

RNPS No. 2142 ISSN 2072-6287 Versión Digital

REPORTE TÉCNICO Reconocimiento de Patrones

Integración semántica de datos espaciales con sistemas de información geográfica

Ing. Rainer Larín Fonseca, Dr. C. Eduardo Garea Llano

RT_011

Noviembre 2009







RNPS No. 2142 ISSN 2072-6287 Versión Digital

REPORTE TÉCNICO Reconocimiento de Patrones

Integración semántica de datos espaciales con sistemas de información geográfica

Ing. Rainer Larín Fonseca, Dr. C. Eduardo Garea Llano

RT_011

Noviembre 2009



Integración semántica de datos espaciales con sistemas de información geográfica

Ing. Rainer Larin Fonseca, Dr. C. Eduardo Garea Llano

Centro de Aplicaciones de Tecnología de Avanzada, 7a #21812 e/ 218 y 222, Siboney, Playa, Habana, Cuba rlarin@cenatav.co.cu

RT_011 CENATAV Fecha del camera ready: 28 de mayo de 2009

Resumen: El desarrollo de sistemas de información geográfica que faciliten la gestión de conocimiento es una de las principales tareas que son llevadas a cabo actualmente por muchas instituciones. Encontrar una forma de aprovechar todo el conocimiento que se encuentra subyacente en los datos espaciales abriría nuevas fronteras en la explotación de sistemas de este tipo. El uso de ontologías como la forma de representación del conocimiento más óptima a utilizar, es debido en gran medida a lo que prometen. Las ontologías proveen una comprensión compartida y consensuada del conocimiento de un dominio, que puede ser comunicado entre personas y sistemas heterogéneos. Por ello, el desarrollo de ontologías se presenta como el instrumento adecuado para alcanzar la integración semántica en el entorno de las infraestructuras de los datos espaciales de una manera mucho más abstracta, en la que el conocimiento juegue un papel fundamental. Es así como se han concebido sistemas de información geográfica (SIG) de nueva generación, SIGs en los que las ontologías participen como un componente activo dentro de su arquitectura. Estos sistemas de información geográfica gobernados por ontologías (SIGGOs) son construidos utilizando clases derivadas de las ontologías, que como consecuencia de esto es posible utilizar todo el conocimiento embebido, estas clases contienen atributos y funciones que darán soporte a todas las funcionalidades del sistema. La nueva generación de los SIG (SIGGOs) deberá ser capaz de resolver la interoperabilidad semántica, en la que un hecho puede tener un valor mayor que solamente su descripción.

Palabras claves: sistemas de información geográfica, semántica, datos espaciales, geo-ontología, integración de datos, sistemas de información geográfica gobernados por ontologías.

Abstract: The development of Geographic Information Systems which make easier the knowledge management is one of the main tasks that are accomplished nowadays by a lot of institutions. Finding a way of making the best use of all the knowledge embedded in the space data, would open new frontiers in suchlike systems. The use of ontologies as the optimum way of representation of knowledge is greatly due to its potentialities. Ontologies supply a joint and consensual understanding of the knowledge of a domain, which can be communicated between persons and heterogeneous systems. For it, the development of ontologies shows up like the adequate tool to reach the semantic integration at the context of the infrastructures of the space data in a lot more abstract way, where it plays a fundamental role in the knowledge one. It is just like the way that the geographic information systems (SIG) of new generation have been conceived, SIGs where the ontologies are an active component within their architecture. These Ontology-Driven Geographic Information Systems (SIGGOs), are constructed using classes derived from ontologies, and as consequences of this, it is possible the use of all the embedded knowledge in it; these classes contain attributes which will give technical support to all the system's functionalities. The SIG's new generation (SIGGOs) will have to be able to solve the semantic interoperability, in which an event can have a bigger value than its description only.

Keywords: Geographic Information Systems, Semantic, Spatial Data, Geo-Ontology, Data Integration, Ontology-Driven Geographic Information Systems.

1 Introducción

En los últimos años, las técnicas de organización y búsqueda de la información geográfica han cobrado gran importancia debido, en gran medida, al gran cúmulo de datos que se generan diariamente y que es una necesidad primordial que sean procesados para poder extraer toda la información útil que sea posible. A pesar de la gran cantidad de datos geográficos disponibles, no existe un formato estándar de almacenamiento capaz de ser utilizado por todo tipo de sistemas geográficos con éxito, dificultando así su integración y uso. Los datos geográficos poseen características específicas que dificultan su manipulación, esas características van generalmente enfocadas en que presentan gran heterogeneidad y volumen de almacenamiento, la ubicación espacial, o sea, estar en una localización sobre la superficie de la tierra referida a un sistema de coordenadas, la temporalidad y las relaciones espaciales con otros objetos o datos. A esto se suman los principales problemas presentados por los usuarios que buscan información geográfica, ya que los datos pueden estar incompletos o ser insuficientes así como también pueden contener errores que hayan sido introducidos durante su recolección. Existe, por tanto, una creciente necesidad de encontrar una solución que permita la integración de datos geográficos de una manera mucho más abstracta, en la que el conocimiento juegue un rol esencial y se explote con mayor efectividad la información semántica existente que se encuentra subvacente en los datos almacenados.

Teniendo en cuenta la importancia que tiene el conocimiento subyacente en los datos es necesario considerar cuidadosamente cómo representarlo para que sea factible su generación y uso, de manera que se reduzcan las problemáticas existentes en la obtención de información geográfica. Existen varias formas de representación del conocimiento y en general, éstas están limitadas o enfocadas a un dominio semántico específico. Entre las primeras y más usadas formas de representación del conocimiento se encuentran las redes semánticas (sistema NUDE, creada en 1956 por Richens [1]) y los marcos conceptuales [2]). A finales de los 80 aparece un nuevo modelo como medio para la compartición y re-uso de conocimiento: las ontologías.

Desde comienzos de los noventa las ontologías se han convertido en un tema de investigación importante en diferentes campos, donde podemos incluir ingeniería del conocimiento, procesamiento del lenguaje natural, representación del conocimiento. Más recientemente, el uso de ontologías ha adquirido gran aceptación en campos como integración inteligente de información, sistemas cooperativos de información, recuperación de información, comercio electrónico, y gestión de conocimiento. Una de las razones por las que las ontologías se han hecho tan populares es, en gran medida, debido a lo que prometen:

Una comprensión compartida y común de algún dominio que puede ser comunicado entre individuos y aplicaciones.[3].

Las formas actuales de organización de la geo-información (catálogos de fenómenos y tesauros) no son eficaces para solucionar los problemas derivados de la integración semántica de conjuntos de geo-datos. Por ello, el desarrollo de ontologías se presenta como el instrumento adecuado para alcanzar interoperabilidad semántica en el entorno de las infraestructuras de datos espaciales[4].

En este sentido se han de utilizar ontologías de tipo geográfico (geo-ontologías) ya que éstas están especialmente diseñadas para obtener una mejor comprensión de la estructura del mundo

geográfico [5]. Las geo-ontologías ofrecen una descripción de entidades en el ámbito geográfico y se diferencian de otras ontologías debido a la presencia predominante de las relaciones topológicas, espaciales y temporales que se añaden y que son muy importantes en el medio ambiente espacial.

Sumado a esto surgen nuevas tecnologías como los sistemas de información geográfica (SIG), que permiten presentar los datos en forma consolidada y ubicados geográficamente. La integración de mapas e imágenes del territorio, de bases de datos asociadas a ellos, de equipos informáticos y el software adecuado para la visualización y análisis, conforman lo que conocemos como un sistema de información geográfica (SIG) [6]. Los SIG son una herramienta que fortalece la gestión de planificación, administración y prestación de servicios de información. Su capacidad para almacenar, recuperar, analizar, modelar y representar grandes extensiones de terreno con grandes volúmenes de datos espaciales es lo que ha determinado su éxito frente a una gran gama de aplicaciones similares. El SIG funciona como un recopilatorio de información geográfica (datos alfanuméricos) que se encuentra asociada por un identificador común a los objetos gráficos de un mapa digital. De esta forma, señalando un objeto se conocen sus atributos e, inversamente, preguntando por un registro se puede saber su localización en la cartografía.

Un SIG que utiliza ontologías para generar nueva información a partir de un conjunto previo de datos es lo que se denomina SIGGO (sistema de información geográfica gobernado por ontologías) o SIGBO (sistema de información geográfica basado en ontologías). De acuerdo con [7], la nueva generación de los SI deberá ser capaz de resolver la interoperabilidad semántica, en la que un hecho puede ser más que una descripción y los SIG no están aparte de este problema. El concepto de SIG basados en ontologías (SIGBO) abre nuevas maneras de pensar sobre las ontologías y los SIG, y cubre las dimensiones estructurales y temporales de los SIG.

Definiciones

Para comprender de una mejor manera los asuntos abordados en este trabajo se definirán algunos términos:

- Sistema de información (SI): Es un sistema (informático o no) que está creado para dar respuestas a preguntas no predefinidas de antemano. Un SI incluye, por tanto, una base de datos, una base de conocimientos y un sistema de interacción con el usuario.
- Sistemas de apoyo a la decisión (SAD): En ellos, los datos y las bases de conocimiento se estructuran para servir de ayuda en la toma de decisiones, dando información de lo que podría ocurrir en caso de tomar una decisión determinada.
- Sistemas de información geográfica (SIG): El término sistema de información geográfica (SIG), se presta a confusión por carecer de una definición mundialmente aceptada. Una definición ampliamente aceptada menciona que: "Es un sistema compuesto de hardware, software, información geográfica y procedimientos específicos para soportar la captura, el análisis, el modelado y el despliegue de datos espacialmente referenciados, para la solución de los problemas complejos del manejo y planeamiento territorial" [8]. Los SIG son, en algunos casos, simultáneamente un sistema de información y un sistema de apoyo a la decisión. Los SIG permiten relacionar información de cualquier tipo con una localización geográfica (mapa). Con un SIG, instituciones gubernamentales o empresas pueden relacionar información demográfica de censos con mapas políticos; médicos y hospitales pueden relacionar mapas de

enfermedades con condiciones de salubridad; autoridades pueden relacionar mapas de lugares donde se cometieron crímenes con patrones de criminalidad; personal de servicios de emergencia pueden relacionar mapas de áreas de riesgo con información sobre inundaciones o incendios forestales. Los ejemplos de utilización de los SIG son virtualmente ilimitados, como ilimitado es el tipo y cantidad de información que se puede asociar con lugares particulares de la tierra.

- Ontologías: Comenzando por la definición clásica de Gruber "Una ontología es una especificación explícita de una conceptualización" [9] Las ontologías como "el estudio metafísico de la naturaleza del ser y la existencia" es tan antigua como la disciplina de la filosofía. Recientemente, la ontología se ha definido como "la ciencia de lo que es, de los tipos y estructuras de objetos, propiedades, eventos, procesos, y relaciones en cada área de la realidad" [10].
- Ontologías genéricas o meta-ontologías: proporcionan términos genéricos reutilizables en diferentes dominios, por ejemplo, los términos: estado, evento, acción, componente, etc. [11].
- Ontologías de dominio: Expresan conceptos que son específicos de un dominio determinado. Los conceptos en este tipo de ontologías son definidos usualmente como especializaciones de conceptos existentes en ontologías genéricas [11].
- Ontologías de aplicación: Contienen todas las definiciones que son necesarias para modelar los conocimientos requeridos por una aplicación particular. Incluyen conceptos tomados de ontologías de dominio y genéricas, a menudo definidas utilizando el vocabulario indicado en ontologías de representación. Pueden contener extensiones de métodos y tareas específicas. Las ontologías de aplicación tienden a ser menos que las anteriores, pues son especificaciones concretas del dominio que se necesita para realizar una tarea particular en ese dominio determinado [11].
- Geo-ontologías: Las geo-ontologías cumplen con todas la características de una ontología convencional, pero tienen además propiedades propias de este dominio, por ejemplo, un par de coordenadas (x,y) que representa la posición geográfica de un objeto. También incluye una serie de relaciones topológicas con una semántica determinada, por ejemplo, en el siguiente planteamiento: "el río cruza el bosque", podemos encontrar los conceptos (objetos) río y bosque, y también encontramos una relación, cruza, que actúa sobre los objetos río y bosque [12]. A continuación se hará referencia a geo-ontologías y ontologías, indistintamente.
- **Interoperabilidad:** Capacidad de un sistema para compartir e intercambiar informaciones y aplicaciones.
- Taxonomía: La palabra taxonomía tiene su origen en dos términos griegos, a saber, taxis (orden) y nomos (tratado) y esta palabra proviene de la Filosofía. Taxonomía es la ciencia que estudia la división en grupos ordenados o categorías, ciencia que trata de los principios, métodos y fines de la clasificación. Se aplica, en particular, para la ordenación jerarquizada y sistemática [13]. Desde un punto de vista ontológico, una taxonomía es una organización ontológica basada en una relación de orden a través de la que se agrupan las entidades y son subsumidas por clases de más alto nivel. En general, las taxonomías han sido importantes para modelar esquemas de bases de datos, sistemas basados en conocimiento y vocabularios semánticos.

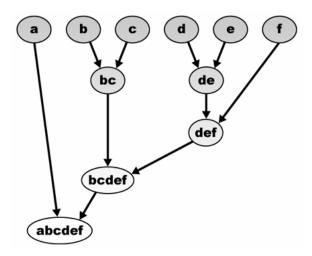


Fig. 1. Ejemplo de taxonomía

Datos geográficos: Es un conjunto de señales o signos con un significado particular. Es la unidad más pequeña de información, "es una colección de hechos considerados de forma aislada"[14].

Los datos pueden ser:

- Numéricos
- Alfanuméricos: letras y números
- Gráficos: mapas, fotografías, etc.
- Geográficos: están orientados espacialmente
- Dato Geoespacial: Es un tipo de dato que hace referencia a un lugar específico, o sea, a un espacio geográfico cuya ubicación es conocida mediante un sistema de coordenadas. Los datos geoespaciales constan de dos componentes: espacial y temática.
 - Componente espacial: Las observaciones tienen dos aspectos en referencia a su localización:
 - **Localización absoluta:** Basada en un sistema de coordenadas. Ejemplo: El centro "X" se encuentra en las coordenadas X, Y.
 - Las relaciones topológicas: Ubicación con respecto a otras entidades. Ejemplo: El centro "X" se encuentra localizado entre centro "Y" y el centro "Z"
 - Componente Temática: Las variables o atributos de las entidades se pueden estudiar considerando el aspecto temático (estadística), su localización (análisis espacial) o ambos.
- Árbol R o R-Tree [15]: Los árboles R son estructuras de datos de tipo árbol, basados en el principio de descomposición recursiva del espacio [16], que se utilizan para métodos de acceso espacial, es decir, para indexar información multidimensional; por ejemplo, las coordenadas (x, y) de un lugar geográfico en especifico. En un R-Tree no se almacenan los objetos espaciales en forma directa sino que se almacena su MBR (Minimum Bounding Rectangle), es decir, el menor rectángulo que contiene al objeto en cuestión. Un problema con aplicación práctica en el mundo real podría ser: "Encontrar todos los asentamientos en un radio de dos kilómetros alrededor de la posición actual".

6 Ing. Rainer Larin Fonseca, Dr. C. Eduardo Garea Llano

La estructura de datos divide el espacio de forma jerárquica en conjuntos, posiblemente superpuestos.

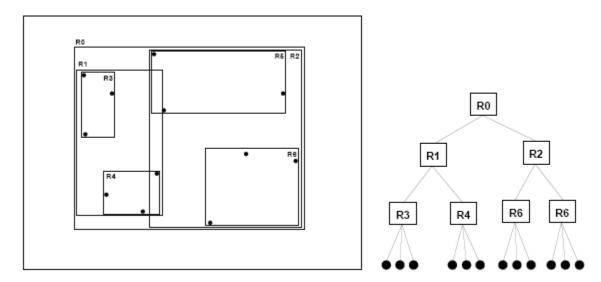


Fig. 2. Estructura de un Árbol-R para rectángulos 2D

• Árbol de cuadrantes o Quad-Tree [17]: El nombre de Quad-Tree viene del inglés quadrant trees (árboles de cuadrantes) y se llaman así porque, justamente, para organizar datos espaciales divide en cuadrantes el espacio en el que se encuentran dichos datos. Por ejemplo, cuando los datos son puntos en el plano, los quad-trees se construyen tomando un punto como raíz del árbol y trazando las líneas perpendiculares a los ejes que pasan por dicho punto. Este proceso divide el plano en cuatro secciones (los cuadrantes), cada una de ellas corresponde a un hijo de la raíz del árbol. Cada una de estas secciones se somete recursivamente al mismo proceso hasta que las regiones consten de un único punto. En la figura que se presenta a continuación podemos ver un ejemplo de la subdivisión recursiva de cada nodo.

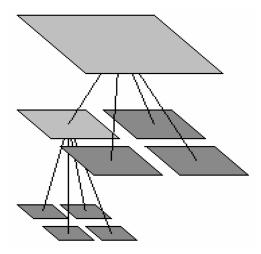


Fig. 3. Ejemplo de subdivisión de cada nodo de un Quad-Tree

Según Samet [18], el término Quad-Tree es usado para describir una estructura de datos jerárquica que está basada en el principio de la descomposición recursiva del espacio en un plano 2D, como se muestra a continuación:

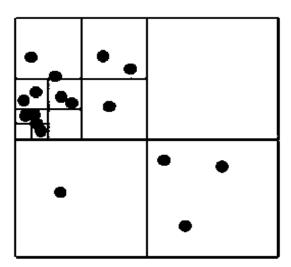


Fig. 4. Esquema de la descomposición recursiva del espacio del Quad-Tree

Actualmente son usados para representar puntos, áreas, curvas, superficies y volúmenes. La descomposición puede ser en partes iguales en cada nivel (llamada descomposición regular), o puede estar regida por los datos de entrada. El tipo de Quad-Tree más utilizado es aquél que se caracteriza porque en cada nivel de descomposición se divide un cuadrante en cuatro cuadrantes iguales. La subdivisión prosigue hasta llegar a un cuadrante que es representable directamente o hasta que se llega a una subdivisión mínima, denominada resolución del Quad-Tree. Existen varios tipos de Quad-Tree. Éstos se diferencian en el tipo de datos que representan: puntos, líneas, regiones,

rectángulos, polígonos, etc. La diferencia entre estos Quad-Tree es la "condición de uniformidad" necesaria para no continuar realizando divisiones de los cuadrantes.

