

Modelos de Datos para Datos Espaciales

Mercedes Vitturini Pablo Fillottrani Silvia Castro
Departamento de Ciencias e Ingeniería de la Computación
Universidad Nacional del Sur
Av. Alem 1253 - 8000 Bahía Blanca
{ mvitturi,prf, smc }@cs.uns.edu.ar

Resumen

El desarrollo de tecnologías de recolección y digitalización de datos geográficos junto con la nueva demanda de visualización, manipulación interactiva y análisis de los resultados, ha generado la necesidad de un nuevo tipo de software dedicado, denominados GIS. Las Bases de Datos Espaciales en sus diferentes modalidades son una parte integral de los GIS: proveen capacidades para representar, almacenar y acceder a los datos, además de resolver los problemas de control de concurrencia, manejo de transacciones, backups y recuperación ante fallas.

En este trabajo de investigación se analizan los diferentes modelos de representación del espacio, los tipos de datos espaciales usados para representar información geoespacial, sus operadores y lenguajes de consultas espaciales.

1 Introducción

En estos tiempos el área de bases de datos recibe siempre nuevos desafíos. Hoy día, una base de datos puede mantener páginas web, diseño de chips, video, imágenes satelitales, música o mapas. Las funciones básicas que debe ofrecer un sistema de base de datos son siempre las mismas: almacenamiento eficiente y posibilidades de acceso y consultas sobre los datos. Sin embargo, se requiere de una amplia variedad de técnicas, integrando aproximaciones clásicas y especializadas para cada uno de los varios tipos de datos que debe manejar [RSV02]. Entre los datos especializados manejados por las bases de datos están los *datos espaciales*. Los datos espaciales son un tipo de dato que tiene asociado información geográfica. Estos datos son centrales en diferentes tipos de aplicaciones, como son los Sistemas de Información Geográfica (GIS), sistemas para Diseño Asistido por Computadora (CAD), robótica, procesamiento de imágenes e Integración en Gran Escala (VLSI). Todas estas aplicaciones coinciden en el requerimiento de objetos espaciales que serán almacenados, consultados y visualizados.

Las *Bases de Datos Espaciales* ofrecen capacidades para el manejo apropiado de datos espaciales, incluyendo modelos de para la representación de objetos espaciales, métodos de recuperación rápidos y lenguajes de consulta específicos. Los datos espaciales están formados por objetos espaciales construidos a partir de puntos, líneas, regiones, rectángulos,

superficies, volúmenes y aún objetos de dimensiones superiores que pueden incluir restricciones temporales. Como ejemplos de datos espaciales se pueden citar ciudades, ríos, rutas, áreas montañosas, entre otros.

2 Bases de Datos Espaciales y GIS

Los *Sistemas de Información Geográfica* (GIS) son sistemas usados para recolectar, analizar y presentar información que describe algunas de las propiedades físicas y lógicas del mundo de la geografía. En un sentido amplio, se denomina GIS a cualquier conjunto de procedimientos, sean manuales o asistidos por computadora, usado para manejar datos que tengan referencia geográfica [SMG⁺97]. Un dato con referencia geográfica es un dato espacial que tiene asociada una locación en la superficie terrestre.

Hace unas décadas atrás el manejo de la información geográfica estaba limitado a procesos manuales no interactivos. Con el desarrollo de nuevas tecnologías para la recolección y digitalización de geográficos, surgió la demanda de manejo y análisis interactivo de los datos. Esto dio lugar a las que hoy se conocen como aplicaciones GIS. Un GIS es más que una herramienta cartográfica para producir mapas, es un tipo de software de aplicación que incluye, entre otras, capacidades para [RSV02]:

- Almacenar datos geográficos.
- Recuperar y combinar datos para crear nuevas representaciones de espacios geográficos.
- Proveer herramientas para análisis espacial.
- Ofrecer mecanismos de visualización e interacción que permitan a los usuarios expertos analizar la información.

En una aplicación GIS típica se pueden identificar cuatro unidades funcionales principales [SMG⁺97]: ingreso de datos, modelo de datos, capacidades de manipulación de datos y facilidades para presentar los resultados. Un modelo de dato conceptual es un tipo de abstracción de dato que oculta detalles sobre la forma en que se almacenan los datos. La manipulación de datos geográficos incluye operaciones de búsqueda espaciales, operaciones geométricas (p. e. área, límite, intersección), operaciones topológicas (p. e. desconexión) y operaciones métricas (p. e. distancia).

