



Georreferenciador de veredas y centros poblados de Colombia para bases de datos e inventarios.

Miguel Ángel Molina Plaza

Trabajo de grado presentado para optar al título de Ingeniero Civil

Asesores

Edwin Fabian García Aristizábal, Doctor (PhD) Ing. Civil.

Roberto José Marín Sánchez, Doctor (PhD) Ing. Civil.

Universidad de Antioquia
Facultad de Ingeniería
Ingeniería Civil
Medellín, Antioquia, Colombia
2022

Cita	(Molina Plaza, 2022)
Referencia	Molina Plaza, M. (2022). <i>Georreferenciador de veredas y centros poblados de Colombia para bases de datos e inventarios</i> [Trabajo de grado profesional].
Estilo APA 7 (2020)	Universidad de Antioquia, Medellín, Colombia.



Grupo de Investigación Infraestructura (GII).



CENDOJ

Repositorio Institucional: <http://bibliotecadigital.udea.edu.co>

Universidad de Antioquia - www.udea.edu.co

Rector: John Jairo Arboleda Céspedes.

Decano/Director: Jesús Francisco Vargas Bonilla.

Jefe departamento: Diana Catalina Rodríguez Loaiza

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

Tabla de contenido

Resumen	7
Abstract	8
Introducción	9
1 Objetivos	10
1.1 Objetivo general	10
1.2 Objetivos específicos.....	10
2 Marco teórico	11
3 Metodología	15
3.1 Recolección y gestión de la información.	15
3.2 Desarrollo del código.	15
3.3 Aplicación, corrección y validación.....	16
4 Resultados y análisis	17
4.1 Recolección y gestión de la información.	17
4.2 Desarrollo del código.	17
4.3 Aplicación, corrección y validación.....	20
4.4 Georreferenciación con Google Maps.....	24
4.5 Interfaz gráfica.	26
5 Conclusiones	29
Referencias	30
Anexos.....	32

Lista de tablas

Tabla 1 Principales librerías utilizadas de Python.....	15
Tabla 2 Datos antes de depurar	18
Tabla 3 Datos después de depurar	18
Tabla 4 Rangos de coincidencia.....	19
Tabla 5 Datos georreferenciados	19
Tabla 6 Cantidad de datos georreferenciados por Rango de coincidencia.....	21
Tabla 7 Distancia entre ubicación real y georreferenciada.	22

Lista de figuras

Figura 1 Sistema de coordenadas geográficas	11
Figura 2 Visualización de datos georreferenciados.	20
Figura 3 Visualización de los puntos georreferenciados.....	21
Figura 4 Gráfico de cajas y bigotes de distancias de todos los datos georreferenciados	22
Figura 5 Gráfico de cajas y bigotes de distancias por rango de coincidencia.....	23
Figura 6 Coordenadas reales contra georreferenciadas.....	24
Figura 7 Gráfico de cajas y bigotes de distancias de datos georreferenciados con Google.....	25
Figura 8 Gráfico de cajas y bigotes de distancias de datos georreferenciados con Google corregido.....	25
Figura 9 Coordenadas reales contra georreferenciadas con Google.	26
Figura 10 Interfaz gráfica, página de inicio.	27
Figura 11 Interfaz gráfica, página georreferenciar.....	27
Figura 12 Interfaz gráfica, página georreferenciar con los datos procesados.	28

Siglas, acrónimos y abreviaturas

CSV	Comma Separated Values
DANE	Departamento Administrativo Nacional de Estadística
ERIC	Education Resources Information Center
ESRI	Environmental Systems Research Institute
Lat	Latitud
Long	Longitud
NCGIA	National Center of Geographic Information and Analysis
PhD	Philosophiae Doctor
SIG	Sistemas de Información Geográfica
UdeA	Universidad de Antioquia

Resumen

El presente trabajo describe el desarrollo de una herramienta que facilita el proceso de georreferenciación de bases de datos e inventarios con datos espaciales de ubicaciones correspondientes a veredas, centros poblados y municipios utilizando bases de datos previamente georreferenciadas, por lo tanto, la herramienta se limita en precisión ya que encuentra únicamente las ubicaciones mencionadas. Se pretende que los datos georreferenciados sean útiles para la toma de decisiones de gestión territorial y el estudio y análisis de fenómenos espaciales-temporales, como los desastres naturales.

