

REPORTE TÉCNICO
**Reconocimiento
de Patrones**

**Métodos de recuperación semántica
de datos espaciales en
formato vectorial**

Ing. Alexis Martínez Terrero,
Dr. C. Eduardo Garea Llano

RT_034

octubre 2010





CENATAV

Centro de Aplicaciones de
Tecnologías de Avanzada
MINISTERIO DE LA INDUSTRIA BÁSICA

RNPS No. 2142
ISSN 2072-6287
Versión Digital

SERIE AZUL

REPORTE TÉCNICO
**Reconocimiento
de Patrones**

**Métodos de recuperación semántica
de datos espaciales en
formato vectorial**

Ing. Alexis Martínez Terrero,
Dr. C. Eduardo Garea Llano

RT_034

octubre 2010



Métodos de recuperación semántica de datos espaciales en formato vectorial

Ing. Alexis Martínez Terrero, Dr. C. Eduardo Garea Llano

Centro de Aplicaciones de Tecnología de Avanzada
7a #21812 e/ 218 y 222, Rpto. Siboney, Playa. C.P. 12200, Ciudad de la Habana, Cuba.
amartinez@cenatav.co.cu

RT_034 CENATAV

Fecha del camera ready: 30 de junio de 2010

Resumen: Desde las últimas décadas la información espacial ha venido creciendo en volumen, diversidad, complejidad e importancia. En consecuencia, los Sistemas de Información Geográfica (SIG) necesitan contar con nuevos procedimientos para acceder a sus datos, en aras de poder identificar y procesar de manera eficaz la mayor cantidad posible de información contenida en esos recursos. Conceptos tales como ontología o semántica, apuntan a convertirse en las soluciones más solventes en esta dirección a la vez que constituyen un nuevo enfoque en la representación y recuperación de la información geo-espacial. El presente reporte pretende incluir los distintos métodos a través de los cuales los SIG han venido recuperando su información, llegando hasta las propuestas que involucran semántica y que constituyen la vanguardia actual, centrándose en el modelo de datos vectoriales donde se pretenden obtener nuevos métodos semánticos que permitan aprovechar al máximo los beneficios de este modelo extendiendo cada vez más sus potencialidades.

Palabras clave: métodos de recuperación, datos espaciales, semántica, modelo de datos vectorial

Abstract: Since the last decades spatial information has been growing in volume, diversity, complexity and importance. Consequently, Geographic Information Systems (GIS) need to have new procedures to access data, in order to be able to identify and efficiently process the largest possible amount of information contained in these resources. Concepts such as ontology or semantics will become more solvent solutions in this direction and at the same time they represent new approaches to the representation and retrieval of the geospatial information. This report seeks to include the various methods by means of which GIS have been recovering their data, reaching proposals that involve semantics. At present they are the most up-to-date methods, focusing on the vector data model which is aimed at acquiring new semantic methods that allow maximizing the benefits of this model, thus increasingly expanding its potential.

Keyword: Retrieval's methods, Spatial Data, Semantic, Vector Data Model

1. Introducción

Los últimos años han sido testigos de una rápida evolución de los sistemas de información geográfica (SIG), debida entre muchas razones, al aumento también progresivo de la posibilidad de ingresar cada vez más datos y permitir análisis más provechosos en su procesamiento.

El nivel de complejidad actual que poseen estos datos posibilitaría obtener y gestionar conocimiento en estas aplicaciones, pero aún los SIG están carentes de contar con métodos efectivos que viabilicen un aprovechamiento óptimo de la información contenida en estos recursos, o dicho en otras palabras, no son capaces de explotar toda su capacidad analítica. En la mayoría de los casos, tras visualizar los resultados de una consulta espacial es al especialista encargado de operar el SIG a quien corresponde la tarea de realizar el “razonamiento” deseado.

Las causas que aseveran esto son diversas, pero sin dudas un primer e importante obstáculo lo constituye la falta de interoperabilidad entre los servicios y datos que hacen ineficientes las posibilidades reales de análisis. Aparejado a esto otras cuestiones significativas también aportan, como la estructuración de las consultas, donde es necesario un conocimiento previo por parte del usuario para formular esta y donde el procesamiento sintáctico de las mismas hacen que la recuperación dentro del SIG no sea del todo eficiente que pudiera llegar a ser, limitándose en gran medida la resolución de problemas de análisis que hoy no se hacen.

Las técnicas convencionales de procesamiento sintáctico han demostrado no ser capaces de resolver estos problemas a pesar que han sido muy provechosas para la solución de otras dificultades, dando paso a un nuevo enfoque que realiza el procesamiento basado en el significado y no en la sintaxis. La disciplina de la Interpretación Semántica de datos espaciales comienza a surgir con esta necesidad imperiosa de lograr una interoperabilidad e integración de los datos geoespaciales [1].

La inclusión de semántica en los SIG posibilita relacionar el contenido y la representación de la información a diferentes niveles con las entidades o conceptos de un contexto dado. El contexto en este sentido se entiende como un espacio dentro del dominio en el cual la semántica es dependiente del entorno. En el caso del dominio geográfico sabemos que es bien amplio, por lo que a veces resulta más útil para el trabajo referirnos solo a una porción de este para detallar mejor ciertos escenarios. Por lo general se describen con más granularidad y permiten orientarnos mejor dentro de determinados aspectos sin tener en cuenta todo el espacio. Así podemos referirnos al contexto de las playas, al contexto ciudadano o al contexto de Venezuela. Veamos el siguiente ejemplo:

Si tuviéramos información referente a la profundidad de todos los ríos del mundo, una consulta en lenguaje natural que pregunte por los ríos más profundos del planeta pudiera resolverse devolviendo los resultados de tal vez los diez ríos que tengan valores más altos en su atributo profundidad. En este caso encuestamos al dominio completo, pero ¿qué sucede si solo encuestamos una parte?

Pensemos ahora en otras dos consultas en lenguaje natural: Una solicitando los ríos más profundos de Cuba y otra solicitando los ríos más profundos de Brasil. Obviamente en lo referente a los ríos, los “profundos” de Cuba no son iguales a los “profundos” de Brasil. Si la solución de esta consulta implica una cota de profundidad a partir de la cual consideramos a un río profundo, probablemente Cuba no tendría ningún resultado que mostrar si la medida fuese brasileña, siendo esto a los efectos de la intención del usuario erróneo.

Por tanto podemos ver como el contexto influye en los valores que pueden tomar las variables a la hora de realizar una consulta., posibilitando análisis más profundos y diversos de los fenómenos asociados a este. La incorporación de semántica a la gestión de la información geográfica es una necesidad para su mejor explotación, así como para vencer los problemas derivados de la heterogeneidad y la poca interoperabilidad existente [2]. En sistemas de información, la semántica relaciona el contenido y representación de la información a entidades o conceptos en el mundo haciendo posible analizar la interacción de diversos fenómenos de un área común, además de poderlos representar adecuadamente a diferentes niveles de detalle de conocimiento.

Los esfuerzos recientes por obtener modelos que permitan resolver estos problemas apuntan hacia una nueva generación de SIG, en la que la información podrá ser procesada automáticamente por estos sistemas, permitiendo resolver consultas aún mucho más complejas que las que hasta hoy procesan. Los SIGGO (Sistemas de Información Geográfica Gobernados por Ontologías) y el desarrollo de ontologías, se presentan como el instrumento adecuado para alcanzar la integración semántica en el entorno de las infraestructuras de datos espaciales, ambiente complejo donde constituirán valiosas herramientas como parte de las redes socio-

técnicas que conforman estas [26], volviéndose entonces muy necesario contar con métodos que permitan hacer la recuperación semánticamente para explotar estas ventajas eficientemente.

Por otro lado, los modelos más recurridos a la hora de representar información espacial son el vectorial y el raster. Nuestra intención es concentrarnos en uno de ellos (vectorial) y explotar sus potencialidades en aras de formar parte de esta solución general, contando con mecanismos lo suficientemente eficientes como para inferir análisis basados en criterios recuperados desde representaciones ontológicas.

2. El modelo de datos vectoriales

En la actualidad existe una diversidad de modelos que representan las propiedades geométricas y las relaciones topológicas con que se caracterizan los datos espaciales. El modelo vectorial es junto al raster, uno de los modelos más utilizados por los sistemas de información geográfica actualmente debido a que brinda buen soporte para representar los fenómenos del mundo geográfico. Su utilización se basa sobre la estructura de representación en correspondencia con las características del tipo de objeto. En el caso vectorial, permitiendo representar más adecuadamente a los objetos que poseen una geometría definida, o sea, a los objetos asociados a fenómenos discretos que puedan ser descritos mediante puntos, líneas o polígonos, mientras que el modelo raster corresponde a objetos que pueden representarse mediante rejillas de valores temáticos, o sea, se asocia a los objetos que ostentan una continuidad espacial como pueden ser los modelos digitales del terreno. A pesar que se menciona el formato raster por su alto grado de utilización en nuestros días no es objetivo de nuestro estudio hacer una comparación entre ambos modelos, sino concentrar la atención en el modelo vectorial.

En este modelo se representan las características geográficas de los objetos a través de datos vectoriales de forma parecida a como lo hacen los mapas, estableciendo que cada tipo de representación geométrica puede estar definida en dos o tres dimensiones y en un sistema de coordenadas determinado. Los vectores posibilitan el almacenamiento explícito de la localización de los nodos y la topología definiendo a las entidades solamente por sus límites y los arcos conectados, pero solo recopila aquellos puntos que delimitan a las entidades y no considera el espacio fuera de éstas. Al almacenar los datos sólo de los elementos digitalizados requiere menos memoria para su almacenamiento y posterior tratamiento.

Los elementos básicos que lo componen son los puntos, líneas o arcos y los polígonos, como muestra la Figura 1. A continuación se presentan una serie de conceptos afines: [3 citando a 4].

Coordenada. Pares de números que expresan las distancias horizontales a lo largo de ejes ortogonales, o tríos de números que miden distancias horizontales y verticales, o n-números a lo largo de n-ejes que expresan una localización concreta en el espacio n-dimensional. Las coordenadas generalmente representan localizaciones de la superficie terrestre relativas a otras localizaciones.

Punto. Abstracción de un objeto de cero dimensiones representado por un par de coordenadas X, Y. Normalmente un punto representa una entidad geográfica demasiado pequeña para ser representada como una línea o como una superficie; por ejemplo, la localización de un edificio en una escala de mapa pequeña, o la localización de un área a la que una instalación da servicio en una escala de mapa media.

Línea. Conjunto de pares de coordenadas ordenados que representan la forma de entidades geográficas demasiado finas para ser visualizadas como superficies a la escala dada (curvas de nivel, ejes de calles, o ríos), o entidades lineales sin área (límites administrativos). Una línea es sinónimo de arco.

Polígono. Entidad utilizada para representar superficies. Un polígono se define por las líneas que forman su contorno y por un punto interno que lo identifica. Los polígonos tienen atributos que describen al elemento geográfico que representan.

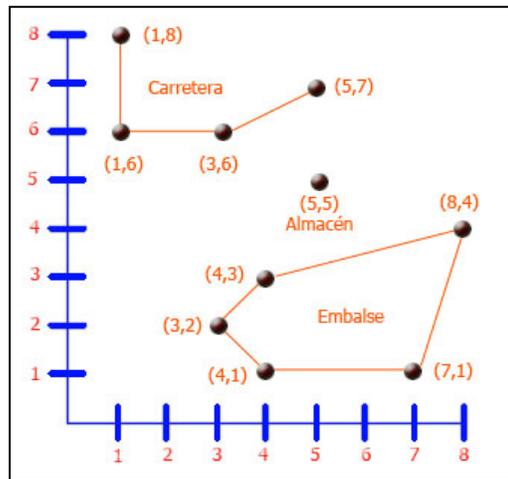


Fig. 1. Representación de entidades geográficas mediante el modelo vectorial

2.1 Estructura de datos vectoriales

La estructura de los datos vectoriales es compacta y compleja, pero por otro lado permite una codificación eficiente de la topología y de las operaciones espaciales. A continuación se presentan las diferentes estructuras con las principales características de cada una de ellas [3, 5].

