

(SEDET) Sistema para la exploración de datos espacio - temporales.

Camilo Andrés Nemocón Farfán
Grupo de Investigación Imagine
Universidad de Los Andes
Bogotá, Colombia
c-nemoco@uniandes.edu.co

Resumen— Este artículo da a conocer el desarrollo del modelo de consultas del sistema SEDET, implementado en una herramienta que facilite la forma de expresión de las consultas Espacio Temporales de forma visual e interactiva, a partir de un modelo que administra la información en tres aspectos de espacio, tiempo y datos, así como de sus niveles de granularidad, para visualizar, navegar y explorar los datos en sistemas, apoyando la tarea de análisis de exploración de datos, para la identificación y ubicación del comportamiento de los datos de movilidad de la ciudad de Bogotá mediante la analítica visual.

Palabras Claves— Analítica visual, sistemas urbanos, movilidad, interacción, visualización, modelo de consultas.

I. INTRODUCCION

El crecimiento de la ciudad de Bogotá en la última década se dio en sentido sur, en la periferia de las localidades de Bosa y también en Ciudad Bolívar en los límites con la localidad de Usme donde se caracteriza por el inicio de nuevos desarrollos de tipo formal e informal. Así mismo, se realizaron proyecciones de la población, calculadas con base en el Censo General de 2005, para Bogotá se estima que en el año 2020 habitan en la ciudad 8'118.782 personas. En cuanto a las características geográficas, se han reglamentado las dos zonas más importantes del suelo de expansión las cuales se ubican en el norte de la ciudad en las localidades de Suba y Usaquén, y en el sur en la localidad de Usme. De otra parte, la densidad urbana de la ciudad, para el año 2010 corresponde a 191 personas / ha [1].

Este crecimiento acelerado de la población y de la ciudad, requiere de una planeación urbana que satisfaga las nuevas necesidades de servicios, infraestructura, vías de acceso de la ciudad, por tanto los expertos en sistemas urbanos necesitan visualizar simultáneamente gran cantidad de datos y explorarlos para descubrir patrones, relaciones, tendencias u otras características que apoyen su posterior proceso de análisis, planteamiento de hipótesis y toma de decisiones.

Sin embargo para realizar la tarea de planeación urbana, uno de los procesos que se debe realizar es identificar y explorar

los datos que se tienen, para ver la calidad de estos y generara un análisis con respecto a datos que puedan ser correlacionados. Los datos que utilizan los expertos en áreas urbanas son datos espacio temporales [2], los cuales permiten a los usuarios identificar y ubicar los factores importantes, para que el experto pueda ver el comportamiento de los distintitos puntos de la ciudad a partir de la segmentación por periodos de tiempo (año, mes, día y hora), permitiendo el análisis, comparación y visualización de la información.

En la actualidad es cada vez es más frecuente realizar análisis sobre datos espacio – temporales en el amplio tema de sistemas urbanos, sin embargo se presentan inconvenientes en la adquisición y estructuración de los datos, ya que estos pueden ser ambiguos, incompletos, con ruido, y al venir de muchas fuentes, terminan siendo datos heterogéneos y no estructurados, lo cual dificulta que éstos se puedan implementar sobre cualquier aplicación y que la aplicación le permita al usuario encontrar patrones, filtrar el ruido, focalizarse en un contexto claro, generando dificultades en el análisis de las situaciones.

Una forma de visualizar la información en sistemas urbanos, es por medio de la computación interactiva, donde se generan representaciones visuales de datos abstractos con el fin de ampliar el conocimiento sobre una situación [3], y esto se puede realizar por medio de la analítica visual.

Este proyecto se centra en proveerle, a los expertos en sistemas urbanos, una herramienta que les permita ver la información correspondiente de datos espacio temporales, es decir, que tenga una vista general de los datos. Así mismo este sistema se enfoca en apoyar el desarrollo de las tareas de análisis por medio de componentes visuales e interactivos que le ayuden al usuario a expresar y realizar consultas complejas y visualizar la información de manera efectiva. Por tanto el sistema se representa en una herramienta interactiva que permite su integración a otra plataforma que se encarga del manejo de datos, lógica del negocio y conexión directa a la base de datos, para su visualización y su manipulación con distintos dispositivos de interacción.