El código se escribió usando Python, se evaluó su precisión utilizando registros reales de deslizamientos de masa, realizando mediciones de la distancia entre las coordenadas obtenidas georreferenciando con el programa y las coordenadas reales. Se comparó la eficiencia y rendimiento del programa con la georreferenciación que ofrece Google Maps, donde se encontró que para zonas rurales como veredas y centros poblados el programa realizado tiene un mejor desempeño y precisión que Google Maps. Finalmente, se describe el desarrollo de una interfaz gráfica de usuario que permite georreferenciar fácilmente archivos.

Palabras clave: Georreferenciación, Sistemas de Coordenadas Geográficas, Python, Inventarios.

Abstract

The present work describes the development of a tool that facilitates the process of georeferencing databases and inventories with spatial data from locations corresponding to sidewalks, Towns and municipalities using previously geo-referenced databases, therefore, the tool is limited in accuracy as it finds only the locations mentioned. Georeferenced data are intended to be useful for territorial management decision-making and the study and analysis of spatial-temporal phenomena, such as natural disasters.

The code was written using Python, its accuracy was evaluated using real records of mass slides, making measurements of the distance between the coordinates obtained georeferencing with the program and the real coordinates. The efficiency and performance of the program was compared with the georeferencing offered by Google Maps, where it was found that for rural areas such as sidewalks and towns the program performed has better performance and accuracy than Google Maps. Finally, it describes the development of a graphical user interface that allows easy georeferencing of files.

Keywords: Georeferencing, Geographic Coordinate Systems, Python, Inventories

Introducción

La localización espacial de un fenómeno que ocurrió en un lugar y momento específico esta descrita por datos espaciales. Estos datos espaciales representan el componente espacial, temporal y cualitativo del fenómeno, lo que le proporciona utilidad a la información y constituyen el núcleo de los Sistemas de Información Geográfica (Fernández, 2021). La estructuración y almacenamiento de información espacial georreferenciada es útil al momento de realizar labores de administración y control sobre bienes de uso público o gestión territorial (Gómez et al. 2022). La georreferenciación nos permite situar espacialmente puntos concretos de la geografía, asignando coordenadas geográficas a puntos de un mapa.

La georreferenciación de inventarios constituye una herramienta eficaz de ayuda a la toma de decisiones en diferentes campos como: Medio ambiente y recursos naturales, catastro de bienes rústicos y urbanos, redes de infraestructura, transporte y articulación regional, planificación urbana y la protección civil por medio de la determinación y evaluación de riesgos, desastres y catástrofes naturales o antrópicas (Lara et al. 2006).

En Colombia ya se ha utilizado sistemas de información geográficos en el aspecto catastral y la planeación municipal, siendo de gran utilidad para la elaboración de los Planes de Ordenamiento Territorial (Granados, 2021). Así mismo, en 1985 en la realización de un proyecto piloto, se efectuó un inventario de los deslizamientos y movimientos de masa más importantes con el fin de efectuar una zonificación de riesgo (Montero et al. 1988). DesInventar posee bases de datos masivas de los movimientos en masa históricos en el territorio colombiano que datan de 1914 a 2018 (DesInventar, 2018), sin embargo, gran parte de estos datos están sin georreferenciar, lo que dificulta realizar análisis espaciales en investigaciones. Recuperar esta información de manera precisa muchas veces no es posible ya que no existe la información suficiente para georreferenciar las ubicaciones en campo, además, si se trata de mucha información a georreferenciar no es viable realizar la georreferenciación en campo debido a los altos costos que esto conlleva (Champman & Wieczorek, 2022).