- Lista de coordenadas "espagueti"
- Diccionario de vértices
- Ficheros DIME ("Dual Independent Map Encoding")
- Arco / nodo

2.1.1 Lista de coordenadas "espagueti"

Comienza anotando según la notación (identificador del objeto, número de vértices) y luego prosigue con la lista de las coordenadas como se aprecia en la Figura 2. Sus principales ventajas radican en que no es una estructura compleja y en su facilidad de manejar. En el caso de la representación gráfica de los polígonos, se repite el primer vértice para indicar que es una figura cerrada. Es muy recurrida por la cartografía automática.

Sus limitantes vienen dadas por que contiene muchas operaciones redundantes que implican gastos de almacenamiento mayores, además no representa de modo explícito las relaciones espaciales entre los distintos elementos y existe el riesgo de que un vértice sea registrado de forma diferente en cada segmento recto al que pertenezca [6].

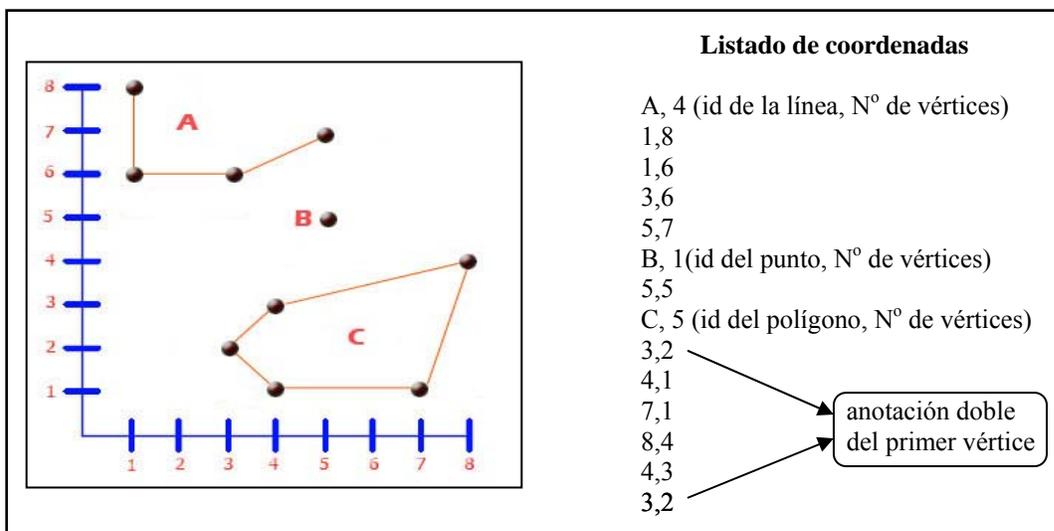


Fig. 2. Representación de la estructura de datos lista de coordenadas

2.1.2 Diccionario de vértices

Se registran una sola vez las coordenadas de cada vértice y se crea un diccionario de los vértices que constituyen cada objeto espacial identificable en el mapa como se observa en la Figura 3. Previamente se identifican con un nombre o etiqueta. En este caso no existen operaciones redundantes, pero aún así la topología de los objetos geográficos no se reconoce suficientemente.

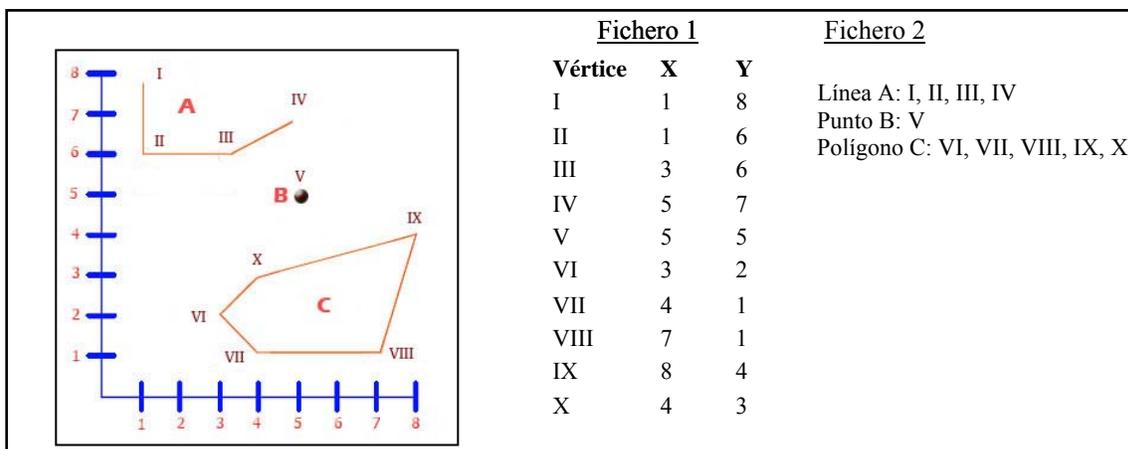


Fig. 3. Representación de la estructura de datos diccionario de vértices

2.1.3 Ficheros DIME ("Dual Independent Map Encoding")

Elaborado por la Oficina del Censo de los Estados Unidos realiza una lista de vértices con nombres y coordenadas de cada objeto, luego los objetos lineales que delimitan polígonos se codifican indicando en que vértices empiezan y terminan, o sea, se asigna un código direccional de la forma "from node" (nodo origen) y "to node" (nodo final) como se aprecia en la Figura 4.

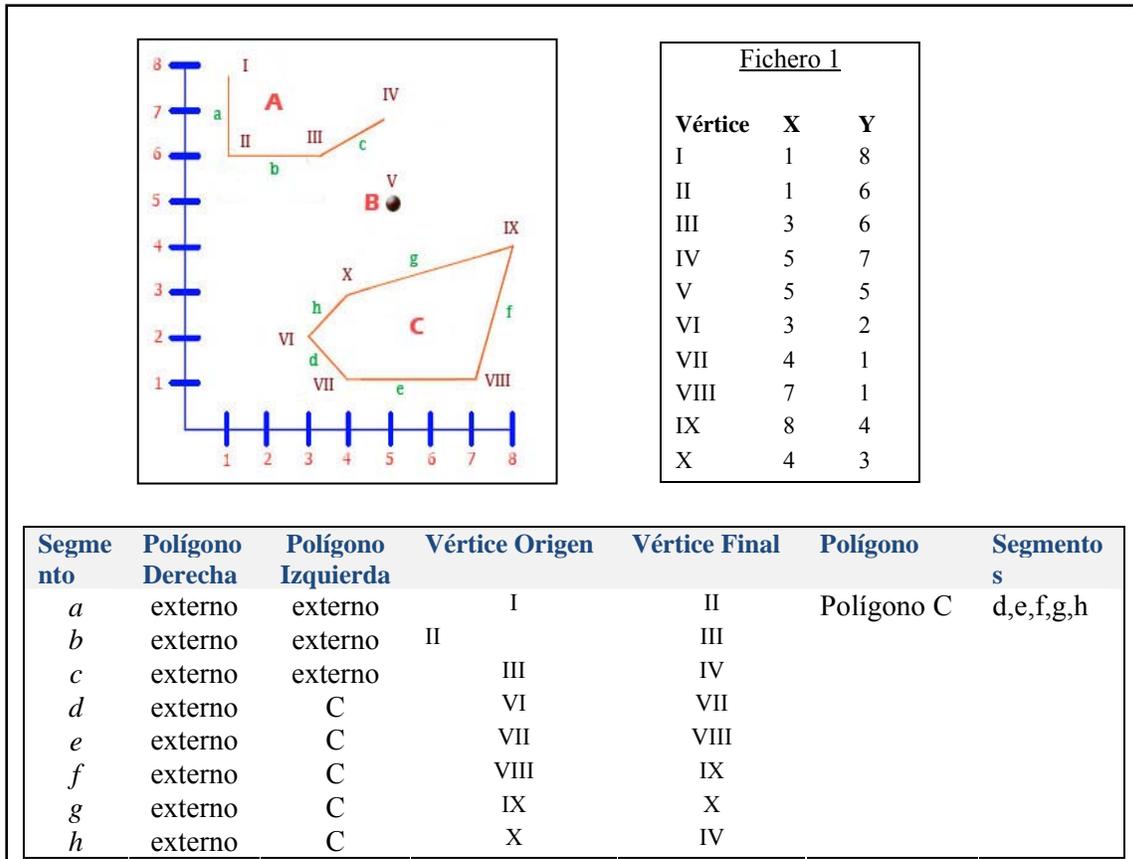


Fig. 4. Representación de la estructura DIME

2.1.4 Estructura arco / nodo

En esta estructura la descripción digital de los aspectos espaciales de los objetos geográficos es más completa, añadiéndose información sobre las relaciones topológicas, complementando así la geometría recogida por las coordenadas. Como lo indica su nombre en la estructura ARCO/NODO el elemento base ya no es el segmento recto sino el arco, mientras que el otro elemento fundamental es el nodo. Se entiende como ARCO al segmento de línea entre dos nodos y como NODO a cada uno de los vértices en que se cruzan tres o más arcos o que es el punto terminal de una línea o arco.[6]. En la Figura 5 se observa como los polígonos se codifican indicando los arcos que les rodean. Podemos apreciar como el arco se registra indicando sus nodos y el polígono de la izquierda y el de la derecha. Entre sus especificaciones tenemos que almacena y referencia a los datos de manera que los nodos forman arcos y los arcos construyen polígonos. Los nodos definen los dos extremos de un arco y pueden o no, conectar dos o más arcos.

Vale destacar que esta estructura se ha convertido en la más característica del mercado para los sistemas vectoriales. Una de sus principales ventajas es que evita repetir las coordenadas para un punto compartido por un número varias líneas, lo que es muy útil a la hora de representar los límites polígonos adyacentes sobre un mapa.

