

Estructuras Espaciales para la Representación de Mapas

Leandro A. Bacic Mauricio Delladio Mariano R. Mignani
Departamento de Ciencias e Ingeniería de la Computación
Laboratorio de Visualización y Computación Gráfica (VyGLab)
Universidad Nacional del Sur
Av. Alem 1253, (8000) Bahía Blanca, Argentina
Tel (0291) 459-5135 – Fax (0291) 459-5136

1 Introducción

Unos pocos años atrás la información geográfica se representaba en mapas de papel y la manipulación de esta información estaba limitada a un proceso manual no interactivo. El rápido desarrollo de la digitalización de la información geográfica junto con una demanda creciente de manipulación y análisis de estos datos ha generado la necesidad de software dedicado. La información geográfica está constituida por grandes volúmenes de información espacial y se hace necesario contar con métodos eficientes para almacenarla y recuperar datos relevantes.

Los datos espaciales consisten de objetos espaciales basados en puntos, líneas, superficies, volúmenes y otros datos de mayor dimensión. Ejemplo de datos espaciales son ciudades, rutas, ríos, etc. En un mapa, estos objetos espaciales combinados con ciertos atributos no espaciales tales como nombres de ciudades, nombres de calles, numeración de rutas, etc. conforman la información geográfica. Las bases de datos espaciales facilitan el almacenamiento y el procesamiento eficiente de información espacial y no espacial, idealmente sin favorecer una sobre otra.

Una propuesta para la representación de datos espaciales es separar estructuralmente los datos no espaciales de los espaciales manteniendo apropiadamente la relación entre ambos. Esto nos permite agilizar las consultas sobre los datos espaciales (las operaciones espaciales son realizadas directamente sobre la estructura de datos espaciales) y nos da la libertad de elegir una estructura de datos espacial más apropiada de las que nos impone la estructura no espacial.

Existen muchos estudios teóricos dedicados al tratamiento de información espacial; no obstante, la mayoría de los trabajos de alcance público se basan en la utilización de bases de datos relacionales las cuales no resultan adecuadas debido a la naturaleza de la información subyacente.

En este contexto, se están estudiando distintas técnicas que permitan representar y manejar adecuadamente datos espaciales, restringiendo la atención a objetos cuyas componentes estén definidas en el plano 2D. Como parte de este estudio se está desarrollando un prototipo para la recuperación y visualización de información espacial que sirva de base para el diseño de distintas aplicaciones de planeamiento urbano.

2 Estructura de Datos Espacial

Existen diversas representaciones que se presentan adecuadas para modelar datos espaciales definidos en el plano 2D con el objetivo de representar información geográfica en 2D.

La elección de una estructura para la representación de esta información espacial depende de los objetos sobre los que se va trabajar y del tipo de operaciones que se desean aplicar sobre los mismos. Los objetos deben poseer una forma geométrica bien definida y pueden contar con ciertos

atributos que los caractericen. Las operaciones a utilizar sobre la estructura de datos espacial deben permitir realizar consultas sobre los objetos o sobre sus atributos, tales como consultas de posicionamiento, proximidad o basadas en atributos.

Entre las distintas formas de representación de datos espaciales en 2D se pueden mencionar las basadas en:

- a) Ocupación del plano, las cuales asumen la existencia de un arreglo de *pixels* que representan implícitamente los objetos geométricos; así se genera una imagen 2D en la que los objetos están representados como subconjuntos de *pixels* del mismo color y son distinguidos unos de otros por el color de los *pixels* adyacentes.
- b) Información explícita de la geometría, donde los objetos están representados por figuras geométricas (líneas, puntos, curvas ,etc) ubicadas sobre el plano 2D.

Debido a que un mapa poligonal se puede pensar como un conjunto de segmentos de línea que representan calles, ríos, vías ferroviarias, tendidos eléctricos, etc., la segunda propuesta de representación aparece como más natural que la primera.

Las estructuras utilizadas en la representación de datos espaciales pueden ser jerárquicas o no jerárquicas. Entre las no jerárquicas se pueden mencionar listas secuenciales, listas inversas, etc [Sam90] las cuales presentan ciertas deficiencias en el manejo de esta clase de datos.

Por su lado, la mayoría de las estructuras utilizadas en la representación de datos espaciales son jerárquicas, estando basadas en el principio de descomposición recursiva. Las estructuras de datos jerárquicas (EDJ) son útiles porque permiten tener distintos niveles de abstracción, enfocando la atención sobre un subconjunto del total de datos, lo que permite construir un conjunto de operaciones mucho más eficientes. Las EDJ son de gran interés por su claridad conceptual y su facilidad de implementación. Entre las distintas representaciones jerárquicas se encuentra el quadtree.

