tomando en cuenta que el lenguaje de ontologías *Ontology Web Language (OWL)*[2] no soporta el tipo no-binario. Las relaciones jerárquicas son generadas a partir de las relaciones de herencia presentes en la base de datos. Las relaciones de cardinalidad son construidas a partir de las restricciones de los atributos en la relación. Las instancias para una clase C_i, son las tuplas de la tabla asociada a ella y las relaciones entre instancias son establecidas utilizando la información en las llaves foráneas de las tablas en la base de datos. Estas reglas son siempre generadas con la restricción de que la base de datos debe estar normalizada en al menos tercera forma normal (*3NF*).

[2] OWL: Lenguaje de marcado para publicar y compartir datos usando ontologías.

REGLAS DE EXTRACCIÓN DE LOS RASGOS ESPACIALES

Como se ha comentado anteriormente en el trabajo de Baglioni y colaboradores (Baglioni *et al*., 2007), fueron añadidas otras reglas redefiniéndose la ontología de la siguiente forma, véase expresión (2)

$$O = \{C, R, H^c, rel, A\} \hookrightarrow O = \{C', R', H^c, rel, A\} \quad (2)$$

Ellos definen:

- *C´*: Contiene los conceptos espaciales que describen a las entidades geográficas (*GeographicObjects, GeoRef, Point, Line, Polygon*).
- *R´*: Contiene las relaciones:
  - *has_georef (Geographic Objects, GeoRef):*Donde A*has_georef* B muestra que A tiene la localización de B.
  - *is_a (GeographicObjects, GeographicObjects):* Donde A *is_a* B muestra que si B *has_georef* C y A *has_georef* C, entonces A tiene la localización de B.
  - *has_geometry (GeographicObjects, Geometry):* Donde A *has_geometry* B muestra que el objeto A tiene la geometría de B, donde B = {*Point, Line, Polygon*}.

Además:

- *Geometry:* Tipos de datos que contienen la geometría del objeto (*línea, punto, polígono*) y sus coordenadas (*X, Y*), las que son almacenadas en el atributo *the_geom.*
- *ObjectLocation:* Posición del objeto respecto a un sistema de coordenadas.
- *Geometric_Information:* La geometría que representa el objeto (*línea, punto, polígono*).

Las reglas introducidas por ellos están basadas en las localizaciones directas e indirectas. Si el objeto contiene el atributo*the_geom*, entonces es llamado *Localización Directa*, de otra manera si no lo contiene pero esta implícitamente contenido en sus atributos temáticos, es llamada *Localización Indirecta.* Aunque estos últimos toman en cuenta la geometría y la posición espacial de los objetos, no es posible modelar las relaciones espaciales y topológicas que están presentes en ellos. La extracción y modelación de estas relaciones en la ontología de aplicación permitirían además, explotar la semántica subyacente en este tipo de relaciones, las cuales son muy propias de los datos espaciales, incrementando así la calidad de los análisis en el proceso de toma de decisiones.

Estas relaciones pueden ser extraídas sobre la base de las relaciones topológicas existentes entre las diferentes entidades geográficas almacenadas en la base de datos. Para ello proponemos el empleo de las relaciones topológicas definidas por Egenhofer (Egenhofer, 1993) para definir los tipos de relaciones topológicas que se van a extraer entre los diferentes objetos espaciales contenidos en la base de datos.

Si tenemos que:

- *A, B*: Regiones de las entidades A´ y B´ respectivamente.
- *R(A, B)*: Relaciones entre las regiones A y B.
  - A^0: Interior del objeto A
  - ∂A: Contorno del objeto A
  - A^-: Espacio exterior al objeto A.
  - B^0: Interior del objeto B
  - ∂B: Contorno del objeto B
  - B^-: Espacio exterior al objeto B.

