

SISTEMA DE GEORREFERENCIACIÓN PARA FINES ESTADÍSTICOS

Jesús Alberto González Yanes

ISTAC

jgonyanp@gobiernodecanarias.org

Rafael Betancor Villalba

ISTAC

rbetvilb@gobiernodecanarias.org

M^a Salomé Hernández García

EVM

salome.hernandez@evm.net

PALABRAS CLAVE

Georreferenciación, registro de portales, estadísticas, base de conocimiento

RESUMEN

La georreferenciación es el proceso de referenciar datos contra un sistema de coordenadas geoespacial conocido, ajustándose a puntos conocidos en el sistema de coordenadas, de manera que los datos puedan ser visualizados, procesados, consultados y analizados junto con otros datos geográficos.

Toda fuente que disponga de una dirección podrá ser georreferenciada es por ello que esto se podrá aplicar a la gestión de urbanismos, a las políticas sociales, a los transportes, etc.

El método empleado para la georreferenciación consiste en la creación de un registro de portales que nos servirá como base de conocimiento, es decir en este registro se almacenará todas las direcciones o portales georreferenciados permitiendo que cualquier dirección que llegue pueda obtener sus coordenadas directamente de este registro.

Una dirección que nos llega de cualquier fuente puede que se encuentre en nuestro registro de manera idéntica o que la podamos asociar a una ya existente por su similitud. Para ello un primer proceso es la normalización de direcciones. Es decir, las direcciones pasan un proceso de limpieza, como puede ser eliminación de caracteres raros, eliminación de puntuaciones, pasar a mayúsculas, etc.

Si la dirección ya existe de manera idéntica en el registro, no se hace nada. En cambio, si la dirección se ha asociado por similitud, esta nueva dirección será almacenada en nuestro registro como una versión diferente de la dirección, permitiendo así mejorar nuestra base de conocimiento.

También puede suceder que no tengamos esa dirección, o ninguna parecida. Para poder georreferenciar esta nueva dirección recurrimos a las APIs de georreferenciación, que previamente hemos estudiado y seleccionado las de mejor calidad para el territorio canario. El proceso consiste en llamar a las diferentes APIs y hacer una selección del

mejor punto devuelto por ellas. Esta nueva dirección y sus respectivas coordenadas serán añadidas al registro de portales. Pudiendo estar disponible para una próxima ocasión.

Este método permite en su proceso de georreferenciación la geocodificación que consiste, por ejemplo, en asignar un código de vía a la dirección. Es decir que toda información adicional que se pueda obtener o incluir en nuestro registro de portales aportará más conocimiento a nuestra base de conocimiento.

Este método se ha implementado en su totalidad en código R, donde las principales librerías utilizadas han sido dplyr, stringr, RPostgreSQL, RecordLinkage, jsonlite, RCurl. Como interfaz de desarrollo se ha utilizado RStudio, como almacenamiento de datos se dispone de una base de datos en PostgreSQL.

Las fuentes que ya han sido georreferenciadas en el Instituto Canario de Estadística son el padrón municipal de habitantes, demandantes de empleo, afiliados a la seguridad social, afiliados a Muface y cuentas de cotización. También se ha hecho uso de parte de este método para la georreferenciación de aeropuertos, restauración, fincas de cultivos, etc.