1 Objetivos

1.1 Objetivo general

Desarrollar una herramienta que permita georreferenciar ubicaciones geográficas de Colombia de manera automática y masiva para bases de datos carentes de información relacionada con la ubicación espacial.

1.2 Objetivos específicos

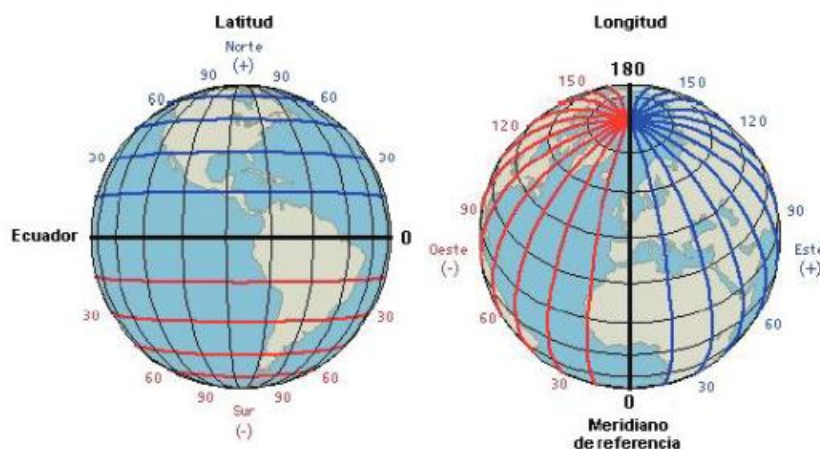
- Recolectar bases de datos con las coordenadas de veredas, corregimientos, y municipios de Colombia para usar como referencia para georreferenciar las ubicaciones.
- Elaborar un código que, a partir de una base de datos ingresada, georreferencie sus ubicaciones estableciendo rangos de exactitud.
- Elaborar una interfaz que facilite al usuario usar la herramienta y le permita obtener la base de datos georreferenciada y visualizar las coordenadas en un mapa de Colombia.
- Aplicar la herramienta en una base de datos de un inventario de movimientos de masa para validar su desempeño y utilidad.

2 Marco teórico

El sistema de coordenadas geográficas es un sistema de referencia que utiliza las dos coordenadas angulares: latitud (norte o sur) y longitud (este u oeste). Sirve para determinar los ángulos laterales de la superficie terrestre; la latitud mide el ángulo entre cualquier punto y el ecuador y la longitud mide el ángulo a lo largo del ecuador desde cualquier punto de la tierra (tomando como 0 el meridiano de Greenwich); tomando estos dos ángulos, se puede expresar la posición de cualquier punto de la superficie de la tierra (Rojas, 2012). La **Figura 1** muestra la representación gráfica del sistema de coordenadas geográficas, a la izquierda los cortes de este a oeste paralelos al Ecuador que representan la latitud y a la derecha los cortes de norte a sur paralelos al meridiano de Greenwich que representan la longitud (Fernández, 2021) (GIS Geography, 2021).

Figura 1

Sistema de coordenadas geográficas



Nota. Fuente (GIS Geography, 2021).

Las proyecciones cartográficas o proyecciones geográficas son un sistema de representación gráfico que establece una relación ordenada entre los puntos de la superficie curva de la tierra y los de una superficie plana. Un datum geodésico es un conjunto de puntos de referencia en la superficie terrestre en base a los cuales las medidas de la posición son tomadas y un modelo asociado de la forma de la tierra para definir el sistema de coordenadas geográfico. El datum WGS84 es el único sistema de referencia mundial utilizado hoy en día, es el datum estándar para coordenadas en los dispositivos GPS (Rojas, 2012).

El National Center of Geographic Information and Analysis (NCGIA) de los Estados Unidos define a los Sistemas de Información Geográfica (SIG) como un sistema de información compuesto por hardware, software y procedimientos para capturar, manejar, manipular, analizar, modelizar y representar datos georreferenciados, con el objetivo de resolver problemas de gestión y planificación (Wang et al. 2011). Los SIG (sistemas de información geográfica) están ampliamente utilizados tanto en la geografía como en otras ciencias, principalmente, aquellas implican la planificación y ordenamiento territorial y la solución de problemas socioeconómicos y ambientales (López et al. 2006). En la administración tanto como en la investigación, se analiza el territorio a través de distintas capas temáticas como el suelo y sus usos, los términos municipales, la red hidrográfica, el sistema de asentamientos, la infraestructura vial, etc.

