

Técnica de extracción de información de bases de datos para su manejo como datos con posición geográfica

Data extraction technique of databases for its management as data with geographical position

MUÑIZ-ELIZALDE, José David^{1†*}, SÁNCHEZ-LARA, Benito², HERNÁNDEZ-GUILLEN, Jesús² y SERRANO-BALTAZAR, Andrea²

¹Instituto Tecnológico de Tláhuac

²Universidad Nacional Autónoma de México, Facultad de Ingeniería

ID 1^{er} Autor: *José David, Muñoz-Elizalde* / ORC ID: 0000-0001-8407-4418, CVU CONACYT ID: 958124

ID 1^{er} Coautor: *Benito, Sánchez-Lara* / ORC ID: 0000-0002-2158-6974, CVU CONACYT ID: 69973

ID 2^{do} Coautor: *Jesús, Hernández-Guillen* / ORC ID: 0000-0002-4665-3010

ID 3^{er} Coautor: *Andrea, Serrano-Baltazar* / ORC ID: 0000-0002-8928-5828

DOI: 10.35429/JUSD.2019.16.5.11.14

Recibido: 30 de Julio, 2019; Aceptado 20 de Septiembre, 2019

Resumen

El uso de sistemas de información geográfica (SIG) es amplio en diversas y diferentes disciplinas; como uno de sus principales atributos, los SIG permiten el manejo de información geográficamente referenciada, esto es, datos con posición geográfica. Este trabajo presenta una técnica de extracción de información a partir de bases de datos y su posterior geolocalización. La técnica se ilustra con un caso de estudio donde datos de demanda de servicios de transporte de una Institución de Educación Superior en la Ciudad de México se manejan considerando su posición geográfica. En la técnica se parte de establecer criterios de búsqueda, selección y agrupación de datos tomando en cuenta qué datos sí es posible geolocalizar, hecho esto se exportan a un SIG que permite un manejo diferenciado, en particular, una visión espacial de los datos.

SIG, Bases de Datos, Técnica de Extracción de Información, Demanda de Transporte

Abstract

The use of geographic information systems (GIS) is wide in diverse and different disciplines; as one of its main attributes, GIS allow the management of geographically referenced information, that is, data with geographical position. This work presents a technique for extracting information from databases and their subsequent geolocation. The technique is illustrated with a case where demand data for transportation services of a higher education Institution in Mexico City are handled considering its geographical position. The technique start with establishing criteria of search, selection and grouping in a data base taking into account what data it is possible to geolocate. After it information is exported to a GIS that allows a differentiated management, in particular, a spatial view of the data.

GIS, Databases, Information Extraction Technique, Transportation Demand

Citación: MUÑIZ-ELIZALDE, José David, SÁNCHEZ-LARA, Benito, HERNÁNDEZ-GUILLEN, Jesús y SERRANO-BALTAZAR, Andrea. Técnica de extracción de información de bases de datos para su manejo como datos con posición geográfica. Revista del Desarrollo Urbano y Sustentable. 2019. 5-16: 11-14

*Correspondencia al autor (Correo electrónico: dj.triple@hotmail.com)

† Investigador contribuyendo como primer autor

Introducción

El análisis de datos se ha vuelto imprescindible en los últimos años, con la era del conocimiento y el llamado internet de las cosas han surgido diversos sistemas dentro de los sistemas que proporcionan suficiente información para conocer los estados de los sistemas. En términos del análisis de datos, surge un concepto llamado minería de datos, el cual busca identificar lo más valioso de la información mediante ciertos métodos para el descubrimiento de datos. Los sistemas analizados mediante la técnica de minería de datos pueden integrar datos espaciales.

Un sistema SIG, consiste de un grupo de algoritmos que permiten acceder y analizar datos espaciales junto con sus atributos asociados. La elección de un sistema SIG, depende de la aplicación y del tipo de análisis que se requiera realizar.

Particularmente se conoce como Información Geográfica al conjunto de datos espaciales georeferenciados requeridos para realizar alguna operación. Actualmente, el 80% de los datos corporativos existentes en el mundo poseen una componente geográfica.

Por otro lado, la georeferenciación es el posicionamiento en el que se define la localización de un objeto espacial en un sistema de coordenadas y un dato asociado. Una de sus ventajas, es la combinación de datos espaciales descriptivos. Esto permite establecer una relación uno a uno entre los elementos y registros en una tabla de atributos de elementos. Es de gran utilidad implementar técnicas para analizar información georeferenciada, ya que sustenta toda la información al mismo tiempo. De esta manera se puede visualizar gran parte del contexto y permite distinguir información diferente.

Objetivos

Identificar la posición geográfica de los datos domiciliarios de una población estudiantil perteneciente a una Institución de Educación Superior, mediante la extracción de información en una base de datos y exportación hacia un sistema de información geográfica, con el propósito de identificar la distribución de viajes de los estudiantes en la Zona Metropolitana del Valle de México (ZMVM).