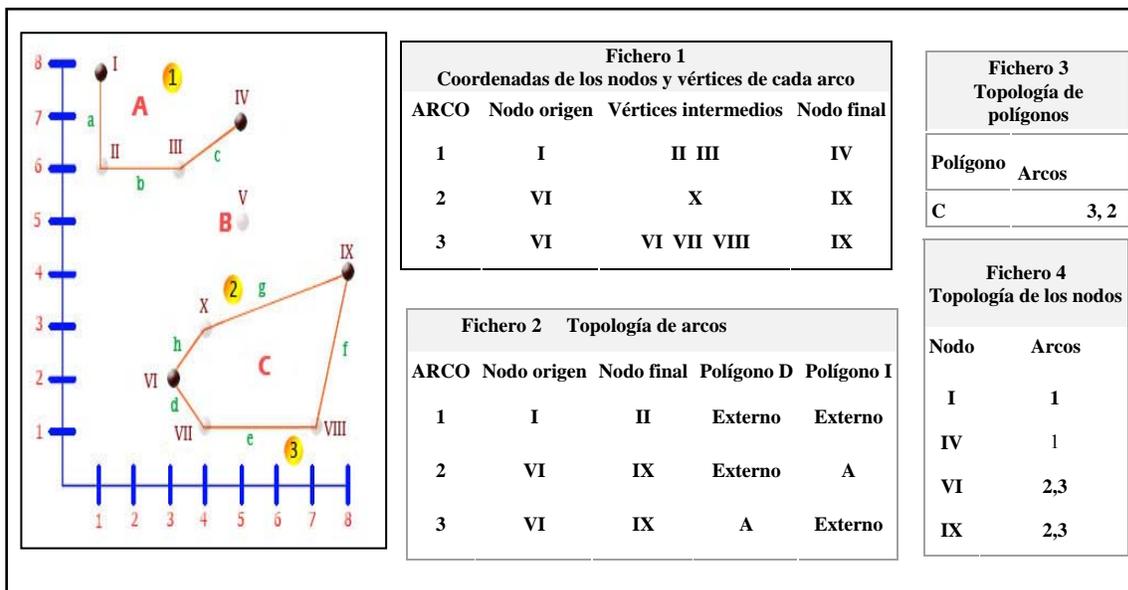


Fig. 5. Representación de la estructura arco/nodo

2.2 Los SIG vectoriales

Los SIG se definen vectoriales precisamente por la representación vectorial de la componente espacial de los datos geográficos [7]. Generalmente implementan una buena salida gráfica ya que los elementos son representados como gráficos vectoriales que no pierden definición si se amplía la escala de visualización.

Normalmente tienen una mayor compatibilidad con entornos de bases de datos relacionales ya que de un lado almacenan los aspectos temáticos de los datos espaciales, pero por otro también tienen que guardar la componente espacial en una base de datos topológica. Respecto a los datos podemos decir que son fáciles de mantener y actualizar, pero cuando su variación es alta reducen la eficacia del SIG.

El punto más importante viene dado por sus funcionalidades ya que permiten una gran capacidad de análisis, sobre todo en redes. Entre las más importantes podemos ver:[8]

Visualización de la información: La representación vectorial es la manera en que se representan los objetos espaciales utilizando vectores definidos por pares de coordenadas relativas a algún sistema cartográfico. Esta tiene varias formas de presentarse según lo que se desee visualizar. En ocasiones los usuarios solo necesitan la parte temática prescindiendo de la componente espacial, por lo que el resultado puede devolverse en valores cuantificados que informen sobre la distribución de la cantidad de objetos geográficos, de acuerdo a algún criterio de clasificación. También es posible la representación cartográfica temática ya que permite la representación simultánea de los aspectos temático y espacial, de hecho también según [8] la presentación de mapas temáticos de estas características es una de las principales contribuciones de los SIG vectoriales.

Reclasificación: Función fundamentalmente analítica resultado de añadir o modificar variables temáticas de las entidades geográficas, provocando un cambio en la clase a la cual pertenece. Las nuevas variables persisten para la elaboración de nuevos mapas temáticos.

Recuperación selectiva: A pesar de que las funciones de recuperación combinan los objetos espaciales con la información temática, solo es esta última es la que puede ser modificada. Como

su nombre sugiere la recuperación selectiva o filtrada permite buscar (en la base de datos, mediante selección espacial o combinación de ambos) de acuerdo a un criterio de clasificación.

Superposición de capas: Constituye una opción muy útil para realizar análisis espaciales porque permite combinar diferentes capas de información, de hecho es una de las funciones más importantes porque su aplicación envuelve a los elementos espaciales del sistema de tal manera que un resultado puede ser la obtención nuevos objetos que se resulten de los existentes.

Incluye análisis como el de *puntos en polígonos* (encuentra si los puntos de una determinada capa están incluidos dentro un polígono perteneciente a otra capa), el análisis de *líneas en polígonos* (igual que el análisis de puntos en polígonos, pero buscando objetos lineales) y el de *superposición de polígonos* (supone la generación de una capa de polígonos de intersección de los preexistentes) Es una de las operaciones de superposición más importantes, ya que supone la identificación de áreas que cumplan dos o varias condiciones a la vez[8].

Análisis de redes: Útil para la solución de problemas prácticos como lo pueden ser de intersección de lugares dentro de un mismo territorio, localización de recursos o encontrar rutas óptimas. Este tipo de análisis es muy común en el modelo vectorial por la gran cantidad de objetos lineales que existen. Además posibilita simular flujos a lo largo de una red lineal y suele ser utilizado comúnmente en la gestión de infraestructuras lineales.

La medición de las magnitudes geométricas: Muchas son calculadas de forma interactiva y ofrecen la posibilidad de conocer los valores de muchas variables como longitud, volumen, superficie, etc., además de que pueden corregir las posibles distorsiones en las proyecciones y el sistema de coordenadas. Luego de su obtención pueden almacenarse permanentemente en una base de datos para futuras operaciones.

Funciones de vecindad: Analizan el área que envuelve a un punto determinado. Permite operaciones tales como la generación automática de polígonos y de cálculo topográfico. En su utilización se deben especificar tres parámetros: la ubicación de referencia, el ámbito de vecindad y la función a realizar.

Agregación de elementos: Permite crear nuevos elementos a partir de la agrupación de elementos espaciales existentes reduciendo su cantidad. Los atributos temáticos de los nuevos objetos se forman teniendo alguna regla lógica y/o matemática y a partir de los valores que tienen los atributos temáticos en los elementos de origen[8].

Como hemos visto anteriormente en los datos vectoriales la utilidad de las representaciones se centra en la precisión de localización de los elementos geográficos sobre el espacio y donde los fenómenos a representar son discretos, es decir, de límites definidos. Este modelo permite además realizar operaciones de búsqueda y recuperación de información a partir de entidades espaciales o de los atributos asociados a ellas y realizar análisis de áreas de influencia con buffers, generar mapas temáticos, tablas y gráficos para una mejor comprensión de la información.

3. Lenguajes de consulta espaciales

En la actualidad las bases de datos espaciales (BDE) han crecido considerablemente en cuanto al volumen y la complejidad de los datos que almacenan. Una consecuencia directa es el alto costo computacional que usualmente traen aparejadas las consultas espaciales, por lo que paralelo al desarrollo de los sistemas de información geográfica ha ido creciendo también el interés por perfeccionar los lenguajes de consultas. Los trabajos siguientes son ejemplos ello:

- **Spatial SQL:** A query and presentation language [9]. En este trabajo se identifican los principales requerimientos que deberían soportar los lenguajes de consulta espaciales y que no poseen los sistemas convencionales. Concebido en dos modelos básicos: Uno un

lenguaje de consultas y dos un lenguaje de presentación. El primero basado en el SQL tradicional mantiene las cláusulas tradicionales SELECT-FROM-WHERE y el segundo nombrado GPL (Graphical Presentation Language) enfatiza en la importancia de brindar a los usuarios la posibilidad de personalizar la forma de presentación de los resultados de la consulta.

- **GeoSQL** [10] basado también en el SQL tradicional y similar al SQL espacial. Fue implementado como un módulo de un SIG prototipo orientado a objetos nombrado YH-GIS.
- **Sql/sda**: Es un lenguaje de consulta con soporte para datos espaciales. Es extendido del SQL espacial tradicional y está basado en OGC SFS4SQL (OGC Simple Feature Specification For SQL) [11].
- Querying Geographical Data Warehouses with **GeoMDQL** [12]. El lenguaje nombrado GeoMDQL es un lenguaje de consulta multidimensional para consultas geográficas, específicamente diseñado para entornos de SOLAP¹ y se basa en estándares bien conocidos, tales como MDX y OGC.
- **GML-QL**: A Spatial Query Language Specification for GML [13] **GML-QL** es una extensión de XQuery² para proporcionar soportes espaciales sobre documentos GML³

Una motivación inicial para el desarrollo de los lenguajes de consulta espaciales lo constituyó la falta de soporte para información espacial que brindaban los lenguajes de consultas convencionales, sin embargo estos ofrecen un gran número de funcionalidades que hacen todo un reto el decidir cuando es necesario extender un lenguaje de consultas o cuando crear uno nuevo.

3.1 Las consultas espaciales

Uno de los aspectos fundamentales a la hora de recuperar eficientemente la información geográfica lo constituye el procesamiento de la consulta. En los SIG se realiza sobre fuentes de datos ya indexadas que realizan este proceso basado en la sintaxis, pero prescinden todavía del procesamiento semántico. Esto requiere que los usuarios deban tener previamente al menos una noción mínima de lo que deseen buscar, para poder acotar la búsqueda lo mejor posible de acuerdo a sus intenciones, porque mientras menos puedan hacerlo, más probable es la obtención de resultados irrelevantes producto de introducir parámetros incorrectos de búsqueda o términos ambiguos.

A veces los términos consulta espacial y consulta geográfica son usados indistintamente para referirse a un mismo aspecto, sin embargo es un punto donde se debería tener cuidado dado que su significado difiere un poco, siendo el término de consulta espacial un tanto más abarcador. En [29,36 citando a 37] se definen ambos términos: Se define consulta espacial como “aquella que involucre relaciones espaciales (intersección, contenido, límites, adyacencia, proximidad) de entidades geoméricamente definidas y localizadas en el espacio sin considerar la naturaleza del sistema de coordenadas”, mientras que las geográficas “asumen que el espacio es delineado por un sistema de coordenadas bien definido del “mundo real”. Más adelante señalan que estas relaciones geográficas impuestas en los sistemas de coordenadas del mundo real son relaciones geométricas, por lo que en un marco geométrico, donde la distancia y la dirección puedan ser medidas en una escala continua muchos tipos de relaciones entre los objetos contenidos en ese espacio pueden ser determinadas a través de la geometría. Las relaciones espaciales sin embargo, pueden ser geométricas o topológicas, y tanto las consultas espaciales como las geográficas combinan ambos componentes. La topología involucra relaciones como (conectividad, adyacencia, contenido), pero no cuantifica distancia o específica dirección.

Por otro lado, otros factores tales como que los distintos sistemas de base de datos geográficos usan diferentes métodos para el procesamiento y optimización de sus consultas y de que un mismo objeto geográfico tiene pluralidad de representaciones, según el sistema de coordenadas, la escala o las proyecciones, son elementos que también contribuyen a la heterogeneidad en el procesamiento. Las consultas pueden ser realizadas en el sistema directamente sobre los datos alfanuméricos y reflejadas en los resultados gráficos, dejando el procesamiento de la consulta en manos del lenguaje o metodología, y de la versatilidad del sistema de base de datos para responder con eficiencia [14]. A continuación se mencionan los distintos tipos de consultas espaciales que fueron encontradas en [29,36], y que además serán objeto de estudio importante para su posible resolución a través de métodos semántico.

3.1.1 Consulta punto en polígono

Esta consulta responde a la pregunta de localización: “¿Qué hay en el punto A con coordenadas (x,y) de este sistema de coordenadas?” como muestra la Figura 6. Su esencia es encuestar objetos georreferenciados que rodeen o refieran a una ubicación específica en la superficie terrestre. Esto permite por ejemplo, dado el punto A encontrar todos los objetos espaciales que lo contienen. Es una de las consultas más simples para procesar.



Fig. 6. Representación de la consulta punto en polígono

3.1.2 Consultas de regiones

Estas consultas recuperan todo lo que quede circunscrito dentro de un polígono cerrado, al cual denominamos región, por lo que es utilizada para obtener la información asociada al área que engloba (contenido en, adyacente a, superposición). Por tanto responde a preguntas tales como: “¿Qué ciudades encontramos en este perímetro?” como muestra la Figura 7, o “¿Qué carreteras cruzan o están contenidas dentro de esta región?”, en fin, puede asociarse cualquier combinación de elementos que demanden información dentro del área seleccionada, selección que dicho sea

de paso, puede responder a polígonos regulares o irregulares, e incluso círculos, a pesar de que esta última clasificaría también como consulta de Zona de Buffer en un punto que comentaremos a continuación.