La georreferenciación es un proceso de aplicación de algún tipo de referencia geográfica (Muñoz, 2009). La georreferenciación se puede realizar desde coordenadas conocidas o desconocidas. La primera se puede efectuar cuando existe un catastro de coordenadas a disposición, pero existe la posibilidad de que la calidad no sea adecuada, o no cumpla con los requerimientos de las bases de datos ya existentes, por lo que se requiere una revisión automática o manual de las bases de datos para luego normalizar las coordenadas. La georreferenciación con coordenadas desconocidas solo se puede realizar a través de la captura de puntos por medio de GPS, o bien, utilizando imágenes satelitales apoyadas de referencias geográficas emitidas por autoridades competentes (Muñoz, 2009).

En Colombia los sistemas de información geográficos son de gran utilidad para la elaboración de los Planes de Ordenamiento Territorial. También, se han desarrollado proyectos de gestión de pavimentos, en los cuales se identificaban las deficiencias estructurales en la red vial de diversos sectores, mediante la recolección, organización, verificación y procesamiento de la información (Granados, 2021).

Los recursos tecnológicos han permitido grandes avances en la creación y disponibilidad de información de cartografía de calidad (Muñoz, 2009). Crear bases de datos con información cartográfica requiere de un proceso sistemático que consiste inicialmente en la captura de datos, que trata de transformar los datos disponibles a un formato idóneo para el tratamiento automático, también es necesario detectar y corregir errores para no perder el valor de los datos, además, se deben procesar los datos dependiendo del propósito cartográfico o de la representación de los datos

y finalmente, almacenar los datos lo que permite asegurar que los datos podrán ser empleados en usos posteriores (Muñoz, 2009).

Los procesos de georreferenciación que cumplen con los estándares (fiable, accesible, interoperables, reusables) son por lo general totalmente georreferenciadas en campo, sin embargo, realizar levantamientos de campo pueden requerir de la realización de proyectos que pueden abarcar altos costos y tiempo de realización que dependerán de si es un proyecto para un investigador individual o para una institución grande. Por lo tanto, si no se requiere una exactitud rigurosa, se puede georreferenciar bases de datos usando registros previamente georreferenciados, utilizando software como Google Maps, Google Earth, bases de datos con localizaciones georreferenciadas por una institución, etc. (Chapman & Wieczorek, 2022). Esta práctica permite georreferenciar bases de datos masivas de manera eficiente que pueden ser utilizadas para análisis (o preanálisis) espaciales en escalas de grandes magnitudes.

El cálculo de la distancia entre dos puntos está resuelto por la ecuación de distancia Euclidiana ilustrada en la **Ecuación 1**. Esta ecuación es válida en ejes de referencia cartesianos X, Y, Z en superficies planas (Casquero, 2022). Al tener en cuenta que la superficie terrestre no es plana, la Ecuación 1 puede ser utilizada para medir distancias geográficas siempre y cuando la distancia sea menor que 20 km (12 mi), esta medición puede tener un error dependiendo la latitud en la que se ubican los puntos, que va desde 30 metros para latitudes menores que 70° a 9 metros para latitudes menores que 30° (Chamberlain, 1996).

Ecuación 1

Distancia Euclidiana

$$d = \sqrt{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2}$$

El cálculo de distancias de lugares lejanos o lugares con altas latitudes mediante la ecuación Euclidiana no es apropiado. La fórmula del semiverseno mide distancias de dos puntos ubicados en una esfera, lo que le otorga gran importancia en la astronomía la náutica y navegación, disponiendo de un método fiable para el cálculo de las distancias en el globo terrestre (Casquero, 2022). En la **Ecuación 2** se ilustra el cálculo de la distancia d entre dos coordenadas geográficas de latitudes φ_1 y φ_2 y longitudes λ_1 y λ_2 y donde R es el radio de la tierra.