II. TRABAJO RELACIONADO

Existen trabajos de investigación para la integración de sistemas de información con fines de simulación y planeación urbana, como es el trabajo de Wang cuya aplicación llamada CAL3QHC [4], integra los sistemas de información geográfica como ArcGis, con simuladores de Trafico y simuladores de calidad de aire, generando visualizaciones pertinentes para los expertos, sin embargo el usuario debe generar los datos en cada simulador para ser integrados en la aplicación, lo cual genera un aumento en el tiempo para la preparación de los datos para la visualización de cada tema a analizar y no permite la visualización web centrado y limitando la información y la toma de decisiones.

Otro proyecto de vinculación de analítica visual y Big data, es VisGets [5], en donde se centran en el desarrollo de una interfaz web para formular consultas sobre los datos, donde muestra un listado con las respuestas de la consultas utilizando widgets de consulta interactivos, sin embargo la visualización de las respuestas no es clara y es limitado para la visualización de datos heterogéneos.

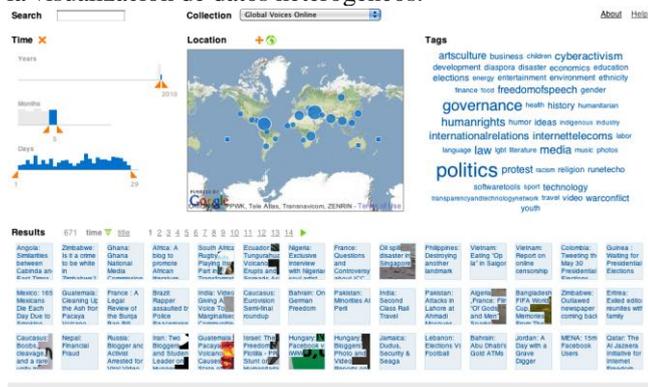


Figura 1. Interfaz de VisGets.

También encontramos otra aplicación llamada Europeana Connect [6], en donde se puede obtener información relevante frente a tópicos históricos dependiendo de la ubicación y el momento en el tiempo consultado, permitiendo visualizar espacial y temporalmente la información de un tópico específico, sin embargo no permite la relación de temas que podrían ser importantes para el usuario.



Figura 2. Interfaz de Europeana Connect.

III. DISEÑO E IMPLEMENTACIÓN

MODELO DE CONSULTAS:

Este sistema de visualización interactivo consta de un modelo que administra la información en una estructura de desacoplamiento en los tres aspectos de espacio, tiempo, y datos, así como de sus niveles de granularidad. De esta manera, los niveles pueden ser mostrados y explorados de lado a lado para facilitar su análisis combinado. Este modelo permite la interacción por medio del semantic zoom y desagregación de los datos viéndolos desde diferentes niveles.

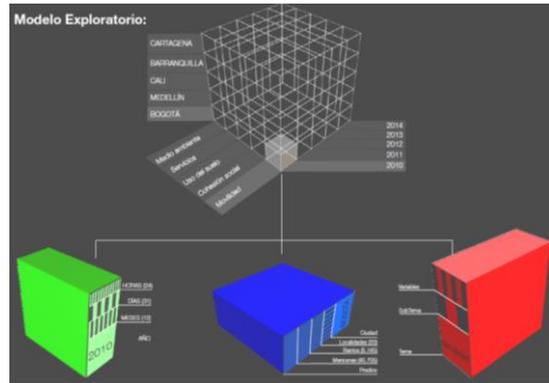


Figura 3. Modelo Exploratorio.

Con base en este cubo puedo generar una estructura de selección Y exploración, ya que permite establecer una navegación para la lectura, búsqueda y filtrado en datos espacio temporales y sus diferentes granularidades, donde cada celda en el cubo representa un objeto espacial asociado a un atributo y a un paso en el tiempo, dentro del cual hay niveles de granularidad.