2.1 Quadrees

Una forma de representación de datos espaciales es hacer uso de estructuras de datos basados en ocupación espacial. Los métodos de ocupación espacial descomponen el espacio de datos en regiones llamadas Buckets. Estos métodos son conocidos como métodos *Bucketing*.. Dentro de las distintas propuestas de descomposición basadas en Buckets se pueden mencionar [Sam90]:

- Propuestas basadas en el concepto de rectángulo abarcador mínimo (ej. R-Tree).
- Propuestas basadas en la descomposición del espacio en celdas disjuntas las cuales son mapeadas dentro de lo buckets (ej: R^+ -tree y Cell tree).
- Propuestas basadas en la descomposición del espacio en bloques de tamaño uniforme (ej: uniform grid).
- Propuestas basadas en la descomposición regular disjunta del espacio en bloques (o Celdas) cuyo anchos están restringidos a ser potencia de dos, y sus posiciones están también restringidas (ej: quadtrees).

El término *quadtree* es usado, en el sentido más general, para describir una clase de EDJ cuya propiedad común es la de basarse en el principio de descomposición recursiva del espacio. Los quadtrees se diferencian de acuerdo a las siguientes características:

- *El tipo de datos que representan*: por ejemplo puntos, rectángulos, regiones, curvas, líneas, superficies, y volúmenes.
- *El proceso de descomposición*: la descomposición puede ser en partes iguales en cada nivel (llamada *descomposición regular*) o puede depender de la entrada.

- *La resolución de la descomposición*: es el número de veces que se aplica en proceso de descomposición y puede ser fija o depender de ciertas propiedades de los datos de entrada.

En general, un *quadtree* puede ser visto como un árbol donde cada nodo interno tiene cuatro nodos hijos también conocidos como cuadrantes (Nw, Ne, Sw, Se), mientras que cada hoja contiene una lista de objetos. La información de la base de datos está distribuida en las hojas. Es decir, cada uno de los nodos hoja, contiene parte de esta información.

Actualmente los *quadtree* son usados para puntos (Point quadtree), líneas y curvas (PM quadtree y sus variaciones), rectángulos, regiones (PR quadtree), superficies y volúmenes, lo que permite cubrir un amplio espectro de necesidades.

2.2 PM Quadrees

PM Quadtree es el término utilizado para describir una estructura de datos utilizada en la representación de mapas poligonales. Existen distintas variaciones del PM quadtree, entre las que se pueden mencionar PM1, y sus posteriores mejoras PM2 y PM3 [Sam90] las cuales se basan en vértices (puntos) o en lados (segmentos de líneas), y PMR que se basa solo en segmentos de línea. Sus implementaciones hacen uso de la misma estructura de datos. Todas son construidas aplicando el principio de “partir” repetidamente el conjunto de vértices o lados hasta obtener subconjuntos que son suficientemente simples para ser organizados en alguna otra estructura de datos y asociados a hojas del PM quadtree.

2.3 PMR Quadtree

El PMR quadtree es una variante del PM quadtree basada en segmentos de línea, el cual hace uso de una regla de división probabilística (probabilistic splitting). A un bloque se le permite contener un número variable de segmentos de línea. El PRM quadtree es construido insertando los segmentos uno a uno en una estructura inicialmente vacía, la cual consiste de un único bloque. Cada segmento de línea es insertado en todos aquellos bloques que intersecta u ocupa. Durante este proceso, cada bloque afectado es controlado para ver si la inserción causa el exceso de un *umbral de división* (splitting threshold). Si se excede este umbral, entonces el bloque es dividido *solo una vez* en cuatro bloques de igual tamaño. Esto evita dividir un nodo del árbol muchas veces cuando hay unas pocas líneas muy cercanas dentro de un bloque, lo que aumenta la velocidad de procesamiento y reduce el tamaño de la estructura.

Durante el proceso de borrado un segmento de línea es removido de todos los bloques que intersecta u ocupa y luego estos bloques se controlan para ver si es necesario mezclarlos (e.i las suma de los segmentos de línea de estos bloques es menor que el *umbral de división*).

2.4 Recuperación de la información a través de consultas - Ventana de consulta

Una operación importante (relacionada a una base de datos espacial) que necesita cierto grado de eficiencia es la recuperación de información contenida en la base de datos espacial.

Si bien existen distintos métodos de consulta para recuperar información de una base de datos espacial tales como consultas basadas en coordenadas espaciales, atributos no espaciales o selección de regiones poligonales, uno de los más adecuados es el conocido como ventana de consulta.

Una ventana de consulta recupera todos los objetos pertenecientes a un área rectangular determinada por un rango de valores de coordenadas x e y , que define la ventana. El método de consulta recupera de la base de datos subyacente todos aquellos bloques de información que son total o parcialmente cubiertos por la ventana de consulta. Luego, se procesan estos bloques para extraer los objetos relevantes.

