

Recolección de direcciones postales oficiales y su servicio web de direcciones: geocodificador de CartoCiudad

REVISTA **MAPPING**
Vol. 29, 203, 32-37
noviembre-diciembre 2021
ISSN: 1131-9100

Official postal addresses capture and the addresses web service: CartoCiudad Geocoder

Itziar Doñate Vadillo, Beatriz Brenes Fernández, Emilio López Romero, Paloma Abad Power

Resumen

En este artículo se presenta la evolución del servicio web, denominado «Geocoder de Cartociudad» del Centro Nacional de Información Geográfica (CNIG), que permite localizar un lugar a partir de su dirección postal o un topónimo. En un principio, este servicio de geocodificación solo publicaba la red viaria de la base de datos de «Información Geográfica de Referencia de Redes de Transporte (IGR-RT)» del Sistema Cartográfico Nacional, completadas con los códigos postales proporcionados por el Grupo Correos y con las poblaciones de la IGR de Poblaciones.

A lo largo del año 2020 se han incorporado nuevas fuentes disponibles en internet de forma abierta, lo que ha permitido incluir 3,7 millones de direcciones nuevas, con lo que se llega a un total de 15 millones de direcciones postales publicadas a través del «Geocoder». Debido a que no solo hay cambios en las direcciones, sino también en elementos puntuales, como son los puntos de interés (POI), se ha realizado un análisis de las fuentes oficiales de las que se dispone y se han obtenido un mayor número de puntos. Teniendo actualmente 89 000 POI.

Las nuevas direcciones obtenidas y los POI se han ido incorporado para las búsquedas de lugares que se realicen utilizando el servicio de geocodificación para la localización de direcciones. Aparte de lo anterior, se pueden localizar y ubicar unidades administrativas, poblaciones, códigos postales y geolocalizar referencias catastrales de la D.G de Catastro.

El geocodificador de CartoCiudad, junto con el servicio de descarga (WFS) de los topónimos del Nomenclátor Geográfico Básico de España (NGBE), se combinan para dar respuesta a través de un *plugging* o componente de búsqueda denominado «IGN_Search» desarrollado por el CNIG y que está disponible en los visualizadores del Instituto Geográfico Nacional (IGN) y en el visualizador de CartoCiudad.

Abstract

CartoCity is a leader in the field of postal addresses in Spain. It is based on the road network of the Reference Geographic Information of Transport Networks (RGI-TN) of the National Cartographic System, and it is supported with the Postal Codes provided by the Post Office Group together with the settlements of the RGI of Settlements.

Regarding to the data, work is continuously done to incorporate the latest version of the RGI-TN, but it has also been decided to incorporate all those addresses that are available on the Internet from opened sources and, if possible, already standardized. In fact, it has been incorporated more than 3.7 million new addresses on data base.

Not only addresses have changed but also point elements, such as points of interest (POIs). An analysis of the official sources available has been made and a greater number of points have been obtained. We currently have 89,000 POIs.

The new addresses and the POIs obtained have been incorporated for the searches that are made using the addresses location service, called "geocoder". In addition to the above, you can search for Administrative Units, villages, postal codes and geolocate cadastral references.

The geocoder service together with the Web Feature Service (WFS) which provided the geographical names of the Basic Geographic Nomenclature of Spain are combined to give answers in the search component called "IGN_search" developed by the National Center of Geographic Information (CNIG). The search engine is incorporated both in the CartoCity viewer and in the rest of the National Geographic Institute's thematic viewers.

Palabras clave: CartoCiudad, Direcciones, Códigos Postales, geocodificador, buscador, puntos de interés, Redes de Transporte, Poblaciones.

Keywords: CartoCity, Addresses, Postal Codes, Geocoder, Search Engine, Points of Interest, Transportation Networks, Village.

Centro Nacional de Información Geográfica
itziar.donate@cnig.es, bbrenes@fomento.es,
elromero@fomento.es, pabad@fomento.es

Recepción 10/04/2020
Aprobación 25/04/2020

1. INTRODUCCIÓN

La georreferenciación es la localización de un lugar que se puede realizar de dos formas, denominadas directa e inversa. La primera georreferenciación utiliza directamente las coordenadas para localizar un lugar y la georreferenciación indirecta no utiliza explícitamente las coordenadas, sino que establece la relación con una localización definida por un objeto u objetos geográficos, como por ejemplo un código postal, una dirección postal, el nombre de una unidad administrativa o un topónimo, denominados objetos direccionables.

