

TRAYECTORIAS CON RUTA PREDEFINIDA EN UNA BODEGA DE DATOS: UN CASO DE ESTUDIO SOBRE TRANSPORTE PÚBLICO

Francisco Javier Moreno Arboleda^[*]
Fernando Arango Isaza^[**]
Jaime Alberto Echeverri Arias^[***]

Recibido: 31/08/2009

Aceptado: 07/05/2010

RESUMEN

Hoy, gracias a los sistemas de posicionamiento global y dispositivos móviles equipados con sensores se puede capturar información acerca de la trayectoria seguida por objetos móviles tales como personas, animales, vehículos, entre otros. En este artículo se examinan en particular las trayectorias seguidas por objetos móviles en una ruta predefinida, por ejemplo un sistema de transporte público en el que los vehículos deben recorrer una ruta preestablecida. Se presenta un caso de estudio donde este tipo de trayectorias se incorporan como un sujeto de análisis en una bodega de datos.

Palabras clave: objetos móviles, trayectorias, bodegas de datos, modelos multidimensionales.

* Ph. D.(c) Ingeniería de Sistemas, Universidad Nacional de Colombia, Escuela de Sistemas. Medellín, Colombia. E-mail: fjmoreno@unal.edu.co

** Ph. D. Ingeniería de Sistemas, Universidad Nacional de Colombia, Escuela de Sistemas. Cra. 80 No. 65 223 Medellín, Colombia. E-mail: farango@unal.edu.co

*** Ph. D.(c) Ingeniería de Sistemas, Líder Grupo de Investigación ARKADIUS. Universidad de Medellín. Medellín, Colombia. E-mail: jaecheverri@udem.edu.co

TRAJECTORY WITH PREDEFINED PATH ON A DATA WAREHOUSE: A CASE STUDY ON PUBLIC TRANSPORT

ABSTRACT

Today, thanks to global positioning systems and mobile devices equipped with sensors, it is possible to capture information about the trajectory followed by moving objects such as persons, animals, and vehicles, among others. In this paper we analyze, in particular, the trajectories followed by moving objects on a predefined route, as in the case of a public transport system, in which vehicles should travel by a predetermined route. We present a case study where this type of trajectories are incorporated as a subject of analysis in a data warehouse.

Key words: trajectories, moving objects, data warehouses, multidimensional model.

INTRODUCCIÓN

En los últimos años, con el avance de tecnologías como sistemas de posicionamiento global y sensores, es posible obtener una gran cantidad de información relacionada con la trayectoria seguida por objetos móviles tales como personas, animales, vehículos, entre otros. “El concepto de trayectoria está asociado con la posición cambiante de un objeto en un espacio dado durante un intervalo dado” [1]. Esta definición evidencia la naturaleza espacio-temporal de una trayectoria.

Por otro lado, aunque muchos objetos móviles se pueden desplazar libremente por el espacio, por ejemplo un pájaro en el aire, una persona o un taxi por la ciudad, existen otros objetos cuyo desplazamiento está restringido [2], por ejemplo, un tren (metro) o un bus los cuales deben seguir una ruta predefinida y se deben detener en algunos sitios específicos (estaciones).

En este artículo se examinan las trayectorias de objetos cuya ruta está preestablecida y se presenta un caso de estudio en el que este tipo de trayectorias se incorporan como un sujeto de análisis en una bodega de datos [3, 4], es decir, como una medida compleja. El caso de estudio evidencia cómo este tipo de información puede conducir al descubrimiento de tendencias y patrones espacio-temporales que de otro modo serían muy difíciles de detectar. Estos descubrimientos pueden ayudar a concebir políticas que mejoren el servicio a los usuarios y optimicen el uso de los recursos (vehículos).

Aunque existen trabajos relacionados con bodegas de datos de trayectorias [5-8] ninguno de ellos trata con trayectorias de objetos cuya ruta está predefinida. La mayoría de estos trabajos se enfocan en aspectos de extracción, transformación y carga de la información o en la concepción de operadores para analizar datos de tipo trayectoria. Por otro lado, [1, 9] examinan el modelamiento conceptual de trayectorias pero en un contexto que no está orientado a las bodegas de datos.

El artículo está organizado así: en la sección 2 se discuten las trayectorias y sus componentes. En la sección 3 se presenta el caso de estudio y se incorpora la trayectoria como una medida compleja en una bodega de datos. Finalmente, en la sección 4 se presentan conclusiones y el trabajo futuro.