Entonces podemos modelar las relaciones entre Ay B como (3)

$$R(A,B)=\begin{pmatrix} A^{\circ}\cap B^{\circ} & A^{\circ}\cap \partial B & A^{\circ}\cap B^{-} \\ \partial A\cap B^{\circ} & \partial A\cap \partial B & \partial A\cap B^{-} \\ A^{-}\cap B^{\circ} & A^{-}\cap \partial B & A^{-}\cap B^{-} \end{pmatrix} \tag{3}$$

Donde cada una de estas 9 intersecciones pueden ser caracterizadas por un valor nulo (Φ) y uno no-nulo (¬ Φ) como se muestra en la expresión (4), de la misma forma es posible obtener las relaciones topológicas entre las dos regiones (A y B), véase la Fig. 1.

$$R(A,B)=\begin{pmatrix} A^{\circ}\cap B^{\circ}=\{\phi,\neg\phi\} & A^{\circ}\cap \partial B=\{\phi,\neg\phi\} & A^{\circ}\cap B^{-}=\{\phi,\neg\phi\} \\ \partial A\cap B^{\circ}=\{\phi,\neg\phi\} & \partial A\cap \partial B=\{\phi,\neg\phi\} & \partial A\cap B^{-}=\{\phi,\neg\phi\} \\ A^{-}\cap B^{\circ}=\{\phi,\neg\phi\} & A^{-}\cap \partial B=\{\phi,\neg\phi\} & A^{-}\cap B^{-}=\{\phi,\neg\phi\} \end{pmatrix} \tag{4}$$

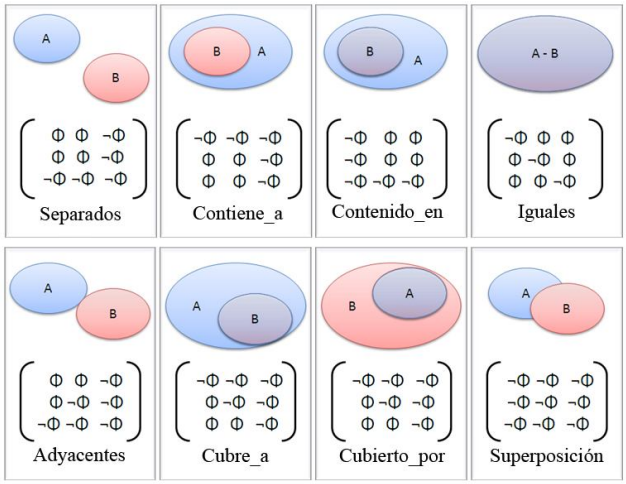


Fig.1. Posibles intersecciones entre las regiones A y B

Sobre la base de los tipos de intersecciones mostrados en la Fig. 1, podemos determinar las reglas para la generación de relaciones del tipo *Separados, Contiene_a, Contenido_en, Iguales, Adyacentes, Cubre_a, Cubierto_por, Superposición.*

A partir de las nuevas relaciones definidas es posible plantear la modificación del esquema de la ontología planteado en la literatura de la siguiente forma (5):

$$O=\{C,R,H^{c},rel,A\} \hookrightarrow O=\{C',R',H^{c},rel,A\} \hookrightarrow O=\{C'',R'',H^{c},rel',A\} \tag{5}$$

Definimos:

- *R´´*:Contendrá además las relaciones topológicas definidas anteriormente: *Contiene_a, Contenido_en, Iguales, Adyacentes, Cubre_a, Cubierto_por, Superposición*existentes entre instancias.

Estructura de la ontología de aplicación generada

En la ontología de aplicación van a estar definidos los términos y las relaciones resultantes del procesamiento de la base de datos, pero estos términos han de representar además las abstracciones de los datos contenidos en la base de datos. Estas abstracciones van a contener las principales componentes de los datos geográficos (*Temática, Espacial y Temporal*) y van a estar definidas como conceptos terminales dentro de la ontología. La estructura de la Ontología generada se define a través de la representación de nodos de representación de datos (NR). Estos NR modelan los conceptos presentes en la Ontología y además las características temáticas y espaciales para cada dato en particular. Esta modelación se especifica en dos tipos NR para representar solo conceptos no terminales (NRC) y los NR para representar las entidades o instancias almacenadas en la base de datos (NRE), véase 2. Esta especificación es debido a que los conceptos terminales constituyen entidades geográficas las cuales presentas características que no están presentes en los conceptos no terminales, (como las presentes en la componente espacial).