I. Metodología**Fase de búsqueda**

Una base de datos se constituye de datos pertenecientes a un mismo contexto almacenados sistemáticamente. La técnica echa mano de un administrador de bases de datos (DBA).

Fase de selección

La selección de la información consideró la filtración de información relevante para permitir consultas.

Fase de agrupación

La agrupación implica formar grupos con un mismo atributo.

Ilustración del uso de la técnica

Para ilustrar la aplicación de la técnica se considera el desarrollo de un estudio para determinar la distribución de viajes de la población estudiantil de una Institución de Educación Superior en la Zona Metropolitana del Valle de México (ZMVM). Lo anterior, en el marco de un proceso de planeación de un sistema de transporte.

La investigación, integra información de cada estudiante de la Institución en donde se realizó el estudio. Respecto a sus datos domiciliarios, se obtuvieron 2,792 registros correspondientes a estudiantes inscritos en el momento del estudio.

Se optó por exportar la información a un administrador de bases de datos (DBA), con la finalidad de filtrar cierta información y generar consultas especializadas para obtener los Códigos Postales de los domicilios de los estudiantes. Posteriormente, se inicia la búsqueda de los Códigos coincidentes y se agrupan con el propósito de identificar la cantidad de estudiantes que habitan en un mismo Código Postal. Al finalizar, estos datos se representaron espacialmente para interpretar la información con una visión diferente.

A continuación, se presenta la Tabla 1 la cual muestra algunos datos colectados y asociados entre sí, como resultado de la agrupación de datos por Códigos Postales, que relaciona cuantos estudiantes viven en la misma zona respecto de su Código Postal.

Código Postal	Total, de estudiantes ubicados
9820	5
9830	14
9839	1
9840	13
9849	5
9850	29
9856	1
9860	9

Tabla 1 Resultado de la agrupación de datos por Códigos Postales

Posteriormente se geocalizó cada uno de los Códigos Postales y se colocó un marcador indicando la cantidad de estudiantes que habitan por Código Postal. Los puntos marcados, son georreferenciados mediante la herramienta My Maps de Google. Gracias a esta, se puede exportar la información en archivos kml. los cuales son una variante de archivos xml. En esencia, se plantea la representación espacial de los datos para identificar los posibles orígenes de viajes de la población de estudio. Una vez obtenida la información con referencia espacial, se exportó a un Sistema de Información geográfica (SIG). La Figura 1 muestra los datos representados espacialmente mediante el uso de SIG, donde se aprecian nodos en color naranja que representan los Códigos Postales que refieren los domicilios de los estudiantes.

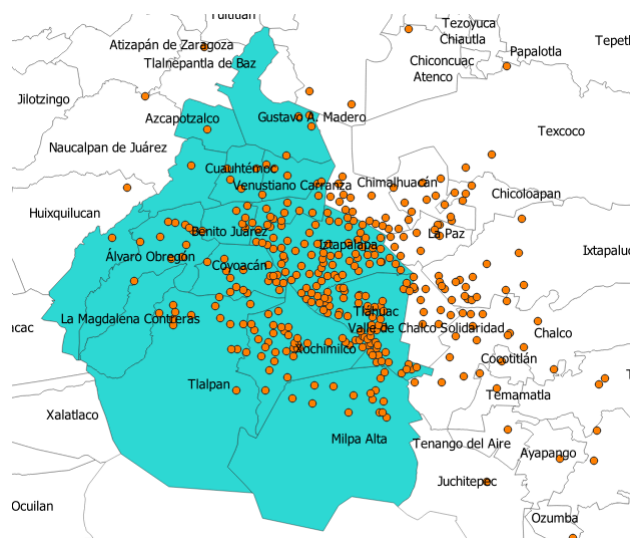


Figura 1 Datos representados espacialmente.

Con lo anterior, se logra percibir como se distribuye una población estudiantil que genera viajes en la Ciudad de México y el área conurbada. Asimismo, se puede estimar la cantidad de viajes en la ZMVM a consecuencia de la demanda que genera una Institución de educación superior.

Utilizando el SIG, se categorizaron los datos con la finalidad de observar los Códigos Postales que concentran una mayor cantidad de estudiantes. La Figura 2 muestra nodos en color blanco, los cuales representan una mínima concentración de estos en intervalos poblacionales de uno a diez habitando en un Código Postal. Por otro lado, el aumento en la tonalidad de los nodos representa Códigos Postales con alta concentración definida en intervalos que van desde los cincuenta hasta ochenta y seis estudiantes ubicados en el mismo Código Postal.