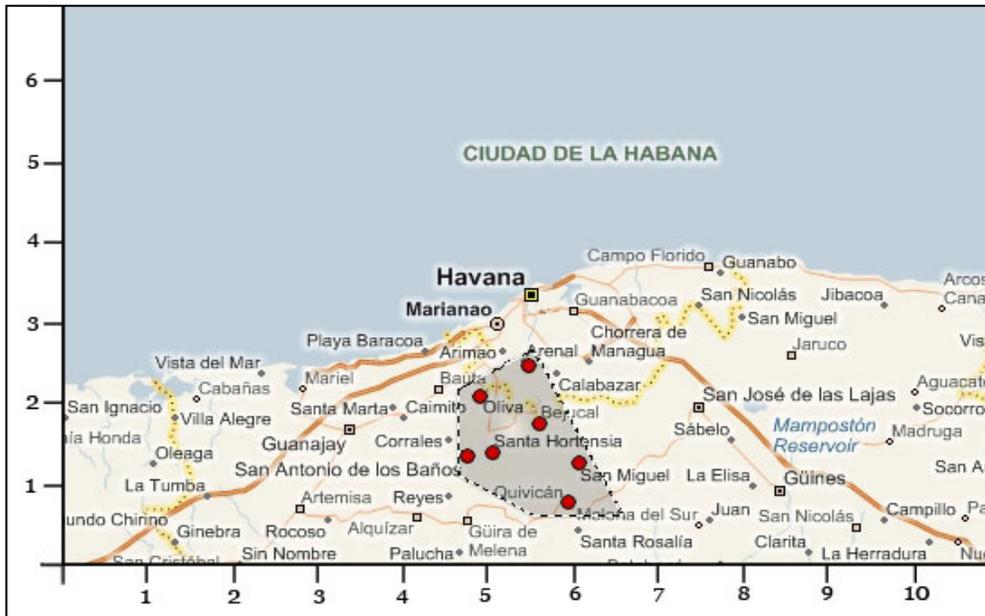


Fig. 7. Representación de la consulta región

3.1.3 Consultas de distancia y zona de buffer

Esta consulta se utiliza para saber qué objetos se encuentran dentro de una distancia de otro objeto, objeto que puede ser punto, línea o polígono. Un tipo de consulta distancia y zona de buffer lo podemos apreciar en la Figura 8 donde mostramos que municipios encontramos a una distancia D de la Habana. Por lo que responde a preguntas como: “¿Cuáles hoteles se encuentran a 1 km de la carretera C?”, “¿Qué hospitales se encuentra a 5 km del punto A?”, “¿Cuales centrales están dentro de un radio de 10 km de la cabecera de provincial?”. Hay casos sin embargo donde la distancia no se especifica claramente, por lo que la zona buffer no es exacta. Cuando la ubicación de conjunto de objetos a devolver es difusa, es mejor retornar una lista ordenada por cercanía de los objetos recuperados.

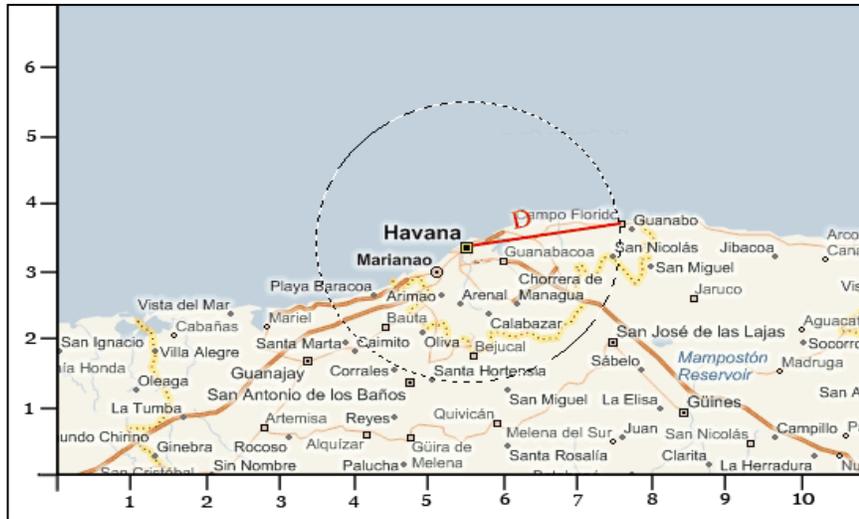


Fig. 8. Representación de la consulta distancia y zona de buffer

3.1.4 Consultas de camino

Las consultas de camino constituyen uno de los problemas de análisis más recurridos por los usuarios actualmente. Entre sus aplicaciones, estas consultas responden al tipo de pregunta tal como muestra la Figura 9: “¿Cuál es la ruta más corta entre la Habana y Guanabo?”, por lo que son utilizadas para hallar la ruta más óptima. Para resolver estas consultas es necesario que los datos espaciales estén conectados mediante una estructura que permita la interconexión de los conjuntos de datos en forma de red para así poder efectuar los análisis pertinentes. De este modo una de sus funcionalidades básicas se basa en encontrar la ruta óptima entre un punto de la red y otro, pero también ofrece información sobre la conexión de los objetos, etc.

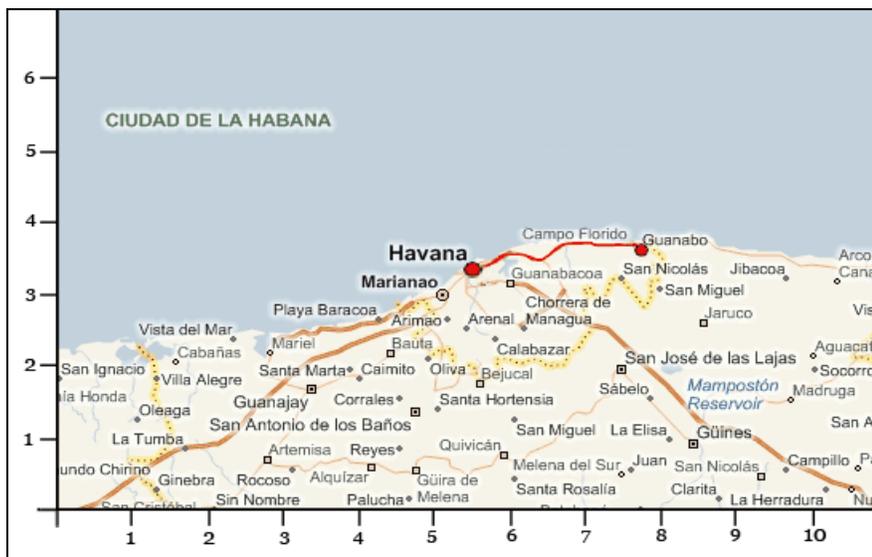


Fig. 9. Representación de la consulta camino

formular la pregunta y expresar la consulta en lenguaje natural, cayendo sobre el sistema la responsabilidad de procesarla, siendo esto totalmente transparente al usuario.



Fig. 11. Ejemplo de interfaz para la realización de una consulta en el SIG ArcView 3.2

El segundo viene dado por la dispersión de los resultados, o sea, que para obtener un resultado general el usuario necesita encuestar varias fuentes de datos. Las consultas espaciales, como por ejemplo las que implican análisis de fronteras, muchas veces no pueden resolverse accediendo a una sola fuente de datos. Aquí también las consultas semánticas posibilitan la integración de las fuentes de datos a través de reglas definidas en las ontologías.

Un ejemplo de esto lo constituye el análisis de la afectaciones de la sequia en un determinado periodo de tiempo. Para ello es necesario para el especialista encuestar manualmente distintas fuentes como son:

- Serie temporal de imágenes satelitales de alta resolución
- Datos tabulares climáticos de precipitaciones y temperaturas tomados de las distintas estaciones meteorológicas locales
- mapas vectoriales de suelo y vegetación
- mapas raster de modelo digital del terreno
- mapa vectorial de la infraestructura de los viales.
- mapa vectorial hidrológico

Un procesamiento semántico posibilitaría encuestar estas fuentes de forma automática extrayendo de cada una el significado de los datos en una capa o nivel semántico que posibilite procesar información entre estos repositorios. Las ontologías como veremos más adelante, son las estructuras que nos permitirán realizar esta conceptualización de los datos y guiar el proceso de recuperación. También en [29] podemos ver un ejemplo claro de cómo recuperar información geográfica desde repositorios no estructurados guiada por ontologías.

El tercero, viene dado porque las consultas actuales no permiten expresar conceptos como criterios de búsquedas, es decir, estas se basan en ideas que puedan enlazar diferentes subconsultas para llegar a un resultado general, en cambio las consultas conceptuales sí permiten buscar a partir de conceptos o abstracciones relevantes para los usuarios.

4. Métodos de acceso a datos

Usualmente dada la naturaleza compleja del dato espacial las bases de datos espaciales contienen grandes volúmenes de datos. Esto desfavorece al proceso de recuperación ya que implican gran consumo de tiempo. La diferencia con las bases de datos convencionales estriba en que toda la información contenida en un SIG está unida a entidades geográficamente localizadas, por lo que en una base de datos espacial (BDE) podemos encontrar tanto datos alfanuméricos como datos espaciales.

Los trabajos de investigación en BDE se han centralizado en el modelado y la resolución de consultas, basándose en la geometría asociada a los objetos almacenados en una base de datos. Resulta muy conveniente a la hora de recuperar un dato geográfico que este haya sido bien representado. Una buena representación sin dudas permite mejorar la efectividad de los resultados en las consultas, ya que para su representación es necesario tener en cuenta aspectos tales como la estructura de los datos, las operaciones sobre ellos y los caminos de acceso requeridos para su representación y recuperación.

4.1 El indizado espacial

Un buen indizado es el que trata de no visitar cada vez que sea requerido cada uno de los elementos de la base de datos. Podemos decir que los índices espaciales nos permiten ir directo al grano, optimizando la recuperación. El indizado espacial es uno de los aspectos claves y principales para mejorar el rendimiento de las consultas en los datos geométricos representado los objetos espaciales a través de índices, además permite descartar rápidamente subconjuntos grandes del conjunto de objetos a recuperar, y buscar en conjuntos reducidos.

Métodos de bucketing

Se le llaman métodos de bucketing a los métodos de ocupación espacial que descomponen el espacio donde se ubican los objetos espaciales en regiones llamadas cubetas o buckets. Una aplicación de los métodos bucketing consiste en ser utilizados sobre un espacio con objetos geométricos. Existen varias aproximaciones principales para descomponer el espacio. Una muy frecuentemente usada consiste en encerrar los elementos bajo el concepto de mínimo rectángulo limítrofe. En este caso los objetos son agrupados en jerarquías, posiblemente por una relación de proximidad, y almacenados en otra estructura como un B-Tree. Los R-Tree emplean esta filosofía [16].

R-Tree [3]

Un R-Tree es un árbol balanceado similar al B-Tree [17] con registros de índices en sus nodos hojas que contienen referencias a los datos espaciales. Los nodos se corresponden con “páginas” en el disco duro si el árbol es persistente, y esta estructura está diseñada de tal manera que las búsquedas espaciales requieren visitar solo un pequeño número de nodos. El índice es completamente dinámico, operaciones de inserción y eliminación pueden ser combinadas con búsquedas y no se necesitan reorganizaciones periódicas.