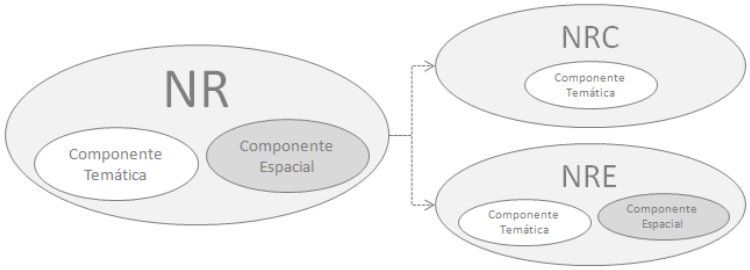


Fig. 2. Nodo de Representación para Clases e Instancias

En la representación de nodos (NR) son descritas todas las características de las clases y las entidades geográficas, entonces otras relaciones como las topológicas o las jerárquicas presentes entre ellas son expresadas en las relaciones existentes entre los diferentes NR (Fig. 3).

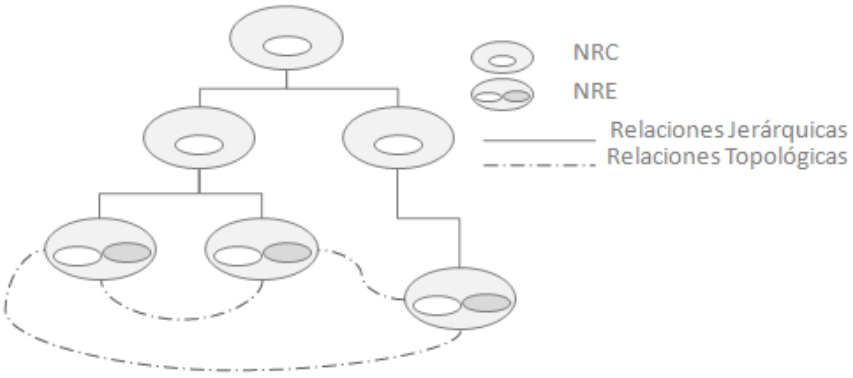


Fig. 3. Representación de los NR y sus tipos de relaciones en la ontología de aplicación

De esta forma la estructura de la ontología de aplicación que se genera automáticamente contendrá además, las relaciones topológicas entre los diferentes objetos geográficos que integran la aplicación lo que permitirá realizar análisis más eficaces al tomar en cuenta estas características.

RESULTADOS DE LA APLICACIÓN DEL MÉTODO PROPUESTO A UN CASO DE ESTUDIO EN LA EVALUACIÓN DE RIESGOS

Las instalaciones generadoras de energía (campos eólicos, paneles solares, mini-hidroeléctricas, etc.), debido a sus características están principalmente localizadas en regiones montañosas, costas y zonas de difícil acceso. Estas

regiones están caracterizadas por diferentes niveles de complejidad en términos de la influencia de fenómenos naturales que pueden afectar su integridad, elementos que tienen que ser tomados en cuenta cuando se realiza la estimación de su factibilidad y costos de instalación.

El objetivo de esta aplicación es realizar una estimación de los niveles de riesgo que los fenómenos atmosféricos representan para nuevas instalaciones eólicas mediante la generación automática de una ontología de aplicación en un SIG de código abierto que permita una evaluación certera y rápida para la localización más segura de este tipo de costosas instalaciones y como combinar diferentes soluciones tecnológicas para mitigar los riesgos de acuerdo a su distribución espacial.

La aplicación se desarrolló en el territorio de la Isla de la Juventud (Fig. 4) tomando en cuenta las estimaciones de su potencial eólico (Garea *et al.*, 2003). Para ello se integraron al SIG una serie de mapas con sus correspondientes bases de datos en las temáticas de *clima, relieve, infraestructura energética y vial*, entre otros que sirven de apoyo al proceso de toma de decisiones.

