

PROCESAMIENTO DE JOIN ESPACIO-TEMPORAL

Anibal Díaz ⁽¹⁾

Edilma Gagliardi ⁽²⁾

Gilberto Gutierrez ⁽³⁾

Norma Herrera ⁽²⁾

⁽¹⁾ Facultad Regional Concepción del Uruguay
Universidad Tecnológica Nacional
anibaljdiaz@yahoo.es

⁽²⁾ Departamento de Informática
Facultad de Ciencias Físico-Matemáticas y Naturales
Universidad Nacional de San Luis
{oli, nherrera}@unsl.edu.ar

⁽³⁾ Departamento de Auditoría e Informática
Facultad de Ciencias Empresariales
Universidad del Bío-Bío, Chile
ggutierr@ubiobio.cl

Resumen

Los sistemas de bases de datos que administran objetos espaciales y temporales han recibido creciente interés en los últimos años. Las bases de datos que guardan objetos espaciales que cambian su tamaño y/o su posición a través del tiempo se llaman *bases de datos espacio-temporales*. Las consultas de *join espacio-temporal* combinan dos conjuntos de objetos espacio-temporales de acuerdo con algún predicado que involucra tanto atributos espaciales como temporales. A pesar de que estos tipos de consultas son muy comunes, en la literatura se les ha prestado menos atención al procesamiento del operador join que a consultas de tipo timeSlice, interval, orientadas a la trayectoria, entre otras. En este contexto, nos proponemos estudiar exhaustivamente los join espacio-temporales con el fin de contribuir a la comprensión de este operador. El objetivo principal de este trabajo es diseñar y evaluar estrategias que permitan procesar eficientemente un join espacio-temporal.

1. Introducción

La mayoría de las bases de datos se conciben para describir algún aspecto del mundo (dimensión espacial), sobre un período de tiempo dado. Cualquier valor que se guarde describe el pasado. Es decir que un hecho definido en un registro de una base de datos es una afirmación sobre el estado del mundo en un momento dado.

Los sistemas de bases de datos que administran objetos espaciales y temporales han recibido creciente interés en los últimos años. Las bases de datos que guardan objetos espaciales que cambian su tamaño y/o su posición a través del tiempo se llaman *Bases de Datos Espacio-temporales*. En estas bases de datos, el presente, el pasado como así también la anticipación de la posición y extensión futura de los objetos, son de frecuente interés. Las aplicaciones que tratan con los objetos espacio-temporales incluyen cambios globales (datos climáticos), transporte (supervisión de tráfico), aspectos sociales (demografía, salud), y multimedia (películas animadas).

Otra característica que surge en las aplicaciones espacio-temporales en campos tales como ciencias de la tierra, cartografía y sistemas de información terrestre, involucra el almacenamiento y análisis de gran cantidad de datos espaciales históricos. Tales datos pueden ser utilizados con diferentes orientaciones, por ejemplo, modelo de datos espacio-temporales y técnicas de indexación.

Los cambios temporales de objetos espaciales inducen modificaciones en sus relaciones topológicas mutuas en el tiempo. Por ejemplo, en un mismo momento dos objetos espacio-temporales podrían ser disjuntos mientras que un tiempo más tarde pueden interceptarse (solaparse). Estas modificaciones normalmente ocurren continuamente en el tiempo pero pueden, por supuesto, también suceder en pasos discretos. Un breve pero motivante ejemplo ilustrará nuestro modelo de aproximación para estos cambios temporales. Para ello consideremos una base de datos con información sobre vuelos de aeroplanos y condiciones meteorológicas. Consultar si un aeroplano cruzó una cierta tormenta tiene una naturaleza espacio-temporal y significa chequear la validez de diversos predicados espaciales durante una serie de eventos y períodos dados en un orden temporal. Esto significa que debemos examinar si, entre los objetos *tormenta* y *aeroplano*, se ha dado la siguiente serie de eventos:

- que el aeroplano y la tormenta eran disjuntos por un momento,
- que más tarde se encuentren en un mismo tiempo,
- luego, que el aeroplano permanezca en la tormenta por un momento,
- que después el aeroplano alcance nuevamente el borde de la tormenta en algún momento, y finalmente
- que el aeroplano y la tormenta vuelvan a ser disjuntos nuevamente.

Podemos observar que durante cierto período de tiempo las relaciones topológicas entre ambos objetos son constantes y que en cierto instante del tiempo cambian. En consecuencia, obtenemos una secuencia alternante de intervalos de tiempo e instantes de tiempo en los cuales las relaciones topológicas son constantes entre ambos objetos. Informalmente hablando, una relación espacio temporal es una secuencia de relaciones espaciales que continúan desarrollándose a intervalos de tiempo o instantes de tiempo; a esto lo llamamos *acontecimiento*.

El operador que permite resolver este tipo de consultas es el **Join**, y es uno de los más comunes en las consultas a bases de datos. Este operador permite, dados dos conjuntos de datos, recuperar todos los pares de objetos que satisfacen un predicado específico del tipo de dato que se trate (relacional, espacial, temporal, espacio-temporal), respondiendo así a la consulta que se haga.