1 TRAYECTORIAS

Una trayectoria es el registro de la posición cambiante de un objeto que se mueve en un espacio durante un intervalo dado $[t_1, t_n]$ [1]. Este intervalo puede ser definido por el usuario o ser inherente a la aplicación. Por ejemplo, se pueden considerar las trayectorias diarias o semanales de un vehículo. Esta definición permite que un objeto realice varias trayectorias durante su existencia, cada una con su intervalo específico. Nótese que las trayectorias de un objeto son disjuntas y no son necesariamente consecutivas en el tiempo.

Una trayectoria T se representa como una secuencia de observaciones (generadas por un sensor), es decir, posiciones con un tiempo asociado: $T = \langle o_1, o_2, \dots, o_n \rangle$ donde cada $o_i = (p_i, t_i)$, es decir, el objeto móvil está en la posición p_i en el tiempo t_i ($t_i < t_{i+1}$).

Si el objeto móvil se representa mediante una región móvil, la proyección en el plano de sus posiciones genera su área recorrida [10]. Por otro lado, si el objeto móvil se representa mediante un punto móvil, la proyección en el plano de sus posiciones genera su ruta [11-13]. Por simplicidad, se restringe la discusión de acá en adelante a puntos móviles. A menos que más información esté disponible, se supone que el objeto se mueve de la posición (x_i, y_i) a la posición (x_{i+1}, y_{i+1}) en línea recta. La figura 1 muestra la trayectoria de un punto móvil con cuatro observaciones y su correspondiente ruta.

Las observaciones pueden incluir información adicional sobre el objeto. Por ejemplo se pueden tener observaciones de un vehículo cuando a) está en movimiento e incluir información sobre su

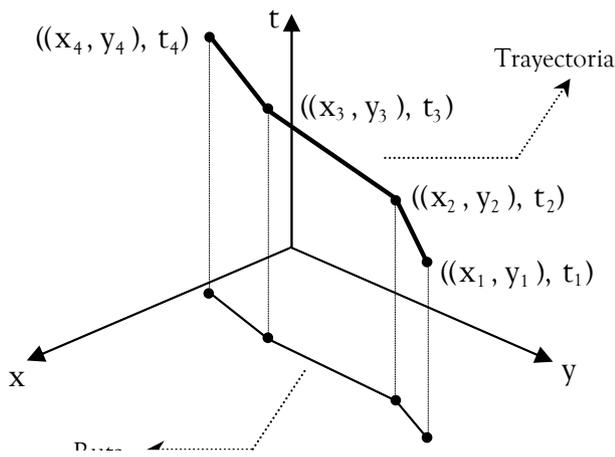


Figura 1. Trayectoria de un punto móvil.
Fuente: Elaboración propia

velocidad y su nivel de combustible, b) se detiene a recoger pasajeros y registrar cuántos pasajeros recogió y c) se detiene a tanquear y registrar cuántos galones de combustible compró. Así se pueden tener *observaciones de movimiento, observaciones de*

recogida de pasajeros y observaciones de tanqueo. Por lo tanto, dependiendo de los requisitos de una aplicación particular, las observaciones de una trayectoria se pueden clasificar en tipos. En la figura 2 se presenta un ejemplo.

2 CASO DE ESTUDIO: SISTEMA DE TRANSPORTE PÚBLICO

Considérese un sistema de transporte público. Existe un conjunto de rutas urbanas y cada ruta posee un conjunto de estaciones predefinidas donde los vehículos se detienen para recoger pasajeros. Las estaciones pueden ser compartidas entre las rutas (esto origina una relación muchos a muchos entre las estaciones y las rutas). Un vehículo puede recorrer varias rutas durante el día y puede repetir algunas de ellas. Acá se puede observar la diferencia entre trayectoria y ruta: una ruta puede ser recorrida por varias trayectorias, por ejemplo, en la figura 3