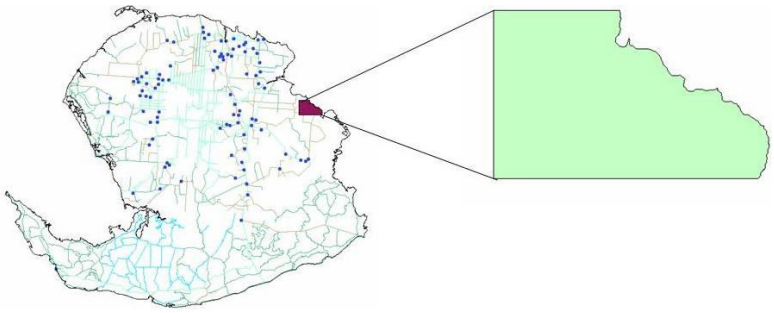


Fig. 4. Región en la que se realizó la evaluación de los riesgos

El análisis se basó en los criterios especializados de expertos y las características técnicas de las instalaciones. Un ejemplo de criterios y clasificación de las áreas analizadas para la instalación de un parque eólico se presenta en la Tabla I. Estos criterios fueron conceptualizados en una ontología de dominio que sirvió de base para el proceso de enriquecimiento de la ontología de aplicación generada.

Tabla I. Criterios de experto para la evaluación de riesgos

Criterios					**Evaluación del riesgo**
Tormenta tropical y huracanes	Velocidad del viento Km/h	Altura del terreno (m)	lluvias (mm/h)	temperatura (C^0)	
Huracán, Categ. 4 y 5	>100	>0	> 600	>30	Muy Alto
Huracán, Categ. 1 a3	>100	200-400	400-600	> 30	Alto
HuracánCateg. 1 a 3 y Tormenta tropical	30-99	>0	> 200	>25	Medio
Depresión tropical	< 30	>0	0- 600	0- 25	Bajo

Generación automática del conocimiento

Como primer paso para la validación del método propuesto se realizó la generación automática de la ontología de aplicación mediante el módulo de extracción de reglas al cual se le habían incorporado las nuevas reglas propuestas, la 5 muestra un esquema de la ontología obtenida. Esta ontología contiene instancias (regiones) de las clases que determinan si una región es apta para la instalación de parques eólicos debido a su bajo riesgo en términos de la influencia de los fenómenos ambiente natural o no.

Para la realización de los análisis fue necesaria además la implementación de un módulo que permite la selección de los conceptos en la ontología de dominio y la visualización de los resultados. Para su implementación se utilizó la librería de código abierto JENA[3].

Una vez generada la ontología de aplicación y enriquecida por la ontología de dominio es posible a través del módulo implementado la navegación y selección de las áreas con diferentes categorías de riesgo a partir de que al estar representadas las relaciones espaciales y topológicas en los nodos de representación de instancias (NRE) las áreas de cada categoría de riesgo son determinadas por las regiones o instancias con los valores de atributos presentes en los criterios de expertos y sus relaciones topológicas como *Contiene_a, Contenido_en, Iguales, Cubre_a,* y *Cubierto_por.* De manera más formal puede ser expresado de la siguiente forma.

Teniendo las siguientes funciones:

- getReg (R_i): Función que obtiene las instancias (regiones) que tienen la relación R_i.
- getEnt (I_i): Función que obtiene las entidades geográficas que pertenecen a la instancia (región) I_i.
- P = {*Tormenta tropical y Huracanes, Velocidad del viento, Altura, Temperatura, Lluvia*}.
- R = {*Contiene_a, Contenido_en, Iguales, Cubre_a, Cubierto_por*}.
- n: número máximo de relaciones.
- m: número máximo de instancias (regiones) de clases.

Asumimos que:

$$I_1=\text{getReg}(R_1), I_2=\text{getReg}(R_2),\ldots, \text{in}=\text{getReg}(R_n)$$

Las áreas de categorías de riesgo son $reg = (I_1 \cap I_2 \cap,...,\cap I_n)$ siempre que:

- $R_i \in R : i = 1..n$
- $getEnt(I_j) \in P : j = 1...m$
- $\bigcup_{j=1}^{m} \{getEnt(I_j)\} = P$