Algunos de estos servicios que necesita un sistema GIS para funcionar se pueden rescatar de las funcionalidades ofrecidas por los *Sistemas de Manejo de Base de Datos Espaciales* (SDBMS), como son: proveer un modelo de dato conceptual con capacidades para representar datos geométricos, incluir un lenguaje de consulta con operadores de consultas sobre datos geométricos, proveer un modo de representación física eficiente de datos espaciales y permitir acceso eficiente a los datos espaciales almacenados [EN00].

A pesar de que la interdependencias que existe entre las tecnologías GIS y SDBMS, sus visiones son diferentes. Un GIS considera al DBMS como una herramienta que provee facilidades para almacenar y recuperar datos, que serán procesadas por otras herramientas. Desde el punto de vista del DBMS, un GIS es una aplicación importante pero no la única para ofrecer la habilidad de manejar datos espaciales.

2.1 Modelado del Espacio Geográfico

Existen distintas maneras de modelar y representar información geométrica y topológica. Cada uno de los procesos de modelado tiene sus fundamentos teórico. A modo de referencia se presentan algunos de los conceptos comunes para el modelado de información geográfica. Se define un *espacio* de interés \mathbb{R}^d y la función distancia Eucladiana al que se denomina *Espacio Eucladiano*. En general, a menos que se indique explícitamente d corresponde con dos. Los *puntos* son elementos de este espacio. Un punto tiene asociadas un par de coordenadas cartesianas x e y .

Según [RSV02] existen dos modelos principales para representar información geográfica: el *modelo basado en entidades* y el *modelo de campo*. El primero se caracteriza por representar objetos en el espacio. Un objeto particular podría tener asociada por ejemplo una superficie que incluya varios puntos del espacio. En la segunda visión el espacio es el foco principal. Los puntos que forman el espacio tienen asociado valores para determinados atributos, como puede ser temperatura. A continuación se describen con más detalle cada uno de los modelos.

2.1.1 Modelos Basados en Entidades

Este modelo también se denomina modelo de objetos. Concibe objetos geográficos embebidos en el espacio. En un objeto geográfico se distinguen dos componentes: (1) una descripción y (2) una componente espacial, que corresponde con la forma y ubicación del objeto en el espacio. Esta vista de la información geográfica reúne dentro de un objeto espacial puntos del espacio subyacente que comparten propiedades similares, esto es, tienen la misma descripción. Para poder distinguir a un objetos de otros, a cada objeto se le asigna una identificación. El conjunto entidad completo (identificación, objeto espacial y descripción común) constituye un objeto geográfico.

La interpretación del espacio depende de la semántica asociada con el territorio geográfico. Por ejemplo si se considera el territorio de la provincia de Buenos Aires, desde un punto de vista administrativo, la provincia se particiona en municipios; desde el punto de vista geológico, el espacio se organiza en áreas geológicas; si lo que interesa es el control del tráfico, el foco se ubica en la red de caminos. Para cada caso se define una nueva interpretación del espacio y una nueva colección de entidades describiendo el espacio. A la colección de entidades se la llama *tema*. La elección de los tipos geométricos que forman los objetos es arbitraria y depende del uso que se le dará a las colecciones de entidades. En la Tabla 1 se presentan a modo de referencia algunos tipos geométricos.

2.1.2 Modelos Basados en Campo

En la aproximación basada en campo cada punto del espacio está asociada con distintos valores de atributos, definidos como una función continua sobre x y y , por ejemplo la altitud sobre el nivel del mar, la temperatura, etc. Las mediciones sobre los distintos fenómenos se reúnen como valores de atributos variando con la ubicación en el plano. La vista del espacio como un campo continuo es lo que contrasta con el modelo el modelo basado en entidad, que identifica como entidad u objeto a un conjunto de puntos.

Objetos 0-Dimensión	<i>punto</i>	Ubican entidades cuya forma no es considerada relevante o la superficie que ocupan es muy chica con relación al espacio. Ejemplos: ciudades, centros comerciales, etc.
Objetos 1-Dimensión	<i>línea</i>	Se usan para representar entidades del con forma de redes Ejemplos: ríos, caminos, etc.
Objetos 2-Dimensión	<i>superficie</i>	Representan entidades con área. Los polígonos son el principal tipo geométrico para tales objetos Ejemplos: regiones, secciones, etc.