Ecuación 2*Distancia semiverseno*

$$d = \sqrt{\sin^2\left(\frac{\varphi_1 - \varphi_2}{2}\right)^2 + \cos(\varphi_1) \cos(\varphi_2) \sin^2\left(\frac{\lambda_1 - \lambda_2}{2}\right)^2}$$

3 Metodología

Para el desarrollo del presente trabajo se comprenden 3 etapas: Recolección y gestión de la información; desarrollo del código; aplicación, corrección y evaluación de desempeño.

3.1 Recolección y gestión de la información.

El tipo de localizaciones a georreferenciar por medio de bases de datos previamente georreferenciadas se delimitó a veredas, centros poblados y municipios. Los datos de referencia se deben recolectar de bases de datos abiertas de instituciones oficiales (DANE, ESRI), una vez recolectada la información, se deben preparar los datos para que puedan ser utilizados, para esto se debe revisar la exactitud y realizar una limpieza de datos incorrectos, repetidos o sin información.

3.2 Desarrollo del código.

El código se desarrolló utilizando el lenguaje Python, ya que es un lenguaje multiplataforma que facilita el cómputo matemático, manejo de grandes volúmenes de datos y uso de funciones para su análisis (Fernández, 2021). Python ofrece múltiples librerías robustas, las cuales están conformadas por gran variedad de métodos y buen material descriptivo de respaldo (Fernández, 2021), además, permiten gestionar los datos ingresados y visualizarlos en un mapa como vista previa.

La siguiente tabla muestra las principales librerías utilizadas y la función que cumplen en el código.

Tabla 1

Principales librerías utilizadas de Python

Librería	Descripción
Numpy	Operaciones matemáticas y cálculo numérico. (Numpy, 2022)
Pandas	Manipulación y gestión de archivos y tablas de datos. (Pandas, 2022)
Folium	Visualización con mapas interactivos. (Folium, 2022)

Entre las funciones y utilidades del programa se encuentra:

Permitir el ingreso de bases de datos. Estas bases de datos deberán estar previamente organizadas en un formato establecido: extensión .csv y con las columnas con los nombres DEPARTAMENTO, MUNICIPIO, ZONA (en el cual estarán la vereda o centro poblado). El código depurará los datos para que coincidan con los datos de referencia.

Georreferenciar la latitud y longitud en coordenadas geométricas y poner un rango que clasifique el tipo de coincidencia; es decir, si logró encontrar la coincidencia en la columna zona o municipio o departamento.

Visualizar las coordenadas encontradas por medio de puntos geométricos en un mapa de Colombia, donde se pueda observar el rango de coincidencia según el color del punto.

Exportar la base de datos georreferenciada con toda la información para que el usuario pueda hacer uso de esta.

3.3 Aplicación, corrección y validación.

Se aplicó el programa en bases de datos reales para realizar pruebas y corregir errores en el código. Se validó el desempeño del código teniendo en cuenta la exactitud de la georreferenciación; para esto, se mide que tanto se alejan las coordenadas georreferenciadas con el código de las coordenadas reales medidas en campo. No es práctico verificar cada registro individualmente, por lo tanto, se usa procedimientos de detección de valores atípicos, se utilizan diagramas de caja para identificar estos valores (Chapman & Wieczorek, 2022). Adicionalmente, se realiza una regresión lineal de las coordenadas reales y las georreferenciadas, para esto se adicionan los valores de la longitud y latitud de las coordenadas reales y se grafican contra la adición de los valores de la longitud y latitud de las coordenadas reales, esto con el fin de buscar la correlación tanto en la longitud como en la latitud en la misma gráfica, ya que se pueden tener ubicaciones con una coordenada muy cercana a la real y la otra muy lejana.