Una de sus mayores aplicaciones está en que se utilizan para manipular de forma eficiente grandes cantidades de información geométrica de N dimensiones, por lo que se ha convertido en uno de los métodos de acceso estándar para las BDE. También es uno de los más estudiados con

respecto a las temáticas del procesamiento y optimización de consultas, modelos de costo, paralelismo, control de concurrencia y recuperación, etc.

Los R-Tree y sus variantes son utilizados para manejar una colección de objetos espaciales arbitraria representados como rectángulos d-dimensionales que encierran a sus nodos hijo. Un nodo hijo es aquel que tiene punteros a los objetos reales de la base de datos. Los objetos están representados por el mínimo rectángulo que los contiene. Sus principales características pueden ser las siguientes:

- Los mínimos rectángulos limítrofes que corresponden a nodos diferentes pueden superponerse.
- Un objeto puede estar contenido espacialmente en varios nodos, pero se asocia únicamente con un nodo.
- El R-Tree de un espacio no es único. Su estructura es fuertemente dependiente del orden en que se adicionan o eliminan los segmentos de líneas.

Un aspecto en contra de este método reside en que resulta en descomposiciones segregadas del espacio. Un objeto se asocia solamente con un rectángulo incluso si este pasa por más de uno. Según [16] esto significa que en el peor de los casos si deseamos determinar el objeto que está asociado con un punto determinado para un espacio de dos dimensiones podría resultar en una búsqueda sobre la base de datos entera. Los otros acercamientos se basan en la descomposición del espacio en celdas separadas que son mapeadas a buckets. Los objetos son divididos en subobjetos disjuntos, cada uno asociado con una celda diferente. Las diferentes aproximaciones varían en el grado de regularidad impuesta por las reglas de descomposición subyacente y en la forma en que las celdas son agregadas.

En detrimento de estos métodos está que puede traer ambigüedad en consultas tales como la de seleccionar todos los objetos que estén contenidos en una misma región, ya que puede devolver un objeto repetidas veces, o cuando se necesita obtener todas las celdas que ocupa un determinado objeto para poder conocer su área. Entre las aproximaciones que utilizan la descomposición disjunta del espacio se pueden identificar los métodos basados en particiones disjuntas de los objetos en subobjetos arbitrarios disjuntos. Los subobjetos son luego agrupados en otra estructura como puede ser un B-Tree. Esta idea es la que implementan los R+Tree y Cell-Tree[16].

Los R+-Tree extienden de los kd-B-trees y manejan colecciones de objetos que son restringidas por rectángulos, de ellos se destaca:

- Están conducidos por el objetivo de evitar superposiciones entre rectángulos limítrofes.
- Cada objeto está asociado con todos los rectángulos que lo intersecan.
- Desde la raíz pueden existir varios caminos para alcanzar un mismo objeto.

Esto resulta en un incremento en el tamaño del árbol, sin embargo el tiempo de respuesta es mejor. Entre las desventajas de este método se puede citar que la descomposición es dependiente de los datos, lo que dificulta las tareas que requieren composición de operaciones y datos como por ejemplo la operación superposición.

Método Quadtree [2]

Quadtree es una estructura de datos jerárquica utilizada usualmente para manejar o representar objetos tales como imágenes raster, puntos, líneas, o áreas. Esta estructura está basada en el principio de descomposición recursiva, similar a métodos divide y vencerás [18]. Cada región a analizar se subdivide en cuatro regiones, por lo que los Quadtrees son árboles cuyos nodos tienen cuatro hijos, también son conocidos como árboles Cuaternarios.

Según [19] los Quadtrees pueden ser diferenciados teniendo en cuenta el tipo de datos que representan, el principio que guía el proceso de descomposición y la resolución de la

descomposición. La división puede ser en polígonos regulares, en este caso se nombra descomposición regular, o en partes desiguales, en cuyo caso las dimensiones generalmente son dadas en la entrada de datos. Se llama resolución de la descomposición al número de veces que es aplicado el proceso de descomposición, éste puede ser fijado de antemano o estar controlado por propiedades de los datos de entrada. En la representación del árbol, la raíz se corresponde con la región completa que se quiere analizar. Cada hijo de un nodo representa un cuadrante de la región representada por ese nodo. Las hojas del árbol representan aquellos bloques para los cuales no son necesarias más divisiones.

La construcción de un Quadtree puede realizarse recursivamente, es decir, se divide el cuadrante actual siguiendo el principio de descomposición en cuatro subcuadrantes y recursivamente construir los Quadtrees para cada subcuadrante. El punto de parada del algoritmo se fija teniendo en consideración la resolución de la descomposición. Una desventaja que pudiéramos citar aquí es que los Quadtrees son sensibles a la posición de los objetos con relación a la descomposición del espacio donde se encuentran embebidos.

Método grilla uniforme [16]

Este método respeta la descomposición disjunta del espacio y tienen un grado mayor de independencia de datos. Realiza la descomposición regular del espacio en bloques de tamaño uniforme. El método de Grilla Regular es ideal para datos con distribución regular.

Métodos espacio-temporales [20]

- El método de acceso 3D R-Tree, considera el tiempo como otro eje junto a las coordenadas espaciales, o lo que es lo mismo considera al tiempo como otra dimensión. Si conocidas las coordenadas iniciales de un objeto en un plano xy y el intervalo de tiempo que estuvo en esa ubicación y más tarde se conoce una nueva posición del objeto en xy y la duración de tiempo que permaneció ahí, este puede ser modelado a través de dos segmentos $[(x_1, y_1, z_1) (x_1, y_1, z_2)]$ y $[(x_1, y_1, z_2) (x_1, y_1, z_3)]$ de líneas en el espacio tridimensional, donde (x_1, y_1) corresponden a las coordenadas iniciales del objeto, (z_1, z_2) el intervalo de tiempo que estuvo en esa posición y (z_2, z_3) el intervalo de tiempo que estuvo en la segunda ubicación, pudiendo ser indexadas por un R-tree, considerando los objetos en tres dimensiones. Sus principales desventajas consisten en que es necesario conocer de antemano los límites finales de los intervalos tiempo y es ineficiente a la hora de procesar consultas del tipo timeslice⁴ [20].
- Otro método que también emplea el tiempo como otra dimensión es el 2+3 R-Tree. Su nombre viene dado a que utiliza dos R-tree, uno para los que tienen la información del tiempo final (de dos dimensiones) y otro (de tres dimensiones) para los objetos que poseen tiempo inicial y final en una ubicación determinada. Esto posibilita acceder a información actual e histórica de los diferentes objetos, incluyendo también la trayectoria.
- El método de acceso RT-Tree permite la recuperación de información histórica. Aquí la información temporal se mantiene en los nodos del Rtree, por lo que el árbol puede crecer mucho. No obstante en [25] se llegan a criterios que demuestran que RT-tree es una de las estructuras más eficientes en la utilización del almacenamiento. Sus desventajas radican una, en la ineficiencia para procesar consultas del tipo timeslice⁴, ya que como dijimos anteriormente al aumentar el contenido que normalmente tienen los

nodos el árbol puede alcanzar gran altura, y segundo al ser regidas las búsquedas por la componente espacial, los predicados temporales quedan relegados a un segundo plano y las consultas con predicados en este dominio no son eficientemente procesadas.

- Otro método que pertenece a esta categoría es el DR-Tree, el cual es un método de acceso espacio-temporal de tipo histórico. La idea básica del DRTree es mantener puntos de referencia con el componente espacial de los objetos para ciertos instantes de tiempo, los objetos “vivos” en esos instantes de tiempo se almacenan en una estructura de datos R-Tree. Los movimientos de los objetos entre puntos de referencia consecutivos se mantienen en una lista de movimientos denominada bitácora, ordenada de acuerdo al tiempo, y que permite reconstruir cualquier estado de la base de datos entre dos puntos de referencia consecutivos [20].

5. Métodos semánticos

Estos métodos proponen volverse un nuevo paradigma en la recuperación espacial. Según [21] muchos lenguajes relevantes en el contexto de los sistemas de información (lenguajes de programación, lenguajes de consulta, lenguajes de modelado de flujo de trabajo, etc.) manifiestan como los seres humanos conceptualizan las cosas con el propósito de representarlas y manipularlas en las máquinas. Sin embargo, añadir significado a las expresiones del lenguaje es un fenómeno conceptual, ya que el lenguaje natural evoca símbolos y expresiones en nuestra mente que son usadas para dar significado a esos conceptos. Lo mismo atañe a los SIG. La semántica geoespacial puede comprender y convertir el contenido del SIG y capturar este conocimiento en teorías formales. Hoy en día contamos con varias estructuras para representar el conocimiento, siendo las ontologías las más eficientes para abarcar la mayor cantidad posible de soluciones a los problemas que conciernen al dominio geográfico.

5.1 Uso de ontologías para la conceptualización del dominio espacial

Ontología como término, procede de la Filosofía y según la Enciclopedia Filosófica [30] “Ontología es el estudio de lo que hay” (“ontology is the study of what there is”), mientras que otra definición un poco más concreta es aportada por el Diccionario de la Real Academia de la Lengua Española expresando que una ontología es la “Parte de la metafísica que trata del ser en general y de sus propiedades trascendentales”. Estas, sin embargo no son las únicas definiciones que se brindan al respecto, existen varias que pudieran dar una idea mejor al lector, pero como no es el aspecto filosófico el que no interesa, sino que sin pretender explicar la analogía de este concepto de la filosofía a la inteligencia artificial, podemos decir que precisamente las definiciones dadas en esta área de las ciencias técnicas fue lo que comenzó a popularizar su uso dentro de los sistemas informáticos. Las más citadas por la literatura son las de Gruber [31], “Una ontología es una representación explícita de una conceptualización” (“an ontology is an explicit specification of a conceptualization”) y la especificación de este concepto realizada por Guarino [32] “una ontología es un conjunto de axiomas lógicos diseñados para alcanzar el significado intencional de un vocabulario, o sea, un comportamiento ontológico hacia una conceptualización específica del mundo” (“An ontology is a logical theory accounting for the intended meaning of a formal vocabulary, i.e. its ontological commitment to a particular conceptualization of the world.”). Entonces, ¿por qué son especialmente útiles las ontologías para resolver los problemas de interoperabilidad entre datos, servicios y sistemas?

Resultado del propio desarrollo de software los sistemas basados en el conocimiento son concebidos desde distintos puntos de vista. Los SIG por ejemplo, asimilan gran cantidad de

formatos de datos espaciales, pero no son las únicas aplicaciones que asimilan información espacial. Actualmente, se puede extraer información espacial de las más diversas maneras, de un código postal (conocido como geocoding), un teléfono, un nombre de lugar, coordenadas espaciales etc., dicha información está muchas veces contenida en archivos que no permiten ser identificados por un SIG haciendo imposible compartir o reutilizar datos. Si tenemos en cuenta una estadística bastante popular que refiere que más del 80% de la información almacenada en sistemas de todo tipo, es información georreferenciada o susceptible de serlo, entonces un buen identificador de la eficiencia de un SIG, vendría dado por la cuantía en alcance de la identificación de esta información en cualquiera de sus formas, en aras de contar con mayor cantidad de datos que posibilitarían procesamientos más completos.

Hoy sin embargo, esto no es posible dadas varias causas. En primer lugar se pudiera argumentar que existe muy poco desarrollo en cuanto a la representación y recuperación de semántica espacial. La mayoría de la información espacial que se publica está pensada para ser precisamente interpretada por expertos y no por agentes. Los razonadores con que se cuenta en la actualidad no permiten realizar estos procesos dejando muchas veces al especialista la tarea de identificar y analizar los datos espaciales.

