

MODELOS DE DATOS ESPACIALES

Mercedes Vitturini Silvia Castro Sergio Martig

VyGLab, Laboratorio de Investigación y Desarrollo en Visualización y Gráfica
Departamento de Ciencias e Ingeniería de la Computación
Universidad Nacional del Sur – Bahía Blanca - Argentina
{mvitturi, smc, srm}@cs.uns.edu.ar

Resumen

En distintos campos de aplicación se presenta la necesidad de manejar datos espaciales, esto es, datos vinculados a un espacio subyacente. Bajo estas circunstancias, generalmente también se requiere de capacidades para interactuar con grandes volúmenes de información, que naturalmente tienen una “representación” visual y relaciones espaciales particulares. En este sentido, la visualización de información espacial es una herramienta poderosa. Tanto los sistemas de Información Geográfica, como los de Visualización han ido evolucionando en paralelo. Los esfuerzos en el desarrollo de estándares para los datos geoespaciales raramente incluían cómo visualizar los mismos. Las librerías y estándares de Computación Gráfica evolucionaron independientemente de los modelos de datos geoespaciales. Esto resultó en ineficiencias asociadas a la geovisualización. En este contexto vemos que es necesario integrar la información espacial en una visualización. Para esto, debemos tener en cuenta los distintos aspectos que abarcan la representación de los datos espaciales y su integración con los métodos emergentes del campo de visualización lo que permitirá el desarrollo de herramientas de visualización geoespaciales efectivas. Con este propósito, es necesario considerar la representación adecuada de los datos espaciales y su integración al modelo de visualización que desarrollamos.

1. Introducción

Un Sistema de Información Geográfica (SIG) resulta de la combinación de un Sistema de Administración de Bases de Datos Espaciales (SDMBS), un conjunto de operaciones para explorar los datos y un sistema gráfico asociado al análisis geoespacial de los mismos.

Los SDBMS tienen el objetivo de proveer la administración efectiva y eficiente de datos espaciales. Este espacio puede modelar el mundo físico (geografía, planificación urbana, astronomía), partes de organismos vivos (anatomía humana), diseños de ingeniería (circuitos integrados de gran escala, diseño de automóviles o estructuras moleculares para drogas farmacológicas), etc. Los requerimientos y técnicas necesarios para poder manipular objetos geográficos, que tienen identidad, extensión, ubicación y relaciones especiales son particulares, lo que hace que los servicios ofrecidos por los sistemas de bases de datos tradicionales no sean suficientes.

La Visualización es el uso interactivo de representaciones visuales de datos soportadas en computadoras con el objetivo de amplificar el conocimiento. Aplicada a la representación y el análisis de grandes cantidades de datos abstractos, espaciales y temporales, constituye una poderosa herramienta gráfica de exploración.

A principios de los 90, ambas disciplinas evolucionaron en paralelo. Los esfuerzos en el desarrollo de estándares para los datos geoespaciales raramente incluían cómo visualizar los mismos. Las librerías y estándares de Computación Gráfica evolucionaron independientemente de los modelos de datos geoespaciales [1]. Esto resultó en ineficiencias asociadas a la geovisualización. Éstas incluyen dificultades con la registración de los datos geoespaciales dentro del software de visualización, secuencias de producciones de animación desorganizadas y confusas dentro de los ambientes GIS, y quizás lo más importante, la falta de conexiones de los ambientes de bases de datos y visualización que soporte el mostrado confiable de información geoespacial.

En este contexto vemos que es necesario integrar la información espacial en una visualización. Para esto, debemos tener en cuenta los distintos aspectos que abarcan la representación de los datos

espaciales y su integración con los métodos emergentes del campo de visualización lo que permitirá el desarrollo de herramientas de visualización geoespaciales efectivas. Con este propósito, es necesario considerar la representación adecuada de los datos espaciales y su integración al modelo de visualización que desarrollamos.

2. Contexto de la Investigación

En el contexto de visualización, hemos propuesto un Modelo Unificado de Visualización [3] que tiene en cuenta las características propias de todas las áreas de la Visualización. Nuestra propuesta consiste en un único modelo que se pueda enfocar tanto en los procesos como en los estados de los datos y que es aplicable a cualquier visualización independientemente del campo particular de origen. En este modelo quedan explícitamente representadas las operaciones provistas, los operandos sobre los que se pueden aplicar, conjuntamente con la secuencia de transformaciones propias del proceso. Un modelo tal constituye un marco conceptual en el que se pueden definir las interacciones necesarias, pudiéndose determinar sobre *quién* se opera, *qué* resultados se obtienen y *cómo* impacta sobre el proceso general.