4 Resultados y análisis

4.1 Recolección y gestión de la información.

Los datos de referencia se recolectaron de bases de datos abiertas de instituciones oficiales. Para las veredas, se usaron los datos abiertos del portal de ESRI (ESRI, 2022), donde se puede acceder a un shapefile con los polígonos que representan todas las veredas de Colombia, utilizando ArcMap se calculó el centroide de cada polígono, se exportaron las coordenadas de cada centroide con su respectiva vereda a un archivo Excel, estos datos son los que se utilizarán como referencia para georreferenciar las veredas. El DANE (DANE, 2021) posee bases de datos abiertas de los municipios y centros poblados de Colombia en las cuales ya se encuentran georreferenciadas sus respectivas coordenadas. Una vez con los datos de las veredas, centros poblados y municipios se procede a revisar manualmente los datos en busca de información incorrecta y datos repetidos. En esta inspección se encontraron y eliminaron varias veredas y centros poblados repetidos, en el caso de las veredas se eliminaron un total de 3443 veredas por falta de información; de estos datos, se observa que en la región del pacífico en departamentos como Chocó, Nariño y Cauca se encontraron muchos datos faltantes por lo que se dificultará georreferenciar coordenadas en estos departamentos. En recuento, las bases de datos finales poseen 30.460 veredas y 7.806 centros poblados y municipios georreferenciados.

4.2 Desarrollo del código.

El código fue desarrollado en Python y está disponible en Google Colaboratory. Adicionalmente, se desarrolló una interfaz disponible como página web para que facilite a el usuario utilizar la herramienta. El usuario debe cargar un archivo de extensión .csv donde únicamente se encuentren tres columnas que contengan el departamento, el municipio y la zona (vereda o centro poblado), estas columnas deben llevar sus nombres en mayúsculas (DEPARTAMENTO, MUNICIPIO, ZONA). Una vez el usuario carga un archivo de extensión .csv, el código realiza las siguientes funciones: depuración de los datos, georreferenciación y visualización.

En la depuración de los datos, se busca que los nombres de las ubicaciones tengan coincidencias exactas entre los caracteres de los datos ingresados por el usuario y los datos de referencia. Las bases de datos de referencia no tienen tildes, espacios y símbolos; y en las columnas se encuentra solamente el nombre propio de la zona por coordenada. Para mejorar la búsqueda, se depura la información cambiando todos los caracteres a mayúsculas y eliminando tildes, espacios y palabras o letras que pueden estar en la base de datos del usuario y alteren la búsqueda. En la **Tabla 2**, se muestra un ejemplo de datos ingresados, en la **Tabla 3** se observan estos datos depurados donde se eliminaron los espacios, se pusieron todos los caracteres en mayúsculas, se eliminaron tildes, símbolos, siglas y abreviaturas.

Tabla 2*Datos antes de depurar*

DEPARTAMENTO	MUNICIPIO	ZONA
Antioquia	Amagá	4 Palos
Antioquia	Peque	El Seno
Santander	Bucaramanga	Honda de Palmas
Antioquia	La Unión	Vereda las piedras
Santander	Bucaramanga	Sitio
Risaralda	Dosquebradas	Acequia Planta Belmonte
Quindío	La Tebaida	San José
Nariño	San Lorenzo	Zona Rural
Huila	La Argentina	I.P. El Pencil
Caldas	Aguadas	V. Arma

Tabla 3*Datos después de depurar*

DEPARTAMENTO	MUNICIPIO	ZONA
ANTIOQUIA	AMAGA	4PALOS
ANTIOQUIA	PEQUE	ELSENO
SANTANDER	BUCARAMANGA	HONDADEPALMAS
ANTIOQUIA	LAUNION	VEREDA LAS PIEDRAS
SANTANDER	BUCARAMANGA	SITIO
RISARALDA	DOSQUEBRADAS	ACEQUIA PLANTABELMONTE
QUINDIO	LATEBAIDA	SAN JOSE
NARIÑO	SANLORENZO	RURAL
HUILA	LAARGENTINA	ELPENCIL
CALDAS	AGUADAS	ARMA