- **Indexación espacial:** Usar una estructura de datos que permita descartar rápidamente subconjuntos grandes del conjunto de objetos, y buscar en conjuntos reducidos.
- Procesamiento del lenguaje natural: El concepto de procesamiento del lenguaje natural (PLN) hace referencia a las técnicas de tratamiento del lenguaje y su aplicación en diversas áreas por medio de métodos computacionales. Paralelamente se hace uso de otros términos próximos, como lingüística computacional o ingeniería lingüística. Eduardo Sosa lo define de la siguiente forma:

El procesamiento del lenguaje natural (PLN) se concibe como el reconocimiento y utilización de la información expresada en lenguaje humano a través del uso de sistemas informáticos[19].

El PLN es un área de investigación en continuo desarrollo, se aplica en la actualidad en diferentes actividades como la traducción automática, sistemas de recuperación de información, elaboración automática de resúmenes, interfaces en lenguaje natural, etc. Aun siendo evidente que los obstáculos a superar en el estudio del tratamiento del lenguaje son considerables, los resultados obtenidos y la evolución en los últimos años sitúan al PLN en posición para liderar una nueva dimensión en las aplicaciones informáticas del futuro: los medios de comunicación del usuario con el ordenador pueden ser más flexibles y el acceso a la información almacenada más eficiente.

- 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. Por lo tanto:
 - X es merónimo de Y si X forma parte de Y.
 - X es merónimo de Y si X es una sustancia de Y.
 - X es merónimo de Y si X es un miembro de Y.
 - azul es merónimo de color.
 - doctor es merónimo de oficio.
 - dedo es merónimo de mano.
- Hipónimo: Se denomina hipónimo (del griego: υπονύμιον, que literalmente significa 'pocos nombres') a aquella palabra que posee todos los rasgos semánticos, o semas, de otra más general, su hiperónimo, pero que añade en su definición otros rasgos semánticos que la diferencian de la segunda. Por ejemplo, descapotable es hipónimo de coche, ya que comparte todos sus rasgos mínimos, a saber [+vehículo], [+con motor], [+pequeño tamaño], etcétera, pero añade a estos el rasgo [+sin capota]. Otro ejemplo podría ser:
 - Hiperónimo: día
 - Hipónimo: lunes, martes, miércoles, etc.

En el mundo de las tecnologías de información se podría encontrar el símil en la orientación a objetos, donde el hipónimo podría ser una clase (u objeto, dependiendo del caso) y el hiperónimo la superclase (o clase).

• **Hiperónimo**: 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. Al redactar un texto conviene utilizar hiperónimos para evitar la repetición de palabras ya empleadas anteriormente, como se hace en el siguiente ejemplo:

• De repente, un descapotable rojo paró frente al banco. Del automóvil salieron dos individuos encapuchados, mientras otro esperaba en el vehículo.

2 Sistemas de información geográfica

Al igual que la propia geografía, es dificil definir el término sistema de información geográfica (SIG) ya que engloba la integración de áreas muy diversas. Podría definirse *a priori* como un poderoso conjunto de herramientas para recoger, almacenar, recuperar, transformar, y mostrar datos espaciales del mundo real [20]. Por esto no existe una única definición de SIG totalmente consensuada [21]. Una definición de SIG bastante aceptada es la redactada por el NCGIA (*National Centre of Geographic Information and Analysis*):

"Un SIG es un sistema de hardware, software y procedimientos elaborados para facilitar la obtención, gestión, manipulación, análisis, modelado, representación y salida de datos espacialmente referenciados, para resolver problemas complejos de planificación y gestión" [22].

Un SIG es, por tanto, un conjunto de mapas de la misma porción de territorio donde un lugar concreto tiene la misma localización en todos los mapas incluidos en el sistema. Lo más característico de un SIG es su capacidad de generar nueva información a partir de un conjunto previo de datos mediante su manipulación y reelaboración. La razón fundamental para utilizar un SIG es la gestión de información espacial. El sistema permite separar la información en diferentes capas temáticas y las almacena independientemente, permitiendo trabajar con ellas de manera rápida y sencilla, y facilitando al profesional la posibilidad de relacionar la información existente a través de la topología de los objetos, con el fin de generar otra nueva que no podríamos obtener de otra forma.

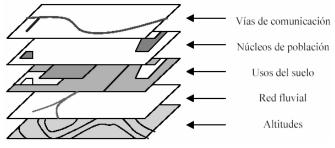


Fig. 5. Diferentes capas temáticas referidas a una misma localización

Por esto se puede decir que para el manejo de datos espaciales, los sistemas de información geográfica son el paso adelante más importante desde la invención del mapa. De manera más simple, un SIG se puede definir como un conjunto de mapas de la misma porción de territorio donde un lugar concreto tiene la misma localización geográfica en todos los mapas incluidos en el sistema, tal como se muestra en la siguiente figura:

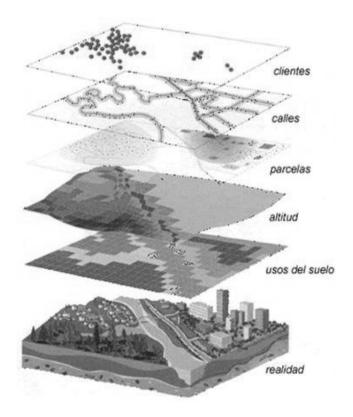


Fig. 6. Diferentes capas temáticas referidas a una misma localización

2.1 Subsistemas o componentes lógicos de los SIG

Un SIG es, entre otras cosas, un programa con capacidades específicas que se pueden resumir en los siguientes subsistemas o componentes lógicos:

- Entrada de información: Son los procedimientos que permiten convertir la información geográfica del formato analógico (generalmente en forma de mapas tradicionales) al formato digital, esta conversión debe realizarse manteniendo todas las características iniciales de los datos espaciales, es por esto que se incluyen además procedimiento de eliminación de ruido o redundancia en la información que se incorpora al SIG.
- Salida y/o representación gráfica y cartográfica de la información espacial: Se refiere a las actividades que sirven para mostrar al usuario los datos y los resultados de

las operaciones analíticas realizadas sobre ellos. Permiten la obtención de mapas, gráficos, tablas numéricas y otros resultados en diferentes soportes (papel, pantalla, etc.)

- **Gestión de la información espacial:** Son las funciones con las que se extraen de una base de datos las porciones de información que se necesitan en cada momento.
- Funciones Analíticas: Son el elemento más característico de un SIG, facilitan el procesamiento de los datos integrados en él, de modo que sea posible obtener mayor información y con ello mayor conocimiento.

2.2 Aplicaciones de los SIG

En los SIG se emplea la tecnología de computadores para integrar, manipular y visualizar una amplia gama de datos capaces de crear una imagen de la geografía, medio ambiente y características socioeconómicas de una zona. Hoy en día es común utilizar los sistemas de información geográfica en una gran diversidad de actividades; desde el trazado básico de mapas hasta la exploración y desarrollo de recursos; desde la ordenación del medio ambiente hasta la planificación y administración de los sistemas de transporte y telecomunicaciones, de la infraestructura de servicios públicos, del desarrollo urbano y de la explotación del terreno. Algunas aplicaciones específicas de los SIG son:

- Cartografía de localizaciones: Los SIG se pueden utilizar en la cartografía de localizaciones. Los SIG permiten la creación de mapas por medio de cartografía automatizada, captura de datos y herramientas de análisis.
- Mapas cuantitativos: Mapas de población que se utilizan para localizar lugares que reúnen ciertos criterios demográficos y tomar decisiones, o para ver las relaciones existentes entre diferentes lugares. Esto proporciona un nivel de información adicional más allá de los simples mapas de localizaciones de entidades.
- Mapas de densidades: Aunque las concentraciones se pueden ver simplemente en un mapa de localización de entidades, en aquellas áreas donde existen muchas de ellas se hace complicado ver qué áreas tienen mayores concentraciones que otras. Un mapa de densidad permite medir el número de entidades en una unidad de área uniforme, tal como el metro o el kilómetro cuadrado, de forma que se puede ver claramente la distribución.
- Cálculo de distancias: Los SIG se pueden utilizar para saber qué está pasando en un radio determinado alrededor de una entidad.
- Cartografía y detección del cambio: Los SIG se pueden utilizar para cartografíar el cambio en una zona para predecir condiciones futuras, tomar decisiones, o evaluar los resultados de una acción o una política concreta.

Un SIG es un mecanismo informático para manejar información y/o datos y facilitar el entendimiento de los fenómenos y objetos espaciales. Estos objetos están dotados de propiedades intrínsecas que se pueden medir y esos datos obtenidos pueden a su vez descomponerse en determinados elementos como:

- **Observación o soporte:** Una entidad de la realidad sobre la que se observa un fenómeno.
- Variable o atributo temático: Puede ser cualquier hecho que adopte diferentes modalidades en cada observación.

La característica fundamental de estos datos es que el soporte (la unidad de observación) está localizado en el espacio. Las unidades de observación se pueden subdividir en dos grandes grupos:

- Naturales: Son aquellas donde la referencia espacial es más intrínseca al propio hecho observado, ej.: la subdivisión del espacio por el uso de los suelos.
- Artificiales: Generalmente creadas por el hombre, donde la referencia espacial es más extrínseca, ej.: la partición del espacio en unidades administrativas, En este caso no existe ninguna razón "natural" para establecer fronteras de separación, la definición humana de límites espaciales es siempre arbitraria.

Los datos para aplicaciones SIG incluyen:

- datos digitalizados y escaneados
- bases de datos
- muestreo de campo con GPS
- imágenes de satélite y fotografía aérea

El mapa tradicional es una representación analógica de la realidad, por tanto, es necesario convertirlo al formato digital para poder procesarlo por un SIG; para una correcta representación se necesita la resolución de estas cuestiones:

- Geo-codificación de los datos: Es el proceso mediante el que un objeto recibe directa o indirectamente una etiqueta que especifica su posición espacial con respecto a un punto común o marco de referencia.
- Descripción en términos digitales de las características espaciales: Descripción de la posición geométrica de cada objeto y de las relaciones espaciales que mantiene con los restantes objetos geográficos.

Los SIG trabajan principalmente con dos tipos de representaciones de los datos espaciales, la representación vectorial y la raster.

2.3 Representación vectorial

El modelo vectorial [23] es una estructura de datos utilizada para almacenar datos geográficos. Los datos vectoriales constan de líneas o arcos, definidos por sus puntos de inicio y fin, y puntos donde se cruzan varios arcos, los nodos. La localización de los nodos y la estructura topológica se almacena de forma explícita. Las entidades quedan definidas por sus límites solamente y los segmentos curvos se representan como una serie de arcos conectados. El almacenamiento de los vectores implica el almacenamiento explícito de la topología, sin embargo, sólo almacena aquellos puntos que definen las entidades y todo el espacio fuera de éstas no está considerado. De acuerdo a las peculiaridades de este tipo de representación de datos, los objetos geográficos se muestran explícitamente y junto a sus características espaciales se asocian sus valores temáticos. Un ejemplo de representación vectorial se muestra a continuación:

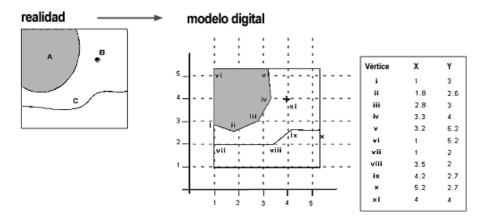


Fig. 7. Representación vectorial

Los datos geográficos se representan en forma de coordenadas. Las unidades básicas de información geográfica en los datos vectoriales son puntos, líneas (arcos) y polígonos. Cada una de éstas se compone de uno o más pares de coordenadas, por ejemplo, una línea es una colección de puntos interconectados y un polígono es un conjunto de líneas interconectadas.

- 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 nnúmeros a lo largo de n-ejes que expresan una localización concreta en el espacio ndimensional. 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.
- Arco: Término específico de ARC/INFO que se utiliza como sinónimo de línea.
- **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.

Existen diferentes estructuras para almacenar los datos en la representación vectorial. Cada una de ellas tiene diferentes ventajas e inconvenientes que serán mostradas a continuación:

- **Lista de coordenadas:** Se almacenan únicamente los dos pares de coordenadas que definen un segmento recto, para los objetos geográficos se registra su nombre, el número de vértices que definen su frontera y las coordenadas X, Y de cada vértice.
 - Ventajas:
 - Sencilla de comprender.
 - Fácil de operar.
 - Muy utilizada en la cartografía automática.

Desventajas:

- Muchos vértices están duplicados, el mismo vértice puede ser registrado de forma diferente en cada segmento recto.
- No almacena topología.
- Tiene muchas operaciones redundantes, que implican espacios de almacenamiento mayores.

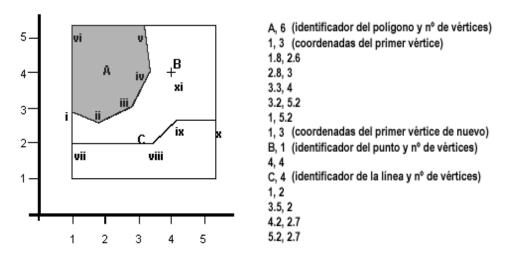


Fig. 8. Representación de la estructura de datos: Lista de coordenadas

- **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.
 - Ventajas:
 - No hay operaciones redundantes.

Desventajas:

 La topología de los objetos geográficos no se conoce del modo suficiente.

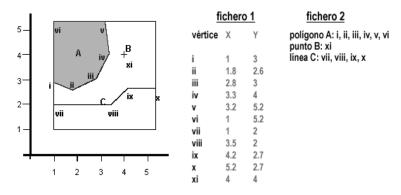


Fig. 9. Representación de la estructura de datos: Diccionario de vértices

• Organización DIME ("Dual Independent Map Encoding"): Se crea una lista de vértices con nombres y coordenadas de cada objeto, luego los objetos lineales se codifican indicando en qué vértices empiezan y terminan. Los segmentos rectos que delimitan polígonos se codifican indicando su nombre en el vértice que se inicia y en el que termina.

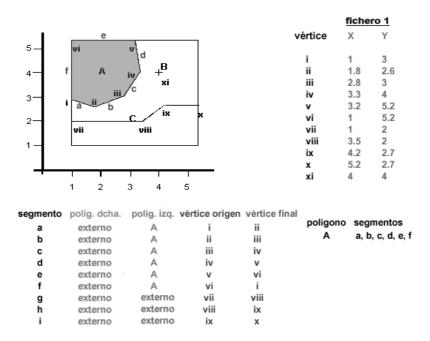


Fig. 10. Representación de la organización DIME

- **Estructura arco / nodo:** El elemento base ya no es el segmento recto sino el arco, otro elemento fundamental es el nodo; esta estructura se ha convertido en la más característica del mercado para los sistemas vectoriales. Los polígonos se codifican indicando los arcos que les rodean. El arco se registra indicando sus nodos y el polígono de la izquierda y el de la derecha.
 - Arco: Formado por una sucesión de segmentos rectos.
 - Nodo: Cada uno de los vértices en que se cruzan tres o más arcos y el punto terminal de una línea o arco.

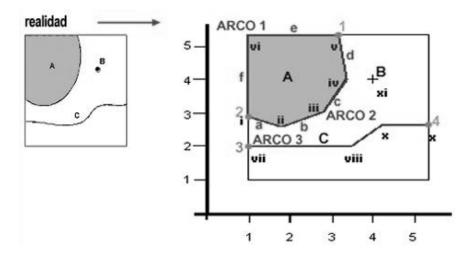


Fig. 11. Representación de la estructura de datos: Arco / Nodo

	Fichero 1. Coo	rdenadas de los nodos	y vértices de cada arco
Arco	Nodo origen	Vértices intermedios	Nodo final
1	3.2, 5.2	1, 5.2	1, 3
2	1, 3	1.8, 2.6 2.8, 3 3.3, 4	3.2, 5.2
3	1, 2	3.5, 2 4.2, ,2.7	5.2, 2.7

Tabla 1. Fichero 1. Coordenadas de los nodos y vértices de cada arco

	F	ichero 2. Top	ología de arcos	
Arco	Nodo origen	Nodo final	Polígono derecho	Polígono izquierdo
1	1	2	Externo	A
2	2	1	A	Externo
3	3	4	Externo	Externo

Tabla 2. Fichero 2. Topología de arcos

Fichero 3. Topo	logía de polígonos
Polígonos	Arcos
A	1, 2

Tabla 3. Fichero 3. Topología de polígonos

Fichero 4. Topo	logía de los nodos
Nodos	Arcos
1	1, 2
2	1, 2
3	3
4	4
5	5

Table 4. Fichero 4. Topología de los nodos

2.4 Representación Raster

El modelo raster [23] es un método para el almacenamiento, el procesado y la visualización de datos geográficos. Cada superficie a representar se divide en filas y columnas, formando una malla o rejilla regular (matriz). Cada celda ha de ser rectangular, aunque no necesariamente cuadrada. Cada celda de la rejilla guarda tanto las coordenadas de la localización como el valor temático. La localización de cada celda es implícita, dependiendo directamente del orden que ocupa en la rejilla, a diferencia de la estructura vectorial en la que se almacena de forma explícita la topología. Las áreas que contienen idéntico atributo temático son reconocidas como tal, aunque las estructuras raster no identifican los límites de esas áreas como polígonos en sí.

Los datos raster son una abstracción de la realidad, representan ésta como una rejilla de celdas o píxeles, en la que la posición de cada elemento es implícita según el orden que ocupa en dicha rejilla. En el modelo raster, el espacio no es continuo sino que se divide en unidades discretas. Esto le hace especialmente indicado para ciertas operaciones espaciales, por ejemplo, las superposiciones de mapas o el cálculo de superficies. Las estructuras raster pueden implicar en ocasiones un incremento del espacio de almacenamiento, ya que almacenan cada celda de la matriz sin tener en cuenta si se trata de una entidad o simplemente de un espacio "vacío".