Por otro lado, un sistema de bases de datos espacial tiene las siguientes características [2]:

- (a) *Un sistema de base de datos espacial es un sistema de base de datos*: un SDBMS debe proveer todos los servicios de los DBMS tradicionales más capacidades adicionales para manejar datos espaciales.
- (b) *Ofrece tipos de datos espaciales (SDTS) y un modelo de datos y lenguaje de consulta*: los tipos de datos espaciales, como son punto, línea, región, proveen la abstracción fundamental para modelar la estructura de entidades geométricas en el espacio y sus relaciones.
- (c) *Soporta tipos de datos espaciales y su implementación, con capacidades para indexado espacial y algoritmos eficientes para recuperar la información*: el SMDBS por lo menos debe tener capacidades para recuperar objetos desde una colección de datos almacenados sin necesidad de realizar una búsqueda secuencial. Además debería soportar posibilidades para conectar objetos diferentes por alguna relación espacial.

El principal objetivo del indexado espacial es poder soportar la selección espacial o join espacial, esto es, recuperar a partir de un conjunto grande de objetos espaciales aquellos que cumplen con una relación particular.

Hasta el momento, se han dedicado esfuerzos a esta integración y, en distintas aplicaciones como AVS/Express Prof. Ed. Y ArcGIS 3D, se han logrado establecer distintos grados de integración. Nuestro objetivo es lograr una integración de los datos geoespaciales en una visualización. Hasta el momento, hemos desarrollado el modelo de visualización y hemos relevado los modelos de datos espaciales, siendo los más representativos los que se detallan a continuación.

3. Modelos Espaciales

Los sistemas de bases de datos espaciales, distinguen dos posibilidades de representación, de acuerdo a lo que se desee ver [4,5]:

1. *Objetos Espaciales*: interesa poder distinguir, manejar y consultar entidades geográficas distribuidas en un espacio de representación. Cada una de estas entidades tiene su descripción geométrica propia.
2. *Espacio*: se describe el espacio en sí mismo, esto es, se quiere decir algo sobre cada punto del espacio.

La primera alternativa permite modelar por ejemplo ciudades, rutas, ríos, etc., la segunda vista se usa para describir mapas temáticos como por ejemplo división política.

Un *objeto geográfico* se corresponde con una entidad del mundo real y tiene dos componentes:

- Una *descripción*. Cada objeto se describe por un conjunto de atributos descriptivos. Por ejemplo el nombre y la población de una ciudad.
- Una *componente espacial*, también denominado *objeto espacial* o extensión espacial, puede incluir geometría (ubicación en el espacio geométrico subyacente, forma, etc.) y topología (relaciones espaciales con otros objetos). Por ejemplo, una ciudad puede tener como valor geométrico un polígono en el espacio 2D. La componente espacial aislada de un objeto geométrico se denomina objeto espacial. Puede darse el caso que se considere separadamente.

El atributo espacial de un objeto geográfico no tiene una correspondencia con los tipos de datos tradicionales. La representación geometría y topología requiere de modelos poderosos, que lleva a los modelos de datos espaciales. Los tipos de datos básicos que usualmente se utilizan en los modelos de datos espaciales incluyen: *punto*, *línea* y *región*.

En lo que respecta a las consultas espaciales y para que éstas brinden la flexibilidad adecuada es necesaria una representación apropiada para los datos espaciales. Entre las operaciones ofrecidas por las álgebras espaciales, las más importantes son las *relaciones espaciales*. Las consultas espaciales pueden clasificarse en:

- *Relaciones topológicas*, como adyacencia, contenido, disjunto. Estas relaciones son invariantes bajo transformaciones topológicas como son: traslación, escalado y notación.
- *Relaciones de dirección*, por ejemplo arriba, abajo, al norte de, etc.
- *Relaciones métricas*, por ejemplo distancia.

De las relaciones presentadas, las relaciones de mayor interés son las relaciones topológicas fundamentales para la mayoría de las aplicaciones GIS.

3. Indexado espacial

La representación adecuada de los datos espaciales permite responder consultas espaciales. Una posibilidad está en utilizar estructuras de datos basadas en la ocupación espacial. Los métodos de ocupación espacial descomponen el espacio donde se ubican los objetos espaciales en regiones llamadas cubetas o *buckets*. Estos métodos se conocen con el como *métodos de bucketing*. Los métodos de tradicionales de bucketing se usan sobre representaciones espaciales transformadas a puntos e incluyen: Grid File, Bang File, LSD Trees y Buddy Tree [2,4,5]. Otra aplicación de los métodos bucketing consiste en aplicar los métodos de bucketing sobre el espacio donde se ubican los objetos geométricos. Existen cuatro aproximaciones principales para descomponer el espacio.

La primera consiste en encerrar los datos basándose en el concepto de mínimo rectángulo limítrofe. En este caso los objetos son agrupados en jerarquías, posiblemente por una relación de proximidad, y almacenados en otra estructura como un B-Árbol. Los R-Árboles usan esta idea.

Los R-Árboles y sus variantes fueron diseñados para organizar una colección de objetos espaciales arbitraria representados como rectángulos d-dimensionales que encierran a sus nodos hijo. Los nodos hijo son los que tienen punteros a los objetos reales de la base de datos. Los objetos están representados por el mínimo rectángulo que los contiene. Entre las características que se pueden destacar de este método:

- Los mínimos rectángulos limítrofes que corresponden a nodos diferentes pueden superponerse.