La navegación por el cubo a nivel general se puede realizar sobre el eje "X", para la exploración de atributos, sobre el eje "Y" para los lugares y sobre el "Z" podemos encontrar el tiempo y permite la selección y filtro de datos.

Este modelo se basa en la estructura planteada por Pequet [7], quien plantea que a partir de la pregunta de 2 componentes de los datos espacio – temporales, el resultado es el tercer componente, el cual es el dato que busca el experto. Lo bueno de este modelo es que me permite seleccionar cada uno de los componentes espacio – temporales, sin embargo no me permite la selección de más de 2 slots de un mismo componente, es decir, no me permite la selección de más de un lugar, de un tiempo o una variable.

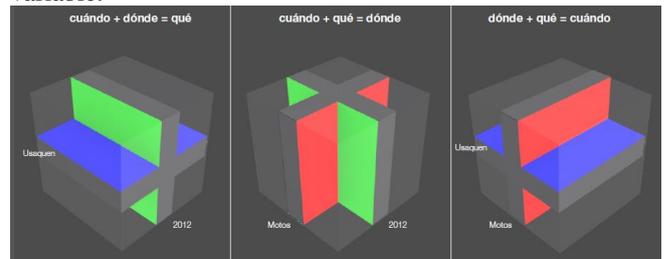


Figura 4. Estructura de consultas.

También se tuvo en cuenta los niveles de información de Bertin [8], ya que él plantea la forma de selección de más de

un elemento de un mismo componente espacio temporal a través de sus niveles, sin embargo, esta estructura no permite la navegación por los diferentes niveles de granularidad de tiempo, espacio y atributos.

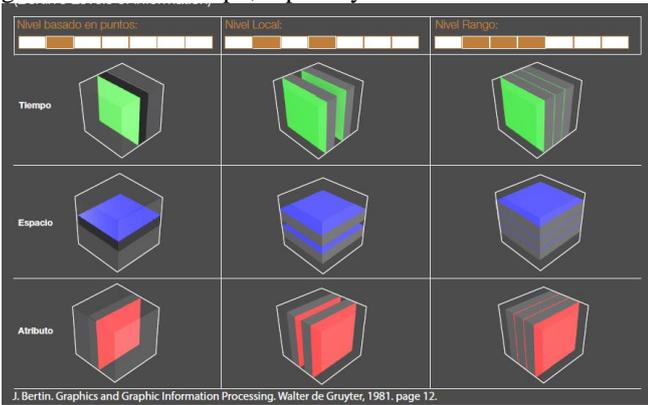


Figura 5. Niveles de información.

Por tanto se realizó una extensión de la estructura de datos de J. Bertin y D. Peuquet, extendiéndolo al dominio del tiempo, el espacio y los atributos mediante un modelo de desacoplamiento de estos tres dominios, teniendo en cuenta los niveles de granularidad y su escalabilidad. A esta extensión se le llamo **modos exploratorios**, el cual apoya la tarea de análisis de exploración de datos mostrando información solicitada y eliminando de la vista del usuario los datos que no se ajusten a las restricciones de consulta.

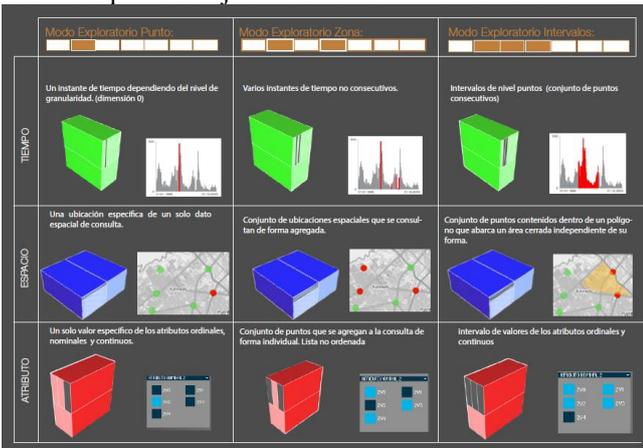


Figura 6. Modos Exploratorios.