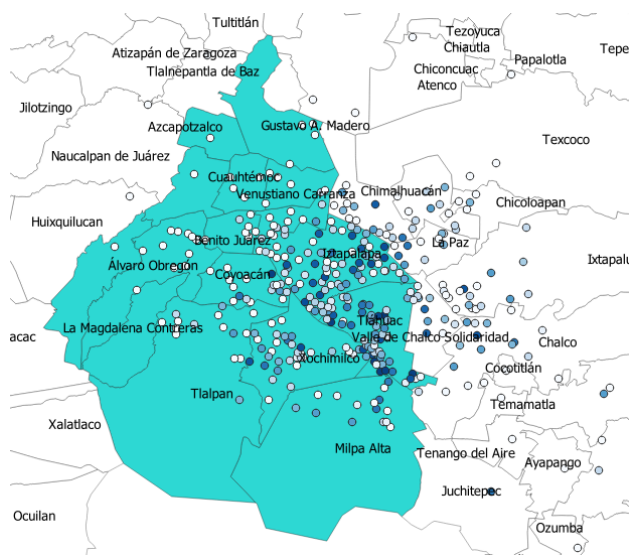


Figura 2 Categorización de nodos por número de estudiantes

Una vez concentrada y localizada la información espacialmente, se determinaron líneas que marcan la dirección de viajes y el porcentaje de distribución en función de la demanda que genera la Institución, la Figura 3 muestra el porcentaje y distribución de viajes de la población analizada.

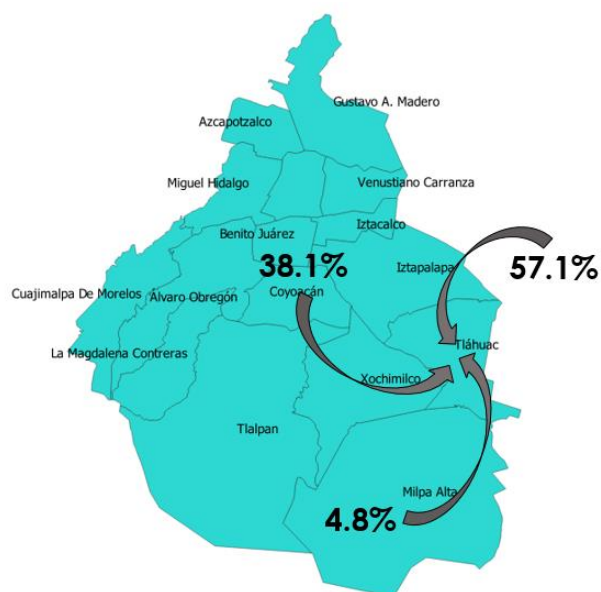


Figura 3 Distribución de viajes hacia la Institución

Después de haber desarrollado esta técnica, el resultado se utilizó para proponer alternativas para mejorar el servicio de transporte público en la zona Sur Oriente de la Ciudad de México, el cual se deriva de la demanda de servicio que ocasiona la población estudiantil de la Institución de Educación Superior como caso de estudio.

Contribuciones

La técnica de manejo de datos espacialmente se aplicó en el análisis de un sistema de transporte para identificar la oferta y demanda que se genera en una zona por origen-destino. El estudio se aplicó en el servicio de transporte público para una población estudiantil, consiguiendo resultados satisfactorios para reconocer líneas de acción como propuesta para mejorar el servicio de transporte público.

Por otro lado, se percibe que la técnica puede ser empleada en diversos contextos ya que ofrece una visión amplia de los datos. Por ejemplo, esta técnica se puede implementar en un centro médico para geolocalizar a los pacientes vulnerables con la finalidad de implementar estrategias de traslados oportunos. También se puede aplicar la técnica para resolver problemas logísticos y de traslados por mencionar algunos ejemplos.

Finalmente, se pueden plantear propuestas significativas en diversos contextos, por ejemplo, en la planeación de un sistema de transporte, en actividades logísticas orientadas a las cadenas de suministro por mencionar algunos campos de estudio.

La técnica podría dar valor agregado al sector que se requiera beneficiar, por la forma de representar la información. Así también, se busca propiciar una mejor toma de decisiones al plasmar la información espacialmente.

Referencias

- Caplan, J. (2010). GIS For Public Safety.
- Graser, A. (2016). Learning QGIS. United Kingdom: Birmingham. Packt Publishing Ltd.
- Graser, A., Mearns, B., Mandel, A., Olaya, V., & Bruy, A. (2017). QGIS: Becoming a GIS Power User Learning Path.
- Heywood, I., Cornelius, S., & Carver, S. (2006). An Introduction to Geographical Information Systems. England. Pearson Education.
- Li, D., Wang, S. & Li, D. (2015). Spatial Data Mining. Berlin. Springer
- Lloyd, C. D. (2010). Spatial Data Analysis. New York. Oxford University Press.
- Olaya, V. (2016). Sistemas de Información Geográfica. CreateSpace Independent Publishing Platform.
- Jiang, Z., & Shekhar, S. (2017) Spatial Big Data Science classification techniques for earth observation imagery. Springer.
- Rodrigue, J. P. (2006) Transport Systems. Springer.
- Shellito, B. (2012). Introduction to Geospatial Technologies. New york. W.H. Freeman and Company.
- The National Academy Press (2006) BEYOND MAPPING. Washington, DC.