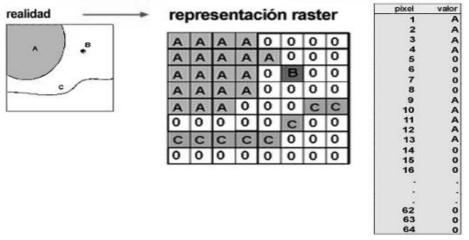


Fig. 12. Representación raster

Existen diferentes estructuras para almacenar los datos en la representación raster. Cada una de ellas será mostrada a continuación:

• **Enumeración exhaustiva:** El valor de cada celda se registra individualmente, lo que supone gran abundancia de información, que es en muchos casos reiterativa.

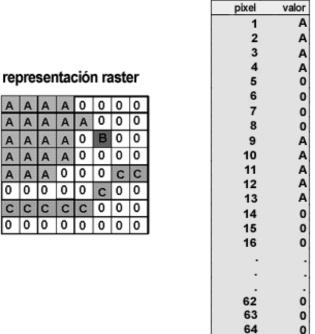


Fig. 13. Representación de la estructura de datos: Enumeración exhaustiva

Codificación <*run-length>*: En un mapa raster, lo usual es que muchos píxeles de una misma fila o columna tengan los mismos valores, por lo tanto, el procedimiento consiste en recoger para cada fila el valor temático que existe y las columnas entre las que se produce.

representación raster

A	A	A	A	Го	0	0	0
Α	Δ	Δ	Δ	Δ	0	n	0
10000				0		١ž	1
Α	Α	Α	Α	٧		ľ	10
Α	A	A	Α	0	0	0	0
Α	A	A	0	0	0	C	C
0	0	0	0	0	С	0	0
С	С	С	C	С	0	0	0
0	0	0	0	0	0	0	0

Fig. 14. Representación de la estructura de datos: Codificación < run-length>

2.5 Ventajas e inconvenientes de cada tipo de representación

• Representación vectorial:

– Ventajas:

- Es más adecuada para le representación de mapas precisos.
- Es más compacta en cuanto a volumen de información.

- Desventajas:

- La organización de los datos es muy compleja.
- Cualquier cambio implica modificar muchos elementos.
- La realización de buenos gráficos exige disponer de aparatos complejos y caros.

• Representación raster:

Ventajas:

- La organización de los datos es muy simple.
- Los gráficos, aunque deficientes, se pueden realizar con dispositivos baratos y sencillos.

Desventajas:

- Presenta gran volumen de almacenamiento.
- No reconoce la existencia de objetos geográficos y por tanto, en aplicaciones donde sea esencial su empleo, este modelo tiene pocas posibilidades de ser utilizado.

Los SIG además poseen procedimientos para llevar de datos rasters a vectoriales y viceversa:

Rasterización de datos vectoriales:

Es el procedimiento a través del cual se convierten datos vectoriales (puntos, líneas y polígonos) a formato raster, formados éstos por celdas (píxeles) con un valor temático discreto en cada una. Es más sencillo que el procedimiento contrario, la conversión de datos raster a formato vectorial.

• Vectorización de datos raster:

Es el procedimiento que convierte una imagen formada por celdas en un archivo vectorial. Puede o no incluir la creación de topología.

La utilidad final del SIG radica en su capacidad para elaborar modelos, es decir, construir modelos del mundo real a partir de las bases de datos digitales y utilizar esos modelos para simular el efecto de un proceso específico en el tiempo para un determinado escenario. La construcción de modelos constituye un instrumento muy eficaz para analizar las tendencias y determinar los factores que influyen en ellas, o para exponer las posibles consecuencias de las decisiones o proyectos de planificación que repercuten en la utilización y ordenación de los recursos. En un SIG, la información espacial y la información descriptiva existen como una sola entidad. Gracias a esta estructura es que un SIG puede manipular grandes volúmenes de información espacial en diferentes niveles de detalles, permitiendo ejecutar muchas funciones de análisis, manipuleo y modelajes. Un sistema de información geográfica (SIG), es un sistema digital capaz de almacenar, manipular y desplegar información referenciada geográficamente y por tanto, provee al usuario las herramientas y operadores especiales necesarios para modelar sobre el mundo real.

3 Estándar en los formatos de datos de los SIG

3.1 Consorcio OpenGis

El *Open Geospatial Consortium* (OGC) fue creado en 1994 y agrupa a más de 250 organizaciones públicas y privadas. Su fin es la definición de estándares abiertos e interoperables dentro de los sistemas de información geográfica. Persigue acuerdos entre las diferentes empresas del sector que posibiliten la interoperación de sus sistemas de geo-procesamiento y facilitar el intercambio de la información geográfica en beneficio de los usuarios. Anteriormente, fue conocido como *Open GIS Consortium*. El estándar propuesto por *OpenGis Consortium* es muy abarcador y reúne todos los elementos que son necesarios para documentar debidamente datos geográficos. Este estándar define una terminología básica e imprescindible para la correcta documentación de la información geográfica.

Las especificaciones más importantes surgidas del OGC son:

- GML [24] Lenguaje de marcado geográfico (no confundir con lenguaje de marcado generalizado, también GML) acrónimo inglés de *Geography Markup Language* (lenguaje de marcado geográfico). Es un sublenguaje de XML descrito como una gramática en XML *Schema* para el modelaje, transporte y almacenamiento de información geográfica. Su importancia radica en que a nivel informático se constituye como una lengua franca para el manejo y trasvase de información entre los diferentes software que hacen uso de este tipo de datos, como los sistema de información geográfica.
- WFS[25] El Web Feature Service Version o servicio de publicación de objetos (WFS) permite recuperar y modificar (consultar, insertar, actualizar y eliminar) datos espaciales en formato vectorial codificados en Geography Markup Language. Cada servicio puede manejar uno o más tipos de fenómenos, cada uno de ellos tiene asociado un XML Schema que describe su estructura. Para acceder y manipular estos fenómenos geográficos, el estándar WFS define interfaces que operan mediante la utilización de HTTP como plataforma de cómputo distribuido. Gracias a esos interfaces, un usuario o servicio web puede combinar, utilizar y gestionar información acerca de los fenómenos que constituyen los mapas, siempre que cumplan los siguientes requisitos:
 - Los interfaces deben definirse en XML.
 - Se debe usar GML para servir la información sobre fenómenos.
 - Un WFS debe ser capaz, como mínimo, de ofrecer fenómenos utilizando GML.
 - El predicado o filtro de lenguaje debe definirse en XML y derivarse de CQL2, como se indica en el OpenGIS Catalogue Interface Implementation Specification.

El almacén de datos utilizado para almacenar fenómenos geográficos puede ser opaco para la aplicación cliente, siendo el único acceso a los datos a través del interfaz del WFS. La única función de un WFS cuando interacciona con el sistema de almacenamiento de fenómenos geográficos, es asegurarse de que los cambios realizados en los datos sean coherentes.

 WMS [26] - Web Map Service o Servicio de mapas en la web que produce mapas en formato imagen a la demanda para ser visualizados por un navegador web o en un cliente simple. Produce mapas de datos espaciales referidos de forma dinámica a partir de información geográfica. Este estándar internacional define un "mapa" como una representación de la información geográfica en forma de un archivo de imagen digital conveniente para la exhibición en una pantalla de ordenador. Un mapa no consiste en los propios datos. Los mapas producidos por WMS se generan normalmente en un formato de imagen como PNG, GIF o JPEG, y ocasionalmente como gráficos vectoriales en formato SVG (Scalable Vector Graphics) o WebCGM (Web Computer Graphics Metafile).

El estándar define tres operaciones:

- 1. Devolver metadatos del nivel de servicio.
- 2. Devolver un mapa cuyos parámetros geográficos y dimensionales han sido bien definidos.
- 3. Devolver información de características particulares mostradas en el mapa (opcionales).
- WCS [27] Web Coverage Service extiende la interfaz WMS para permitir el acceso a coberturas geo-espaciales que representan valores o propiedades de áreas geográficas, en lugar de los mapas generados por WMS. Un ejemplo del tipo de información ofrecida a través de un servicio WCS podría ser una imagen en la que cada píxel representa un valor de temperaturas, la cota de altura en el terreno, etc.

4 Beneficios y Problemas de los SIG

4.1 Beneficios:

Los sistemas de información geográfica presentan los siguientes beneficios [28]:

- Los SIG ayudan al planificador a tomar decisiones de compromiso en la asignación de los recursos; utilizando una interfaz interactiva puede simular y manipular instantáneamente diversos supuestos y criterios con objeto de generar diferentes escenarios finales o ensayar distintas hipótesis, lo que le permite adoptar su decisión sobre la base de un abanico de opciones.
- 2. Los SIG permiten introducir una serie de datos -a veces de fuentes espaciales sumamente distintas- en un escenario de análisis que también procede, en buena parte, de diferentes ramas del saber. Por ejemplo, en la selección de lugares para la acuicultura costera es posible integrar rápidamente en el proceso de adopción de decisiones datos de la labor de campo, imágenes tele percibidas, mapas secundarios y datos tabulares de los sectores de la recreación, la extracción de recursos, la diversidad biológica, la valoración del paisaje, las formas actuales de aprovechamiento de la tierra, etc.
- 3. Los SIG están orientados hacia las aplicaciones. Utilizan la tecnología espacial para tratar directamente los problemas del mundo real. Pueden ilustrar en un instante el papel crucial que desempeñan el emplazamiento y la diferenciación espacial en la mejora del bienestar socioeconómico.
- 4. Los SIG han acelerado mucho la ejecución de toda una gama de funciones. Brindan al usuario un rápido acceso a enormes volúmenes de datos y aseguran que las decisiones o resultados tengan un mayor grado de objetividad; y esto se realiza, en algunas situaciones, en tiempo casi real.
- 5. Los datos digitalizados del SIG pueden actualizarse rápida y eficientemente, lo que significa que es posible efectuar revisiones más frecuentes.

- 6. Los cambios en las series cronológicas pueden observarse con prontitud y los cálculos estadísticos correspondientes se pueden cuantificar, lo que permite hacer proyecciones para el futuro.
- 7. Su eficacia en función de los costos es tal, que el SIG contribuye a la competitividad de las empresas de distintas formas, por ejemplo, seleccionando el emplazamiento óptimo. También aumenta la eficiencia y rentabilidad de los servicios públicos.
- 8. Los SIG permiten producir a bajo costo mapas especiales (a veces para un solo propósito) y difundir una gran variedad de mapas u otros datos de salida que de otra forma no sería posible producir.
- 9. La tecnología de los SIG permite alcanzar y normalizar una producción de alta calidad a muchas personas que tal vez no tengan ningún conocimiento cartográfico ni de dibujo.
- 10. Para el usuario investigador o académico, el SIG no sólo aumenta las posibilidades de describir, explicar y predecir modelos y procesos espaciales, sino que también permite formular y ensayar modelos complejos y realistas. Los SIG pueden estimular el desarrollo de una nueva filosofía que integre la labor de los geógrafos humanos y físicos y de los expertos en varias otras disciplinas conexas.

4.2 Problemas:

Los sistemas de información geográfica presentan los siguientes problemas [28]:

- 1. Muchos de los beneficios de los SIG no tienen un valor monetario; es el caso, por ejemplo, de las ventajas intangibles de la "adopción de decisiones más acertadas", la "mejor planificación" o la "mejor información". Estos beneficios podrían ser incluso mayores que los aumentos mensurables, como los costos de la producción de mapas o los gastos en consultorías. El valor que realmente tiene la adquisición de un SIG es sumamente dificil de demostrar.
- 2. El proceso extremadamente lento de la digitalización de los mapas existentes supone que el usuario ha de incurrir en unos gastos considerables para la conversión digital: en algunos casos hasta el 75 por ciento de la suma total gastada en un SIG está destinado a la adquisición o conversión de los datos. Esto significa que muchos datos potencialmente valiosos permanecerán "bloqueados" por muchos años.
- 3. Muchos de los datos de consulta, gráficos o tabulares existentes en copia impresa son de mala calidad, por lo que, al introducirlos en el SIG, se propagan errores. Algunos errores se deben al carácter "borroso" de la entidad; por ejemplo, los Alpes: ¿cómo adquirir datos precisos sobre esta vaga región? Otros son el resultado del uso de ciertas fuentes de datos, la mala digitalización, el muestreo deficiente, errores de procesamiento, conversiones incorrectas de los datos, equivocaciones en la clasificación, etc. Hay que incorporar algoritmos que determinen los límites de confianza estimados de los datos de salida, además de análisis del arrastre de errores para evaluar los posibles resultados de su propagación.
- 4. ¿Quién será responsable de la actualización de los datos digitales y de decidir cuáles datos institucionales han de actualizarse regularmente?
- 5. Los mecanismos y derechos de acceso a la información: en la actualidad varían mucho de un país a otro y además, con frecuencia, el marco jurídico está muy mal definido o es innecesariamente restrictivo. Hay legítimos problemas de derechos de autor que resolver antes de poner los datos a disposición del público, y también están los problemas de

- propiedad de los datos digitales que surgen, por ejemplo, cuando se genera un nuevo conjunto de datos a partir de material de propiedad de otra persona o entidad.
- 6. Hay una escasez crítica de expertos en los SIG que conozcan a fondo la tecnología, los métodos de análisis espacial, el diseño de aplicaciones, etc. Hay demasiado pocos lugares de capacitación e instructores, faltan investigaciones y financiación, e incluso falta la conciencia cabal del problema que representan estas carencias. Esta situación persistirá probablemente en muchos países en tanto que los gobiernos no muestren un mayor interés por la manipulación de los datos geográficos.
- 7. Hay que buscar soluciones para enfrentar la enorme afluencia de datos, que aumenta a un ritmo exponencial; por ejemplo, para 1994, el satélite de observación de la tierra (EOS) generará, por sí solo, 10 billones de bitios de datos por día, y una simple lámina de la *Ordnance Survey* del Reino Unido ocupa alrededor de 240 millones de bitios. Hay que encontrar mejores métodos para almacenar, archivar, estructurar y procesar todos esos datos. ¿Y cómo evaluar si los datos adquiridos son realmente necesarios?
- 8. No existe un formato internacional, ni un conjunto de normas acordadas para los datos de los SIG, que facilite la transferencia de datos de calidad reconocida entre los diferentes usuarios. A esto se suma el hecho de que muchos datos son reunidos por grupos que no tienen interés en aprovechar sus aspectos espaciales, por lo que no proporcionan los datos en un formato apropiado para el análisis espacial.
- 9. Numerosos sistemas aún carecen de interfaces que faciliten su uso y permitan un mayor acceso por parte de usuarios no expertos en los sistemas. Estas personas pueden tener claro lo que quieren obtener del SIG, pero no tienen ganas o tiempo de leer extensos manuales de instrucción o aprender las complejidades del funcionamiento del ordenador. En muchos casos, esto puede haber generado en los usuarios una resistencia a la nueva tecnología.
- 10. También está el problema del acceso a los datos al encontrar efectivamente lo que se necesita. Es difícil conseguir información sobre los archivos digitales (y más aún lo es conseguir los datos en ellos contenidos), y la consulta rápida de las bases de datos a través de redes locales o regionales en conexión directa puede ser lenta y presentar problemas de referenciación defectuosa. Y cuando se encuentran los datos requeridos, puede haber difícultades con las escalas, el procesamiento o el formateo.

"Estamos convencidos de que ya se ha iniciado un cambio fundamental hacia el procesamiento ordinario de datos geográficos en los sistemas computarizados. Las consecuencias de ello sólo se pueden predecir en términos muy generales y por analogía. La introducción de los sistemas de contabilidad informatizados hace 25 años tuvo al principio sus altibajos, y se cometieron muchos errores. No obstante, los costos disminuyeron, se fue descubriendo el valor de los nuevos tipos de información obtenidos y, a su debido tiempo, esos sistemas se volvieron herramientas comunes incluso en organizaciones bastante pequeñas. El retorno a los sistemas manuales sería ahora inconcebible. Creemos que la utilización de los sistemas de información geográfica tendrá una evolución parecida." (Dept. of Environment, 1987).

5 Generación y uso del conocimiento

El uso de los sistemas de información en general tienen gran aceptación a nivel mundial debido a la gran gama de aplicaciones a los que pueden ser dirigidos y entre estos, los sistemas de información geográfica no han sido la excepción. Éstos han dejado de ser una herramienta en la que sólo se muestran los distintos fenómenos y objetos espaciales relacionados con una localización y han evolucionado hasta convertirse en una nueva forma de integración de la información geoespacial no sólo para brindar un mejor servicio en la obtención de información sino también porque juegan un papel fundamental para la toma de decisiones.

El desarrollo de los sistemas de información geográfica se realiza generalmente enfocado a un contexto particular, o sea, restringido a un dominio específico, pero debido a esto se pueden generar problemas de interpretación ya que diferentes comunidades pueden definir los mismos objetos de diversas maneras pues tienen distintos puntos de vista y suposiciones acerca del dominio de estudio, lo que provoca problemas de comunicación por falta de entendimiento compartido. Por lo tanto, es necesario usar representaciones tan generales como sea posible con el propósito de abarcar con un nivel de abstracción mayor los objetos geoespaciales pertenecientes al dominio en el que se está trabajando y así aumentar la posibilidad de rehusar al máximo toda la información anteriormente incluida. También es necesario que estas representaciones se correspondan lo más estrechamente posible con las definiciones y conceptos que estos representan, es de esta manera que se presenta la necesidad de integrar "conocimiento", con la información geoespacial. Es así como surge la premisa de encontrar una manera de gestionar el conocimiento de forma eficiente.

Unas de las formas más usuales de representación del conocimiento que han sido usadas son, entre otras, las redes semánticas y los marcos conceptuales y actualmente se han venido utilizado con más aceptación las ontologías.

5.1 Redes semánticas

Las redes semánticas (sistema NUDE, creada en 1956 por Richens [1]): Una red semántica o esquema de representación en red es una forma de representación de conocimiento lingüístico en que las interrelaciones entre diversos conceptos o elementos semánticos se les da la forma de un grafo. Estas redes pueden ser visualizadas como grafos, aunque algunas veces pueden ser también árboles. Las redes semánticas pueden ser mapas conceptuales y mentales.