Este modelo se puede representar en forma de grafo, en donde el núcleo del grafo está compuesto por los componentes de Espacio, Tiempo y Atributos, los cuales poseen sus propios niveles de granularidad, generando una estructura basada en un árbol por cada componente, representando de esta forma la jerarquía de las granularidades de los objetos del modelo con un sistema no direccional entre los componentes, lo que permite la relación libre del núcleo del grafo y la producción de consultas a partir de las respuestas de otras consultas, permitiendo una exploración de los datos de acuerdo al interés de los expertos.

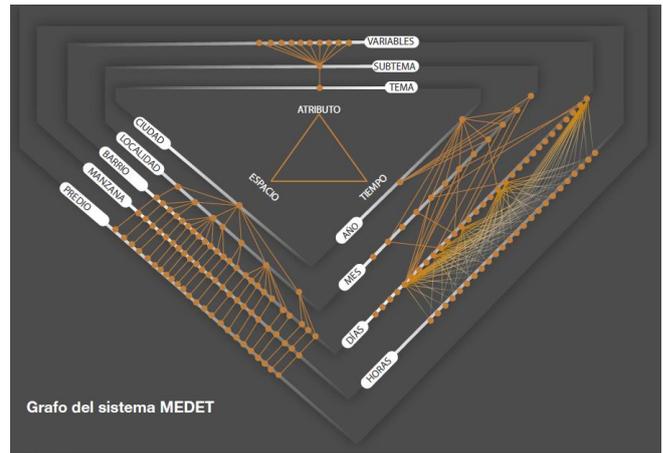


Figura 7. Grafo del sistema SEDET.

El modelo permite generar respuestas sobre los filtros de los datos realizados por el usuario, y sobre estos resultados se pueden realizar nuevas consultas, permitiéndole al usuario analizar los resultados con otros parámetros, planteando nuevas hipótesis y generar un análisis sobre cada una. Este sistema de grafos y consultas permite devolverse a anteriores consultas para retomarlas y plantear otras hipótesis con otros parámetros, generando finalmente un historial de las consultas realizadas por cada usuario.

IMPLEMENTACIÓN:

A partir del modelo exploratorio definido anteriormente, se construye un sistema de visualización e interacción para la formulación de consultas en datos espacio – temporales, llamado SEDET, el cual posee componentes que se encargan de construir y manejar los núcleos del grafo del sistema y la relación entre ellos permitiendo generar consultas y responder a estas de forma visual e interactiva.

Con este modelo como base se diseña y desarrolla una herramienta, que facilita la forma de expresión de las consultas Espacio Temporales, permitiendo su integración a otra plataforma de analítica visual que administra los datos, mediante una estructura cliente – servidor, donde la plataforma de administración de datos y que tiene conexión directa a la base de datos es VAFUS [9], y SEDET se encarga de la visualización de los controles de consulta (espacio, tiempo y atributos) dentro de una aplicación cliente y de la visualización de los resultados de las consultas dentro de una aplicación servidor conectada a VAFUS, así mismo, SEDET se encarga de enviarle las consultas realizadas por el experto desde la aplicación cliente a la plataforma VAFUS mediante unos vectores de información, para que la aplicación servidor pueda mostrar los resultados de la consultas; así mismo, el usuario puede profundizar en la exploración de los resultados, seleccionando uno o un conjunto de datos de los resultados para así generar nuevas consultas a partir de las respuestas de otras consultas. La selección de los datos de resultados son enviados por los vectores de información a la aplicación cliente, para mostrarle al usuario con mayor detalle en que parte se encuentra ese dato, evidenciando todos los niveles del grafo que se navegó para encontrar la respuesta.