- Un objeto puede estar contenido espacialmente en varios nodos, pero se asocia únicamente con un nodo.
- El R-Árbol de un espacio no es único. Su estructura es fuertemente dependiente del orden en que se agregan o borran los segmentos de líneas.

La principal desventaja de este método radica en que resulta en descomposiciones no disjuntas del espacio. Un objeto se asocia únicamente con un rectángulo aunque pase por varios. Esto significa que en el peor de los casos si deseamos determinar el objeto que está asociado con un punto determinado para un espacio de dos dimensiones podría resultar en una búsqueda sobre la base de datos entera.

Las otras aproximaciones se basan en la descomposición del espacio en celdas disjuntas que son mapeadas a buckets. Los objetos son descompuestos en subobjetos disjuntos, cada uno asociado con una celda diferente. Las distintas aproximaciones varían en el grado de regularidad impuesta por las reglas de descomposición subyacente y en la forma en que las celdas son agregadas.

La principal desventaja de estos métodos está en que para poder determinar el área ocupada por un objeto particular, se deberán recuperar todas las celdas que ocupa. Por otra parte, cuando la consulta es conocer todos los objetos que están ubicados en una región particular, puede resultar en recuperar el mismo objeto en repetidas oportunidades.

Entre las aproximaciones que utilizan la descomposición disjunta del espacio se pueden identificar los métodos basados en particiones disjuntas de los objetos en subobjetos arbitrarios disjuntos. Los subobjetos son luego agrupados en otra estructura como puede ser un B-Árbol. Esta idea es la que implementan los R+-Arbol y Cell- Tree.

Los R+-Arbol son una extensión de los *kd-B-trees* y manejan colecciones de objetos que son limitadas por rectángulos, de ellos se destaca:

- Están conducidos por el objetivo de evitar superposiciones entre rectángulos limítrofes.
- Cada objeto está asociado con todos los rectángulos que lo intersectan.
- Desde la raíz pueden existir varios caminos para alcanzar un mismo objeto. Esto resulta en un incremento en el tamaño del árbol, sin embargo el tiempo de respuesta es mejor.

Entre las desventajas de este método se puede citar que la descomposición es dependiente de los datos, lo que dificulta las tareas que requieren composición de operaciones y datos como por ejemplo la operación superposición.

Los siguientes métodos de indexado espacial incluyen las aproximaciones: *grilla uniforme* y métodos *quadtree*. Entre las características que se pueden citar:

- Ambos respetan la descomposición disjunta del espacio y tienen un grado mayor de independencia de datos.
- Están basados en una descomposición regular.
- El espacio se descompone en bloques de tamaño uniforme (en el caso de los métodos de grilla uniforme) o adaptar la descomposición a la distribución de los datos (como en el caso de quadtree).

El método de grilla regular es ideal para datos con distribución regular, mientras que las aproximaciones quadtree son más flexibles en ese sentido. La desventaja de los métodos quadtree está en que son sensibles a la posición de los objetos con relación a la descomposición del espacio donde se encuentran embebidos.

En esta sección se presentaron algunos de los métodos para indexado espacial. Diferentes variantes de estos se adaptan para poder armar índices espaciales, según sea el tipo de dato espacial a representar: punto, línea o región, para el caso de espacios de dos dimensiones.

Conclusiones

Hasta el momento se ha desarrollado un Modelo Unificado de Visualización que apunta esencialmente a brindar un único marco conceptual de manera independiente del dominio de aplicación de origen. En este modelo se describen las transformaciones y los estados intermedios de los datos, lo que determina que se pueda enfocar sobre los operadores o sobre los operandos favoreciendo así el entendimiento de interacciones posibles.

El trabajo que hemos realizado hasta el momento, con el objetivo de lograr una integración de los datos geoespaciales en una visualización, es un relevamiento de los métodos de representación de datos espaciales y en este artículo se presenta un resumen de los más importantes, analizando las características de cada uno de ellos.

Actualmente, y en el contexto del modelo de visualización, estamos dedicados a la integración de los diferentes métodos de representación espacial a nuestro modelo de visualización.

6. Bibliografía

- [1]. Brail R. & Klosterman, R., *Planning Support System – Integrating Geographic Information Systems, Models and Visualization Tools*. ESRI Press. Redlands California. 2001
- [2]. Güting, Ralf G., *An Introducción to Spatial Database System*. VLDB Journal (Vol. 3) Octubre 1994.
- [3]. Martig, S., Castro, S., Fillottrani, P. & Estévez, E., *Un Modelo Unificado de Visualización*. Proceedings, pp. 881-892, 9º Congreso Argentino de Ciencias de la Computación. 6 al 10 de Octubre de 2003. La Plata. Argentina.
- [4]. Rigaux, P., Scholl M. & Voisard, A., *Spatial Databases with Application to GIS*. Editorial Morgan Kaufmann Publishers. 2002.
- [5]. Samet Hanan, *Spatial Data Structures*. Addison Wesley /ACM Press, MA, 1995.