En [33] los autores argumentan que la diversidad de formatos espaciales y la diversidad de conceptualizaciones temáticas que se utilizan para los datos integrados a los SIG hacen que a la mayoría de las aplicaciones no tengan carácter sostenible, además de que no puedan compartir datos y ser reutilizados por toda comunidad geoespacial. También podemos agregar que estos sistemas son desarrollados en diferentes contextos y utilizando vocabulario propio, lo que muchas veces trae ambigüedad a la hora de intercambiar información.

Las ontologías se presentan precisamente como una alternativa potente para resolver muchos de estos problemas de heterogeneidad de las fuentes de información, permitiendo además desambiguar los términos de búsquedas porque en ellas el conocimiento se especifica a través de conceptos. Estos a su vez, están vinculados y regidos mediante relaciones y axiomas que enriquecen el contexto que conceptualizan. En [34] podemos ver otras ventajas de las ontologías. La recuperación por conceptos permite además de que los sistemas puedan identificar mejor los datos, procesarlos en consecuencia y realizar análisis más eficientes con independencia del especialista, lo que constituiría una gran ventaja porque la información no solo será comprendida por las personas sino también por las máquinas incidiendo favorablemente sobre la reutilización del conocimiento, haciendo al sistema más potente y eficaz. También la posibilidad de interactuar semánticamente con otras ontologías constituye un aspecto de avance, ya que posibilita obtener información desde otras fuentes. Esto significaría que pudiera obtenerse datos desde otros sistemas incluso independientemente del lenguaje que tengan, dado que las consultas serían interpretables en un ámbito donde la información contenida en una ontología estaría en correspondencia con otra representación tal vez de la misma información, lo que es muy útil en el campo de la información geográfica.

En [35] se señalan 3 ventajas muy significativas que suponen preeminencias sobre las otras formas de organización de la componente semántica.

I) Reducen la confusión semántica. Disminuye la ambigüedad terminológica al tener en cuenta sinónimos y polisemias, repercutiendo sobre la comunicación.

Un sencillo ejemplo de la reducción de términos usando sinónimos con el uso de tesauros lo constituye la consulta: *Obtenga los hospitales próximos al municipio Plaza*

Aquí próximo nos indica cercano, inmediato, contiguo. Con cualquiera de estas palabras que se hubiese formulado la pregunta el sistema debiera ser capaz de identificar que es lo mismo y responder igual.

II) Permite la posibilidad de reutilizar el conocimiento. Esta ventaja es la referida a la interacción entre distintas ontologías del dominio geográfico, ya que resultado del propio desarrollo de ontologías hay diversas formas de conceptualizar una misma realidad.

III) La traducción e intersección semántica a través de mappings (cotejos) utilizados para describir correspondencias entre fenómenos (ej.: río, river, rivière y fleuve) y entre diferentes ontologías (ej.: ontología de fenómenos hidrográficos y ontología de las ciudades)

Por tanto, podemos concluir que las ontologías se presentan como el componente indicado para solucionar los problemas derivados de la interoperabilidad semántica, que a su vez constituye una de las brechas más extensas que separan a los SIG de poder realizar análisis más completos e independientes, sin la supervisión constante del usuario. Para una consulta que implique el acceso a repositorios heterogéneos siempre está presente el riesgo de que no pueda interpretar el dato en la fuente por la diversidad de formatos que tienen actualmente los datos geográficos. Hoy en día muchos de estos datos se acceden de forma “ad hoc” y la solución a través de estándares no parece ser la más óptima. Ahí es donde vuelven a ser especialmente útiles las ontologías porque permiten extraer la información independientemente del formato del dato apuntando a ser la mejor respuesta a esta dificultad, así como también a muchas otras. De esta manera, se dota a los sistemas de la capacidad de interactuar por si solos en consecuencia de la información que procesen convirtiéndolos en herramientas capaces de buscar e inferir conocimiento.

5.2 Métodos basados en semántica

Uno de los trabajos más interesantes en este sentido lo constituye [22], que aunque no es precisamente un trabajo de recuperación, sí constituye un aporte valioso en este sentido porque posibilita una descripción semántica de un contexto geográfico. En el los autores proponen hacer la conceptualización haciendo uso de las ontologías, destacando la importancia de estas últimas para la solución de los problemas de heterogeneidad e interoperabilidad de los datos geográficos. Enfocados en la representación, refieren que el análisis espacial con frecuencia demanda una descripción muy precisa de los objetos para poder obtener buenos resultados, por lo que en su propuesta utilizan una ontología que describe los datos espaciales de acuerdo a un contexto determinado (en su caso orientado a describir mapas cartográficos vectoriales.)

Para comenzar definen un mapa como una partición espacial Ω dentro del universo de los objetos geográficos a_i , que consisten en un grupo de representaciones primitivas [22 citando a 27] (ecuación 1).

$$\Omega = \alpha_i \cup \{Rp_1, Rp_p, Rp_a\}, i=1, \dots, n, \quad (1)$$

Donde:

Rp_1 es la primitiva de la representación “lineal”,

Rp_p es la primitiva de la representación “puntual”,

Rp_a es la primitiva de la representación “superficie”,

i representa el número temático que involucra la partición espacial

Luego definen dos tipos de conceptos (C) en la ontología: terminal (C_t) y no terminal (C_n)

Donde el concepto terminal no utiliza otros conceptos para definir su significado (pueden estar definidos por valores simples) y los conceptos no terminales que usan otros conceptos para definir su significado (ya sean terminales o no terminales) (ecuación 2).

$$C = C_N \cup C_T \quad (2)$$

Plantean que cada concepto posee un conjunto de características, ya que como precisamente sucede en la realidad, los objetos geográficos las tienen. Ellas son las propiedades y las

relaciones entre las entidades geográficas. En el trabajo se utiliza el término “relación” para denotar una relación/propiedad unaria. Haciendo desde este punto de vista simple todas las características de un concepto terminal, como por ejemplo, el tipo de todas las características que pertenecen al conjunto de primitivas tipos. (T_p) (ecuación 3).

$$T_p = \{\text{número, carácter, cadena, enumeración, estructura}\}, \quad (3)$$

$$A = \{a_i \mid \text{type}(a_i) \in T_p\},$$

Donde:

T_p es el conjunto de primitivas tipos

A es el conjunto de características

Entonces el conjunto de concepto terminal está definido por la ecuación 4.

$$C_T = \{c(a_1, a_2, \dots, a_n)\}, \exists a_i \in A, i=1, \dots, n \quad (4)$$

En el mismo modo, el concepto no terminal posee al menos una característica que no pertenece a T_p Como lo ilustra la ecuación 5.

$$C_N = \{c(a_1, a_2, \dots, a_n)\}, \exists \exists a_i \notin A \quad (5)$$

Donde c es un concepto

Finalmente el conjunto de relaciones R está definida por los pares que están asociados a Γ y Φ , donde Γ y Φ son relaciones no reflexivas, no simétricas y transitivas. (ecuación 6).

$$R = R_\Gamma \cup R_\Phi = \{(a,b) \mid a \Gamma b, a \in C_N, b \in C\} \cup \{(a,b) \mid a \Phi b, a \in C_N, b \in C\} \quad (6)$$

Esta descripción mapea los datos espaciales en la ontología. Una vez que los conceptos son definidos en la ontología, pasan a definir los conceptos no terminales. Este proceso continua hasta que encuentren un concepto terminal. Una vez que este es encontrado es necesario seleccionar un par de objetos geográficos y verificar si la relación entre ellos existe, de lo contrario se necesita generar una parte de su descripción. El concepto terminal es definido por el tipo de relación entre los dos objetos. Usando estos dos conceptos y relaciones se logra construir su mapa semántico que permite describir un mapa cartográfico vectorial.

A favor de este trabajo podemos decir su mayor importancia radica en que logra describir semánticamente el comportamiento de los objetos geográficos, permitiendo así sentar las bases para mecanismos que permitan inferir y recuperar conocimiento a través de las ventajas que brinda la conceptualización. Por otro lado, dos aspectos que se deben tener en cuenta es que con las dos relaciones que utilizan (“tiene”; “es un”) no es posible representar todos los fenómenos del dominio geográfico, y otro que no permite la recuperación de otros elementos geográficos que puedan estar en ese contexto si no están conceptualizados en la ontología. Sin embargo una ampliación de este trabajo tiene lugar años mas tarde devenida en una tesis doctoral [28], que junto a [29] (el cual será comentado posteriormente) y a criterio de este autor, son los trabajos más recientes y completos respecto al tema.

En [28] se propone una metodología “GEONTO – MET”, que permite conceptualizar formalmente el dominio geográfico caracterizando semánticamente el comportamiento de los objetos geográficos a través de sus propiedades y relaciones, apoyados en una ontología de dominio (Kaab-Ontology) elaborada con este objetivo; y aunque el objetivo principal de la tesis no es recuperar, aquí se propone un mecanismo de recuperación a través de conceptos donde se

asocian conceptos representados en la ontología con alguna fuente de información geográfica. Sin embargo, a pesar de que las consultas utilizadas son estáticas, sirven para ilustrar como acceder a la ontología.

Utilizando la API Jena de java es posible extraer las instancias de conceptos clases de Kaab-Ontology, ya que dicha ontología fue elaborada en Protégé utilizando OWL. De aquí se obtiene un modelo persistente que funciona como enlace entre Jena y Protégé de donde se recuperan instancias de conceptos geográficos en OWL. Por otro lado las consultas implementadas en SPARQL posibilitan el acceso a datos de recursos OWL, dado que previamente las clases de entidades abstractas y los conceptos que heredan de estas clases fueron llevados a una forma tabular. Así SPARQL a través de Jena puede acceder a todos los recursos de la ontología.

Al ser el modelo de consultas orientado a datos, una vez definido el significado de un determinado elemento en un glosario dado y representado en un lenguaje, es posible entonces obtener instancias de conceptos de forma semántica que estén relacionados dentro del contexto de la ontología. [28] Esto representa una gran ventaja.

Otro método es propuesto en el trabajo [23]. Parten también de que actualmente la comunidad de las ciencias de información geográfica reconoce que los sistemas modernos de recuperación (GIR en Inglés) deben soportar el procesamiento de datos imprecisos distribuidos en repositorios heterogéneos. Esto significaría que incluso cuando una consulta geográfica no encuentre resultados que correspondan con la solicitud del usuario sería útil dar una aproximación de los resultados. Por eso los GIR deben estar centrados en la naturaleza y esencia del dato espacial (sus relaciones y propiedades) y teniendo en consideración el perfil del usuario.

Usualmente los rasgos semánticos están presentes en diferentes fuentes de datos. En este trabajo se escogieron fuentes heterogéneas como son: los datos vectoriales, las ontologías geográficas y los diccionarios geográficos, ya que usualmente estos repositorios almacenan relaciones topológicas, conceptos y descripciones de los objetos bajo ciertos escenarios, información esta, que en trabajos anteriores había sido tratada de manera aislada. Esta integración espera ser una solución mejor para extraer los objetos espaciales mediante la semántica, ya que con el uso de la misma y la integración de varias capas de información los GIR permitirían adecuar mejor los parámetros de recuperación.

Se designo un mecanismo de extracción de información para cada fuente de datos, mientras que el proceso de integración se llevo a cabo por el algoritmo de exploración de ontologías. El proceso de mostrar los resultados está basado en “muestras similares”, usando una teoría ya antes desarrollada (Confusion) [23 citando a 24]. Finalmente, presentan un caso de estudio donde comparan un grupo de resultados de un GIR integrador (iGIR) con algunos de Google en una forma tabular.