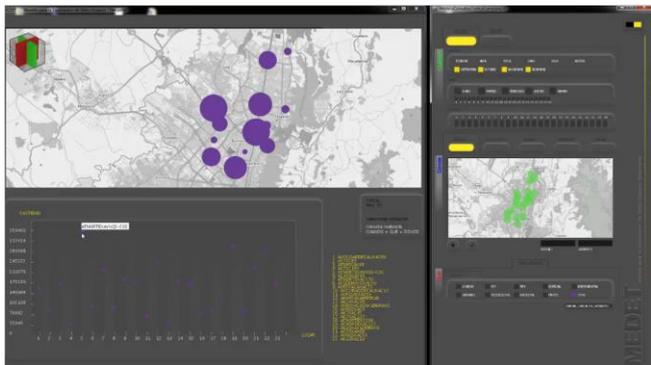


Figura 8. Interfaz cliente – servidor SEDET.

Al costado derecho de la Fig. 5, encontramos la aplicación cliente, el cual es el área de consulta, compuestas por los componentes de tiempo, espacio y atributos, con sus respectivas granularidades. En el lado izquierdo se encuentra la aplicación servidor, la cual tiene en la sección inferior una vista general de los resultados de las consultas, representando estos resultados por medio de un scatterplot o un diagrama de barras. En la parte superior de esta aplicación encontramos la georreferenciación de los resultados, y permite ver un mayor nivel de detalle de aquellos resultados que son de interés para el usuario.

Finalmente el sistema SEDET consta de un componente de interacción que permite el uso de distintos dispositivos de interacción, permitiendo que la aplicación cliente se pueda utilizar por medio de tablets o celulares y la aplicación servidor se visualice en Large Display y se interactúe por medio del uso de laser pointers, lo cual permite un ambiente colaborativo entre varios expertos, los cuales pueden seleccionar y señalar los resultados de las consultas.

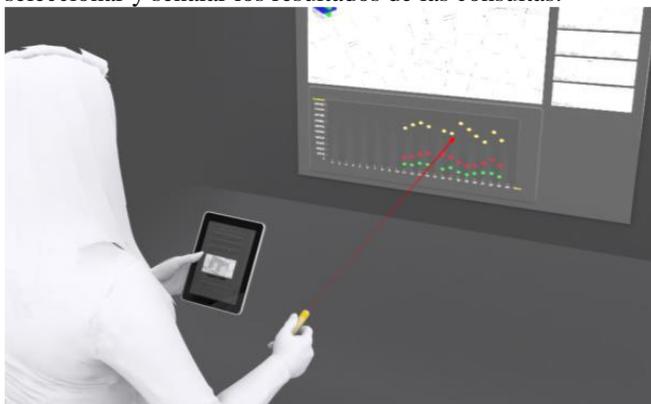


Figura 9. Interacción frente al Large Display.

IV. SISTEMA DE EVALUACIÓN

PROTOCOLO DEL EXPERIMENTO:

Las pruebas se realizaron con 20 personas en donde antes de empezar con la prueba, se les explico a los participantes el funcionamiento de la herramienta y se realizó una tarea práctica con el fin de familiarizarse con la interfaz y el modelo de consulta.

Luego se realizaron 2 tipos de pruebas, uno frente a un computador con las interfaces de consulta y respuestas en

una misma pantalla y otra prueba con los dispositivos de interacción de laser y Tablet frente a un Large Display.

Las tareas prácticas consistían en encontrar los lugares con mayor y menor tráfico en la ciudad, la época en que se presentaban mayor congestión y determinar el tipo de vehículo que más se usa de acuerdo a la hora y el lugar.



Figura 10. Pruebas de usuario frente al PC y al Large Display.

RESULTADOS:

Se evaluó los tiempos en generar las consultas y en resolver las tareas de análisis frente a un computador (prueba 1) y frente a un Large display (prueba 2), donde el tiempo promedio en la prueba 1 fue de 2 minutos y medio en promedio, mientras en la prueba 2 hubo un aumento del tiempo a 8 en promedio, esto se debe a que los usuario al utilizar 2 dispositivos de interacción simultáneamente (Laser y Tables) para realizar las tareas de análisis generaba un aumento en la complejidad en la experiencia del usuario, por tanto el 80% de los participantes aunque les gusto el tipo y la técnica de interacción prefirieron utilizar el sistema SEDET con interacción básica de mouse y teclado, ya que el objetivo es realizar las tareas de análisis de forma rápida y efectiva.