En el caso particular de las bases de datos espacio temporales, las consultas de **Join Espacio Temporal (JET)** combinan dos conjuntos de objetos espacio-temporales de acuerdo con algún predicado que involucra tanto atributos espaciales como temporales. Por ejemplo, una consulta de join espacio-temporal podría ser “identificar qué rutas de colectivos atravesaron una ciudad en el 2001”.

A pesar de que estos tipos de consultas son muy comunes, en la literatura se les ha prestado menos atención a los algoritmos de JET que, por ejemplo, a modelos de datos espacio-

temporales. Sin embargo, algunas áreas relacionadas han sido ampliamente estudiadas, como por ejemplo, algoritmos de join relacional, join espacial y algoritmos de join temporal, entre otros.

En este contexto, nos proponemos estudiar exhaustivamente el procesamiento del operador join Espacio-temporal con el fin de contribuir a la comprensión de este operador. El objetivo principal de este trabajo es diseñar y evaluar estrategias que permitan procesar eficientemente un JET.

2. Join Espacio-Temporal

Administrar datos espaciales que cambian su posición en el tiempo requiere de una administración de datos espacio-temporal efectiva y eficiente y ha dado lugar a un amplio campo de investigación en el tema. Sin embargo, pocos sistemas DBMS soportan datos espacio-temporales y, en general, se desarrollan sistemas a medida utilizando plataformas generales espacio-temporales.

En este sentido, nuestra línea de trabajo pretende contribuir en la comprensión de los JET's, introduciendo y evaluando estrategias para el procesamiento de consultas de este tipo. Con esto pretendemos que los resultados sean de interés tanto en el ámbito de desarrollo de DBMS espacio-temporales, como así también, en el de aplicaciones de usuarios.

Debido a la componente de tiempo, las bases de datos espacio-temporales necesitan administrar grandes volúmenes de datos acumulados a través de un gran período de tiempo. Una consulta sobre estos datos puede ser resuelta por medio de una búsqueda exhaustiva, es decir, accediendo a todos los objetos de la base y retornando aquellos que pertenecen a la respuesta. Obviamente esta aproximación es ineficiente debido al tamaño de estas bases de datos. Una solución más apropiada consiste en construir índices sobre los datos de manera tal que las consultas puedan responderse accediendo solamente a una pequeña parte de la base de datos.

En general, un índice es una forma de organizar un conjunto de datos en páginas de disco con el objetivos de responder un tipo específico de consulta eficientemente, es decir, accediendo solamente un pequeño número de páginas de disco. La indexación por claves primarias se utiliza para optimizar la recuperación de datos almacenados en memoria secundaria y el tiempo de procesamiento de las consultas.

La recuperación de datos y la forma en que ésta se realiza en las bases espacio-temporales es nuestra principal área de interés. Un objeto espacio-temporal es una abstracción de una entidad que tiene un identificador (clave primaria), una localización y forma espacial, una propiedad temporal y alguna otra característica que lo describe. En consecuencia, es posible una indexación de este tipo de objetos.

Teniendo en cuenta que un JET involucra dos conjuntos de datos espacio-temporales, consideraremos todos los casos que pueden presentarse durante la resolución de este operador:

- ambos conjuntos sin indexar
- ambos conjuntos indexados
- sólo uno de los conjuntos indexados.

Puede suceder que existan algoritmos para el operador en el contexto espacial y temporal que se adapten para resolver el JET o bien, que debamos diseñar nuevos métodos de indexación que permitan su resolución.

Los resultados experimentales sobre procesamiento de JET presentados en permiten realizar las siguientes observaciones que serán la base de nuestro trabajo:

- Los JET especializados ejecutan mejor que los join espaciales seguidos de una selección temporal y los join temporales seguidos de una selección espacial.
- Los join que utilizan algoritmos donde ambos conjuntos mantiene índices que se barren en forma conjunta y sincronizada, ejecutan mejor que algoritmos donde sólo uno de los conjuntos ha sido indexado.
- Los join espaciales y los join temporales ejecutan mejor si se usan índices espaciales y temporales respectivamente, que si se usan índices espacio-temporales,

No hay estudios comparables de algoritmos JET, por lo tanto no existe una base para determinar mejores diseños para resolver JET en diversas aplicaciones. Para tomar decisiones sobre cómo resolver los JET hay que considerar las aplicaciones en sí.

3. Trabajo futuro

Como trabajo futuro se pretende ahondar en el estudio de los JET's, con el fin de comprender y conocer el estado del arte del tema.

En este trabajo nos abocaremos al estudio de métodos de indexación espacio-temporal para la resolución de los JET's, con el objetivo de diseñar estructuras de datos y algoritmos que permitan resolver eficientemente este tipo de consultas y realizar trabajos de experimentación que permitan hacer comparaciones con métodos existentes.

Estos trabajos están enmarcados dentro del Proyecto "Tecnologías Avanzadas de Bases de Datos" 22/F314, Departamento de Informática, UNSL y en el marco de la Red Iberoamericana de Tecnologías del Software (RITOS2), financiado por CYTED, por lo que se ha establecido un grupo de interés en el tema conformado por docentes investigadores de la UNSL, de la UTN y de la UBB (Chile).