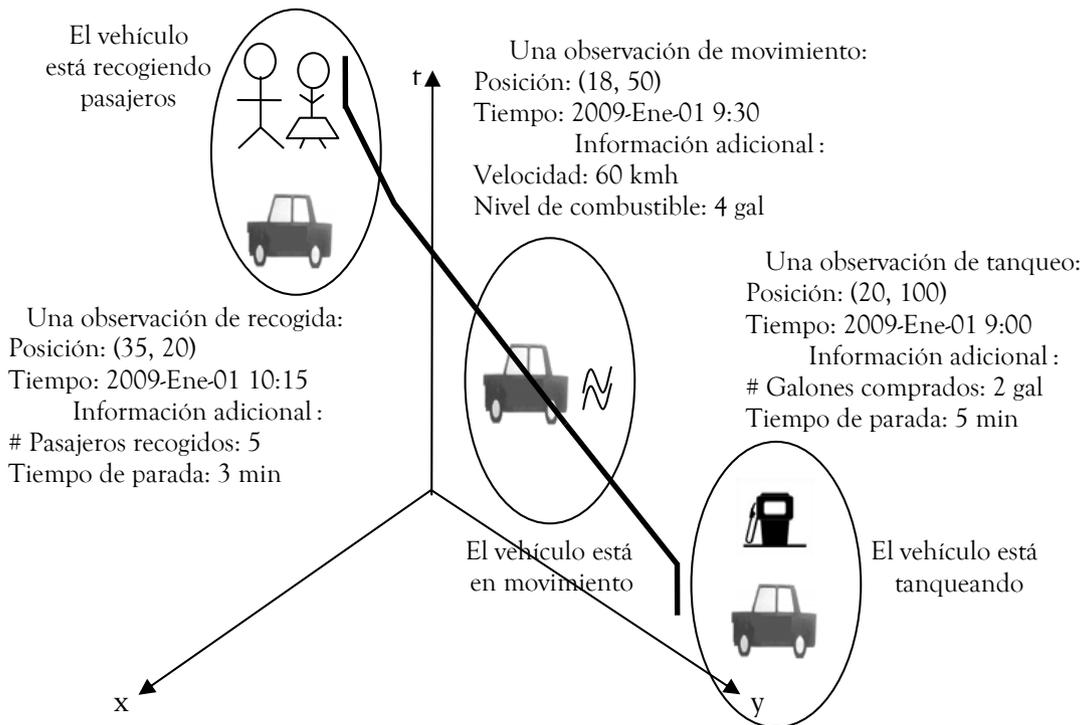


Figura 2. Tres tipos de observaciones para la trayectoria de un vehículo.
Fuente: Elaboración propia.

se muestran dos trayectorias que recorren la misma ruta R_i . Un modelo multidimensional correspondiente a una bodega de datos para representar este escenario se muestra en la figura 4. Para representar el modelo se usa la notación de [13], véase figura 5, la cual está basada en la notación del modelo entidad relación. También se usan iconos[13, 14] para representar tipos de datos espaciales como puntos, regiones y líneas, véase figura 6.

Por simplicidad, en este ejemplo se consideran sólo observaciones de recogida de pasajeros. En

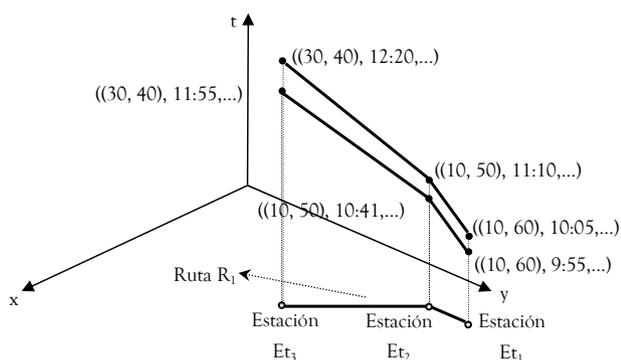


Figura 3. Dos trayectorias que recorren la misma ruta (R_i). Todos los tiempos corresponden a un mismo día.

Fuente: Elaboración propia.

las tablas 1 y 2 se presentan muestras de datos de las relaciones de hechos de recorridos y recogidas, respectivamente. La medida Veh_trayectoria es una *medida compleja* que representa la trayectoria del vehículo. Esta medida está formada por las observaciones de recogida. Para indicar esta composición en la figura 4, se coloca entre paréntesis el nombre de los hechos componentes (recogidas). Por otro lado, la medida galones consumidos indica el total de galones consumidos por un vehículo durante su recorrido.

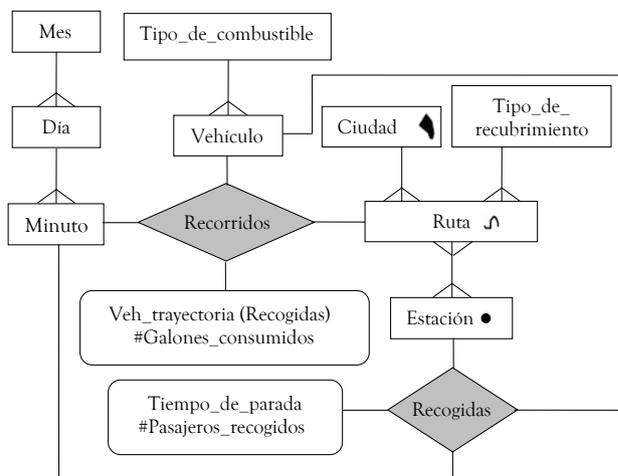


Figura 4. Un modelo multidimensional para analizar recorridos de vehículos sobre rutas preestablecidas.

