

Geocodificación: Caso de Estudio de la Implementación de un SIG para el Centro Comercial de Concepción del Uruguay

Andrés Pascal¹, Anabella De Battista¹, Norma Edith Herrera², Alejandra Díaz¹,
Adrián Planas¹

¹Dpto. de Sist. de Información, Univ. Tecnológica Nacional, FRCU, Entre Ríos, Argentina
{pascala, debattistaa, adiaz, planasa}@frcu.utn.edu.ar

²Dpto. de Informática, Universidad Nacional de San Luis, Argentina
nherrera@unsl.edu.ar

Abstract. Un sistema de información geográfica (SIG) permite capturar, almacenar, manipular, analizar y mostrar información geográficamente referenciada a fin de resolver problemas complejos de planificación y gestión. En el marco de un estudio del sector comercial de la ciudad de Concepción del Uruguay se presentó el problema de modelado e implementación de un SIG, y particularmente la problemática de la geocodificación de los datos relevados. Si bien existen métodos genéricos de geocodificación, cada caso posee características particulares. En este trabajo se presenta la solución de geocodificación de puntos que se aplicó dentro del Estudio del Perfil y Características del Sector Comercial de la ciudad, para la ubicación espacial de los comercios relevados.

Keywords. Bases de datos espaciales, Sistemas de Información Geográfica, georeferenciación, geocodificación, comercios.

1 Contexto

El Centro Comercial, Industrial y de la Producción (CECOM) de la ciudad de Concepción del Uruguay, Entre Ríos, tiene como misión ejercer la representación en defensa del Comercio, la Industria y la Producción y auspicia toda iniciativa, estudio, planificación o proyecto que tenga como finalidad el fomento de la radicación industrial y el mejoramiento tecnológico y de servicios de las actividades que representa [3].

Con el fin de dar soporte a las actividades realizadas por el CECOM y de obtener una herramienta que fundamente y oriente la planificación industrial y comercial de la ciudad, el CECOM firmó un convenio con la Vicegobernación, la Honorable Cámara de Senadores de la Provincia de Entre y la Facultad Regional Concepción del Uruguay de la Universidad Tecnológica Nacional, a fin de llevar a cabo un Estudio del Perfil y Características del Sector Comercial de la ciudad de Concepción del Uruguay.

Los objetivos principales del estudio fueron describir y analizar el sector comercial y de servicios de Concepción del Uruguay; obtener una herramienta de planificación para el sector; y realizar un georreferenciamiento digitalizado que permita ilustrar la ubicación e identificación de los locales comerciales relevados.

Comprender la estructura y la dinámica del sector comercial y dimensionar su peso en la economía local resulta fundamental para el planteo de líneas estratégicas para el desarrollo del mismo en el mediano y largo plazo. En este contexto, la aplicación de una herramienta SIG, que permite la caracterización y análisis de variables espaciales, le será de utilidad al CECOM durante el proceso de asesoramiento a sus socios sobre las posibilidades de desarrollo comercial de la ciudad.

Este proyecto se desarrolló entre los meses de Mayo y Agosto de 2011. La coordinación del mismo estuvo a cargo del Grupo de Estudios de Calidad y Medio Ambiente (GECAL) de la Facultad Regional Concepción del Uruguay de la Universidad Tecnológica Nacional, que contó con el apoyo de becarios de la carrera de Licenciatura en Organización Industrial para el trabajo de campo, y del Grupo de Investigación en Bases de Datos (GIBD) para el procesamiento y geocodificación de los datos y la implementación del SIG.

En este trabajo se presenta el proceso de geocodificación desarrollado en el marco de la vinculación Universidad - Sector Industrial - Gobierno que dio sustento a la implementación de un SIG como herramienta para la planificación y toma de decisiones en el sector comercial de Concepción del Uruguay.

Este artículo se organiza de la siguiente manera: en la Sección 2 se presenta la introducción a la temática de los SIG y al caso de estudio. En la Sección 3 se plantea el problema de la geocodificación y la solución abordada para el caso particular. Finalmente en la Sección 4 se exponen las conclusiones.