Bibliografía consultada

[AMW01] Hee-Kap Ahn, N. Mamoulis, Ho Min Won; "A Survey on Multidimensional Access Methods". May 2001.

[BKO 99] M. de Berg, M. van Kreveld, M. Overmars y O. Schwarzkopf: "Computational Geometry: Algorithms and Applications". Springer. 1999.

[BKS93b] T. Brinkhoff, H.P. Kriegel y R. Schneider; "Comparison of Approximations of Complex Objects Used for Approximation-based Query Processing in Spatial Database Systems", International Conference on Data Engineering (ICDE'93). pp. 40-49. 1993.

- [**BKS 94**] T. Brinkhoff, H.P. Kriegel, R. Schneider y B. Seeger; “*Multi-Step Processing of Spatial Joins*”. ACM SIGMOD Conference (SIGMOD'94). pp. 197-208. 1994.
- [**BMK99**] p.a. Boncz, S. Manegold, and M. Kersten. Database Architecture Optimized for the New Bottleneck: Memory Access. In Proceedings of the 25th VLDB, pages 54-65, Sept.1999.
- [**Cor02**] Antonio L. Corral “*Algoritmos para el procesamiento de consultas espaciales utilizando R-trees*”. Enero 2002 . Ameria .España. 2002.
- [**ChAG**] S. Chen, A. Ailamaki, P. B. Gibbons, T.C. Mowry; “*Improving Hash Join Performance through Prefetching*”.October 2003. Carnegie Mellow University. Pittsburgh.
- [**ChZ00**] C.X. Chen, C. Zaniolo. “*SQLST: A spatio-Temporal data Model and Query Language*”. In Proceedings of ER2000, 19th International Conference on Conceptual Modeling, pags. 96-111, Salt Lake City, Utah, USA, October 2000.
- [**EIN03**] R. Elmasri y S.B. Navathe; “*Sistemas de Bases de Datos: Conceptos Fundamentales*”. Addison-Wesley Iberoamericana. 2003.
- [**FMT02**] A. Frihida, D.J. Marceau, and M. Theriault.”*Spatio-Temporal Object-Oriented Data Model for Disaggregate Travel Behavior*”. Transactions in GIS, 6(3):277{294, 2002.
- [**GaG97**] V. Gaede and O. Guenther; “*Multidimensional Access Methods*”. ACM Computer Surveys, Vol. 30, N°2. pp. 170-231, 1997.
- [**GBE00**] R.H. Güting, M.H. Bühlen, M. Erwig, C.S. Jensen, N.A. Lorentzos, M. Schneider, and M. Vazirgiannis. “*A foundation for representating and querying moving objects*”. ACM Transactions on Database Systems, 25(1):1{42, March 2000.
- [**Gün93**] O. Günther; “*Efficient Computation of Spatial Joins*”. International Conference on Data Engineering (ICDE'93). pp. 50-59. 1993.
- [**GuR00**] C. Gurret and P. Rigaux : “*The Sort/Sweep Algorithm: A new method for R-tree Based Spatial Joins*”. Paris. Francia 2000.
- [**Gut84**] A. Guttman; “*R-trees: A Dynamic Index Structure for Spatial Searching*”. ACM SIGMOD Conference (SIGMOD'84). pp. 47-57. 1984.
- [**HJS98**] G.R. Hjaltason y H. Samet; “*Incremental Distance Join Algorithms for Spatial Databases*”. ACM SIGMOD Conference (SIGMOD'98). pp. 237-248. 1998.
- [**JPF04**] S.H. Jeong, N.W. Patton, A.A. Fernandez, T. Griffiths; “*An Experimental Performance Evaluation of Spatio-Temporal Join Strategies*”. Manchester. May 8, 2004.
- [**MGA03**] M. F. Mokbel, T. M. Ghanem, W. G. Aref; “*Spatio-temporal Access Methods*”.
- [**Ram96**] Raghu Ramakrishnan: “*Database Management System*”. Beta Edition, 1996, ISBN 0-07-052522-6. McGraw Hill Companies, inc.
- [**Rob81**] J.T. Robinson; “*The K-D-B-Tree: A Search Structure For Large Multidimensional Dynamic Indexes*”. ACM SIGMOD Conference (SIGMOD'81). pp. 10-18. 1981.
- [**Sam90a**] H. Samet; “*The Design and Analysis of Spatial Data Structures*”. Addison-Wesley.1990.
- [**Sam90b**] H. Samet; “*Applications of Spatial Data Structures: Computer Graphics, Image Processing, and GIS*”. Addison-Wesley. 1990.
- [**SW88**] H. Six, P. Widmayer: “*Spatial Searching in Geometric Databases*”. In Proceedings of 4th IEEE Int. Conf. On Data Engeeniering, pp. 496-503. 1988.
- [**TVS96**] Y. Theodoridis, M. Vazirgiannis, T. Sellis; “*Spatio-temporal Indexing for Large Multimedia Applications*”. In proceedings of the 3rd IEEE Conference on Multimedia Computing and Systems, pages 441-448, Hiroshima, Japan, June 1996.