Fuente: Elaboración propia.

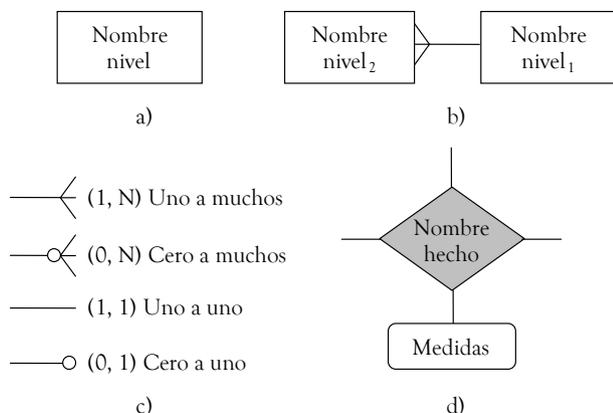


Figura 5. Notaciones esenciales para el modelo conceptual multidimensional: a) nivel, b) jerarquía, c) cardinalidades y d) hecho.

Fuente: [13]

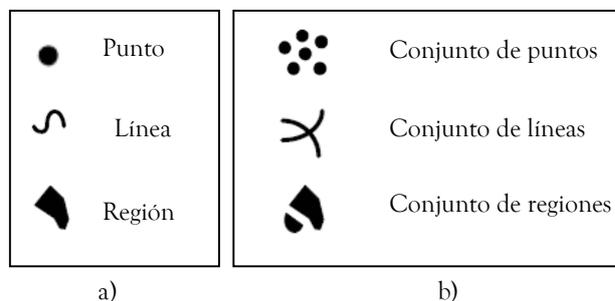
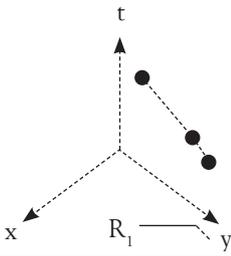
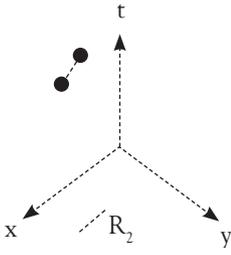
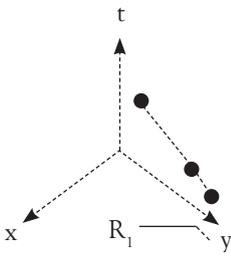


Figura 6. Notaciones para: a) geometrías simples y b) geometrías complejas.

Fuente: [13, 14]

Tabla 1. Muestra de datos de recorridos

Dimensiones			Medidas	
Minuto_inicio	Vehículo	Ruta	Veh_traectoria (Recogidas)	#Galones_consumidos
2009-ene-01 13:00	V_1	R_1		3.9
2009-ene-01 16:05	V_1	R_2		2
2009-ene-01 10:05	V_2	R_1		4.3

Fuente: Elaboración propia

Tabla 2. Muestra de datos de recogidas

Dimensiones			Medidas	
Minuto	Vehículo	Estación	Tiempo de parada (min)	#Pasajeros_recogidos
2009-ene-01 13:00	V_1	Et_1	5	19
2009-ene-01 13:55	V_1	Et_2	8	10
2009-ene-01 14:58	V_1	Et_3	3	14
2009-ene-01 16:05	V_1	Et_8	4	12
2009-ene-01 17:00	V_1	Et_9	5	20
2009-ene-01 10:05	V_2	Et_1	10	25
2009-ene-01 11:10	V_2	Et_2	12	16
2009-ene-01 12:20	V_2	Et_3	8	21