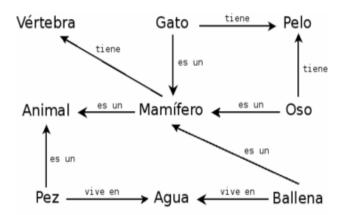


Fig. 15. Un ejemplo de red semántica

En un grafo o red semántica, los elementos semánticos se representan por nodos. Dos elementos semánticos entre los que se admite se da la relación semántica que representa la red,

estarán unidos mediante una línea, flecha, enlace o arista. Cierto tipo de relaciones no simétricas requieren grafos dirigidos que usan flechas en lugar de líneas.

Existen diversos tipos de relaciones semánticas como la hiponimia, hiperonimia, la meronimia, etc. Dado un conjunto de conceptos, elementos semánticos o términos relacionados semánticamente mediante alguna relación semántica, una red semántica representa estas relaciones en forma de grafo. Explícitamente, dado un conjunto de términos (t1, t2,..., tn) y cierta relación semántica simétrica entre ellos, se construye un grafo G = (V,A) cumpliendo las siguientes condiciones:

El conjunto V es el conjunto de vértices o nodos del grafo. Este conjunto estará formado por n elementos (tantos vértices como términos relacionables). A cada uno de los vértices del grafo representará uno de los términos, por tanto, los vértices del grafo se llamarán: t1, t2,..., tn.

El conjunto A es el conjunto de aristas o líneas del grafo. Dados dos vértices (términos) del grafo ti y tj, existirá una línea aij que une los vértices ti y tj, si y sólo si los términos ti y tj están relacionados.

Si la relación no es simétrica, entonces se usan grafos dirigidos para representar la relación.

5.2 Marcos conceptuales

Los marcos conceptuales [2]: El primer proponente de la noción de marco 25 (*frame*) fue el psicólogo cognitivo Marvin Minsky, en cuyas palabras podemos resumir la esencia de los marcos.

Here is the essence of frame theory: When one encounters a new situation (or makes a substantial change in one's view of a problem), one selects from memory a structure called a frame. This is a remembered framework to be adapted to fit reality by changing details as necessary.[2]

Cuando nos enfrentamos con una situación determinada, intentamos ajustarla a otra parecida de la que ya tenemos experiencia previa y esperamos que aparezca un número de elementos comunes y se sucedan algunas situaciones. Por ejemplo, si entramos en una habitación de hotel, esperamos encontrar una cama, un armario, un baño, etc. Nuestra mente reconocerá las instancias específicas de la nueva habitación y los acomodará al estereotipo que ya poseemos. La base de la teoría la conforman, por tanto, las situaciones estereotipadas.

Dejando a un lado esta fundamentación psicológica, la representación basada en marcos constituye, en gran medida, la base del modelo orientado al objeto actual, ya que contiene casi todos los conceptos que éste presenta, aunque estas ideas fueron aplicadas en principio a los lenguajes de programación más que a lenguajes de representación y consulta.

Informalmente, un marco es una estructura de datos compleja que representa una situación estereotipada, por ejemplo hacer una visita a un enfermo o acudir a una fiesta de cumpleaños. Cada marco posee un número de casillas (*slots*) donde se almacena la información respecto a su uso y a lo que se espera que ocurra a continuación. Al igual que las redes semánticas, podemos concebir un marco como una red de nodos y relaciones entre nodos (arcos). Una base de conocimiento basada en marcos es una colección de marcos organizados jerárquicamente, según un número de criterios estrictos y otros principios más o menos imprecisos tales como el de similitud entre marcos. A nivel práctico, podemos considerar los marcos como una red semántica con un número de posibilidades mucho mayor, entre las que destacan, especialmente, la capacidad de activación de procesos y de herencia.

Formalmente, un marco M es una lista de atributos (*slots*) y valores (S, V). El valor V de un *slot* S puede ser:

- Un valor simple, cuyo tipo se halla determinado por el *slot* mismo, por una función que devuelve un valor de ese tipo, o por un puntero al valor de otro *slot* de la jerarquía.
- Una lista de atributos o valores, donde los atributos son funciones booleanas y los valores pueden ser funciones que devuelven valores del tipo del valor del slot, o constantes del tipo del valor del slot.

Como es evidente, la representación de conocimiento basada en marcos debe mucho a las redes semánticas. Sin embargo, los esquemas de representación basados en marcos insisten en una organización jerárquica de éstos, mientras que las redes semánticas no requieren tal organización.

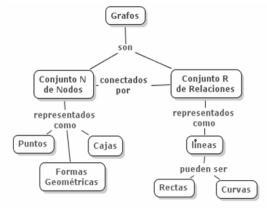


Fig. 16. Ejemplo de un marco conceptual en la definición de un grafo

5.3 Ontologías

"... la palabra 'significado' es probablemente la palabra cuyo significado es, de todo el lenguaje, el más difícil de hallar. ¿Qué significa 'significar'? Me parece que la única respuesta que podemos dar es 'significar' significa la capacidad de cualquier clase de dato para traducirse a un lenguaje diferente".

Claude Lévi- Strauss

Estos paradigmas que fueron mostrados anteriormente no son suficientes para abordar los fenómenos y las situaciones problemáticas que surgen frente a los avances de las tecnologías de la información y la comunicación, por lo que el uso de ontologías se ha popularizado de manera creciente.

Una de las definiciones de ontología más conocidas y citadas en la literatura es la propuesta por Gruber "Una ontología es una especificación explícita de una conceptualización"[9]. El concepto de ontología apareció para designar un vocabulario común de un área y definir, a diferentes niveles de formalismo, el significado de los términos y las relaciones entre ellos. Éstas proveen una comprensión compartida y consensuada del conocimiento de un dominio, que puede ser comunicado entre personas y sistemas heterogéneos. Fueron desarrolladas para el intercambio y uso del conocimiento de una manera eficiente. La Real Academia Española define ontología como: (Del gr. ὄν, ὄντος, el ser, y -logía) parte de la metafísica que trata del ser en general y de sus propiedades trascendentales [29].

Las ontologías son cruciales para la interoperación de conocimiento y pueden servir como cristalización de un consenso alcanzado por cierta comunidad. Tener bien definidas y compartir

las mismas ontologías es una condición previa para el intercambio e integración de la información. Mientras sigue siendo un área fecunda de investigación en el campo de la filosofía, la ontología es actualmente materia de investigación, desarrollo y aplicación en disciplinas relacionadas con la computación, la información y el conocimiento.

Cualquiera que sea el entorno de desarrollo de una ontología, es importante especificar que la base para su construcción es la conceptualización junto con un lenguaje para referirse a las entidades de un dominio en particular. Las ontologías precisan de los siguientes componentes para representar el conocimiento [9]:

Conceptos: son las ideas básicas que se intentan formalizar. Los conceptos pueden ser clases de objetos, métodos, planes, estrategias, procesos de razonamiento, etc.

Relaciones: representan la interacción y enlace entre los conceptos, formando la taxonomía del dominio. Las relaciones básicas son: sub-clase-de, parte-de, conectada-a.

Funciones: son un tipo concreto de relación donde se identifica un elemento mediante el cálculo de una función que considera varios elementos de la ontología.

Instancias: se utilizan para representar objetos determinados de un concepto.

Axiomas: son teoremas que se declaran sobre relaciones que deben cumplir los elementos de la ontología. Especifican las definiciones de los términos en la ontología y las restricciones de sus interpretaciones. Los axiomas deben proveerse para definir la semántica o el significado de los términos.

Los dos usos más importantes de las ontologías (Chandresekaran et al, 1999), son como:

- Vocabulario: Se puede ver la ontología como un vocabulario de representación. Bajo este punto de vista, las ontologías están generalmente especializadas en algún dominio. No es el vocabulario lo que califica a una ontología, sino las conceptualizaciones capturadas por los términos. La representación de un vocabulario proporciona un conjunto de términos con los que describimos hechos del dominio, mientras que el cuerpo de conocimiento que usa este vocabulario es una colección de hechos sobre el dominio.
- Como teoría de contenidos: Uno de los intereses principales en ontologías es la relación entre teoría de contenidos y de mecanismos en la inteligencia artificial. Las ontologías son esencialmente teorías de contenidos porque su contribución principal es identificar clases específicas de objetos y relaciones de un dominio.

Un indicador de la complejidad de una ontología es el conjunto de relaciones conceptuales. [30] sostienen que una ontología se define como un vocabulario más una especificación del significado de dicho vocabulario. En informática y ciencias de la información, se entiende por ontología un conjunto estructurado de conceptos que permiten otorgar un sentido a la información. La ontología constituye un modelo de datos que representa un conjunto de conceptos en un campo de conocimiento y las relaciones entre estos conceptos. Esta estructura de datos es tremendamente útil para optimizar el conocimiento y la utilidad de la información.

Un uso común tecnológico actual del concepto de ontología, en este sentido, lo encontramos en la inteligencia artificial y la representación del conocimiento. En algunas aplicaciones se combinan varios esquemas en una estructura de datos que contiene todas las entidades relevantes y sus relaciones dentro del dominio. Los programas informáticos pueden utilizar así la ontología para una variedad de propósitos, incluyendo el razonamiento inductivo, la clasificación y una variedad de técnicas de resolución de problemas.

El término ontología en informática hace referencia a la formulación de un exhaustivo y riguroso esquema conceptual dentro de un dominio dado, con la finalidad de facilitar la comunicación y la compartición de la información entre diferentes sistemas. Aunque toma su nombre por analogía, ésta es la diferencia con el significado filosófico de la palabra ontología.

El universo físico es el mundo real con todo lo que la gente es capaz de percibir, incluyéndonos a nosotros mismos (medio ambiente). La implementación se realiza a través de la conversión de los componentes de ese universo a las estructuras de un lenguaje de programación entendible por la máquina. A continuación se muestra el paradigma de los cinco universos [31], en el que representa al mundo dividido en cinco partes interrelacionadas, aquí se pone de manifiesto que a través de la visión, las imágenes que se corresponden con los objetos del mundo real se forman dentro de la mente de las personas, estas imágenes en el universo cognitivo son representaciones de las entidades en el universo físico, pero estas imágenes no son simplemente almacenadas en la mente de una manera desordenada, sino que están organizados en un marco lógico [32]. Cuando este marco se hace explícito mediante métodos lógicos, obtenemos ontologías [33]. Por tanto, las ontologías son las representaciones formales de la lógica de la mente humana que existen en el universo lógico definidas en un lenguaje de máquina explicito.

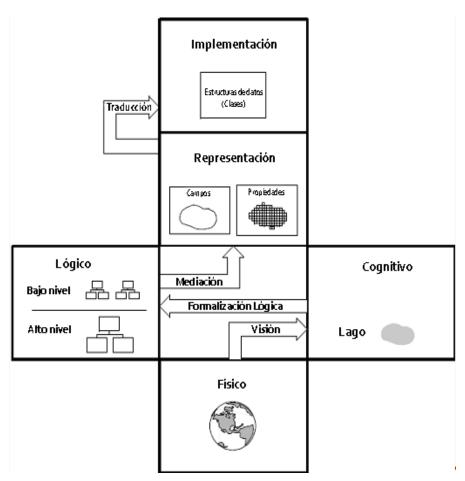


Fig. 17. Paradigma de los cinco universos, propuesto por Federico Fonseca

El universo lógico está conectado con el universo de las representaciones a través de los mediadores semánticos. La implementación del universo de las representaciones puede incluir

elementos tales como algoritmos, estructuras de datos raster, vectoriales y clases. El universo lógico contiene dos tipos de ontologías [31]:

Ontologías de alto nivel: Contienen las teorías más generales del mundo, tales como los conceptos generales de la teoría de la geografía natural.

Ontologías de bajo nivel: Son especializaciones de las ontologías más generales (Ontologías de alto nivel). Pueden ser descripciones detalladas de dominios específicos y las tareas que tratan esos dominios.

Las geo-ontologías, se dividen en dos tipos:

- Ontología del dominio fenomenológico (PDO, por sus siglas en inglés): Esta
 ontología capta las diferentes dimensiones y propiedades internas de los fenómenos
 geográficos.
- Ontología del dominio de aplicación (ADO, por sus siglas en inglés): Esta ontología se refiere a la descripción de los temas específicos y las tareas que los GIS científicos usan como fuente de información.

El PDO se refiere a cómo el fenómeno geográfico puede ser capturado y representado por los sistemas computacionales y está localizado en el universo de representación; el ADO es parte del universo lógico porque trata de la descripción del fenómeno en sí, de cómo encaja éste en el mundo y de cómo puede ser mejor descrito. La conexión entre la PDO y ADO es realizado por los mediadores semánticos, como se muestra en la fig.

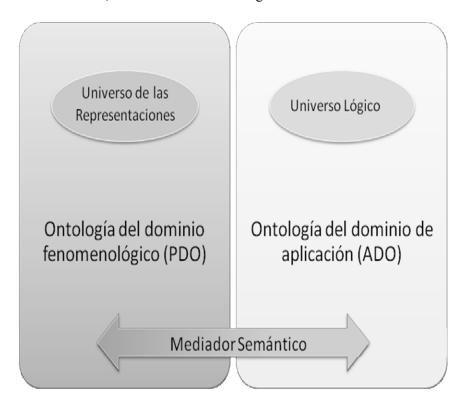


Fig. 18. Esquemas de las ontologías del dominio fenomenológico y de aplicación propuesto por Frederico Fonseca

Cada objeto geográfico es único como concepto en el universo lógico aunque se elijan diferentes conceptualizaciones para representar sus campos y propiedades [34], ya pueden existir diferentes maneras de representar un objeto determinado y no por eso el objeto se va a modificar semánticamente, por ejemplo, un yacimiento es un yacimiento, ya sea representado por una fotografía aérea, una representación vectorial o un modelo digital del terreno, cualquiera que sea el tipo de representación, la concepción semántica implícita en lo que define un yacimiento no va a cambiar, va a ser siempre la misma.

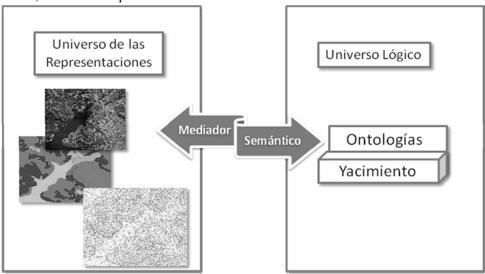


Fig. 19. Esquema de la relación entre el universo de la representación y el lógico por Frederico Fonseca

Las representaciones están localizadas en el universo de representaciones y aquí, tal como se muestra en la figura anterior, se pueden realizar de disímiles maneras, mientras que los conceptos y las descripciones formales están localizadas en el universo lógico, aquí la información semántica que define al objeto en cuestión es invariante.

5.3.1 Ontología del dominio fenomenológico

Los fenómenos geográficos en el nivel de la representación presentan características especiales tales como la dependencia del muestreo con que se haya realizado determinado fenómeno, sus propiedades intrínsecas y la reutilización de conocimiento algorítmico empleado. Los objetos recogidos, por ejemplo: una colección de puntos, pueden acarrear muchas diferencias, incluso si la recogida se hiciera con un GPS, por lo que se debe realizar un procesamiento de los datos para eliminar posibles errores o ruidos en los datos capturados.

Los datos adquiridos por diferentes sensores pueden requerir diferentes tipos de procesamientos y brindar diferentes resultados sobre el mismo fenómeno. La ontología de dominio fenomenológico es dependiente del muestreo y tiene dos componentes muy distintos pero bien interrelacionados:

Ontología de medición: Que describe el proceso físico de la recogida de un fenómeno geográfico. Este proceso de la recogida genera campos y objetos.

- Los objetos son aquellos puntos, líneas y/o polígonos que representan los diferentes objetos de la geografía, como ríos, lagos, etc.
- Los campos son todas aquellas propiedades y características que presentan los objetos.
- Ontología de método: Consiste en un conjunto de algoritmos y estructuras de datos que representan el conocimiento reutilizable para operar en el fenómeno medido. A veces puede ser usado en forma de técnicas de procesamiento para transformar el fenómeno medido desde el nivel de representación a la lógica, por ejemplo:
 - Mediante la selección y/o la mejora de las imágenes y los códigos geográficos de las líneas o polígonos o para llevar a cabo la extracción de características y/o segmentación.

Los algoritmos son esa parte de la ontología de método que representan las transformaciones desde el nivel de las representaciones al nivel lógico a través de un proceso llamado identificación estructural. Cuando se aplica a una imagen (o un conjunto de imágenes), este proceso da lugar a un conjunto de estructuras estrechamente relacionadas con las propiedades del dispositivo de medición y su de interacción con el paisaje físico. Estas estructuras pueden ser geométricas (por ejemplo, las regiones extraídas por un procedimiento de segmentación). Cuando se aplica a los objetos, el resultado es la identificación del objeto y la asociación de este objeto a un significado, por ejemplo, lo que era sólo un polígono se convierte en un lago, si se asocia a la clase lago en una ontología presente en un nivel lógico.

5.3.2 Ontologías del dominio de aplicación

En un dominio de científicos, éstos, para hacer su trabajo, usan conceptos derivados del dominio de conocimiento que ellos manejan. Dentro de ontologías del dominio de aplicación podemos distinguir dos tipos de ontologías:

- Ontologías de tipo: describen el vocabulario relacionado a un dominio genérico, por ejemplo, la geología o la ecología.
- Ontologías de tarea: describen las tareas o actividades dentro de un dominio, por ejemplo, la evaluación de la contaminación del agua para estudios ecológicos.

Los conceptos en este contexto son capaces de hacer frente a los fenómenos, independientes de cómo estén representados.

5.3.3 Integración bi-direccional

Uno de los principales objetivos es integrar la información geográfica de distintas fuentes. Las diversas comunidades de la información geoespacial tienen diferentes visiones acerca del mundo. Estas opiniones pueden ser formalizadas en distintas ontologías. Por lo tanto, es necesario acomodar múltiples ontologías, que en el modelo presentado por Fonseca se encuentran tanto dentro del universo lógico como en el universo de la representación.

Dado que las ontologías en su forma natural están estructuradas de forma jerárquica en una taxonomía, es posible realizar su integración de forma vertical y horizontal:

• Navegación vertical: Se produce cuando una instancia de una clase inmediatamente superior o inmediatamente inferior se genera a partir de una determinada clase. La

navegación vertical implica un cambio de nivel de detalle, porque se produce una nueva instancia más detallada o una con menos detalles que la original.