Dada la naturaleza de los datos geográficos, estos se encuentran usualmente almacenados en repositorios heterogéneos que hacen la tarea de recuperarlos un tanto más difícil. Existen varias propuestas que han sido citadas como métodos para realizar esta tarea, pero ninguna de las metodologías actuales es capaz de manejar la variedad de fuentes como para resolver adecuadamente este problema.

Este trabajo propone recuperar la información geográfica procesando las consultas, las cuales son divididas en tres para su procesamiento <qué, relación, dónde>, donde el “qué” denota un objeto geográfico, el “dónde” puede ser la referencia espacial ó un objeto geográfico y la “relación” denota la relación espacial que vincula a “qué” y “dónde”. El enfoque está basado en una estrategia de recuperación que utiliza tres tipos de coincidencia:

1. **Topológica** .Son las relaciones topológicas extraídas de la superposición de capas de datos, tales como {en, contienen}
2. **Geográfica**. Están limitadas a partir de restricciones obtenidas a partir de los diccionarios , ejemplo {Aeropuertos representados por puntos o polígonos}

3. **Conceptual.** Viene dado por una ontología geográfica, como puede ser, {tipos de aeropuertos}

De esta manera se usan tres tipos de fuentes de datos heterogéneas, archivos vectoriales, diccionarios y ontologías.

Todas almacenan diferentes relaciones y propiedades que representan la naturaleza del dato espacial. Estas ya habían sido tratadas también en trabajos anteriores, pero separadas, por lo que en contraste a eso una propuesta nueva es un sistema integrador de recuperación de información geográfica (iGIR) que agrupa los resultados utilizando unos pocos procesos como , consulta, recuperación, integración y orden jerárquico de los resultados.

Coincidencia conceptual

Es el primer paso de la estrategia de recuperación. La ontología permite guiar la recuperación indicando “qué” datos deben ser buscados y “donde”, o sea, describe la manera de recuperar resultados relevantes de acuerdo con las relaciones semánticas entre objetos geográficos a través de un algoritmo de exploración.

Ejemplo: $Q_{GI} = \{\text{Hoteles cerca del aeropuerto Benito Juárez}\}$

En este caso el objeto geográfico es el hotel, la ontología describe sus relaciones y propiedades con otros objetos, describe que es un hotel y sus tipo(s) de representación(es). De aparecer otros objetos y relaciones son procesados en la misma manera.

En el ejemplo la ontología fue elaborada manualmente utilizando artículos de Wikipedia. Las categorías y los vínculos contenidos en cada artículos han sido consideradas como parámetros para definir las relaciones y los conceptos de la ontología. Las relaciones semánticas fueron clasificadas de acuerdo a su Hiperónimo⁵ y Meronimia⁶.

Se dan 4 posibles clasificaciones para las coincidencias obtenidas como resultados del algoritmo en esta fase:

- *atómica* (encuentra solo uno de los elementos que conforman la tripleta).
- *parcial* (encuentra parejas de los elementos que conforman la tripleta).
- *completa* (encuentra todos los elementos que conforman a una consulta de tipo tripleta).
- *nulo* (no encuentra ninguno de los elementos presente en la ontología, se recurre al algoritmo de confusión [23 citando a 24] y se procesa un Geo_{obj} por un concepto lo más cercano posible al deseado. Ejemplo de “ríos”, con “cuerpos de agua” como resultado más alejado y “lagos” como mas próximo).

Coincidencia geográfica

Constituye el próximo paso del proceso. Utiliza como fuente de datos a los diccionarios geográficos. Son especialmente importantes porque contienen información como la escala, las propiedades, restricciones, relaciones, etc. Inicialmente vienen en pdf, pero seguidamente son transformados semiautomáticamente a archivos XML.

Para describir el macheo conceptual se utiliza el ejemplo de la consulta anterior.

Primero se procesa la consulta segmentándolas en sus tres partes.

(“qué” = hoteles, “relación” =cerca, “donde” = Aeropuerto Benito Juárez)

Se procede a buscar los objetos geográficos, sus relaciones, restricciones, propiedades, etc.

$Geo_{obj} = \{\text{Hoteles}\}$, relaciones = {conectado, comparte}, propiedades= {Tipo (“Local”, “Nacional”, “Internacional”)}, Restricciones = {primitivas de representación (puntos, polígonos)}

Se aplican reglas de asociación.

$R_1(\text{cerca}) = \{X \text{ cerca } Y, \text{ si } X \text{ y } Y \text{ están conectados por } Z\}$ donde X, Y son objetos geográficos (Geo_{obj}), representados por puntos o polígonos, mientras que Z es un objeto geográfico representado por un Arco. Además su longitud es menor que un 1km (z es un camino), entonces finalmente “conectado” es una relación entre X y Y .

Entonces, la recuperación en este paso consistiría en buscar los documentos que cumplan con la regla:

$P_{\text{busqueda}_1} = \{ \text{Geo}_{\text{obj}} X \text{ conectado } \text{Geo}_{\text{obj}} Y \}$ donde $\text{Geo}_{\text{obj}} Y$ puede ser un punto o un polígono.

$P_{\text{busqueda}_2} = \{ \text{Geo}_{\text{obj}} X \text{ comparte } \text{Geo}_{\text{obj}} Y \}$ donde $\text{Geo}_{\text{obj}} Y$ puede ser un punto o un polígono.

Estos parámetros representan la búsqueda ejecutada usando archivos vectoriales.

Coincidencia topológica

Este proceso está basado en las relaciones topológicas entre los objetos geográficos y utiliza archivos vectoriales como fuente de datos. Los datos son procesados y llevados al formato de un Archivo Topológico, que contiene las relaciones entre dos objetos geográficos como muestra la tabla 1.

Tabla 1. Estructura de un archivo topológico

Id	Id_GeoObj_1	Capa_Pertenece	Id_GeoObj_2	Capa_Pertenece	Relacion
1	5	Aeropuerto.geo	2	Caminos.geo	conectado
2	10	Hoteles.geo	2	AreaUrbana.geo	comparte

Luego se comprueba el macheo entre los objetos. En el id_GeoObj_2 vemos que hay una coincidencia por lo que indica que un objeto (2, Avenida “Circuito Interior”) que pertenece a la capa caminos tiene la relación “conectado” que involucra al objeto 5 (aeropuerto Benito Juárez) con el objeto 10 (hotel “Holiday Inn”). Eso significa que el Hotel Holiday Inn está conectado a la Avenida Circuito Interior y que la Avenida esta conecta al Aeropuerto Benito Juárez.

Es entonces cuando los resultados son enviados al modulo de integración y luego al de devolución de resultados (ranking). Es aquí donde el proceso de recuperación termina y comienza el proceso de devolución. Este proceso está basado en medidas de similaridad entre conceptos y relaciones, llamada confusión [23citando a 24].

Este trabajo, también tiene continuidad en una tesis doctoral [29], propiamente de recuperación de información geográfica. En ella se muestra una metodología integral para la recuperación de información geográfica y también propone un método para ponderar los resultados de la recuperación. El aporte está en que integra tres criterios de búsqueda para tres fuentes de datos geográficas heterogéneas: Geo-ontologías, diccionarios de datos y archivos vectoriales. Las cuales describen a los objetos geográficos utilizando la topología, los atributos geográficos y la semántica espacial.

Según plantea el autor hoy la línea de recuperación de información tiene los siguientes retos:

- Extensión del enfoque sintáctico al semántico y espacial.
- Heterogeneidad en las fuentes de información.
- Intención del usuario expresada en las consultas.
- La extracción de términos geográficos desde fuentes estructuradas, y de reto aún mayor, desde fuentes no estructuradas de acuerdo a criterios espaciales.
- La identificación y eliminación de ambigüedades en el ámbito geográfico para los procedimientos de extracción.

- Tipos de consultas, de acuerdo a los siguientes criterios:
 - Tipo de parámetros de entrada: espaciales, geográficas, topológicas, etc.
 - Cardinalidad: número de elementos u objetos que conforman la consulta.
 - Estructura y formato: texto libre, operadores espaciales o geográficos, etc.
- Metodologías para el almacenamiento eficiente de ubicaciones y sus relaciones.
- Técnicas de ponderación.

Su propuesta basada en la integración de fuentes heterogéneas de información geográfica procesa las búsquedas orientadas a localización, ubicación o con algún criterio geográfico-espacial con una alternativa de solución semántica en el procesamiento, lo que le posibilita eliminar las ambigüedades propias de los términos introducidos por los usuarios en las consultas, haciendo la recuperación más precisa y obteniendo resultados de mucha mayor relevancia para el usuario. Como mencionamos anteriormente unos de esos repositorios son las Geontologías, donde la extracción del conocimiento se realizara mediante el algoritmo de exploración OntoExplora como se muestra en la Figura 12 encargado de encontrar los conceptos asociados al criterio de búsqueda de la consulta. Para ello inicia buscando el concepto y relaciones que representan un objeto, y si encuentra coincidencias extrae las relaciones asociadas y sus clases o instancias respectivas, ocurriendo durante este proceso 4 tipos de coincidencias como mismo ocurren en [23]. Luego de extraer las relaciones vecinas y conceptos asociados estos son almacenados en un vector de vecindad el cual contiene varios elementos que son procesados para que sirvan como base para buscar en los demás repositorios (diccionarios geográficos y archivos topológicos). Este trabajo constituye un gran paso en lo que respecta a la recuperación de información geográfica dado entre varias cosas, a que es capaz de aportar una metodología de recuperación que integra tres repositorios geográficos heterogéneas y que brinda un algoritmo para el procesamiento de semántica espacial en las geo-ontologías resolviendo en gran medida uno de los retos antes planteados. Por tanto si brinda la posibilidad de que a través del procesamiento de las relaciones topológicas se puedan recuperar objetos según las relaciones espaciales que tengan con otros objetos entonces esto constituye una base para desarrollar mecanismos de recuperación enfocados a la resolución de problemas de análisis en los Sistemas de Información Geográfica.

Inicio

Identifica y procesa cada elemento de la consulta geo-espacial (QGN) {que, relación, donde}

Asigna la relación topológica, geográfica, o semántica correspondiente a relación (Tr)

Recuperación guiada por GeoOntología:

Busca el concepto correspondiente para cada objeto de la consulta en la GeoOntología

If (hay_coincidencias) Then

Extrae el contexto usando relaciones semánticas de meronimia e hiperonimia

Extrae las propiedades del concepto (Pconcept)

Se llena el vector de vecindad (VectorVec) con las relaciones y propiedades

Se generan las consultas (QueryNuevo) usando VectorVec

Else

Envía QGN a las fuentes vectoriales y de diccionarios

Recuperación de Diccionarios de Datos:

Busca datos en diccionarios de acuerdo a la consulta y VectorVec

Utilizando RespOnt, recupera datos geográficos por sus propiedades y relaciones

If (hay_ coincidencia) Then

Recupera objetos de acuerdo a las relaciones propiedades geográficas

End If

Regresa respuestas (RespGeo)

Recuperación de datos topológicos:

Busca datos en topologyfiles de acuerdo a la consulta y VectorVec

Utilizando RespOnt, solicita los datos vectoriales (capas)

If (hay_ coincidencia (datos vectoriales)) Then

Recupera objetos de acuerdo a las relaciones topológicas

End If

Regresa respuestas RespTopological

Integración y ponderación:

Integra RespTopological , RespOnt y RespGeo en un conjunto final (FVector)

Pondera FS geobj usando iRank y lo muestra al usuario

Recuperación aislada o en dos fuentes

Else

Recupera los datos considerando los objetos de QGN

If (hay_coincidencias) Then

Recupera los datos de acuerdo a propiedades y relaciones espaciales

Regresa las coincidencias encontradas

Else

Recupera los datos vectoriales de acuerdo a relaciones topológicas

Regresa las coincidencias

Fig. 12. Algoritmo OntoExplora [29]

6. Conclusiones

En este trabajo se presentaron los diferentes métodos que han sido utilizados por los SIG para acceder a sus datos. Los enfoques más recientes apuntan hacia las ontologías como estructuras integradoras que permitan solucionar estos problemas basados en que los métodos orientados hacia la conceptualización ofrecen ventajas que permiten hacer búsquedas más completas y a la vez más fáciles aprovechando las bondades de este tipo de consultas. Sin embargo, los trabajos propuestos se encuentran a nivel de laboratorio y están enfocados hacia la representación, que a pesar de que realizan recuperaciones, aún no cuentan con mecanismos para hacer consultas semánticas dinámicamente en los SIG. También podemos ver como los métodos tradicionales utilizados actualmente posibilitan acceder a los datos de manera separada, cuando en los procesos de análisis es necesario encuestar diversos repositorios. La propuesta que se analiza en este sentido permite recuperar datos de tres fuentes heterogéneas y constituye un paso de avance en la recuperación de información geográfica, siendo una herramienta útil para implementar mecanismos de recuperación semánticos enfocados hacia la solución de problemas de análisis para los SIG, que es de hecho la dirección hacia donde se propone avanzar.