2 Introducción

La comprensión de la distribución de datos espaciales para fenómenos que ocurren en el espacio en un momento de tiempo determinado constituye actualmente un gran desafío para la resolución de cuestiones centrales en muchas áreas de conocimiento como salud, medioambiente, prevención y atención de desastres, seguridad, marketing, planificación urbana y comercial, entre otras [1].

Estos estudios son cada vez más comunes gracias a la disponibilidad y gran difusión de Sistemas de Información Geográfica (SIG) [1, 2] con interfaces de usuario cada vez más amigables. La tecnología de los sistemas de información geográfica ha evolucionado notablemente, pasando de ser cartografías temáticas muy simples a herramientas altamente sofisticadas gracias a la creciente capacidad de cómputo y las refinadas técnicas de análisis, además del mayor interés existente en cuestiones cuya resolución se ve altamente favorecida por las bases de datos geográficas.

Un SIG permite la visualización espacial de variables asociadas a datos georreferenciados, además de la posibilidad de realizar consultas sobre los mismos, lo que los convierte en una herramienta indispensable como soporte para la toma de decisiones en los procesos de planificación. A partir de la construcción de una base de datos y

una base geográfica, un SIG es capaz de presentar mapas estratificados que permiten visualizar patrones espaciales del fenómeno en estudio.

En este caso de estudio se tomó como insumo el relevamiento de los locales comerciales de la ciudad de Concepción del Uruguay realizado por el GECAL. Las actividades del GIBD fueron: normalización y verificación de los datos relevados, generación de un mapa base de la ciudad, geocodificación de los comercios, selección de una herramienta SIG para la visualización y el análisis espacial, definición de capas e implementación del SIG. Las capas generadas representan categorías de comercios que fueron provistas por el GECAL. Finalmente se procedió a la instalación y configuración de la herramienta SIG en las oficinas del CECOM.

3 Geocodificación: el Caso de Concepción del Uruguay

3.1 Marco Teórico

En el lenguaje de las ciencias de la información geográfica, el concepto general de asignar información espacial explícita a objetos, se denomina georreferenciación, esto es, la transformación o incorporación de información geográfica válida a objetos que no la poseen, para ser utilizadas en el análisis espacial [4]. La Geocodificación, constituye un caso particular de georreferenciación, y se define como la transformación de un texto descriptivo con ubicación espacial implícita, en coordenadas geográficas expresadas como puntos en un mapa [4, 14].

Las coordenadas que se obtienen permiten ubicar los elementos espacialmente y posteriormente realizar análisis sobre los mismos en función de variables geográficas como área de influencia, cercanía, relación con otros elementos, solapamiento de elementos, etc.

3.2 Planteo del Problema

Como parte de la segunda etapa del proyecto, un grupo de alumnos coordinados por el GECAL realizó el relevamiento del Sector Comercial de Concepción del Uruguay, registrando datos de 2032 comercios en formularios, que fueron trasladados luego a una planilla digital de cálculo. Los datos relevados fueron: razón social, producto o servicio principal y secundario, domicilio, rubro, año de inicio, cantidad de empleados y otros, dentro de los cuales los primeros cuatro fueron utilizados como insumo para el sistema de información geográfica. Hay que observar que el GIBD se incorporó al proyecto cuando éste ya estaba en marcha, y no tuvo tiempo previo al relevamiento suficiente como para normalizar la forma de registro de los datos.

En términos generales, se plantearon dos problemas importantes: obtención de un mapa de la ciudad de Concepción del Uruguay y selección o creación de herramientas de geocodificación de los datos. Para resolver el primero se analizaron los mapas provistos por Yahoo [6], OpenStreetMap (OSM) [7] y Bing [8]. Google maps [9] en ese entonces no poseía mapa de Concepción del Uruguay. En los tres casos, los planos de la ciudad estaban desactualizados e incompletos, por lo cual se tomó como base el de OSM, por ser libre y gratuito, se migraron los datos al DBMS PostgreSQL-

PostGIS [10, 11] y se hicieron las correcciones y agregados necesarios para poder representar la información correctamente utilizando herramientas de edición de mapas.