Cabe destacar que en algunos casos, un mismo objeto, el cual es indivisible, puede estar representado en más de un bloque. Por ejemplo, una línea L puede estar contenida en más de un bloque a la vez. Esto introduce redundancia en la base de datos y en consecuencia, ineficiencia. En estos casos se debe encontrar una manera de recuperar cada objeto sólo una vez.

3 Prototipo en Desarrollo

En base a lo dicho y con el objetivo de contar con un modelo que sirva de base para el diseño de distintas aplicaciones que sean útiles para el planeamiento urbano, se desarrolló un prototipo que permite la generación de bases de datos espaciales, la recuperación de información de estas bases de datos y la visualización de la información recuperada.

3.1 Generación de la Base de Datos Espacial

La base de datos espacial es generada usando un método que combina el mecanismo de división recursivo de PM quater y el concepto de umbral de división de PMR quater. El valor del umbral de división es ingresado por el usuario al comienzo de la generación y define una cota superior a la cantidad de segmentos de líneas almacenados en cada hoja.

3.2 Recuperación de la información

La recuperación de información se realiza usando el método de ventana de consulta el cual permite extraer información de un subconjunto de la base de datos subyacente mediante la selección de un área rectangular.

La selección de la ventana se lleva a cabo utilizando el mouse, que permite una forma sencilla y práctica de selección rectangular. Después de la selección se procede a la recuperación de la información espacial mediante la intersección de la ventana de consulta con la base de datos subyacente.

3.3 Visualización de la información

El prototipo presenta dos vistas del mapa, una vista pequeña del mapa completo (*mapa de referencia*) y otra vista donde se muestra la información recuperada del mapa después de una consulta (*mapa de navegación*); también incluye el control del desplazamiento dentro del mapa de navegación, la visualización de las coordenadas reales del mapa, y la posibilidad de ver los bloques (buckets) en que fue descompuesto el espacio de datos.

Por último cabe destacar que debido a que AutoCAD es una herramienta muy utilizada en la creación de mapas, surge la necesidad de realizar conversiones desde estos formatos a algún otro más conveniente. Por tal motivo se incluye en el prototipo un conversor de formato DXF (AutoCAD) [Rul96] a un formato intermedio que sea útil en la generación de la BD espacial.

4 Trabajos Futuros y Posibles Extensiones

Como ya se ha mencionado, el prototipo de recuperación y visualización de información espacial se está desarrollando con la finalidad que sirva de base para el diseño y construcción de aplicaciones que sean útiles para el planeamiento urbano.

El desarrollo de una aplicación debe integrar información espacial y no espacial. Hasta ahora se desarrolló lo relativo a geometría y topología. En lo relacionado con la generación de la BD

espacial, como trabajo futuro, se plantea la necesidad de estudiar la manera de incorporar esta información no espacial a la espacial ya existente. Por su parte, se plantea la incorporación de nuevos métodos de consulta basados en atributos y la generalización del método de la ventana de consulta que permita la selección de regiones de forma poligonal. Con respecto a la visualización, se plantea incorporar mayor funcionalidad y flexibilidad. Todo esto constituye, sin duda, la base para el desarrollo de un amplio espectro de aplicaciones.

Referencias

[AS97] Walid G. Aref and Hanan Samet. *Efficient Window Block Retrieval in Quatree-Based Spatial Databases*. *GeoInformatica*, 1(1):59-91, April 1997.

[BK2001] Richard K. Brail and Richard E. Klosterman, editors. *Planning Support Systems: Integrating Geographic Information Systems, Models, and Visualization Tools*. Esri Press, first edition, June 2001.

[HS91] Erik G. Hoel and Hanan Samet. *Efficient Processing of Spatial Queries in Line Segment Databases*. In *Proceeding of the 2nd Symp. on Large Spatial databases (SSD'91)*, pages 238-255, Zürich, Aug. 1991.

[HSS97] Gísli R. Hjaltason, Hanan Samet, and Yoram J. Sussmann. *Speeding up Bulk-Loading of Quadtrees*. In *Proceedings of 5th Intl. ACM Workshop on Advances in GIS*, pages 50-53, Las Vegas, NV, November 1997.

[Rul96] Keith Rule. *3D Graphics File Format: A Programmer's Reference*, chapter 7: DXF AutoCAD R13, pages 195-240. Addison-Wesley Developers Press, first edition, Sept 1996.

[Sam90] Hanan Samet. *The Design and Analysis of Spatial Data Structures*. Addison-Wesley, 1990.

[Sam95] Hanan Samet. *Spatial Data Structures*. Appears in *Morden Database Systems: The Object Model, Interoperability, and Beyond*, pages 361-385. Addison-Wesley/ACM Press, Reading, MA, 1995.

[Tim98] Sabine Timpf. *Hierarchical Structures in Map Series*. PhD, Technical University Vienna, 1998.