Según la Especificación de Datos de Direcciones de la Directiva INSPIRE una «dirección» es una identificación de la localización fija de una propiedad por medio de una composición estructurada de nombres geográficos e identificadores, que contiene entre otros atributos un nombre legible para personas (nombre de la vía) y la posición de un punto característico que representa la localización de la dirección de conformidad con cierta especificación, incluyendo información sobre el origen de la posición (número del portal, punto kilométrico...), por lo que estas posiciones geográficas permiten localizar la dirección espacialmente.

En la misma documentación, también se define una «dirección completa» como una jerarquía que consta de los siguientes componentes:

- Los nombres geográficos, con un nivel de detalle cada vez mayor, por ejemplo, la ciudad, y a continuación el nombre de la calle, por ejemplo: Madrid, Paseo del Prado.
- A continuación, el número o el nombre de la casa. Por ejemplo: Museo del Prado.
- Y también puede incluir un código postal u otros descriptores postales.

Las direcciones tienen varios propósitos, como:

- a) La ubicación (por ejemplo, para visitas o entrega de correo),
- b) la identificación (por ejemplo, en el contexto de la ins-

cripción de un edificio),

- c) la identificación de una jurisdicción (por ejemplo, la autoridad responsable de los bienes identificados por la dirección),
- d) la clasificación y ordenamiento,
- e) la respuesta ante una emergencia.

Hay varios tipos de objetos diferentes que pueden estar relacionados con la propiedad. Los tipos más reconocidos que tienen direcciones son las parcelas de tierra y los edificios (incluyendo pisos o apartamentos). En algunos países, otros objetos tienen una dirección, como el mobiliario urbano, las estaciones de bombeo de agua, los lugares de amarre, los estacionamientos y los graneros agrícolas. Aunque no reciben correo, pueden necesitar una dirección para otras funciones. Esto es así tanto en las zonas rurales como en las urbanas.

Una dirección tiene otros atributos como:

- La validez de la dirección dentro del ciclo de vida (versión) del objeto espacial de dirección.
- Fecha y hora en que se insertó o modificó y en la que se reemplazó o retiró en el conjunto de datos espaciales esta versión del objeto espacial.
- Identificador temático externo del objeto espacial de dirección que permite la interoperabilidad con los sistemas o aplicaciones heredados existentes.
- Fecha y hora en que esta versión de la dirección fue o será válida en el mundo real y dejó o dejará de existir.

Y una dirección está asociada con los siguientes elementos:

- Edificio al que está asignado o con el que está asociada.
- Parcela catastral a la que está asignada o con la que está asociada.
- Dirección (madre) principal con la que está estrechamente relacionada esta (sub)dirección.
- Componente de una dirección.

Por lo que, la georreferenciación indirecta, ciñéndonos a



Figura 1. Ejemplo de georreferenciación indirecta a través del topónimo «Teide» y de la dirección postal «Rambla Pulido, 18»



Figura 2. Georreferenciación indirecta del Museo del Prado a partir de su dirección postal

la definición de una dirección, es la identificación de la ubicación fija de un bien o propiedad, por ejemplo, un terreno, un edificio, una parte de un edificio, una vía de acceso u otra construcción, mediante una composición estructurada de nombres e identificadores geográficos.

2. BASE DE DATOS DE CARTOCIUDAD

Desde hace años, CartoCiudad recolecta las direcciones de fuentes oficiales, junto con otros datos, para publicarlas en internet y ser una base de datos de referencia de las direcciones postales de España con representación puntual procedente de distintos organismos estatales, autonómicos y locales, que son complementadas con la asignación del código postal y la población a la que pertenecen.

El objetivo del proyecto de CartoCiudad es tener una base de datos lo más completa y actualizada posible, para que el usuario pueda geolocalizar cualquier propiedad sobre el territorio español con información oficial y así, que dicha información sea utilizada también en los diferentes servicios que tiene CartoCiudad, como son los servicios de descarga (*Web Feature Service, WFS*) y de visualización (*Web Map Service, WMS*) conformes al Reglamento de Servicios de Red y el servicio de procesamiento (*Web Processing Service, WPS*) de OGC y los servicios REST: *services* y *geocodificador* que se describirán en los siguientes puntos.

2.1. Fuente de datos

Las fuentes oficiales de las cuales se obtienen los datos son las siguientes:

1) La red viaria procedente de la Red de Transporte de la Información Geográfica de Referencia (IGR-RT) del Sistema Cartográfico Nacional (SCN), que está constituida por las vías interurbanas (autopistas, autovías, carreteras, caminos, pistas y sendas) con sus puntos kilométricos y los portales ubicados a borde de parcela.