Por otro lado, el problema de la geocodificación planteaba mayores desafíos. Se sabe que las herramientas de geocodificación tienen un margen de error significativo [12] y este proyecto requería que la totalidad de los comercios fuesen ubicados en su posición exacta, con precisión mínima de una cuadra. Por esta razón, se decidió construir un algoritmo de geocodificación propio y particular para este caso.

Los problemas que se necesitó resolver para poder realizar la geocodificación fueron:

1. Correspondencia entre los nombres de las calles: los nombres de calles registrados en la planilla de cálculo en muchos casos no eran los mismos que los definidos en el mapa. E inclusive dentro de la planilla había distintas formas de escribir el mismo nombre. Además, en algunos casos la misma calle posee más de un nombre. Por ejemplo, los nombres “CLEMENTINA C. DE ALIO”, “CONTE DE ALIO C.” y “26 DEL OESTE SUR” corresponden a la misma calle.

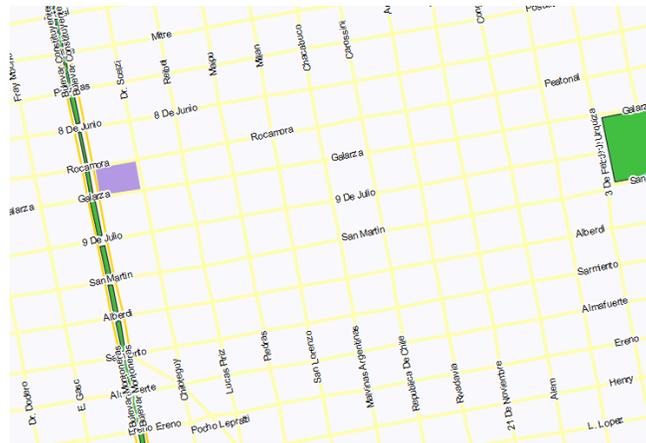


Fig. 1. – Calle 9 de Julio y numeración

2. La numeración de las calles: en la ciudad de Concepción del Uruguay los números se organizan de a 50 por cuadra en la mayoría de los casos, partiendo del número cero (0) desde la zona portuaria (zona Este de la ciudad) para las calles de dirección Este – Oeste (que llamaremos “horizontales”) y desde la calle 9 de Julio para las calles de dirección Norte – Sur (que llamaremos “verticales”). Estas últimas poseen un nombre distinto dependiendo de si se encuentran al norte o al sur de la calle 9 de Julio. Es decir que, en la intersección con 9 de Julio, todas las calles verticales poseen el número 0 y se incrementa su numeración hacia Norte o hacia el Sur según corresponda. Los números pares e impares se ubican del lado Este y Oeste

respectivamente para las calles verticales y del lado Norte y Sur respectivamente para las horizontales [Figura 1]. Los tamaños de las cuerdas no son uniformes y los bulevares constituyen espacios sin correspondencia numérica por lo cual no se puede realizar el cálculo de la posición geográfica por interpolación directa. Además existen cuerdas diagonales donde se desvirtúa la proporción entre el rango de números y las distancias reales [Figura 2].

3. La representación de las calles en el mapa: existen algunas calles que están definidas en el mapa por más de una polilínea como si fuesen calles distintas, ya que se ven interrumpidas por parques, plazas, el ferrocarril, edificaciones, etc.
4. Respecto a la información que indicaba la dirección de un comercio, nos encontramos con tres representaciones diferentes:
 - (a) domicilio del comercio indicado por su calle y su número
 - (b) domicilio de comercio que se ubica en una esquina, definido mediante la intersección de dos calles
 - (c) domicilio del comercio indicado de manera imprecisa mediante la calle sobre la cual se encuentra y las calles de las esquinas para indicar su altura.