Tabla 1: Tipos Geométricos

2.2 Servicios del DBMS para datos Geoespaciales

Cualquier sistema GIS necesita realizar operaciones sobre datos geoespaciales y alfanuméricos. Los primeros GIS fueron construidos directamente sobre el sistema operativo propietario. A continuación se muestran dos aproximaciones que combinan tecnología DBMS y GIS.

2.2.1 Aproximación Levemente Acoplada

Algunos GIS separan el manejo de datos descriptivo del manejo de datos espaciales, por ejemplo ArcInfo (ESRI), MGE, TiGRis (Intergraph). Con esta arquitectura coexisten dos sistemas: un DBMS relacional para los datos descriptivos y un módulo específico para manejar los datos espaciales. Esta organización tiene algunas desventajas: el uso de dos modelos de datos heterogéneos dificulta las actividades de modelado, uso e integración además de perder parcialmente funcionalidades básicas de los DBMS, como por ejemplo recuperación frente a fallos.

2.2.2 Aproximación Integrada basada en la Extensibilidad de DBMS

Recientemente ha aumentado el interés por este tipo de DBMS. El concepto básico está en dar la posibilidad de definir nuevos tipos básicos (punto, línea, región) y sus operaciones a un sistema relacional. Para el caso de aplicaciones geoespaciales, la extensión al DBMS incluye: el lenguaje de consulta SQL para manejar datos espaciales y datos descriptivos y adaptación de las funciones, como el optimizador de consultas, para manejar datos geoespaciales de manera eficiente. Entre los DBMS con extensiones espaciales se pueden citar Oracle8i y Postgres

2.3 Lenguajes de Consulta

Entre los servicios que ofrece un DBMS Espacial se encuentra el Lenguaje de Base de Datos. Existen dos tipos de lenguajes diferentes: un lenguaje para especificar el esquema de base de datos, denominado *Lenguaje de Definición de Datos* (DDL) y otro para expresar las consultas y actualizar los datos, *Lenguaje de Manipulación de Datos* (DML). Entre las motivaciones que guiaron el desarrollo de sistemas GIS se encuentra la posibilidad contar con capacidades de manipulación y consulta sobre datos almacenados. Algunas de las

consultas más simples son del tipo: distancia entre dos lugares, sitios que cumplen con determinadas características. También interesar poder responder preguntas analíticas del tipo: caminos alternativos que unen un punto con otro, cómo se verá afectado el tráfico si se construye una carretera.

3 Area de Investigación

Un número importante de aplicaciones han incluido el uso de bases de datos espaciales. Entre ellas se pueden citar Sistema Observación Terrestre de la NASA, Data Warehouses y Sistemas de Información Multimedia (MMIS). Sin embargo, aún existen temas abiertos a la investigación en área de bases de datos espaciales, que están siendo estudiados en el marco de este proyecto, entre ellos se puede citar:

- Taxonomía del espacio y el error acarreado en el proceso de discretización de espacios continuos con distintas escalas o la operación inversa de interpolación para estimar campos continuos a partir de discretizaciones.
- Modelo de dato espacial en relación con las operaciones, la mayoría de los modelos de datos han sido desarrollados para espacios Eucladianos topológicos, métricos y referidos a coordenadas. Generalmente sólo se consideran operadores topológicos (límite, interior, está dentro de, conecta, etc). Se necesita trabajar para incorporar operaciones que involucren dirección (norte, arriba, izquierda, etc.) y propiedades métricas (como por ejemplo ángulo, área, perímetro)
- Modelos de dato de campo, que sustenten lenguajes de consulta y sus operadores basados en campo.
- Modelos y nuevo tipos de dato, como son arco, nodo o camino para espacios que modelen redes.

Referencias

- [EN00] R. Elmasri and S. Navathe. *Fundamentals of Database Systems*. Addison Wesley 3d. Ed., 2000.
- [RSV02] P. Rigaux, M. Scholl, and A. Voisard. *Spatial Databases*. Morgan Kaufmann, 2002.
- [SMG⁺97] S. Shekhar, M.Coyle, B. Goyal, D. R. Liu, and S. S. Sarkar. Experiences with data models in geographic information systems. *Communications of the ACM*, 40(4), 1997.