3) Los códigos postales procedentes del Grupo Correos

con el fin de establecer una relación con las direcciones y núcleos de población del IGN. Además, el Grupo Correos aporta la geometría, a partir de la cual se generan los polígonos de los códigos postales a través del servicio de visualización (*WMS*). Los códigos postales no están disponibles para su descarga, solo se pueden visualizar.

4) El nombre oficial y tipo

de vía urbana (calle, avenida, plaza, etc.), procedente del Instituto Nacional de Estadística recopilados a partir de los datos de los ayuntamientos. Anualmente también aporta los distritos y secciones censales cuya información se incorpora directamente a la base de datos de CartoCiudad. Estos, al igual que los códigos postales, son visibles y consultables a través de los servicios *WMS*, pero no se pueden descargar sus geometrías.

5) Las líneas límite, poblaciones y puntos de interés procedentes del IGN. Desde el servicio «geocodificador», además de consultar las direcciones se pueden consultar la localización y geometría superficial de los municipios, provincias y comunidades autónomas, puntos de interés procedentes de la Base Topográfica Nacional a escala 1:25.000 (BTN25) y el Nomenclátor Básico de España (NGBE), y las poblaciones.

2.2. Actualización de direcciones

Como una dirección es una información que tiene una frecuencia de actualización muy alta y están en continuo cambio, ya sea porque cambian los nombres de las direcciones o porque nacen nuevas, se trabaja continuamente incorporando la última versión vigente de la IGR-RT, pero además, como hay organizaciones que mantienen sus direcciones actualizadas de forma prácticamente continua, se ha decidido incorporar aquellas direcciones que estén en internet de forma abierta y, a ser posible, normalizada.

Anteriormente se estaba trabajando, como se ha mencionado, con la última versión vigente de IGR-RT, a ello se le sumaban las direcciones de la Dirección General de Catastro, Eusko Jaurlaritza - Gobierno Vasco y el Gobierno de Navarra, y así haciendo cruces masivos se obtenían el mayor número de direcciones actualizadas.

Pero además, durante este año 2020, se ha trabajado en la incorporación de direcciones obtenidas de las Comunidades Autónomas de Madrid, Illes Balears, Comunidad Valenciana, La Rioja y Andalucía (CDAU), y las direcciones locales procedentes de la Diputación de Barcelona y del Ayuntamiento de Cáceres.

2.2.1. Metodología

Los datos anteriormente mencionados se han obtenido de formas diferentes, desde ficheros *shape*, *csv*, o *gml* desde los servicios de descarga WFS y ATOM Feed. También algunos de ellos han sido aportados por el propio gobierno autonómico como es el caso de La Rioja y el Eusko Jaurlaritz - Gobierno Vasco. Todas estas nuevas incorporaciones permiten actualizar, modificar e insertar direcciones en la base de datos de CartoCiudad. Esto se realiza mediante procesos automáticos creados con herramientas ETL (*Extract, Transform and Load*).

Para ello se sigue dos fases, una primera donde se normalizan los nombres de las direcciones tanto de los datos de IGR-RT como de Catastro y los datos locales; y una segunda fase de comparativa de datos, haciendo una comparación de los datos ya existentes en la base de datos (IGR-RT) con las direcciones procedentes de las

fuentes locales y de Catastro. Así, si en un radio de 25 m de los puntos locales, se encuentra algún otro punto de IGR-RT, se hace una comparativa de los nombres de las direcciones normalizadas, y si tienen una similitud mayor o igual al 75 %, significa que ese dato no es nuevo, ya estaba en la base de datos de CartoCiudad, y por ello no se va a incorporar. En cambio, si no encuentra ningún punto o si dicha similitud es inferior a ese 75 %, significará que esa información es nueva o hay una modificación y por ello se incorporará en la base de datos.

Este proceso se hace exactamente igual con los datos de Catastro, primero enfrentándolos a los datos de IGR-RT, y posteriormente, los que hayan pasado se enfrentarán a los nuevos datos locales obtenidos.

Cabe destacar que hay dos grandes diferencias entre los datos de IGR-RT y los datos locales/Catastro. La primera de ellas, es que los datos de IGR-RT tienen un identificador de vial que, gracias a él, cuando se realizan búsquedas de direcciones se tiene toda la geometría lineal de dicha dirección (Figura 4); en cambio los datos locales/Catastro no tienen dicho identificador y por ello solo se tienen elementos puntuales, es decir los portales de las direcciones (Figura 5). La otra diferencia importante es que estos últimos datos incorporados no están disponibles para descarga, son actualizados solo para su inclusión en el «geocodificador».