Fuente: Elaboración propia

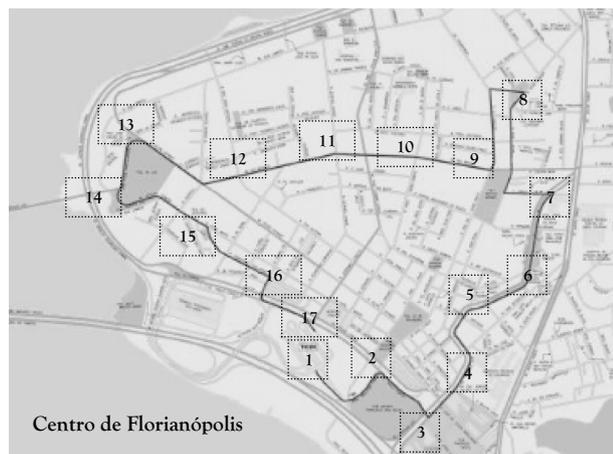


Figura 7. Ruta Circular Centro y sus estaciones.
Fuente: [15]

Nótese que algunas reglas de consistencia se deben aplicar al modelo de la figura 4. Primero, el tiempo de inicio de un recorrido debe corresponder al tiempo de inicio de su primera observación. Segundo los intervalos de dos recorridos (trayectorias) del mismo vehículo deben ser disjuntos. Tercero, si p es una observación de recogida de un recorrido r , entonces la estación asociada con p debe estar localizada en la ruta asociada con r , es decir, las estaciones visitadas en un recorrido deben ser las correspondientes a las estaciones de la ruta asociada con el recorrido.

Aunque sería posible enriquecer el modelo con notaciones adicionales para expresar estas reglas, incluso con tales mejoras se debe tener presente que un modelo conceptual es un diseño parcial que se suele complementar con documentación adicional.

Este modelo puede ayudar a los analistas a identificar, por ejemplo, las estaciones donde más pasajeros son recogidos, los vehículos que más combustible consumen de acuerdo con el tipo de recubrimiento de las rutas, y las rutas más recorridas. Los resultados podrían sugerir la creación de estaciones o rutas, ayudar a escoger los vehículos para recorrer una determinada ruta con el objetivo de ahorrar combustible, e indicar mantenimiento preventivo de rutas muy recorridas.

Con el fin de mostrar la conveniencia de la propuesta en un escenario real, se analizó la ruta “Circular Centro”, de la empresa Transol [15] de la ciudad de Florianópolis, Brasil. Esta ruta pasa por 17 estaciones, véase figura 7, y es recorrida nueve veces en un día. En la tabla 3 se presenta una muestra de datos correspondiente a la trayectoria de un ómnibus que cubre dicha ruta. Se analizaron los datos correspondientes a la operación de esta ruta durante una semana. El número de pasajeros que subió y bajó en cada estación se muestra en la tabla 4.

Tabla 3. Muestra de datos para análisis de la ruta circular centro.

Minuto	Vehículo	Ruta	Datos sobre pasajeros en las estaciones		
			Estación	#Pasajeros_recogidos	#Pasajeros_bajados
2009-nov-01 12:30	183	Circular Centro	1 TICEN	10	-
2009-nov-01 12:33	183	Circular Centro	2 Lobo Acessórios	6	0
2009-nov-01 12:35	183	Circular Centro	3 Próximo nº 930	4	2
2009-nov-01 12:38	183	Circular Centro	4 Próximo nº 1232	5	1
...					
2009-nov-01 13:01	183	Circular Centro	17 Casas D'Água	0	1
2009-nov-01 13:05	183	Circular Centro	1 TICEN	-	19

Fuente: [15]

Tabla 4. Número de pasajeros que subió y bajó en cada estación de la ruta circular centro entre 2009-nov-01 y 2009-nov-07.

Estación	1	2	3	4	5	6	7	8	9
#Pasajeros recogidos	660	400	345	380	411	345	592	780	550
#Pasajeros bajados	-	93	220	341	400	330	501	710	600
Estación	10	11	12	13	14	15	16	17	1
#Pasajeros recogidos	400	345	380	411	345	392	210	80	-
#Pasajeros bajados	388	401	321	453	312	452	420	118	966

Fuente: [15]

Aunque se requiere un análisis más exhaustivo, los resultados sugieren, por ejemplo que la estación 17 (Casas D'Água) podría ser eliminada de esta ruta, ya que muy pocos pasajeros suben y bajan allí; la mayoría bajan en la estación 1, que es la terminal TICEN, de donde también parte la ruta. Nótese que esta es una situación análoga a la que sucede con la estación Alpujarra del metro de Medellín [16] en los horarios de menor congestión del sistema. De hecho, en algunas ocasiones el metro no se detiene en esa estación. Por otro lado, la estación 8 (Casa das Tortas) evidencia bastante afluencia de pasajeros; una causa podría ser su cercanía con el litoral. Este aspecto podría sugerir el paso de la ruta por estaciones más cercanas al litoral. Sin embargo, se debe analizar si otras rutas de transporte pasan por allí.