Para comprobar la efectividad del método realizamos diferentes operaciones de búsqueda y análisis de la información espacial integrada en la aplicación, mediante la navegación por la ontología generada y comparamos los resultados con los obtenidos utilizando los métodos típicos de superposición de mapas y recuperación basada en el lenguaje SQL, los mismos demuestran que la incorporación de semántica al SIG mediante el método propuesto aumenta de forma sustancial la calidad de los análisis pues evita los posibles errores en las operaciones de análisis que debe realizar el especialista al estar los mismos conceptualizados en la ontología de dominio y la ontología de aplicación generada de forma automática al mismo tiempo se logra ahorro de tiempo de realización de los análisis espaciales

[3] http://jena.sourceforge.net

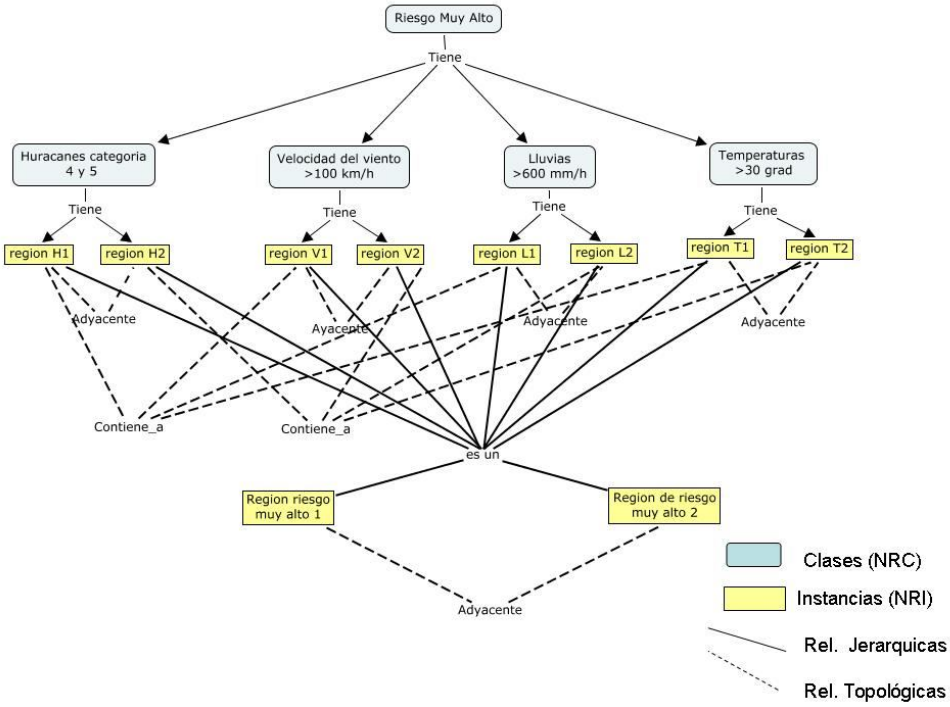


Fig. 5. Esquema parcial de la ontología de aplicación generada

Estos resultados demuestran la efectividad del método propuesto tomando en cuenta de que el uso de la ontología generada y enriquecida con la ontología de dominio como una capa intermedia entre el sistema de recuperación y la base de datos espacial permite el manejo de los datos a través de su representación conceptual (abstracción).

Tabla I. Parámetros fisicoquímicos y métodos de análisis empleados

Operaciones realizadas	Tiempo consumido por los métodos de recuperación y análisis clásicos (seg.)	Tiempo consumido por el método propuesto (seg.)
Recuperación de regiones con valores de lluvias altos.	53	0,122
Recuperación de regiones con valores de lluvia altos superpuestas a regiones con Alturas del terreno mayores de 400 m	64	0,125
Recuperación de regiones con valores de lluvias altos cubiertas por a regiones con Alturas mayores de 400m y regiones afectadas por huracanes de alta intensidad	74	0,140
Recuperación de zonas con bajo nivel de riesgo para la ubicación de parques eólicos combinando todos los criterios de expertos.	72	0,285

CONCLUSIONES Y TRABAJOS FUTUROS

En este trabajo presentamos un método para la generación automática de una ontología de aplicación en la que se agregan las relaciones topológicas existentes en los datos (instancias). El método fue aplicado en un caso de uso para la estimación de riesgos y los resultados obtenidos demuestran su efectividad.