• Navegación horizontal: Se produce cuando uno de los roles que desempeña el objeto se extrae de una instancia. De esta manera, una nueva instancia es generada produciendo un nuevo objeto que pertenece a la clase del rol. La navegación horizontal no implica un cambio en el nivel de detalle y puede ser a cualquier nivel en la jerarquía de clases.

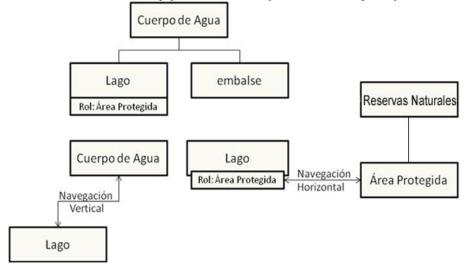


Fig. 20. Representación de la integración horizontal y vertical de ontologías

Cuando una nueva ontología se especifica, es necesario disponer de un conjunto de operaciones que permiten la reutilización de ontologías anteriores o partes de ellas para así poder aprovechar todo el conocimiento insertado en las ontologías anteriores y así generar nuevas ontologías más robustas y eficientes. En un entorno SIGGO, las operaciones básicas que deben estar disponibles son: la herencia, la inclusión, y el uso de roles.

• **Herencia**: Se utiliza para la integración vertical.

Las clases se definen en los SIGGOs jerárquicamente; aprovechando las ventajas de la herencia es posible definir clases más generales, que contengan una estructura genérica de un tipo de objeto y éstas se especialicen en subclases. Las subclases heredan todas las propiedades de la clase padre y añaden características propias.

• **Roles**: Se utilizan para la integración horizontal.

Utilizamos roles para sortear los problemas con la herencia múltiple.

Por ejemplo, un accidente geográfico puede ser, al mismo tiempo, un lago y una atracción turística. En los SIGGOs se representa esta entidad como un lago, que desempeña un rol de una atracción turística o quizás también juegue el rol de área protegida ecológicamente, ya que una entidad puede tener muchos roles.

• **Inclusión**: Puede ser utilizado para ambas integraciones (horizontal y vertical).

La inclusión es una operación en la que una entidad de una ontología se utiliza para especificar cualquier parte de una entidad en una nueva ontología.

Por ejemplo, una ontología que se refiere a las representaciones de objetos espaciales incluirá muchas partes de una ontología geométrica.

Las operaciones de integración son utilizadas en las diferentes etapas del proceso de especificación de la ontología. Esta separación ocurre porque los niveles de detalle son diferentes en muchas etapas de la especificación de la ontología. [31] sugiere la utilización de la herencia en la integración de ontología de alto nivel y los roles en la integración de las ontologías de bajo nivel. La inclusión se utiliza en todos los niveles de integración. Este enfoque multinivel genera un modelo muy flexible y con el fin de aprovechar esta flexibilidad, es necesario un modelo específico para la navegación entre las distintas entidades. Se escoge desarrollar un modelo de navegación en el universo de implementación. Ya que las clases extraídas de las ontologías están en este nivel y el modelo de navegación se basa en el intercambio de clases, o sea, ir de clases más generales a clases más específicas y viceversa.

5.4 Ontologías como un componente más de los SIG

Ontología es una antigua disciplina que, en sentido filosófico, se define como un esquema específico de categorías que refleja una visión específica del mundo. Desde el punto de vista informático, ontologías son teorías que especifican un vocabulario relativo a un cierto dominio. Este vocabulario define entidades, clases, propiedades, predicados, funciones y las relaciones entre estos componentes. Las ontologías toman un papel clave en la resolución de interoperabilidad semántica entre sistemas de información y su uso.

Filósofos e ingenieros de software tienen puntos de vista diferentes sobre ontologías. Para [35], ontología describe una cierta realidad con un vocabulario específico, usando un conjunto de premisas de acuerdo con un sentido intencional de palabras del vocabulario. [9] define una ontología como una especificación explícita de una conceptualización, es decir, que proporciona una estructura y contenidos de forma explícita que codifica las reglas implícitas de una parte de la realidad; estas declaraciones explícitas son independientes del fin y del dominio de la aplicación en el que se usarán o reutilizarán sus definiciones.

De esta forma, se puede decir que para un ingeniero de software existen diversas ontologías, mientras que para un filósofo existe apenas una, la ontología. Para resolver esta cuestión, Smith [5] sugiere una distinción terminológica entre una ontología basada en una realidad (ontología-R) y una ontología epistemológica (ontología-E). Ontología-R es una teoría que explica cómo el universo está organizado y corresponde al mundo de los filósofos.

Mientras que por otro lado podemos definir ontología como una teoría que explica cómo un individuo, grupo, lenguaje o ciencia entiende un determinado dominio. El desarrollo de sistemas basados en el conocimiento y de software en general, normalmente se realiza en diferentes contextos, puntos de vista y suposiciones acerca de su materia de estudio. Cada uno usa su propio vocabulario y por ello pueden tener diferentes conceptos con significados que, a veces, se solapan; pueden tener diferentes métodos y estructuras, por tanto, se crean problemas de comunicación por falta de entendimiento compartido que limita la interoperabilidad y, por tanto, el potencial de reutilizar y compartir información. De acuerdo con Sheth [7], la nueva generación de sistemas de información (SI) deberá ser capaz de resolver la interoperabilidad semántica, en la que un hecho puede ser más que una descripción, para poder hacer un buen uso de las informaciones disponibles como la llegada de internet y la computación distribuida.

Esto ha llevado a pensar en sistemas de información geográfica gobernados por ontologías (SIGGO[31]), un concepto que, aunque en una fase preliminar de desarrollo, abre nuevas maneras de pensar sobre las ontologías y los SI, y cubre las dimensiones estructurales, las dimensiones temporales de los SI e involucra tanto a los desarrolladores como a los usuarios de

los SI [36]. Las ontologías pueden ser usadas para ayudar en la solución de conflictos entre fuentes de información. [31] propone un *framework* para el desarrollo de aplicaciones geográficas usando ontologías. El *framework* usa ontologías como la base para la integración de información geográfica, además, ha creado un mecanismo que permite a la información geográfica una integración basada primeramente en este sentido. Puesto que la integración puede ocurrir a través de diferentes niveles, él también ha creado un mecanismo básico para seleccionar el nivel de detalle. El uso de una ontología, trasladado a un componente de un sistema de información, es la base de un sistema de información geográfica gobernado por ontologías SIGGO [31].

Los SIGGOs no son más que sistemas de información geográfica en los que se incorporan el uso de ontologías como un componente activo. Los SIGGOs son construidos utilizando clases derivadas de las ontologías; como consecuencias de esto se extrae el conocimiento insertado en éstas para aportar mayor eficiencia y robustez al sistema. El mecanismo de intercambio de clases permite que una instancia pueda ser generalizada como un sinónimo de la representación más abstracta del dominio o ser especializada, pudiendo así utilizarse en función del grado de profundidad de la descripción, lo que permite la integración de la información en diferentes niveles de abstracción. En los SIGGOs, las ontologías son una componente más, como lo es la base de datos temáticos o espaciales que interviene y coopera de la misma manera para alcanzar los objetivos para los cuales fue creado el SIG.

El uso de ontologías en la construcción de sistemas de información geográfica gobernados por ontologías permite establecer correspondencia y relaciones entre los diferentes dominios de entidades de información. [37] acredita que el uso de ontologías en el desarrollo de sistemas contribuye con una mejora en la calidad del producto final, ya que ellas pueden ayudar a evitar problemas como:

- Inconsistencia entre ontologías implícitas en SI.
- Conflictos entre conceptos ontológicos e implementaciones.
- Conflictos entre ontología de sentido común y conceptos básicos no incluidos en el software

Cuando nos adentramos en lo que es la construcción de un SIGGO, se debe tener en cuenta, primeramente, el dominio donde se va a desarrollar el sistema, con el objetivo de definir, a través de un editor, todas las ontologías con las que se va a trabajar. Este editor es capaz de englobar una representación formal de las ontologías con las que trabaja el sistema y además, proporcionar un mecanismo de entendimiento entre las ontologías definidas y la máquina. Este mecanismo puede traducirse, en esencia, en algún tipo de lenguaje, por ejemplo: "Java", desarrollado por *Sun Microsystems* o "C", creado en 1972 por Ken Thompson y Dennis M. Ritchie. A través del sistema de navegación ontológica, el usuario puede tener acceso a toda la información contenida en el SIGGO, este sistema de navegación debe estar basado en algoritmos de búsqueda inteligente con el objetivo de que dichas búsquedas sean eficientes y eficaces, de tal forma que pase desapercibido ante el usuario todo este engranaje que da como resultado la información que ha sido solicitada. Una vez que han sido definidas las ontologías, éstas son traducidas y estarán disponibles en estructuras de datos como las "Clases", que contienen atributos y funciones que darán soporte a todas las funcionalidades del sistema.

La arquitectura de un SIGGO también está centrada en dos aspectos fundamentales: la generación y el uso del conocimiento. La generación del conocimiento comprende la especificación de las ontologías utilizando el editor de ontologías, la generación de nuevas ontologías a partir de las ya existentes y la traducción de ontologías a las estructuras de datos. La fase de uso del conocimiento se apoya en los productos obtenidos en la fase anterior: una serie

de ontologías especificadas en un lenguaje formal y en una serie de clases. Las ontologías están disponibles para ser navegadas por el usuario final, para su utilización en la generación de aplicaciones, análisis y finalmente, las posibles alternativas de decisión, como se muestra en el siguiente gráfico:

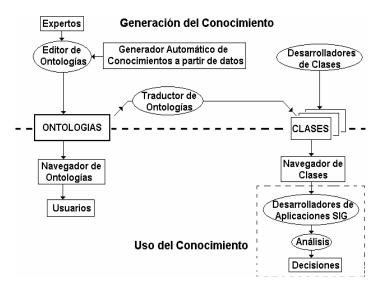


Fig. 21. Esquema de la generación y uso del conocimiento (Eduardo Garea Llano, José Luis Gil Rodríguez[38])

5.4.1 Fase de generación del conocimiento

En esta fase, los SIGGOs han de estar basados en dos criterios básicos; en primer lugar, la creación de ontologías de forma explícita antes de que sea concebido el sistema, ya que éstas contribuyen a mejorar el rendimiento, puesto que cada SIGGO se basa en una ontología implícita. El acto de crear ontologías explícitas evita los conflictos entre los conceptos de las ontologías y su implementación. Además, las ontologías de alto nivel pueden ser utilizadas como base para la integración de los sistemas, ya que éstas representan el vocabulario más general que comparten las diferentes comunidades. El otro criterio es precisamente la organización desde el punto de vista jerárquico de los criterios y conceptualizaciones acerca de determinado dominio de las diferentes comunidades de información geoespacial (CIG o GIC, por sus siglas en inglés). Estas comunidades han sido definidas así por varios autores, puesto que, según esquemas filosóficos, el mundo se encuentra subdividido en diferentes grupos de personas y cada grupo tiene una visión particular acerca del mundo; de aquí parte la concepción de las diferentes comunidades que interactúan entre sí en el ámbito geográfico. En la construcción de una arquitectura SIGGO se asume que una CIG no es más que un grupo de productores y usuarios de datos espaciales que comparten una ontología de un fenómeno del mundo real. Por ende, las ontologías constituyen el conocimiento particular base que describe hechos que son siempre verdaderos para una comunidad de usuarios particular.

Para sortear los problemas generados por estos criterios, se pueden tomar varias alternativas entre las que se encuentran, en primer lugar, el considerar que los pequeños CIG se reúnan con otros CIG con los mismos intereses y traten de construir ontologías de alto nivel, que engloben

los puntos de vista de cada uno de ellos hasta el nivel más bajo, o sea, hasta el nivel más detallado de las ontologías que cada uno de ellos presentan. En segundo lugar, estas GIC se podrían reunir antes de especificar sus propias ontologías con el fin de especificar una ontología de alto nivel común para ambos y a partir de aquí, ir generando sus propias ontologías. Lo más importante aquí es que la arquitectura de un SIGGO permite la reutilización y la integración de ontologías, basado en la reutilización de las clases a través del uso de la herencia y roles.

Es necesario destacar la diferencia entre los términos de ontologías y esquemas de bases de datos, ya que a veces se hace un mal uso de estos conceptos. Un esquema de base de datos representa lo que en ella se almacena, o sea, solamente los datos que contiene y las relaciones entre ellos. Una ontología representa conceptos existentes en determinado dominio, por lo que son semánticamente mucho más ricas y por tanto, están más cerca del modelo cognitivo de los usuarios. Por tanto, al crear la arquitectura del SIGGO, se debe centrar en que las personas alcancen un acuerdo sobre qué son las entidades geográficas básicas de su mundo. No importa si las entidades están almacenadas o no en una base de datos. La información que existe en las bases de datos tiene que ser adaptada para llenar las clases de las ontologías. Por ejemplo, el concepto de suelo puede ser representado de forma diferente en diversas bases de datos, pero el concepto es sólo uno, al menos para el punto de vista de una comunidad. Este punto de vista esta expresado en la ontología que esta comunidad ha especificado [38]. Una vez que las ontologías han sido especificadas, éstas pueden ser traducidas a estructuras de datos y contienen el conocimiento disponible a ser incluido en los nuevos sistemas basados en ontologías.

5.4.2 Fase de uso del conocimiento

Como resultado de la fase de generación del conocimiento se obtienen una serie de ontologías especificadas y un conjunto de estructuras de datos derivadas de estas ontologías, que se utilizan en esta fase y pueden ser navegadas por el usuario final de manera que pueda acceder a todo el conocimiento insertado en ellas. En un entorno SIGGO se destacan principalmente un conjunto de componentes como son:

- El servidor de ontologías: Tiene un papel central en un SIGGO ya que proporciona la conexión entre todos los componentes. También se encarga de hacer que las ontologías estén disponibles para las aplicaciones. La conexión con las fuentes de información se realiza a través de mediadores.
- Los mediadores: Son componentes de software con conocimiento incorporado construidos por expertos, éstos son los encargados de traducir la información geográfica a un formato comprensible por el usuario final.
- Las ontologías: Están representadas por dos tipos de estructuras:
 - Las especificaciones: Son hechas por los expertos y almacenadas según sus características distintivas, su estructura proporciona información sobre el significado de la información disponible. Puede ser utilizado por el usuario para saber que se dispone de información y que su concepción del mundo coincide con otras concepciones disponibles almacenadas por el servidor de ontologías.
 - Las estructuras de datos: Surgen como resultado de la traducción de las ontologías y por lo tanto, contienen todo el conocimiento que se encuentra insertado en las ontologías.
- Las fuentes de información: Éstas, en un SIGGO, pueden ser cualquier tipo de base de datos geográficos, siempre que estén vinculadas con un mediador. El mediador tiene la

función de extraer los elementos de información necesarios para generar una instancia de una entidad perteneciente a una ontología.

• Las aplicaciones: La principal aplicación de un SIGGO es la recuperación de la información. Los mediadores proporcionan instancias de las entidades disponibles en el servidor de ontología. El usuario puede navegar por la información en diferentes niveles de detalle en función del nivel de ontología utilizado. Se pueden desarrollar otros tipos de aplicaciones como la actualización de bases de datos y otro tipo de procesamiento de datos geográficos, incluyendo análisis estadístico y el procesamiento de imágenes.

Un ejemplo de cómo podría usarse un SIGGO sería: un usuario desea recuperar información acerca de los cuerpos de agua de una región determinada. En primer lugar, el usuario navega por el servidor de ontología en busca de las correspondientes clases, como se muestra en la figura:

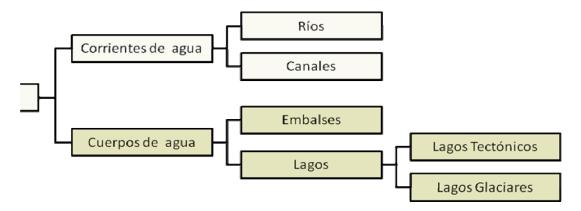


Fig. 22. Porción de ontologías definidas en el navegador de ontologías

Las ontologías están estructuradas en forma jerárquica. Este tipo de organización permite hacer consultas a diferentes niveles de abstracción. Después de eso, el servidor de ontología inicia los mediadores que buscan la información que van a retornar en un conjunto de objetos de la clase especificada, estos objetos se corresponden con localizaciones específicas existentes en el mapa. Los resultados obtenidos se pueden mostrar o pueden someterse a cualquier operación válida, como el análisis estadístico.

Por ejemplo, una consulta para Lagos presenta el siguiente resultado:

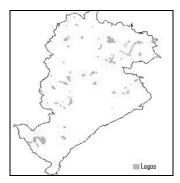


Fig. 23. Consulta para Lagos

La consulta de Embalses es similar a la anterior consulta y presenta el siguiente resultado:

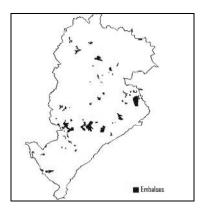


Fig. 24. Consulta para Embalses

El resultado de la consulta para el cuerpo de agua tiene como resultado:



Fig. 25. Consulta para Cuerpos de Agua

Hay que destacar que mientras se hagan búsquedas de objetos definidos en ontologías de mayor nivel, los resultados obtenidos serán más generales. Como se muestra en el ejemplo anterior, la consulta para Cuerpos de Agua no es más que el resultado de la suma de las consultas para Lagos y Embalses, ya que la ontología que define los Cuerpos de Agua se encuentra en un nivel mayor de abstracción en la jerarquía y por tanto, se define como un concepto más general que los de Lago y Embalse.

6 Integración

Desde un inicio, todas las sociedades que han gozado de un grado de civilización han almacenado de alguna manera su información espacial; con el desarrollo de nuevas tecnologías, el monto de estos datos almacenados ha crecido notablemente, dando como resultado grandes volúmenes de datos que en todo su conjunto hacen difícil poder procesarlos. A pesar de esto, no existe un único formato estándar de almacenamiento capaz de ser utilizado por todo tipo de

sistemas de información geográfica (SIG) con éxito, dificultando así la interoperabilidad entre estos sistemas. Una importante iniciativa para lograr acceso uniforme a los datos lo constituye el Consorcio OpenGIS, que es una asociación que busca definir un conjunto de requisitos y normas para la estandarización de los datos en el ámbito geográfico, sin embargo, la normalización de los datos no constituye toda la solución, ya que existen disímiles formatos recogidos en distintos estándares, dando como resultado heterogeneidades en la estructura y el tipo de almacenamiento de los datos. Por tanto, existe una creciente necesidad de encontrar una solución que permita la utilización de éstos, independientemente de los problemas que presentan.