3 CONCLUSIONES Y TRABAJO FUTURO

En este artículo se examinaron las trayectorias de objetos móviles sobre una ruta predefinida y se presentó un caso de estudio sobre transporte público donde las trayectorias se incorporan como un sujeto de análisis en una bodega de datos, es decir, como una medida compleja. El análisis de este tipo de medidas puede dar lugar al descubrimiento de

tendencias y patrones que ayuden a los analistas de una organización en la toma de decisiones.

Como trabajo futuro se planea implementar el modelo presentado. Desde un nivel físico se debe abordar el problema de cómo almacenar de forma eficiente los datos de una trayectoria en un contexto multidimensional. Igualmente se deben concebir y formalizar operaciones de agregación para trayectorias, por ejemplo, ¿qué significa sumar o promediar trayectorias? Un punto de partida para desarrollar este aspecto son los trabajos de [5, 7-9, 17].

REFERENCIAS

- [1] S. Spaccapietra *et al.*, "A conceptual view on trajectories," *Data & Knowledge Engineering* vol. 65, no. 1, pp. 126-146, 2008.
- [2] G. Kellaris *et al.*, "Trajectory compression under network constraints," en 11th International Symposium (SSTD), Aalborg, Dinamarca, 2009, pp. 392-398.
- [3] W. H. Inmon, *Building the data warehouse*, 4th ed., Indianapolis, Ind.: Wiley, 2005.
- [4] R. Kimball, *The data warehouse lifecycle toolkit*, 2nd ed., Indianapolis, IN: Wiley Pub., 2008.
- [5] F. J. Braz, "Trajectory data warehouses: proposal of design and application to exploit data," en 9th Brazilian Symposium on GeoInformatics (GeoInfo), Campos do Jordão, Brasil, 2007, pp. 61-72.
- [6] S. Orlando *et al.*, "Trajectory data warehouses: design and implementation issues," *Journal of Computing Science and Engineering* vol. 1, no. 2, pp. 211-232, 2007.
- [7] S. Orlando *et al.*, "Spatio-temporal aggregations in trajectory data warehouses," en 9th International Conference on Data Warehousing and Knowledge Discovery (DaWaK), Regensburg, Alemania, 2007, pp. 66-77.
- [8] G. Marketos *et al.*, "Building real world trajectory warehouses," en 7th International ACM SIGMOD Workshop on Data Engineering for Wireless and Mobile Access (MobiDE'08), Vancouver, Canada, 2008, pp. 1-8.
- [9] S. Brakatsoulas *et al.*, "Modeling, storing, and mining moving object databases," en 8th International Database

- Engineering and Applications Symposium (IDEAS), Coimbra, Portugal, 2004, pp. 68-77.
- [10] R. H. Güting, y M. Schneider, *Moving Objects Databases*, San Francisco, CA: Morgan Kaufmann Publishers Inc, 2005.
- [11] M. Vazirgiannis, y O. Wolfson, "A spatiotemporal model and language for moving objects on road networks," en 7th International Symposium in Advances in Spatial and Temporal Databases (SSTD), Redondo Beach, USA, 2001, pp. 20-35.
- [12] E. Frenzos *et al.*, "Nearest neighbor search on moving object trajectories," en 9th International Symposium in Advances in Spatial and Temporal Databases (SSTD), Angra dos Reis, Brasil, 2005, pp. 328-345.
- [13] E. Malinowski, *Advanced data warehouse design : from conventional to spatial and temporal applications*, 1st ed., New York: Springer, 2008.
- [14] C. Parent *et al.*, "Spatio-Temporal conceptual models: data structures + space + time," en 7th ACM International Symposium on Advances in Geographic Information Systems (ACM-GIS), Kansas, USA, 1999, pp. 26-33.
- [15] TRANSOL. «Transol Transporte Coletivo Ltda,» agosto, 2009; <http://www.transoltc.com.br>.
- [16] METROMED. «Empresa de Transporte Masivo del Valle de Aburrá Limitada,» agosto, 2009; <http://www.metrodemedellin.gov.co>.
- [17] H. Jeung *et al.*, "Convoy queries in spatio-temporal databases," en 24th IEEE International Conference on Data Engineering (ICDE), Cancún, México, 2008, pp. 1457-1459.