Referencias bibliográficas

1. Garea-Llano, Eduardo "Estado Actual de la Interpretación Semántica de datos Espaciales".RT_001, Serie Azul. Reconocimiento de Patrones, versión digital. RPNS_2113, ISSN 2072-6260. 2009
2. Oliva, R., *Anotaciones Semánticas de Datos Geográficos*, in *Departamento de Ciencia de la Computación. Facultad de Matemática y Computación*. 2009, Universidad de La Habana: Ciudad de La Habana. p. 126.
3. Mendoza, L., *Anotación Semántica de Datos Vectoriales en Sistemas de Información Geográfica. Implementación en gvSIG*, in *Departamento de Ciencia de la Computación.Facultad de Matemática y Computación*. 2008, Universidad de La Habana: Ciudad de la Habana.
4. Rodríguez, Andrea Curso Bases de Datos Espaciales. 2008. U. de Concepción, Chile
5. Larin Fonseca, R. y Garea Llano, E. RT_011. "Integración Semántica de datos Espaciales con sistemas de Información Geográfica", RNPS_2142 ISSN 2072-6287, Noviembre 2009.
6. Miranda, M., Edwards, W. (2007) *Representación de elementos espaciales: Modelo de datos VECTORIAL*.
7. Bosque, J., *Sistemas de Información Geográfica*. Ediciones Rialp ed. Geografía y Ecología. 1992, Madrid. 447.
8. OCW. *Los SIG Vectoriales*. 2008 [cited; Available from: http://ocw.innova.uned.es/ocwuniversia/analisis_%20geografico_%20regional/sistemas-de-informacion-geografica/contenidos/tema6.htm. fecha de acceso: enero 2010
9. Egenhofer, M., *Spatial SQL: A Query and Representation Language*. 1994.
10. H. Chen F. Wang, J.S.a.S.Y. *Geosql: A spatial query language of object-oriented gis*. in *Proceedings of the 2nd International Workshop on Computer Science and Informationn Technologies*. 2000.
11. Hui, L., Huang, Bo (2001) "*SQL/SDA: A Query Language for Supporting Spatial Data Analysis and Its Web-Based Implementation*". IEEE Transactions on Knowledge and Data Engineering **Volume**, pp 671-682 DOI: 10.1109/69.940739
12. Joel da Silva, A.S.C.V., Anjolina G. de Oliveira,Robson do N. Fidalgo,Ana C. Salgado,Valeria C. Times, *Querying Geographical Data Warehouses with GeoMDQL*. XXII Simpósio Brasileiro de Banco de Dados, 2007.
13. Vatsavai, R.R. *GML-QL: A Spatial Query Language Specification for GML*. 2002 [cited; Available from: <http://www.cobblestoneconcepts.com/ucgis2summer2002/vatsavai/vatsavai.htm>. fecha de acceso : enero 2010
14. Méndez, I., Báez, G., *Las Bases de Datos Espaciales (BDE) como nueva componente en los Sistemas de Información*. Mapping Revista internacional de ciencias de la tierra, 2009.
15. Kammersell, W., Dean, M., *Conceptual Search: Incorporating Geospatial Data into Semantic Queries*. 2006.
16. Vitturini, M., Castro, S. & Martig, S., *MODELOS DE DATOS ESPACIALES*, in *VII Workshop de Investigadores en Ciencias de la Computación (WICC2005)*. 2005: Departamento de Computación, Facultad de Cs. Exactas, Físico-Químicas y Naturales.Universidad Nacional de Río Cuarto. 22 y 23 de Mayo de 2005. Córdoba. Argentina. p. 193-197.

17. Comer, D., 1979. *The ubiquitous B-Tree*. Computing Surveys 11, pp. 121–137
18. Aho, A.V., J.E Hopcroft, y J.D Ullman, *The design and Analysis of Computer Algorithms*, R. (Addison-Wesley, MA), Editor. 1974.
19. Samet, H., *Applications of spatial data structures: computer graphics, image processing, and GIS*, R. (Addison-Wesley, MA), Editor. 1990.
20. Gagliardi, E.O., Dorzán, M., Carrasco, F., García Sosa, J., Gómez Barroso, J., Gutiérrez Retamal, G. *Métodos de Acceso Espacio-Temporal:nuevas propuestas basadas en métodos existentes*. in *Workshop de Investigadores en Ciencias de la Computación 2005*. WICC 2005. Río Cuarto.
21. Kuhn, W., *Geospatial Semantics: Why, of What, and How?*, in *Journal on Data Semantics III*, S.B. Heidelberg, Editor. 2005. p. 1-24.
22. Torres, M., Quintero,R.,Moreno, M., Fonseca,F., *Ontology-Dirven Description of Spatial Data for Their Semantic Processing*, in *GeoSpatial Semantic*. 2005, Springer-Verlag Berlin Heidelberg 2005: First International Conference, GeoS 2005, Mexico City. p. 242-249.
23. F. Mata, *Geographic Information Retrieval by Topological, Geographical and Conceptual Matching*, in *GeoSpatial Semantic*. 2007, Springer-Verlag Berlin Heidelberg 2007: Second International Conference, GeoS 2007, Mexico City. p. 98-113.
24. Levachkine, S., Guzman-Arenas, A, *Hierarchy as a New Data Type for Qualitative Variables*. Expert Systems with Applications: An International Journal, 2007. **Volume 32**: p. 899-910
25. Zimbrão, G., Moreira de Souza, J., Chaomey, R., and Teixeira, V de Almeida. Efficient processing of spatiotemporal queries in temporal geographical information systems, 2000.
26. Delgado, T. y Cruz, R. 2009. Construyendo Infraestructuras de Datos a Nivel Local. Habana, Cujae
27. Torres, M. and Levachkine, S. Generating Spatial Ontologies based on Spatial Semantics. Proceedings of International Workshop on Semantic Processing of Spatial Data, México City, México, pp. 163-178, 2003.
28. Torres, M., "Representación ontológica basada en descriptores semánticos aplicada a objetos geográficos". Instituto Politécnico Nacional, Ph. D. (2007)
29. F. Mata. "Recuperación y ponderación de información geográfica desde repositorios no estructurados conducidas por ontologías". Instituto Politécnico Nacional, Ph. D. (2009)
30. Hofweber, Thomas, Logic and Ontology, en Edward N. Zalta (en inglés), Stanford Encyclopedia of Philosophy (Spring 2009 Edition)
31. Gruber, T. (1992). A Translation Approach to Portable Ontology Specifications (Technical Report No. KSL 92-71). Stanford, CA: Knowledge Systems Laboratory, Stanford University.
32. Guarino, N. (1998). Formal Ontology and Information Systems. In N. Guarino (Ed.), Formal Ontology in Information Systems (pp. 3-15). Amsterdam, Netherlands: IOS Press.
33. Garea-Llano, E., Oliva.-S., R, Costales-Llerandi, C. (2009). Integración Ontológica de Datos Metadatos y Conocimiento en Sistemas de Información Geográfica como Herramienta para la Interpretación Semántica de la Información Espacial. Geomatica.
34. Vilches Blázquez, L.M.; Rodríguez Pascual, A. F.; Bernabé Poveda, M. A. (2006) Aplicaciones de la ingeniería ontológica a la gestión y análisis de la información geográfica. El acceso a la información espacial y las nuevas tecnologías geográficas: 545-556.
35. A. F. R. P. L. Vilches Blázquez, M. A. Bernabé Poveda, "Ingeniería Ontológica: El camino hacia la mejora del acceso a la información geográfica en el entorno web" Congreso JIDEE 2006. III Jornadas de la IDDE, Castellón, España, 18 de Octubre de 2006.
36. Ray R. Larson, Geographic Information Retrieval and Spatial Browsing, University of California, Berkeley. [cited; Available from: https://sherlock.sims.berkeley.edu/geo_ir/PART1.html, fecha de acceso: abril 2010
37. De Floriani, L., Marzano, P. & Puppo, E (1993). Spatial Queries and Data Models. In A. Frank & I. Campari (Eds). *Spatial Information Theory: A Theoretical Basis for GIS. (Lecture Notes in Computer Science # 716)*. Berlin: Springer-Verlag

Glosario de Términos

[1] **SOLAP**: Herramienta con soporte para el análisis y exploración de datos multidimensionales espacio-temporales.

[2] **XQuery**: Estadar proporcionado por el W3C como estándar para el lenguaje de consultas XML , pero carente de soporte espacial.

[3] **GML** : Geographic Markup Language (Lenguaje de marcado Geográfico).

[4] **TimeSlice**: Tipo de consulta espacio-temporal en la que el resultado de la consulta consiste de los objetos que se encuentran en una determinada área en un instante de tiempo dado.

[5] **Hiperónimo**: En semántica lingüística se denomina hiperónimo a aquel término general que puede ser utilizado para referirse a la realidad nombrada por un término más particular. Semánticamente, un hiperónimo no posee ningún rasgo semántico, o sema, que no comparta su hipónimo, mientras que éste sí posee rasgos semánticos que lo diferencian de aquél. Por ejemplo, coche posee sólo los semas [+vehículo], [+con motor] y [+pequeño tamaño], que comparte con descapotable, mientras que descapotable posee además el rasgo [+sin capota], que lo diferencia de coche.

[6] **Meronimia**: La meronimia es una relación semántica no-simétrica entre los significados de dos palabras dentro del mismo campo semántico. Se denomina merónimo a la palabra cuyo significado constituye una parte del significado total de otra palabra, denominada ésta holónimo. Por ejemplo, dedo es merónimo de mano y mano es merónimo de brazo; a su vez, brazo es holónimo de mano y mano es holónimo de dedo.

RT_034, octubre 2010

Aprobado por el Consejo Científico CENATAV

Derechos Reservados © CENATAV 2010

Editor: Lic. Lucía González Bayona

Diseño de Portada: Di. Alejandro Pérez Abraham

RNPS No. 2142

ISSN 2072-6287

Indicaciones para los Autores:

Seguir la plantilla que aparece en www.cenatav.co.cu

C E N A T A V

7ma. No. 21812 e/218 y 222, Rpto. Siboney, Playa;

Ciudad de La Habana. Cuba. C.P. 12200

Impreso en Cuba