6.1 Datos y datos geográficos: características

Desde el punto de vista de los datos, consideramos la hipótesis planteada en [39]: "Las entidades en el mundo geográfico son diferentesde muchas maneras". En la Real Academia de la Lengua Española se define la palabra "dato" como un "Antecedente necesario para llegar al conocimiento exacto de algo o para deducir las consecuencias legítimas de un hecho" [40]. En el ámbito geográfico, los datos también se definen como antecedentes para entender un hecho, en este caso, un fenómeno geográfico. Éstos tienen características particulares [41] que los diferencian de los datos que no son de tipo geográfico y que dificultan su manipulación, algunas de ellas son:

- La localización, que está generalmente basada en un sistema de coordenadas (posición absoluta).
- La temporalidad, que se refiere a que un hecho puede estar presente en un determinado espacio de tiempo (por ejemplo, los ríos, que en determinadas épocas desaparecen).
- Las relaciones espaciales con otros objetos o datos: (posición relativa).
- La gran heterogeneidad y volumen de almacenamiento que presentan.

A éstos se suma que los datos pueden estar incompletos o ser insuficientes, así como también pueden contener errores que hayan sido introducidos durante su recolección. Dado que "Un modelo de datos es un conjunto de conceptos que sirven para describir una estructura" [42], podemos decir que la estructura de los datos depende mucho de cómo se han modelado las entidades en el mundo real, en este contexto podemos especificar que los datos geográficos pueden ser representados en varios modelos, por ejemplo: modelo vectorial, modelo ráster y modelo de bases de datos espaciales.

6.2 Integración de los datos geográficos

Los datos geográficos se encuentran recogidos en disímiles formatos, ejemplo de éstos pueden ser los formatos vectoriales, ráster, bases de datos espaciales, etc., por lo que se hace muy difícil el poder trabajar con estos datos indistintamente, principalmente por los problemas de compatibilidad existentes entre estos formatos. El problema de obtener un acceso uniforme desde fuentes variadas de datos y en formatos heterogéneos es conocido como *integración*.

En [43] se presenta una vista de los principales esquemas de integración, que se representan en la Fig.4. En la Fig.4 (a) se presenta un panorámica de integración en la que se traducen las fuentes de información (FI) de un formato hacia el otro. Este tipo de integración tiene como deficiencia que no se asegura que la traducción sea posible para todas las entidades en las FI, además de que es costoso computacionalmente. La panorámica presentada en la Fig.4 (b)

consiste en representar las diferentes FI usando un formato común a ambas fuentes, que también puede contener las deficiencias anteriormente mencionadas. Por último, en la Fig.4 (c) se presenta un esquema basado en la integración a través de conceptos abstractos generados a partir de las FI.

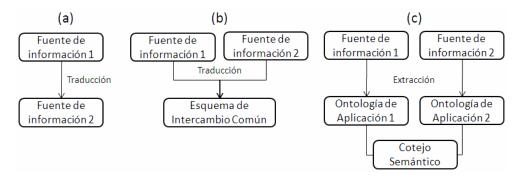


Fig. 26. Modelación de bases de datos espaciales

La integración puede tratarse a varios niveles, desde la integración física que aborda la integración sintáctico-estructural, hasta llegar a la denominada integración semántica. Los esfuerzos realizados en este sentido van enfocados principalmente estos dos tipos de integración:

- Integración sintáctico-estructural: Plantea la existencia de una conexión técnica entre datos que pueden estar en diferentes sistemas de referencia o que se encuentren en diferentes formatos. Está enfocada principalmente en la traducción hacia un esquema de intercambio común, generalmente basado en XML y garantizando que las estructuras de los datos sean homogéneas.
- Integración semántica: Plantea la integración de la información proveniente en formatos heterogéneos no basados en lo que son sino en lo que significan, o sea, se refiere a la integración de las diferencias en las convenciones de nomenclatura y conceptualizaciones en los diferentes grupos [44], garantizando un entendimiento mutuo en un contexto determinado entre los diferentes sistemas, incluyendo además a aquellos humanos que puedan interactuar con ellos.

En [45] se plantea un enfoque para la integración de información geográfica proveniente de fuentes heterogéneas en el que se transforman los datos que están en el SIG a un modelo común elegido anteriormente (GML), o sea, en este trabajo se presentan algunos estándares y formatos de datos utilizados en los SIG's, se analiza la integración de la información proveniente de estas fuentes, y se plantea la definición de un esquema de intercambio basado en XML junto con las transformaciones necesarias para llevar uno de los estándares a dicho esquema. En particular, en este trabajo se establecen correspondencias entre los datos y los conceptos geográficos, y se muestran las transformaciones necesarias para llevar información en formato ESRI (Shapefile) a GML V2.0 de OpenGis y viceversa. Sin embargo, se plantea que la transformación de GML a SHP no es completa pues no todo polígono GML puede transformarse a formato Shapefile, no toda multi geometría tiene un elemento correspondiente, válido en el dominio shapefile y no todas las transformaciones de GML&SHP son biyectivas en el sentido de poder recuperar el objeto original aplicando la transformación correspondiente. Aunque en este trabajo se plantea un enfoque de integración, éste no resuelve los problemas que se generan, sobretodo por la heterogeneidad sintáctica existente en los datos geográficos, por lo que el uso de la información semántica sigue siendo una vía mucho más confiable para la integración de la geo-información, pues todos los geodatos están basados en entidades geográficas en el mundo real, entidades que pueden ser conceptualizadas e integradas de una forma mucho más robusta a nivel semántico.

Entre las ventajas que podemos encontrar al hacer uso de la integración sintácticoestructural, tenemos que presentan una cómoda conversión entre estándares que estén basados en los mismos formatos (ej.: XML) y que al basarse el esquema de intercambio común en XML, es posible aprovechar todas bondades de este lenguaje para el intercambio de información.

Entre las desventajas podemos mencionar que es necesario que se use el mismo modelo de datos para representar a la misma información, o sea, idénticos nombres, estructuras y representaciones, no siempre todas las transformaciones son biyectivas y que no se procesan los datos de una manera consistente con su significado, perdiéndose muchas veces el sentido en las transformaciones de las entidades.

En la literatura existen trabajos recientes [46-51] que abordan el tema de la integración semántica de la información geográfica, en ellos se propone el uso de ontologías como mecanismo de representación del conocimiento para el proceso de integración, de ellos en se extraen una serie de aspectos que se deben tener en cuenta a la hora de realizar este proceso [52]:

- Representación de la información geográfica: Referido a que las entidades geográficas y
 no geográficas son diferentes en un número de maneras [39], o sea, está referido a como los
 diferentes grupos o comunidades representan las entidades en el dominio geográfico a nivel
 semántico.
- Representación formal de las ontologías: Referido a los tipos de lenguajes utilizados para formalizar y manipular la semántica y las ontologías.
- **Proceso de integración:** Referido al conjunto de pasos para encontrar correspondencias útiles entre las fuentes de información.

En [52] se plantea además la preocupación de que, debido a la gran gama de problemas en las que se hace uso de las ontologías, se hace necesario velar por aspectos como: ver si éstas están representando verdaderamente la realidad del dominio geográfico, ya que, al ser ésta una vista simplificada y abstracta del mundo que percibimos y deseamos representar explícitamente, es posible que tales representaciones no refleien tal cual es la realidad del fenómeno.

En la literatura se han definido varios enfoques para la integración de fuentes heterogéneas de información geográfica, como las propuestas en [53]: *Enfoque de ontología única* (Fig. 27), *Enfoque de múltiples ontologías* (Fig. 28) y *Enfoque híbrido de ontologías* (Fig. 29).

En el enfoque de ontología única se hace uso de una ontología global para integrar las fuentes de información, o sea, éstas son relacionadas por medio de una única ontología.



Fig. 27. Enfoque de ontología única

En el enfoque de múltiples ontologías, las fuentes de información son relacionadas por medio de ontologías independientes que representan a cada fuente de información por separado.

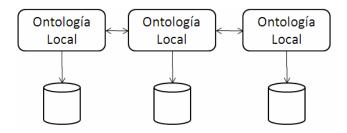


Fig. 28. Enfoque de múltiples ontologías

En el enfoque híbrido de ontologías, las fuentes de información son relacionadas por medio de ontologías independientes que las representan por separado y a su vez, estas ontologías independientes son relacionadas por medio de un vocabulario en el que se compartan los términos de las ontologías locales, el cual podría ser relevado por una ontología global o de domino.

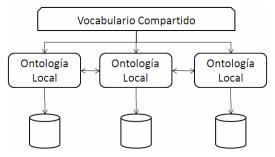


Fig. 29. Enfoque híbrido de ontologías

En la comparación de los diferentes enfoques mostrados anteriormente, los autores concluyen que los requisitos concernientes al lenguaje y estructura de la ontología dependen del tipo de datos que va a integrarse y el uso que se le vaya a dar a la ontología y además, plantean que existe una necesidad de desarrollar una metodología más general que incluya un análisis de la tarea de integración y apoye el proceso de definición del rol de las ontologías con respecto a estos requisitos.

Las soluciones para la integración semántica generalmente se basan en ontologías, ya que éstas proporcionan una especificación formal del modelo mental en el que están basados los datos. Para llevar a cabo dicha integración, es necesario generar un modelo conceptual a partir de un esquema físico (datos geográficos).

El modelo conceptual se ha de desarrollar usando una ontología de aplicación (según los términos para la clasificación de ontologías propuesto por [35]. En este proceso se pueden distinguir dos fases:

- Fase de extracción: Donde se acceden a los datos para recuperar las entidades almacenadas en el esquema físico, dando como resultado un esquema lógico.
- **Fase de conceptualización:** Sobre el esquema lógico obtenido en la fase de extracción se aplican diferentes procesos que permiten generar un esquema conceptual.

En la literatura existen trabajos recientes [54-56] que abordan la integración semántica de los datos geográficos en función de sus principales componentes:

- Componente temática: Referida a las características físicas o abstractas de las entidades.
- Componente espacial: Referida a la localización en el espacio (En un sistema de coordenadas).
- Componente temporal: Referida a las características que son variantes en el tiempo.

7 Consideraciones

Con el objetivo de explotar al máximo todas las ventajas que trae el uso de conocimiento incorporado en los sistemas de información geográfica, es necesario subdividir en tareas más específicas todas las funcionalidades que lleva implícitas un SIGGO. Existen diferentes técnicas inteligentes de búsqueda y gestión de información provenientes de campos como la inteligencia artificial y el reconocimiento de patrones que pueden ser útiles a la hora de concebir una arquitectura SIGGO, por lo que se presentan a continuación algunas consideraciones a tener en cuenta:

7.1 Editor de ontologías:

Este es, quizás, uno de los componentes más importantes dentro de un SIGGO puesto que en él se va definir toda la estructura semántica que va a dar soporte al conocimiento que se va a incluir en el sistema, porque no sólo basta con definir y ponerse de acuerdo en el marco ontológico con el que se va a trabajar sino también hay que llevarlo a un lenguaje explícito y entendible para que luego, a través del uso de mediadores, sean llevados a clases. El editor de ontologías permite al usuario trabajar en la especificación de ontologías, y una vez que éstas han sido especificadas, el usuario puede realizar consultas y actualizaciones de ellas.

7.2 Identificar y definir ontológicamente los objetos geográficos contenidos en un mapa:

Esto podría realizarse de diferentes formas, en datos rásters podría ser a través de técnicas de subdivisión recursiva del espacio y almacenando los objetos en estructuras de datos como los R-Trees o Quad-Trees; dentro de este ámbito es posible encontrarse objetos que no estén claramente delimitados, o sea, que la región que ocupa determinado objeto dentro del mapa sea difícil de determinar. Como ejemplo de esto podemos encontrar mapas sobre los diferentes niveles de salinidad de los suelos o la temperatura de determinada región, como se muestra en la figura:



Fig. 30. Ejemplo de un mapa de temperaturas extremas con regiones difusas

Para el indexado de objetos espaciales de este tipo o particularmente de regiones vagas, se pueden utilizar técnicas de indexado para regiones difusas como por ejemplo [57]:

- Rectángulo de mínimo acotamiento: Que se define como el rectángulo más pequeño que encierra completamente un objeto.
- Rectángulo de mínimo acotamiento difuso: Encierra todos los puntos del mapa de espacios donde una característica del objeto de interés es encontrada.

Otra forma consistiría en la segmentación del mapa y a partir de la adquisición de los objetos geográficos segmentados, el usuario puede proceder a clasificarlos mediante relaciones complejas entre los mismos, forma, textura etc., o el sistema pudiera hacer una pre-definición de cada objeto, dando la posibilidad al usuario de redefinir o aprobar las clasificaciones recomendadas.

En datos vectoriales podría procederse a la conceptualización de cada capa, una vez que los objetos se hayan obtenido y clasificado, éstos han de insertarse o vincularse con una estructura robusta capaz de contener tanto ontologías como objetos espaciales. Estas estructuras deben ser de tipo jerárquicas debido a que este tipo de organización permite la navegación entre los diferentes niveles de especialización y generalización que presentan los conceptos semánticos, o sea, tendríamos una estructura jerárquica tipo árbol en la que estarían definidas los conceptos que se van a integrar al SIGGO y en los nodos terminales insertarles los objetos geoespaciales quedando éstos como nodos hojas dentro de la estructura, por ejemplo:

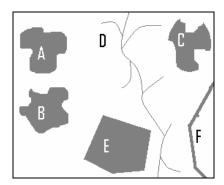


Fig. 31. Mapa de la hidrografía de un determinado lugar

A, B	Lagos Tectónicos
С	Lago Glaciar
D	Rio
E	Embalse
F	Canal

Tabla 5. Objetos espaciales definidos semánticamente

Y la siguiente taxonomía ontológica:

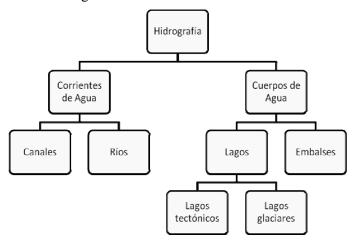


Fig. 3230. Taxonomía Ontológica.

En este caso se ha de recorrer el árbol hasta encontrar el concepto más especializado, o sea, más detallado, que defina el objeto geográfico que se va a introducir, realizando un alineamiento de ontologías en cada caso. Tanto la búsqueda como la inserción se deben realizar utilizando algoritmos adecuados, que permitan descartar rápidamente subconjuntos grandes del conjunto de ontologías y buscar en conjuntos reducidos, o sea, ir podando el árbol, quedándonos siempre con aquellas ramas de interés. Quedando finalmente como se muestra a continuación:

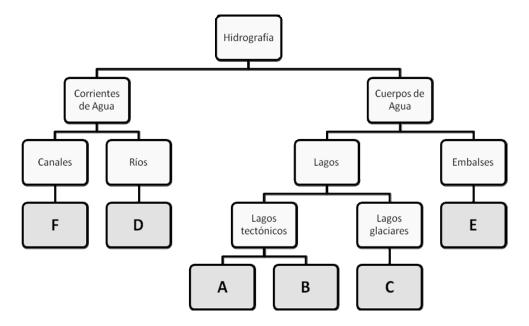


Fig. 33. Taxonomía ontológica con los objetos espaciales insertados

Los árboles de decisión son una opción tentadora debido a la naturaleza de su estructura. Éstos no son más que guías jerárquicas multivía donde los valores de los nodos, adaptándolos a nuestro problema, pueden ser ontologías que se utilizan, según el umbral de similitud, para tomar las decisiones de inserción. Las ramas que salen de los nodos corresponden a las posibles definiciones en las que puede ser semánticamente especializado nuevamente el objeto. Un árbol de decisión clasifica a un objeto, filtrándolo de manera descendente, hasta encontrar el concepto de bajo nivel mas especializado que mejor lo defina.

En general, un árbol de decisión representa una disyunción de conjunciones de restricciones en los posibles valores de los atributos de los ejemplares. Cada rama que va de la raíz del árbol a una hoja, representa una conjunción de tales restricciones y el árbol mismo representa la disyunción de esas conjunciones [58].

A continuación se muestra un tipo de árbol de decisión que podría ser de utilidad, el árbol KD. No es necesario utilizar esta estructura de datos de manera rígida sino más bien especializarla, o sea, adaptarla a la estructura ontológica que se necesita puesto que en esencia se ajusta a los propósitos que se persiguen.

Un árbol kd es un árbol de decisión que sirve para hallar una respuesta o curso de acción, dadas k variables o entradas. Cada nodo suyo hace una pregunta sobre una variable; para k variables, puede haber más de k preguntas antes de emitir una respuesta. Las hojas (nodos finales) no hacen pregunta, sino que "dan la respuesta" o curso a tomar. En el término k-d, la letra k denota la dimensionalidad del espacio que está siendo representado. En principio, es un árbol binario de búsqueda, con la diferencia de que en cada nivel un atributo diferente (o varios atributos a la vez) también llamados llave (key), es comparado cuando se quiere determinar la dirección en la que se debe buscar. El árbol k-d se ha usado para hallar rápido cierta información. Tradicionalmente, los árboles k-d han sido utilizados para hallar la información asociada a ciertos valores de k variables. Por ejemplo, ¿qué información tenemos para Tipo=avión, Destino=Habana, Fecha=miércoles? La respuesta puede ser: el vuelo Cubana 465.

A veces la respuesta es negativa, no hay o no sé, por ejemplo, si tomamos un árbol de decisión para clasificar los días de la semana de acuerdo a si son adecuados para fútbol o no y con las siguientes entradas de información **Estado** = lluvioso, **Humedad** = alta, **Viento** = verdadero, como se muestra a continuación:

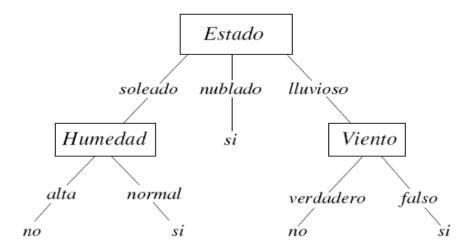


Fig. 31. Un árbol de decisión para el ejemplo de jugar fútbol

La respuesta que daría este árbol de decisión para este caso en particular sería **Jugar Fútbol** = no, puesto que es el curso natural a seguir según las variables de entrada. Este tipo de estructura hace el papel, aunque no utiliza las técnicas, de un archivo cuya llave es la concatenación de las k variables. El K-d Tree garantiza una optimización de búsquedas espaciales en sus datos.