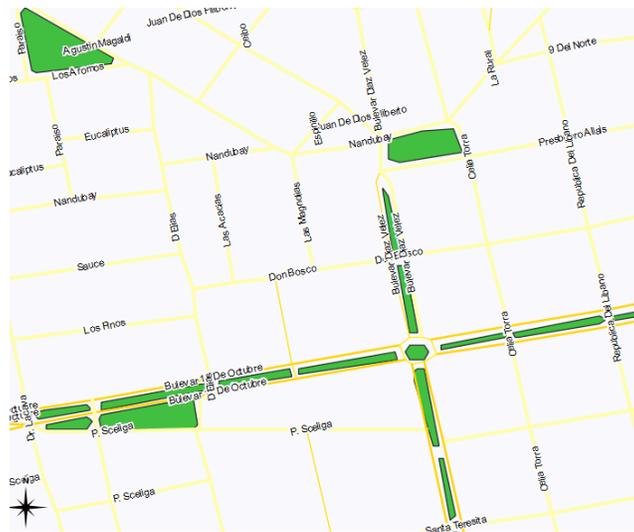


Fig. 2. Problemas cuerdas: calles diagonales, cuerdas de distintos tamaños, bulevares que ocupan espacio pero no tienen numeración.

3.3 Proceso de Geocodificación Desarrollado

Para asegurar la correctitud y completitud del proceso de geocodificación de los comercios, se adoptó como estrategia asignar a cada esquina relevante del mapa, las

alturas correspondientes a las calles que se intersectan y calcular la posición del comercio por interpolación entre las dos esquinas numeradas más cercanas cuyo rango contenga al número del domicilio del comercio. Este proceso se describe a continuación:

1. Asignación de números a las esquinas del mapa, mediante los siguientes pasos:
 - (a) Selección de una calle horizontal (Mitre) y dos calles verticales (Leguizamón y Artigas) como calles de referencia. Para esta selección se tuvo en cuenta que dichas calles son las que tienen las cuadras más regulares de toda la ciudad (cuadras del mismo tamaño y que no son interceptadas por calles diagonales, entre otros aspectos). Sobre estas calles se corrió un algoritmo que numeró las esquinas partiendo desde el punto 0 e incrementando de a 50 unidades por cuadra.
 - (b) Asignación de números a las demás calles en base a las calles de referencia de acuerdo al siguiente algoritmo:

```
Para cada calle C hacer
  Si C es horizontal, Ref:=Calle Mitre
  sino
    Si C está al Norte de la calle 9 de Julio,
      Ref:=Calle Artigas
    sino
      Ref:=Calle Leguizamón
  Fin Si
  Para cada punto PRef de la polilínea de Ref hacer
    Perp:=Recta perpendicular a Ref en el punto PRef
    PCercano:=Punto de la polilínea de C, más
      cercano a la recta Perp
    Si la distancia de PCercano a Perp <
      radio de aceptación (error permitido)
      Asignar a PCercano el mismo número que PRef
  Fin Para
```

Nota: para el caso de las diagonales, cuando el ángulo entre una calle horizontal y la diagonal era mayor a 45° dicha calle se numeraba como una calle horizontal y en caso contrario como vertical.

- (c) Una vez generado el mapa con la numeración se realizó un proceso manual y exhaustivo de verificación de la numeración generada, para realizar los ajustes necesarios.
2. Geocodificación: en una segunda etapa, con el mapa completo y preparado para este proceso, se desarrolló un algoritmo de geocodificación interactivo con las siguientes características:
 - (a) Como entrada se obtuvieron el domicilio y código de cada comercio registrado en una base de datos en PostgreSQL
 - (b) Como salida se actualizó dicha base de datos incorporando las coordenadas geográficas del comercio expresadas en latitud y longitud.

- (c) Se definieron sinónimos para los nombres de las calles, de tal manera de tener en cuenta las distintas formas usuales de denominar algunas calles.
- (d) Para encontrar la calle del mapa correspondiente al domicilio del comercio, se realizaron búsquedas por similitud de cadenas (distancia de Levenshtein) [15, 16] con radio igual a 2, para considerar posibles errores en la escritura de las calles.
- (e) En el caso de los domicilios especificados por calle y número, luego de encontrar la calle correspondiente en el mapa, se calculó por interpolación el punto geográfico del comercio tomando como intervalo el número mayor y el menor de la calle del comercio, más cercanos al número del domicilio. Luego se hizo un ajuste de acuerdo a la paridad del número y el lado de la calle.
- (f) En el caso de los domicilios definidos por una intersección, se buscó el par de puntos más cercano entre ambas calles y luego se hizo un ajuste en base a la ubicación de la esquina (NE, NO, SE y SO).
- (g) En el caso de domicilios especificados mediante el nombre de la calle en la que se encuentra el comercio y las dos calles de las esquinas, luego de encontrar las tres calles se calculó la posición del comercio como el punto medio entre las esquinas.