En el proceso de georreferenciación, el programa busca coincidencias en cada campo, crea 3 columnas adicionales, latitud y longitud con las coordenadas encontradas y un campo que indica el rango de coincidencia en el que fue encontrada la coordenada. Para esta última columna, el programa busca coincidencia entre las columnas de los datos del usuario y las bases de datos de referencia partiendo de la columna zona, al no encontrar coincidencia en esta columna, busca la coordenada correspondiente al centro del municipio y finalmente en el departamento como se describe en la **Tabla 4**. Estos rangos permiten al usuario estimar una exactitud de las coordenadas georreferenciadas.

Tabla 4
Rangos de coincidencia

Rango	Descripción	Color en Mapa
1	Coincidencia en la zona (Centro Poblado o vereda)	Verde
2	Coincidencia en el municipio	Amarillo
3	Coincidencia en el departamento	Rojo
4	No se encontró coordenada	No Aplica

Tabla 5
Datos georreferenciados

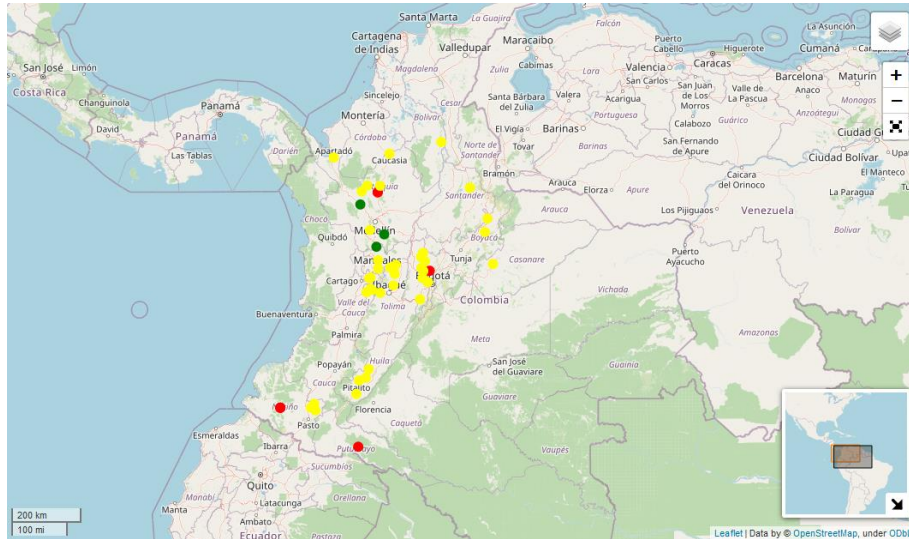
DEPARTAMENTO	MUNICIPIO	ZONA	LATITUD	LONGITUD	RANGO
Antioquia	Amagá	4 Palos	6.03880	-75.70195	2
Antioquia	Peque	El Seno	7.02101	-75.91035	2
Santander	Bucaramanga	Honda de Palmas	7.11626	-73.13260	2
Antioquia	La Unión	Vereda las piedras	5.92955	-75.33458	1
Santander	Bucaramanga	Sitio	7.11626	-73.13260	2
Risaralda	Dosquebradas	Acequia Planta Belmonte	4.82718	-75.68247	2
Quindío	La Tebaida	San José	4.45365	-75.78646	2
Nariño	San Lorenzo	Zona Rural	1.50334	-77.21523	2
Huila	La Argentina	I.P. El Pencil	2.19824	-75.97977	2
Caldas	Aguadas	V. Arma	5.61265	-75.53308	1

En la **Tabla 5** se observa los datos georreferenciados, se regresan los datos de las columnas de departamento, municipio y zona a como estaban previamente a la depuración y se agregaron las columnas de las coordenadas y del rango de coincidencia.

Finalmente, se muestra una visualización donde se ubican todas las coordenadas en un mapa en puntos geográficos donde cada color es el rango de coincidencia descrito en la **Tabla 4**. En la **Figura 2** se observa la visualización de los datos georreferenciados.