2.2.2. Problemas detectados y subsanados en la actualización

Todo proceso de actualización conlleva realizar una serie de comprobaciones y ver que todo funciona correctamente. Por ello, una vez que se han subido las direcciones nuevas a la base de datos de CartoCiudad, se procede a realizar una batería de peticiones y así se puede comprobar si los servicios que publican las direcciones funcionan correctamente.

Una vez realizado este chequeo, se llegó a la conclu-



Figura 3. Procesos automáticos con herramientas ETL



Figura 4. Geometría lineal



Figura 5. Geometría puntual

sión de que había una serie de errores sistemáticos en las nuevas direcciones introducidas en la base de datos. A continuación, se muestran unos ejemplos:

- Errores con los nombres con apóstrofes: cuando se introducía una calle con un apóstrofe, lo que ocurría es que el «geocodificador» encontraba la dirección pero no la geolocalizaba, dando el servicio un *time out*. Este error actualmente se ha solucionado, geolocalizando dicha dirección a partir de sus coordenadas geográficas. Véase un ejemplo de una calle con apóstrofe en la Figura 6.
- Errores con similitud de nombre: Se apreciaban errores cuando existía una calle tanto en los datos de IGR-RT como en los datos locales/Catastro, pero que se diferenciaban la una de la otra en que, por ejemplo, el tipo de vial fuese distinto o que tuviesen distinto número de portal.

Con este error, el servicio REST que proporciona las coordenadas de una dirección, «geocodificador», daba como resultado una calle, la cual podía existir o no, y posteriormente la geolocalizaba incorrectamente. Este error se ha subsanado cambiando la codificación del geocoder para las nuevas direcciones.

En la Figura 7 se aprecia que el «geocodificador» ha encontrado la Plaza de la Iglesia nº 2, pero la geolocaliza o posiciona en la Calle Iglesia, 2 (imagen de la izquierda). Una vez corregido el error, el «geocodificador» sitúa correctamente la dirección, y como no existe en la base de datos el portal número 2 devuelve el portal más próximo; es decir el número 1 de la Plaza Iglesia.

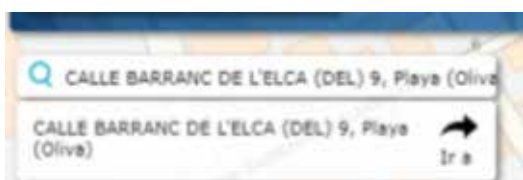


Figura 6. Error de direcciones con apóstrofes



Figura 7. Dos geolocalizaciones diferentes a partir de la dirección postal: Plaza de la Iglesia 2, Madrid

2.3. Actualización de Puntos de Interés (Point of Interests, POI)

También hay que tener en cuenta que no solo hay cambios continuos en las direcciones, sino también en elementos puntuales como los paradores, colegios, hospitales, campings, etc., denominados puntos de interés, que en el día a día son tan importantes como las direcciones postales, ya que se tiende a localizar una posición mediante el nombre reconocido o conocido del edificio, en vez de usar su dirección postal. Por ello, estos últimos meses, se ha hecho un profundo análisis de las fuentes oficiales que se tenían anteriormente y de las fuentes que se tienen ahora y de cómo están clasificados.

Con todo este análisis, se ha llegado a la conclusión de que la Base Topográfica Nacional 1:25.000 (BTN25), es la mayor fuente de estos elementos ya que tiene un periodo de actualización muy alto y usa como fuente oficial la información de IGR-RT para los elementos puntuales como aeropuertos, puertos, y estaciones de ferrocarril. Junto con la BTN25, se tiene también el Nomenclátor Geográfico Básico de España (NGBE) para los barrios y distritos, Naturaleza Cultura y Ocio para los albergues juveniles y paradores, y los datos del Ministerio para la Transición Ecológica y el Reto Demográfico (MITECO) para las Estaciones de Servicio. Hay que destacar que actualmente se están analizando los elementos contenidos en la base de datos del NGBE, ya que al incorporar nuevos POI se va a poder prescindir de muchos elementos de esa fuente.

3. RESULTADOS

Una vez que se ha realizado dicha actualización, se pueden comprobar los resultados en el componente de búsqueda o *plugin*, denominado «IGN_Search», que combina el servicio REST «geocodificador» de direcciones con el servicio de descarga INSPIRE (WFS) que publica los topónimos del NGBE. El «IGN_Search» es una caja de búsqueda incrustada en los visualizadores que permite localizar un lugar a través de una dirección, un nombre geográfico o un lugar de interés publi-

cados a través de los dos servicios mencionados.