Esta estructura puede ser utilizada además en otras áreas, por ejemplo, representar una subdivisión recursiva de un universo en subespacios por medio de hiperplanos de K-Dimensiones. Los hiperplanos son iso-orientados, y sus direcciones se alternan entre las d posibilidades, por ejemplo, para d = 2, la división de los hiperplanos es alternadamente perpendicular a los ejes X, Y.

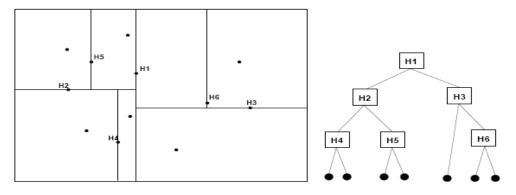


Fig. 32. Ejemplo de subdivisión del espacio utilizando un árbol KD

En [59-60] se utilizan estructuras de datos basadas en el Quadtree: el MX-CIF Quadtree [61] y el R-Tree [15] para representar los datos geográficos en formato ráster y vectoriales,

respectivamente; estas estructuras garantizan la navegación de los datos al conocimiento, o sea, que seleccionado una entidad geográfica se pueda obtener una descripción semántica de sus propiedades y relaciones a través de una estructura ontológica, dando así un nuevo enfoque al problema de incluir conocimiento en los datos espaciales y garantizar una nueva manera de tratar la integración semántica de los datos, mostrando la utilidad de los algoritmos de subdivisión recursiva del espacio para la gestión semántica de datos geográficos o geoespaciales.

7.3 Integración, navegación y/o búsqueda de información geográfica:

El desarrollador debe ser capaz de navegar por las ontologías. El navegador de ontologías tiene dos funciones importantes, puede ser usado durante la especificación de la ontología por usuarios que deseen colaborar en elaborar una ontología compartida y una vez que la ontología se ha especificado, el navegador se usa para mostrar las entidades geográficas disponibles a los diferentes usuarios. El proceso de búsqueda de información puede realizarse utilizando consultas SQL que es como comúnmente se realiza en los SIG, donde el usuario debe insertar dicha consulta para realizar las búsquedas teniendo como condición que el usuario debe tener conocimiento de dicho lenguaje, otra técnicas que podría explotarse es el uso del procesamiento de lenguaje natural. Las técnicas más usadas actualmente en esta disciplina y especialmente para la recuperación de información involucran la búsqueda por palabras clave, que en este contexto, podrían buscarse las definiciones semánticas de las palabras introducidas en el criterio de búsqueda para realizar luego un cotejo semántico en la estructura ontológica y así a través de técnicas de búsqueda inteligente brindar al usuario los resultados de una forma eficiente. Los algoritmos de búsqueda inteligente siempre comienzan en un estado inicial y la meta es llegar a un estado final u objetivo, el proceso de evaluación de las alternativas para llegar desde el estado inicial al objetivo se designa como búsqueda. El conjunto de pasos posibles para llegar desde un estado inicial al objetivo, es llamado el espacio de búsqueda. Las principales diferencias que pueden aparecer en las diferentes técnicas de búsqueda, son:

- La dirección en la que se conduce la búsqueda (hacia adelante o hacia atrás).
- La estrategia de control o forma de seleccionar las reglas que pueden ser aplicables. Los principales requerimientos de una buena estrategia de control son: que cause desplazamiento en el espacio de estado y que sea sistemático.
- La forma de representar cada nodo del proceso de búsqueda (representación del conocimiento en forma de ontologías).

Un ejemplo de cómo podría utilizarse el uso del procesamiento del lenguaje natural en un proceso de búsqueda podría ser como se muestra a continuación:

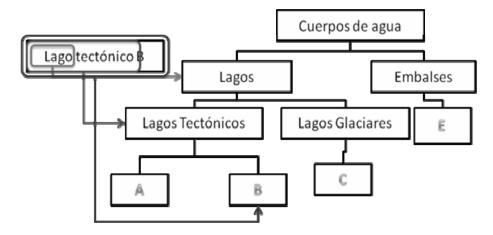


Fig. 33. Esquema de búsqueda utilizando procesamiento de lenguaje natural

Dado que el usuario está buscando a "B" e introduce como criterio de búsqueda "Lago Tectónico B", se irían cotejando las diferentes definiciones semánticas de las palabras introducidas e ir realizando la búsqueda hasta encontrar finalmente el resultado esperado.

Anteriormente se mostró que en [59-60] se utilizaban algoritmos de subdivisión recursiva del espacio para navegación de los datos al conocimiento y la propuesta va incluso más allá abordando los procesos de búsqueda y navegación en los dos sentidos: del conocimiento a los datos y de los datos al conocimiento. La navegación de los datos al conocimiento se garantiza mediante una estructura de datos de indizado de datos espaciales, mientras que la navegación del conocimiento a los datos es un problema que se inserta ya dentro de la Ingeniería Ontológica [62].

Por otro lado, al interactuar con los geodatos, podemos encontrarnos con varios problemas tales como, obtener un conjunto de datos expresados en el dominio semántico que se desea o el poder trabajar con diferentes datos independientemente del tipo de almacenamiento en el que se encuentre. En la literatura existen muchas extensiones del modelo entidad - relación para la captura de la semántica de los datos; por ejemplo, el modelo de datos orientados a objetos [63]. Sin embargo, estos modelos aún presentan limitaciones tales como considerar un solo punto de vista del mundo y una sola posible interpretación de las instancias de interés.

Para hablar de la integración de datos espaciales almacenados en fuentes heterogéneas partiremos del estudio que hicieron [64] para las bases de datos tradicionales, pero esta vez planteado desde el punto de vista de bases de datos espaciales [65], mostrando a continuación las principales heterogeneidades que se exponen en el trabajo anterior:

Heterogeneidad del modelo de datos (esquema): No todas las bases de datos en una arquitectura heterogénea son representadas por el mismo modelo de datos. Por tanto, para lograr una integración de datos es necesario que estos modelos conceptuales sean integrados. Pero integrar diferentes modelos, además, significa saber qué tipo de información es posible integrar, no será lo mismo integrar un modelo de datos para redes que integrar modelos para geografía y geometría. Por esto, será necesario una subclasificación antes de dicha integración, o medir el nivel de integración posible, pues también es cierto que podría ser necesario juntar modelos de representación de redes con modelos de representación de tipos de terreno. Para lograr esta integración será necesario

- un sinnúmero de operaciones en los esquemas (transformaciones) hasta llegar a esquemas que sean posibles de ser integrados.
- Heterogeneidad en el procesamiento de transacciones: Los diferentes SGBDs o SGBD espaciales, podrían utilizar diferentes algoritmos para el procesamiento de transacciones. Ya que, en general, las transacciones a realizar serán entre sistemas GISs, se añaden, por tanto, características espaciales que deben guardar consistencia. El problema en el procesamiento de transacciones debería subdividirse en dos niveles, uno que mantenga consistencia entre datos alfanuméricos y otro que mantenga consistencia entre datos geográficos. Por lo que el trabajo deberá ser dirigido hacia la integración de varios mecanismos de procesamiento de transacciones. Aquí es donde se puede tener mayores problemas, en el caso de sistemas geográficos, pues al tratarse de representación discreta de realidad continua, sería probable que se introduzca error tanto numérico como topológico. Esto repercutiría a la hora de comprobar la consistencia en el procesamiento de la transacción completa, ya que podría ser consistente por separado, tanto de lado geométrico como alfanumérico, pero no si se considera el conjunto. Por ejemplo, en la representación de una ciudad, en un caso, puede ser hecha como un punto en el plano a gran escala, mientras que en uno de escala pequeña se puede representar como un polígono, y al realizar transacciones en las que intervengan datos del tipo distancias, superficies, límites, se introducirá cierto error debido a la diferente representación. Por tanto, el procesamiento de transacciones contribuye a la heterogeneidad y es necesario que se tome en cuenta a la hora de integrar.
- Heterogeneidad en el procesamiento de consultas: Diferentes SGBDs podrían utilizar diferentes procesamientos y estrategias de optimización de consultas. Debido a la diversidad en la representación de elementos geográficos por dependencia de escalas, sistemas de coordenadas, sistema de proyecciones, geoides, y otros similares, el procesamiento de consultas contribuirá a esta heterogeneidad. Es claro que los sistemas de procesamiento de consultas pueden ser diferentes en cada una de las bases de datos espaciales componentes, por lo que el costo de desarrollo de una operación podría ser diferente y por tanto, el tiempo que tomaría realizarla también. Se debe considerar además que, al tener sistemas SIGs, las consultas a ser realizadas pueden estar tratadas en unos sistemas directamente sobre los datos alfanuméricos y reflejadas en los gráficos o sobre el mismo gráfico. Aquí dependerá de la versatilidad de la base de datos espacial componente así como del lenguaje y la metodología para el procesamiento de consultas.
- Heterogeneidad en el lenguaje de consulta: Los diferentes SGBDs utilizarán diferentes lenguajes de consulta. Aun si los SGBDs estuvieren basados en el modelo relacional, uno podría usar SQL y otro un lenguaje basado en Cálculo Relacional. Además, contando con la presencia de modelos para sistemas geográficos, los lenguajes de consulta variarán aún más y el espectro se verá ampliado en número de posibilidades para ocupar un lenguaje de consulta apropiado a cada sistema.
- Heterogeneidad Semántica: Los datos podrían ser interpretados de forma diferente en los diversos componentes. Es justamente esta heterogeneidad la más dificil de manejar (García-Solaco, Saltor, Castellanos [66]). Y cuando se consideran datos de tipo espacial, este problema vuelve a verse multiplicado. Considerando que el inicio de las bases de datos que mantienen datos espaciales fue muy diferente, la posibilidad de diferencia en la interpretación para la representación de datos espaciales crece aún más. Por ejemplo, la diferencia en la especialización: uno especializado en demarcación territorial dará más prioridad a los límites y por tanto, las muestras topográficas serán tomadas con mucha

más precisión escalar que otro en uso de suelos, ordenanzas o estatutos que gobiernan a la entidad que desarrolló el sistema; idioma de los datos representados: un sistema de navegación móvil seguramente deberá sobrepasar límites de países, por tanto, la pluralidad en el idioma será inminente; unidades de medida: en general, el país en el que se haya desarrollado el SIG mantendrá las unidades de medida establecidas para ese país, así podría referirse al sistema americano, el europeo o cualquier otro; sistemas de coordenadas: puede uno estar referenciado a un sistema diferente al nuevo estándar Universal Tranverse Mercator (UTM), de la misma manera podría existir diferencia para la localización de un punto en el espacio al referirse al eje z de coordenadas, pues dependerá del geoide; concepción del objeto a representar: por ejemplo, no es lo mismo hablar de lo que se mantiene con el nombre de carretera en Europa que en Sudamérica o en Asia, quizás la posibilidad de reclasificación sea necesaria para una adecuada interpretación semántica de las distintas bases de datos componentes de la federación; representación gráfica: la peor de las dificultades al momento de integrar datos geométricos de distintas fuentes será el hecho de que las instituciones encargadas hayan seguido o no normas internacionales para representación de dichos datos si no será una pérdida semántica notable; topología: existirán grandes problemas al tratar de integrar información sobre las relaciones topológicas de los objetos.

Como se ha mostrado anteriormente, para tener una buena representación ontológica del mundo real, todos los conceptos en el universo ontológico deben estar representados por un modelo del universo real y para poder realizar estas transiciones de manera efectiva se hace necesario referirse a criterios mesurables relacionados con las propiedades de los objetos o entidades, tales como el tipo y rango de valores que éstos puedan tener. Este enfoque cuantitativo es en realidad un mapeo de las características del mundo real a un mundo que está un poco más cerca del mundo digital. Una forma de afrontar estos problemas podría ser, inicialmente, construir ontologías de aplicación a partir de los datos y luego asociar estas ontologías de aplicación al dominio semántico que se desea. Al expresar la información geográfica en un espacio semántico, ésta puede ser definida, representada e integrada no sólo por lo que es, sino por lo que significa. Se puede proponer entonces tratar los geodatos en dos niveles por separado pero muy correlacionados entre sí, en el que cada uno juega un papel importante en la integración de la geo-información y en su conjunto brindarían un nuevo enfoque mucho más robusto para resolver los problemas de integración existentes hoy en día en este ámbito

Estos niveles se muestran a continuación por separado para una mayor comprensión:

• Nivel Inicial o Bajo Nivel de Abstracción:

En este nivel se representa la geo-información tal y como es almacenada y/o recogida en su formato SIG original, dígase imágenes, bases de datos, formatos vectoriales, rasters, etc.

• Nivel Semántico o Alto Nivel de Abstracción:

En este nivel se representa la geo-información a través de su conceptualización, partiendo de que los geodatos tienen un significado, éstos pueden ser definidos y representados no por lo que son sino por lo que significan logrando así una interoperabilidad sintáctica.

En [67] se presenta un método de generación automática de ontologías a partir de bases de datos geográficas que está basado en la arquitectura mostrada a continuación:

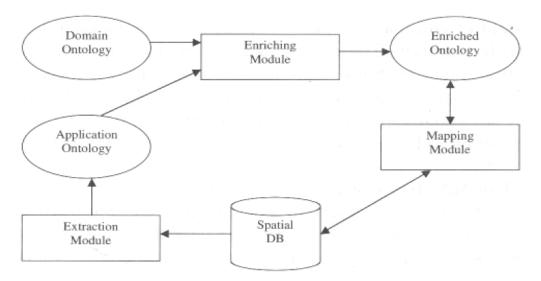


Fig. 34. Arquitectura propuesta [67]

En la figura 37 se comienza construyendo una ontología de aplicación a partir de la base de datos con la que se desea trabajar, por medio del módulo de extracción; esta ontología está compuesta por un conjunto de conceptos y las relaciones existentes entre estos conceptos, los que están explícitamente representados en la estructura de la base de datos, luego, esta ontología es enriquecida con la ontología de dominio para proveer del dominio semántico en el que se va a trabajar. Esta ontología enriquecida representa la vista semántica y la taxonomía de los datos, los que son mapeados para posibilitar consultas espaciales semánticas.

En este sentido, dado que la ontología de aplicación debe ser capaz de representar a la abstracción de los datos en la base de datos, es necesario definir explícitamente qué tipos de relaciones pueden estar presentes en ellos para poder expresar fielmente sus propiedades espaciales. Por lo que se podría definir la ontología sobre la base de dos tipos de conceptos ("terminal" y "no terminal") y dos tipos de relaciones ("tiene" y "es_un") [68].

Conceptos:

- (CT) conceptos terminales: No utilizan otros conceptos para definir su significado (que se definen por "valores simples").
- (CN) conceptos no terminales: Conceptos que utilizan otros conceptos (terminal o no terminal) en sus definiciones.

Relaciones:

- es_un: Define que un objeto A puede definirse, en un nivel de abstracción mayor, como un objeto B, por ejemplo:
 - ■Un Hospital es un Edificio.
- tiene: Define que un objeto A tiene un objeto B, por ejemplo:
 - •Un Árbol tiene Ramas.

Aquí podemos observar cómo tratar el problema de integración semántica de los geodatos que se encuentran almacenados en una base de datos geográfica, pero no todos éstos (como se dijo anteriormente) se encuentran en una base de datos geográfica, por lo que a continuación se presenta una vista arquitectónica definida en varias capas de abstracción para la generación e integración automática de conocimiento a partir de datos, independientemente del formato de

almacenamiento, mostrando así una arquitectura más robusta a diferentes tipos de almacenamiento de datos, basándose en la arquitectura anterior:

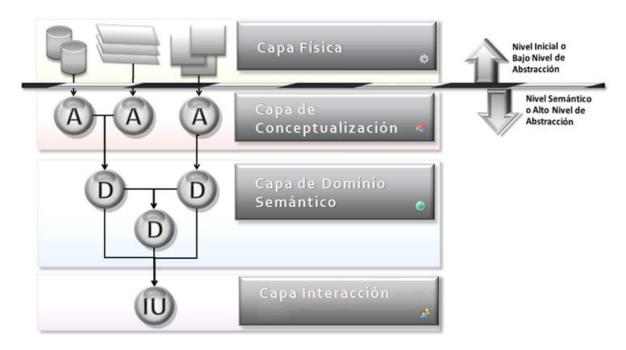


Fig. 35. Vista arquitectónica definida en varias capas de abstracción para la generación e integración automática de conocimiento a partir de datos independientemente del formato de almacenamiento

En la Capa Física de la figura se encuentran los geodatos en su forma primitiva (formato en el que han sido almacenados dígase imágenes, bases de datos, formatos vectoriales, rasters, etc), esta capa se encuentra en el Nivel Inicial o Bajo Nivel de Abstracción, luego estos datos en la Capa de Conceptualización son interpretados ubicándose así en el Nivel Semántico, o sea, aquí serán construidas las Ontologías de Aplicación, las que en la Capa de Dominio Semántico serán vinculadas a ontologías de dominio en dependencia del dominio semántico que se esté tratando; estas ontologías de dominio en esta capa podrán, si se desea, ser generalizadas en un dominio de abstracción mucho más amplio y/o ser presentadas para ser utilizadas en la Capa Interacción.

Esta vista es una generalización de pasos para la integración semántica de datos geográficos en la que se plantea una aproximación a la solución de los problemas generados por la heterogeneidad existente en este ámbito. Hay que destacar que la Capa de Conceptualización es una de las más complicadas ya que es donde los datos geográficos se llevan al nivel semántico, por lo que interpretar la gran gama de formatos en el que éstos se encuentran y conceptualizarlos es un problema aún abierto y en el que se empiezan a dar los primeros pasos.

Este diseño ha sido basado en las propuestas para la integración semántica existentes para la integración de fuentes heterogéneas de información geográfica, como las propuestas en [53].