Fig. 3. Puntos que representan los comercios geocodificados

Posteriormente a todo este proceso, se realizó una etapa manual de verificación de los puntos geocodificados, tomando una muestra aleatoria de comercios. No hubo correcciones relativas al proceso de geocodificación, pero se encontró un porcentaje mínimo de comercios cuyos domicilios fueron registrados erróneamente, que fueron corregidos oportunamente.

4 Conclusiones

Este artículo se presentó el caso de estudio de implementación de un SIG para visualizar y analizar espacialmente datos del sector comercial de Concepción del Uruguay, y en particular, el problema de la geocodificación de los datos obtenidos. Si bien el proceso de geocodificación fue desarrollado para este caso específico, la estrategia utilizada de asignar alturas a todas las esquinas relevantes y el método de asignación, se pueden generalizar para ser utilizada en otros casos donde se requiera que todos los puntos sean geocodificados con precisión y exactitud. Por otro lado, es de destacar que a partir de este trabajo se cuenta con un mapa preparado para geocodificar cualquier otra base de datos que posea domicilios correspondientes a la ciudad, tales como datos de clientes de una empresa, contribuyentes de impuestos municipales, etc.

Las actividades de vinculación entre el Centro Comercial de la ciudad de Concepción del Uruguay, la Vicegobernación de la Provincia de Entre Ríos, llevadas a cabo mediante el trabajo colaborativo entre los grupos de investigación de la Facultad Regional Concepción del Uruguay de la Universidad Tecnológica Nacional, Grupo de Estudio sobre Calidad y Medioambiente (GECAL) y Grupo de Investigación sobre Bases de Datos (GIBD) permitieron dotar al CECOM de una herramienta SIG, con la que podrán obtener información muy valiosa para orientar futuras inversiones en la ciudad en el marco industrial y comercial.

5 Referencias

1. Spatial Analysis and GIS: A Primer. Gilberto Câmara, Antônio Miguel Monteiro, Suzana Druck Fucks, Marília Sá Carvalho. 2004
2. Sistemas de Información Geográfica. Víctor Olaya. http://wiki.osgeo.org/wiki/Libro_SIG Versión 1.0. Rev. 25 de noviembre de 2011
3. Sitio web CECOM, <http://www.cecom-web.org.ar>
4. A geocoding best practices guide. Daniel W. Goldberg. University of Southern California. GIS Research Laboratory. 2008.
5. Relevamiento de comercios y servicios de Concepción del Uruguay. Informe Final. 2011.
6. Mapas Yahoo. <http://maps.yahoo.com>
7. Mapas OpenStreetMap. <http://www.openstreetmap.org>
8. Mapas Bing. <http://www.bing.com/maps>
9. Mapas Google. <http://maps.google.es>
10. DBMS PostgreSQL. <http://www.postgresql.org>
11. Extensión espacial PostGIS para el DBMS PostgreSQL. <http://postgis.refractor.net>
12. Improving Geocoding Practices: Evaluation of Geocoding Tools. Duck-Hye Yang, Lucy Mackey Bilaver, Oscar Hayes, Robert Goerge. Journal of Medical Systems, Vol 28, Nro 4, 2004.
13. Spatial Database Case Study: A GIS Based Metal Contamination Application. Wendy Zhang, Ke Yang, Theresa Beaubouef, Ju Chou, and Ghassan Alkadi. 2010.
14. A Hybrid Approach to Segment-Type Geocoding of New York City Traffic Data. Jianting Zhang, Simin You, Li Chen, Cynthia Chen. COM. Geo 2010, Washington, DC, USA. 2010

15. A guided tour to approximate string matching. Gonzalo Navarro. *ACM Comput. Surv.* 2001
16. Searching in metric spaces. . Chávez, Edgar and Navarro, Gonzalo and Baeza-Yates, Ricardo and Marroquín, José Luis. *ACM Comput. Surv.* 2001