Igualmente se evaluó el tipo de dispositivo que se prefiere utilizar frente a la interacción en Large Display, donde el 90% de los participantes prefirieron el laser, ya que éstos facilitan el trabajo con múltiples usuarios, apoyando de esta forma el trabajo en grupo, mientras que las tablets, aunque le permiten a cada usuario generar su consulta de forma individual sin afectar las consultas realizadas por los demás usuarios, al tener 2 pantallas (Large diplay y tablet) el usuario se distrae.

I. COCLUSIONES Y TRABAJO FUTURO

El diseño de una visualización permitió expresividad visual, ya que las interfaces de consultas, de resultados y controles fueron intuitivos y fáciles de utilizar. Así mismo la interface apoyó la expresión de las consultas complejas y detalladas.

El sistema es versátil y liviano, por tanto necesita de poca infraestructura, bajo presupuesto, bajo tiempo de montaje y

desarrollo. Así mismo las aplicaciones clientes y servidor no mezclan la interfaz con lógica de negocio, corroborando la independencia y su fácil integración de los componentes de visualización e interacción en sistemas de manejadores de datos.

La extensión al modelo de consultas existentes permitió ver una vista general de los datos en todas sus granularidades, sin tener que entrar a explorarlos, permitiéndole al experto una introducción de los datos que posee para realizar su análisis.

Se implementó el modelo a un caso de estudio, el cual era la movilidad y transporte de Bogotá, donde la herramienta permitió evidenciar con datos espacio temporales el desarrollo de tareas de análisis con expertos en una de las áreas de sistemas urbanos, en donde ellos podían explorar y analizar los datos entorno a la planeación urbana.

Se realizaron propuestas preliminares para interacciones grupales, ya que las pruebas realizadas fueron una primera aproximación para generar un trabajo colaborativo, ya que el Large Display permite que los usuarios pueden dialogar en torno a la visualización y se realizó una evaluación previa de 2 dispositivos de interacción, lo cual da pie para generar un estudio posterior enfocado a dispositivos de interacción pertinentes para visualizar información de sistemas urbanos frente a un “Large Display”.

Como trabajo se propone aumentar las funcionalidades gráficas para la representación de los resultados a parte del diagrama de barras y el scatterplot.

Generar un historial en donde se puede evidenciar todas las decisiones de consulta y sus respectivas respuestas para que el usuario pueda realizar nuevas consultas a partir de las consultas realizadas por otro experto.

- [6] M. Dork, S. Carpendale, “VisGets: Coordinated Visualizations for Web-based Information Exploration and Discovery,” *Visualization and Computer Graphics*, IEEE Transactions on, vol. 14, no.6, Oct 2008.
- [7] D. Peuquet, “It’s about time: Framework for the representation of temporal dynamics in geographic information systems”, *Annals of the Association of American Geographers* Vol. 84, No. 3, Sep. 1994.
- [8] J. Bertin, “Graphics and Graphic Information Processing”, *Morgan Kaufmann Publishers Inc. San Francisco, CA, USA*, p.62-65. 1999.
- [9] J. Ibarra, “TaCAT: Task-Centered Analysis Tools in Urban Planning”. 2010.

REFERENCIAS

- [1] J. TÉLLEZ, “Crecimiento urbano,” *Instituto de estudios urbanos*, No. 23, p. 12, Apr. 2008.
- [2] G. Andrienko, N. Andrienko, “Space, time and visual analytics” *International Journal of Geographical Information Science*, p. 577, 2010.
- [3] C. Card, Mackinlay, Shneiderman, “Information Visualization: Using Vision to Think (Interactive Technologies)”, *Courier Corporation*, p. 128. Feb 1999.
- [4] Xinhao Wang, “Integrating GIS, simulation models, and visualization in traffic impact analysis”, *Computers, Environment and Urban Systems*, pp. 496, 2005.
- [5] Max Kaiser, “EuropeanaConnect,” *eContentplus programme, a multiannual Community program to make digital content in Europe more accessible*, pp. 65, 1994.