El trabajo realizado muestra como haciendo uso de las potencialidades que ofrece el uso de Ontologías como mecanismo de representación del conocimiento para el procesamiento de los datos geoespaciales y en especial la inclusión de las relaciones topológicas existentes en estos datos es posible aumentar la eficacia y la eficiencia en las diferentes actividades que involucran este tipo de información ya que por un lado se habilitan a los sistemas de herramientas semánticas que les da la posibilidad de no solo procesar los datos sino también comprenderlos y por el otro lado se hace mucho más flexible el trabajo ya que este puede ser realizado además por aquellos usuarios no especialistas.

Aunque queda mucho por hacer todavía el procesamiento semántico de la información tiende indudablemente a mejorar las tareas de análisis y/o toma de decisiones y a disminuir la brecha semántica existente entre Humanos y Máquinas.

Como trabajos futuros continuaremos el desarrollo de nuevas reglas de generación automática de ontologías mediante la representación semántica de todos los rasgos de bajo nivel que modelan el significado de un objeto o entidad geográfica.

REFERENCIAS

An, Y., A. Borgida y J. Mylopoulos (2006): *Discovering the semantics of relational tables through mappings.* Paper presented at the LNCS 4244 - Journal on Data Semantics VII.

Baglioni, M., M. V. Masserotti, Ch. Renso y L. Spinsanti (2007): Building geospatial ontologies from geographical databases. *LNCS 4853, Proceedings of the Second International Conference on Geospatial Semantics, Mexico City*.

Bishr, Y. (1998): Overcoming the semantic and other barriers to gis interoperability. *International Journal of Geographical information Science. Sepecial Issue on Interoperatibility in GIS, Vol. 12* 299-314

Egenhofer, M. J. (1993): A model for detailed binary topological relationships. *Geomatica Vo. 47, No. 3 y 4.*

Fonseca, F. y M. Egenhofer (1999): Ontology-driven geographic information systems. *Proc. 7th ACM Symposium on Advances in Geographic Information Systems*

Garea, B, A. Curbelo, O. Jiménez, E. Garea y G. Martin (2003): Swera. Cuba progress report. *Informe de los resultados del proyecto Internacional SWERA en Cuba.*

Harvey, F, W. Kuhn, H. Pundt, Y. Bishr y C. Riedemann (1999): Semantic interoperability: A central issue for sharing geographic information. *Annals of Regional Science.*

Klien, E., U. Einspanier, M. M. Lutz y S. Hübner (2004): An architecture for ontology-based discovery and retrieval of geographic information. *Toppen, F. & P. Prastacos (Eds), 7th Conf on Geographic Information Science (AGILE).*

---------- y M. Lutz (2005): The Role of Spatial Relations in Automating the Semantic Annotation of Geodata. In: Cohn, A. and Mark, D (eds.). Proceedings of the Conference of Spatial Information Theory (COSIT'05), Ellicottville, NY, USA. Lecture Notes in Computer Science, Vol. 3693, pp. 133-148.

Larin-Fonseca, R. y E. Garea-Llano (2009): Combination of spatial analysis methods and semantics extraction in a sdi environment for agro-ecological zonation. *Proceedings of The 12th AGILE International Conference on GIScience en Hannover, Germany.*

---------- y ---------- (2010): Representación semántica automática de datos geoespaciales. *Segundo Taller Cubano Eureka 2010. Red Iberoamericana de Descubrimiento de Conocimientos (EUREKA IBEROAMERICA)*

Leung, Y. (2010): Knowledge discovery in spatial data Manfred M. Fischer, Geoffrey J.D. Hewings, Peter Nijkamp y Folke Snickars (Eds.), doi:10.1007/978-3-642-02664-5

Li, M., D. Xiao-Young, y S. Wang (2005): Learning ontology from relational database. Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou.

Lutz, M. y E. Klien (2008): Overcoming differences of meaning during the discoverty and retrieval of geospatial information in spatial data infrastructures. Computers & Geosciences, Special Issue on Geoscience Knowledge Representation in Cyberinfrastructure, forthcoming.

Raubal, M. (2005): Mappings for cognitive semantic interoperability. *Toppen and M. Painho (Eds.), AGILE 2005 - 8th Conference on Geographic Information Science, IGP, Lisboa, Portugal.*