Con todo ello se van a tener más de 3 700 000 millones de direcciones aproximadamente procedentes de las fuentes locales/Catastro. Todas éstas, fusionándolas con las de IGR-RT darán un total de unas 15 700 000 de direcciones que se tendrán en la base de datos de CartoCiudad.

Respecto a los POI, actualmente, CartoCiudad publica unos 89 000 puntos de interés respecto a los 59 000 puntos que se tenían anteriormente.

4. CONCLUSIONES

El objetivo del CNIG es publicar, a través de servicios interoperables, la información geográfica lo más actualizada posible, como son los mapas, las ortoimágenes, los fotogramas, las bases de datos o topográficas o en este caso las direcciones y puntos de interés de CartoCiudad cuyo mayor escaparate es a través del *plugin IGN-Search* que está incrustado en todos los visualizadores del IGN y del CNIG para localizar un lugar. Para ello, en 2020 se han mejorado los procesos de actualización y detección de errores de direcciones duplicadas y obtención de nuevas direcciones de todas las fuentes como IGR-RT, Catastro, Gobierno de Navarra, Eusko Jaurlaritz - Gobierno Vasco y otras fuentes locales. El número de nuevas direcciones que se han incorporado ha sido de 3 700 000 aproximadamente y el número de puntos de interés de 89 000.

Otro de los objetivos del CNIG es ir aumentando el número de fuentes locales, y si las direcciones procedentes de las fuentes locales se publican a través de WFS o *ATOM Feed* y los datos siguen la Especificación de Datos de Direcciones de la Directiva INSPIRE facilitará mucho la labor, ya que se podrán reutilizar los procesos ya creados para la D. G. de Catastro y la Diputación de Barcelona.

Esta información estará a disposición del público a través del Centro de Descargas del CNIG y se publicará además mediante servicios de datos espaciales de visualización (WMS), descarga (WFS), abiertos y gratuitos, de cobertura nacional con licencia CC BY 4.0 y reconociendo las fuentes de datos utilizadas.

REFERENCIAS

- Centro de Descargas del CNIG: <http://centrodedescargas.cnig.es/CentroDescargas/index.jsp>
- INSPIRE Data Specification on Addresses – Technical Guidelines. <https://inspire.ec.europa.eu/file/1728/download?token=K1Jh4B5h>
- Proyecto CartoCiudad. <https://www.cartociudad.es/porta/web/guest/que-es-cartociudad>

Sobre los autores

Emilio López Romero

Ingeniero en Informática por la Universidad de Málaga. Ha trabajado en la empresa privada y en el 2003 ingreso en el Cuerpo Superior de Sistemas y Tecnologías de la Información. Actualmente es el Director del Centro Nacional de Información Geográfica y Presidente del Consejo Directivo de la Infraestructura de Información Geográfica (CODIIGE) del Ministerio de Transportes, Movilidad y Agenda Urbana, donde lidera entre otros proyectos, el proyecto colaborativo de la Infraestructura de Datos Espaciales de España (IDEE) y es punto de contacto de la Directiva INSPIRE.

Paloma Abad Power

Ingeniera en Geodesia y Cartografía. En el 2003 ingresó en el cuerpo de Ingenieros Geógrafos y desde entonces participo en el equipo de la IDEE que lidera el CNIG. Actualmente es la Subdirectora Adjunta del Centro Nacional de Información Geográfica del Ministerio de Transportes, Movilidad y Agenda Urbana, es miembro Consejo Directivo de la Infraestructura de Información Geográfica (CODIIGE) como responsable del Seguimiento e Informe de la implementación de la Directiva INSPIRE.

Itziar Doñate Vadillo

Ingeniera en Geodesia y Cartografía e Ingeniera Técnica en Topografía por la Universidad Politécnica de Madrid. En 2020 ingresa en el Cuerpo de Ingenieros Topógrafos del Estado y desarrolla sus funciones en el Centro Nacional de Información Geográfica en el Área de Infraestructura de Información Geográfica en el proyecto de Cartociudad. Actualmente es la responsable de los servicios web de Cartociudad y de la actualización de los datos de direcciones.

Beatriz Brenes Fernández

Ingeniera en Geodesia y Cartografía e Ingeniera Técnica en Topografía. Responsable de la supervisión de trabajos técnicos de producción de la Información Geográfica de Referencia de Redes de Transporte (RT), de la programación de los automatismos para el control de calidad y la mejora y completitud de los datos de proyectos SIG de transportes y direcciones y de la coordinación e interlocución con empresas y supervisión de los trabajos realizados externos en el proyecto de CartoCiudad.