8 Conclusiones

Los sistemas de información geográfica (SIG) son una nueva tecnología que ha revolucionado el mundo de la información en el ámbito geográfico y se han transformado hoy en las herramientas más comunes de miles de usuarios y organizaciones oficiales o privadas, utilizadas para encontrar la solución a disímiles interrogantes planteadas. El contexto general en el que surgen es el de "la sociedad de la información" en la que resulta esencial la disponibilidad rápida de información para resolver problemas y contestar preguntas de manera inmediata. Para resolver todos los problemas derivados de la integración de los datos espaciales se ha propuesto el uso de las ontologías como el elemento básico en la arquitectura de un SIG, revolucionando así la forma de pensar sobre éstos y las ontologías, y abriendo nuevos horizontes para aplicaciones de este tipo. Las ontologías, generalmente, se usan para especificar y comunicar el conocimiento del dominio de una manera genérica y son muy útiles para estructurar y definir el significado de los términos. Es así como el uso de ontologías en el desarrollo de los SIG permite establecer correspondencia y relaciones entre los diferentes dominios de entidades de información, estableciendo una nueva forma mucho más eficiente de tratar la integración semántica de los datos espaciales. Las ontologías están llegando a ser una herramienta fructífera no sólo dentro del ámbito geográfico sino también en muchas esferas como la Integración Inteligente de Información, Comercio Electrónico y la Gestión de Conocimiento.

En este trabajo se plantearon algunas consideraciones a tener en cuenta para un mejor aprovechamiento del uso de las ontologías en los sistemas de información geográfica, en estas consideraciones se ha recogido una primera aproximación a los problemas a tratar para la integración de ontologías en los SIG, por lo que es posible profundizar en ellas y en otras que no hayan sido definidas con el objetivo de conocer más a fondo todas las particularidades de los sistemas con estas características para hacerle frente a los problemas derivados de la integración semántica de los datos espaciales. Es importante destacar que en el mundo geográfico se están dando pasos sólidos para resolver los problemas de integración de los datos geográficos y aún falta mucho por hacer, pero el hecho es que el tratar los geodatos no por lo que son sino por lo que significan es el primer y más importante paso para lograr una buena interoperabilidad en el ámbito geográfico. También es importante destacar que las geo-ontologías, aunque heredan las características de las ontologías del mundo filosófico, presentan características propias e intrínsecas que son mucho más importantes en el dominio geográfico; en ellas, la topología, el contorno y las relaciones espaciales, entre otras propiedades, son más importantes debido a la naturaleza de los geodatos.

El uso de geo-ontologías como parte fundamental en la creación y desarrollo de los sistemas de información geográfica (SIG) hace que se realice un análisis mucho más amplio de la información geográfica ya que al nivel conceptual es posible integrar cualquier dato independientemente del formato en el que haya sido almacenado, resolviendo así muchos problemas que se generan con heterogeneidades tanto sintácticas como de escalas, entre otras que hacen difícil un aprovechamiento eficiente de la información geográfica existente. Crear métodos y algoritmos que sean capaces de llevar la información espacial del nivel primitivo o básico (formato en el que se encuentran inicialmente) a un nivel semántico eficientemente es una de las grandes tareas que deben ser afrontadas y en las que ya se vienen haciendo aportes interesantes.

Sin duda alguna, el aprovechamiento de la semántica existente en los geodatos hará que los sistemas de información geográfica sean mucho más eficientes, mucho más confiables y que puedan abrirse a nuevos retos en una gran gama de aplicaciones.

Referencias bibliográficas

- 1. Lehmann, F., Semantic networks, Computers Math. Applic, 1992. Vol. 23: p. pp. 1-50.
- 2. Minsky, M., *A framework for representing knowledge*. P.H. Winston (ed.), The Psychology of Computer Vision, McGraw-Hill, 1975.
- 3. Breis, J.T.F., Un Entorno de Integración de Ontologías para el Desarrollo de Sistemas de Gestión de Conocimiento, in Departamento de Ingeniería de la Información y las Comunicaciones. marzo 2003, UNIVERSIDAD DE MURCIA.
- 4. Blázquez, L.V., A.F.R. Pascual, and M.A.B. Poveda, *Ingeniería Ontológica: El camino hacia la mejora del acceso a la información geográfica en el entorno web.* 18 de Octubre de 2006: Congreso JIDEE 2006. III Jornadas de la IDEE, Castellón, España.
- 5. Smith, B. and D.M. Mark, Ontology and geographic kinds. 1998.
- 6. Redondo, A.M.S., Integración de las Ciencias de Información de la Tierra en un Sistema de Información Geográfica. 24-26 de abril de 2006.
- 7. Sheth, A.P., Changing focus on interoperability in information systems: from system, syntax, structure to semantics. http://lsdis.cs.uga.edu/library/download/898-changing.pdf. Fecha de acceso: 20 de Abril de 2008., 1999.
- 8. Clarke, K.C., *Getting Started with Geographic Information System*. Prentice Hall. Upper Saddle River, New Kersey., 1997.
- 9. Gruber, T., Ontolingua: A mechanism to support portable ontologies. Stanford University, 1992.
- 10. Smith, B., "Ontology and Information Systems". Disponible en: < http://ontology.buffalo.edu/ontology%28PIC%29.pdf>, accedido 28 de Abril 2008
- 11. Cibulsky, I.F.B., Ejemplos de Ontologías Ontologías aplicadas a la Información Geográfica. 2008.
- 12. Voronisky, F.V. and E.G. Llano, *Ontology Alignment in Geospatial Domain*. VI Congreso Internacional de Geomática 2009.
- 13. ESPAÑOLA, R.A. REAL ACADEMIA ESPAÑOLA: http://buscon.rae.es/draeI/SrvltGUIBusUsual?LEMA=taxonom%C3%ADa. Accedido el 5 de mayo 2008. Available from: http://buscon.rae.es/draeI/SrvltGUIBusUsual?LEMA=taxonom%C3%ADa.
- 14. Benthley, W. and W. Barlow, Datos geográficos. http://www.igac.gov.co:8080/igac_web/UserFiles/File/ciaf/TutorialSIG_2005_26_02/paginas/com_datos.htm, Accedido el 10 de mayo 2008, 1996.
- 15. Beckmann, N., et al., The r-tree: an efficient and robust access method for points and rectangles. 1990.
- 16. Guttman., A., *R-trees: A dynamic index structure for spatial searching.* In SIGMOD Conference, pages 47–57, 1984.
- 17. Samet., H., Spatial Data Structures: Quadtrees, Octrees, and Other Hierarchical Methods. Addison-Wesley, Reading, MA,, 1989.
- 18. Samet, H., *The design and analysis of spatial data structures*. Addison-Wesley Publising Company, Inc. USA, 1990.
- 19. Sosa, E., Procesamiento del lenguaje natural: revisión del estado actual, bases teóricas y aplicaciones Revista Internacional Científica y Profesional, Enero 1997(ISSN: 1386-6710).
- 20. BURROUGH, P., Principles of Geographical Information Systems for land resources assessment. Oxford University Press, Oxford. P 80, 1988.
- 21. deMers, M.N., Fundamentals of Geographic Information Systems, 1997.
- 22. NCGIA, *Introduction to GIS*. National Center for Geographic Information and Analysis. University of California, 1990. **Vol. 1**.
- Sendra, J.B., Sistemas de Información Geográfica, ed. M. Ediciones Rialp. Vol. 2º edición corregida, . 1997.

- 24. Open Geospatial Consortium, I. http://www.opengis.net/gml. accedido el 21 de abril 2008; Available from: http://portal.opengeospatial.org/files/?artifact_id=4700.
- 25. Open Geospatial Consortium, I., http://www.opengeospatial.org/standards/wfs accedido el 10 de abril del 2008.
- 26. Open Geospatial Consortium, I., http://www.opengeospatial.org/standards/wms accedido el 17 de abril 2008.
- 27. John D. Evans Open Geospatial Consortium, I., Web Coverage Service (WCS). 2002. 0.7.
- 28. Meaden, G.J. and J.M. Kapetsky, Los sistemas de información geográfica y la telepercepción en la pesca continental y la acuicultura. 1992.
- 29. ESPAÑOLA, R.A. REAL ACADEMIA ESPAÑOLA http://buscon.rae.es/drael/SrvltGUIBusUsual?LEMA=ontolog%C3%ADa, Accedido el 05 de mayo 2008. 2008.
- 30. Gruninger, M. and M.S. Fox, *The logic of enterprise modelling*. AI Magazine, 19(3):109-121. 1998, 1998
- 31. Fonseca, F.T., *ONTOLOGY-DRIVEN GEOGRAPHIC INFORMATION SYSTEMS*. May, 2001, The University of Maine.
- 32. Bryant, D. and B. Tversky, *Internal and External Spatial Frameworks for Representing Described Scenes*. Journal of Memory and Language 31: 74-98, 1992.
- 33. Guarino, N.a.W., C., A Formal Ontology of Properties.in: R. Dieng and O. Corby, (Eds.), Proceedings of EKAW-2000: The 12th International Conference on Knowledge Engineering and Knowledge Management. Lecture Notes in Computer Science 1920, 97-112. 2000.
- 34. Hornsby, K., Identity-Based Reasoning about Spatio-Temporal Change. 1999, University of Maine, Orono.
- 35. Guarino, N., Formal Ontology and Information Systems". Proceedings of FOIS '98. National Research Council, LADSEB-CNR. 1998.
- 36. Barchini, G., et al., El Rol de las Ontologías en los SI. Revista Ingeniería Informática, 2007.
- 37. FRANK, A.U., *Spatial Ontology: A Geographical Point of View.* O. Stock , Spatial and Temporal Reasoning, Dordrecht: Kluwer Academic Publishers, 1997.
- 38. Llano, E.G. and J.L.G. Rodríguez (2007) Los Sistemas de Información Geográfica Gobernados por Ontologías como Herramienta para la Interpretación Semántica de la Información Espacial y su integración a la IDERC. Convención Internacional Informática 2007, V Congreso Internacional de Geomática 2007. CD, ISBN 978-959-286-002-5.
- 39. Mark, D.M., A. Skupin, and B. Smith, *Features, objects, and other things: Ontological distinctions in the geographic domain.* Lecture Notes in Computer Science, editor, Spatial Information Theory: Foundations of Geographic Information Science, Berlin: Springer-Verlag, 2001.
- 40. Española, R.A., *definicion de Dato*. http://buscon.rae.es/draeI/SrvltConsulta?TIPO BUS=3&LEMA=dato, mayo 2009.
- 41. Bosque, J., Sistemas de Información Geográfica. Madrid: Rialp. S.A., 1992.
- 42. Morant, T. and M. Martín, *Modelos de Datos Espaciales*. Reuniones de Profesores de Cartografía, 2003.
- 43. Stuckenschmidt, H., et al., *Ontologies for Geographic Information Integration*. Pergamon Press, Inc. Tarrytown, NY, USA, 2002.
- 44. Bishr, Y., Overcoming the Semantic and Other Barriers to GIS Interoperability. International Journal of Geographical information Science, 1998.
- 45. Guzman, J. and R. Motz, Integración de la Información Geográfica. 2002.
- 46. Visser, U., Intelligent Information Integration for the Semantic Web. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2004.
- 47. Kavouras, M., M. Kokla, and E. Tomai, *Comparing categories among geographic ontologies*. Computers & Geosciences, Special Issue, Geospatial Research in Europe: AGILE 2003.
- 48. Schwering, A. and M. Raubal, *Spatial relations for semantic similarity measurement*. In Lecture "Notes in Computer Science, editor, Proceedings of the ER'05 Perspectives in Conceptual Modeling, 3770, pages 259-269. Springer Berlin/Heidelberg, 2005.

- 49. Hakimpour, F., Using Ontologies to Resolve Semantic Heterogeneity for Integrating Spatial Database Schemata. PhD thesis Zurich University, 2003.
- 50. Hess, G.N. and C. Iochpe, *Ontology-driven resolution of semantic heterogeneities in gdb conceptual schemas*. Proceedings of the GEOINFO'04: VI Brazilian Symposium on GeoInformatics, 2004.
- 51. Fonseca, F.T., Ontology-Driven Geographic Information Systems. 2001, The University of Maine.
- 52. Buccella, A., A. Cechich, and P. Fillottrani, *Integración de Sistemas de Información Geográfica*. IX Workshop de Investigadores en Ciencias de la Computación (WICC), 2007.
- 53. Wache, H., et al., *Ontology-based integration of information—a survey of existing approaches*. Proceedings of the IJCAI'01: 17th International Joint Conferences on Artificial Intelligence, Seattle, WA, pp.108–117, 2001.
- 54. Navarrete, T., Semantic integration of thematic geographic information in a multimedia context. 2006, Universitat Pompeu Fabra.
- 55. Worboys, M. and M. Duckham, *Integrating spatio-thematic information*. Springer-Verlag London, UK, 2002.
- 56. Ram, S., et al., GeoCosm: A Semantics-Based Approach for Information Integration of Geospatial Data. Springer-Verlag Berlin Heidelberg, 2002.
- 57. Coob, M.A., F.E. Petry, and K.B. Shaw., Fuzzy spatial relationship refinements bases on minimum bounding rectangle variations. Elsevier Science, 2000.
- 58. Argentina.UNSL, D.d.I.U.N.d.S.L.U.S.L., Aprendizaje de árboles de decisión Curso: Aprendizaje Automático y Minería de Datos. Octubre de 2006.
- 59. Llerandi, C.C., Anotación Semántica de Datos Ráster en Sistemas de Información Geográfica. Implementación en gySIG. . 2008.
- 60. Robaina, L.M., Anotación Semántica de Datos Vectoriales en Sistemas de Información Geográfica. Implementación en gvSIG. 2008.
- 61. Kedem, G., *The quad-CIF tree: a data structure for hierarchical on-line algorithms.* Proceedings of the 19th Design Automation Conference, pp. 352{357. Las Vegas, NV, 1982.
- 62. Pérez, A.G., O. Corcho, and M.F. López, Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web. Springer Verlag, 2004.
- 63. FRANK, A.U., *Spatial Ontology: A Geographical Point of View.* O. Stock, Spatial and Temporal Reasoning, Dordrecht: Kluwer Academic Publishers, 135–153., 1997.
- 64. Sheth, A.P. and J.A. Larson, Federated Database Systems for Managing Distributed Heterogeneous and Autonomous Databases. ACM Computing Surveys, 1990.
- 65. Zurita, V.M., F. Saltor, and L.P. Vidal, *Hacia la Integración de Bases de Datos Espaciales*. Proceedings of Jornadas de Sistemas de Información Geográfica, pages 69–84, Almagro, Spain., 2001.
- 66. Bukhres, O. and A.K. Elmagarmid, *Object-Oriented Multidatabase Systems: A Solution for Advanced Applications*. Prentice Hall. Upper Saddle River, New Kersey., 1996.
- 67. Baglioni, M., et al., *Building Geospatial Ontologies from Geographical Databases*. LNCS 4853, Proceedings of the Second International Conference on Geospatial Semantics, Mexico City, 2007.
- 68. Torres, M., et al., Ontology-Driven Description of Spatial Data for Their Semantic Processing. 2005.

Anexo 1 Principales sistemas de información geográfica actualmente utilizados.

Software SIG 📕	Windows 🗵	Mac OS X ⋈	GNU/Linux ⋈	BSD 🖂	Unix 🗵	Entorno Web 🔟	Licencia de software 🗵
ArcGIS	Sí	No	No	No	Sí	Sí	Software no libre
utodesk Map	Sí	No	No	No	No	Sí	Software no libre
Caris	Sí	No	No	No	No	Sí	Software no libre
CartaLinx	Sí	No	No	No	No	No	Software no libre
Geomedia	Sí	No	No	No	Sí	Sí	Software no libre
GeoPista	Java	Java	Java	Java	Java	Sí	Libre: GNU
GeoServer	Sí	Sí	Sí	Sí	Sí	Java	Libre: GNU
GRASS	Sí	Sí	Sí	Sí	Sí	Mediante py\\PS ፟፟	Libre: GNU
jvSIG	Java	Java	Java	Java	Java	No	Libre: GNU
DRISI	Sí	No	No	No	No	No	Software no libre
LWIS	Sí	No	No	No	No	No	Libre: GNU
Generic Mapping Tools	Sí	Sí	Sí	Sí	Sí	Sí	Libre: GNU
IUMP	Java	Java	Java	Java	Java	No	Libre: GNU
osmo	Java	Java	Java	Java	Java	En desarrollo	Libre: GNU
ocalGIS	Java	Java	Java	Java	Java	Sí	Libre: GNU
/lanifold	Sí	No	No	No	No	Sí	Software no libre
MapGuide Open Source	Sí	Sí	Sí	Sí	Sí	LAMPAVAMP	Libre: LGNU
Mapinfo	Sí	No	Sí	No	Sí	Sí	Software no libre
MapServer	Sí	Sí	Sí	Sí	Sí	LAMP/WAMP	Libre: BSD
Naptitude	Sí	No	No	No	No	Sí	Software no libre
fap₩indow GIS	Sí (ActiveX)	No	No	No	No	No	Libre: MPL
AicroStation Geographics	Sí	Abandonado	No	No	Abandonado	Sí	Software no libre
uantum GIS	Sí	Sí	Sí	Sí	Sí	Sí	Libre: GNU
AGA GIS	Sí	Sí	Sí	Sí	Sí	No	Libre: GNU
E Smallworld	Sí	?	Sí	?	Sí	Sí	Software no libre
avGIS	Sí	No	No	N0	No	Integración con Google Maps	Software no libre: Freeware
EXTANTE-gvSIG	Java	Java	Java	Java	Java	No	Libre: GNU
ITAL	Sí	No	No	No	No	Integración con Google Maps	Software no libre
SPRING	Sí	No	Sí	No	Solaris	No	Software no libre: Freeware
FatukGIS	Sí	No	No	No	No	?	Software no libre

RT_011, noviembre 2009

Aprobado por el Consejo Científico CENATAV

Derechos Reservados © CENATAV 2009

Editor: Lic. Lucía González Bayona

Diseño de Portada: DCG Matilde Galindo Sánchez

RNPS No. 2142

ISSN 2072-6287

Indicaciones para los Autores:

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

CENATAV

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

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

Impreso en Cuba