Figura 2

Visualización de datos georreferenciados.



4.3 Aplicación, corrección y validación.

El programa se aplicó en una colección de datos de diferentes inventarios de eventos de deslizamientos de masa (DesInventar, SIMMA, 2018), donde se tienen un total de 34.095 datos de los cuales 10.994 se encuentran sin georreferenciar. Del total se seleccionaron 99 datos manualmente, asegurándose de que estos valores hayan sido tomados en campo y la información este completa para poder georreferenciar con el programa.

Los 99 datos fueron georreferenciados utilizando el programa. La **Tabla 6** muestra cuantos datos se encontraron por cada rango de coincidencia que arrojó el código al georreferenciar los datos de prueba. Para este ejemplo, la mayoría de los datos se encuentran en el rango de coincidencia 1, al buscar manualmente los datos con rango 3, se encontró que los datos pertenecían a departamentos de la región del pacífico (Chocó y Cauca) como se mencionó en la sección de

recolección y gestión de la información esta región tiene datos faltantes en la base de datos, por lo tanto, se espera que los datos con coordenadas menos exactas sean de esta región.

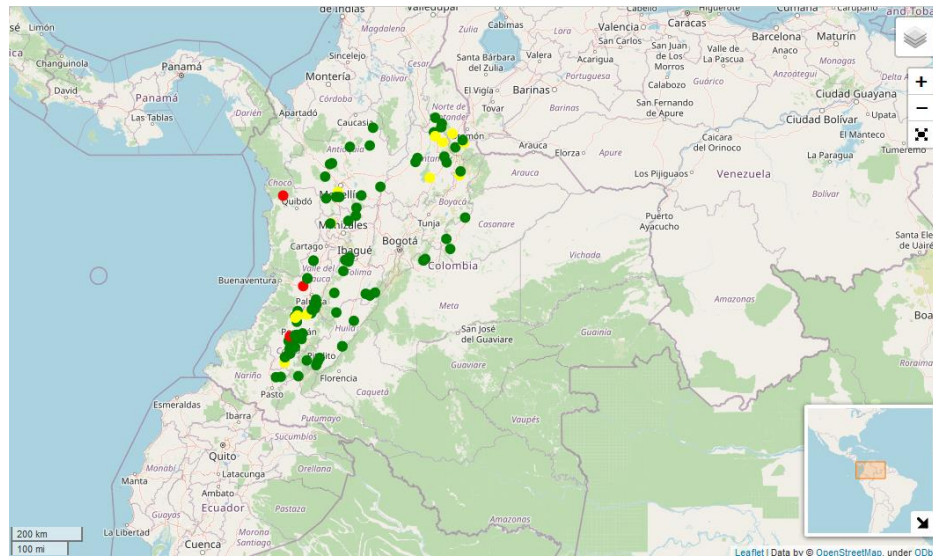
Tabla 6

Cantidad de datos georreferenciados por Rango de coincidencia.

Rango	Cantidad
1	82
2	14
3	3
4	0

Figura 3

Visualización de los puntos georreferenciados



En la **Figura 3** se muestran los puntos geográficos georreferenciados, donde se observan que la mayoría de estos son verdes, es decir, tienen rango de coincidencia 1.

Para poder medir el desempeño, se realizó un código que implementa la ecuación del semiverseno (Ecuación 2) para cada uno de los datos. Para esto, el código recibe un archivo que posee las coordenadas de los puntos a medir repartidos en 4 columnas con nombres Long 1 y Lat1 para la longitud y latitud del primer punto y Lon 2 y Lat 2 para la longitud y latitud del punto a medir, una vez calculada la distancia con la **Ecuación 2**, se ingresa una nueva columna con el resultado para cada fila. La **Tabla 7** muestra algunos de los resultados que arroja el código, donde la longitud y

latitud 1 corresponden a los datos originales y la longitud y latitud 2 corresponden a las coordenadas georreferenciadas, la distancia está en kilómetros.

Tabla 7

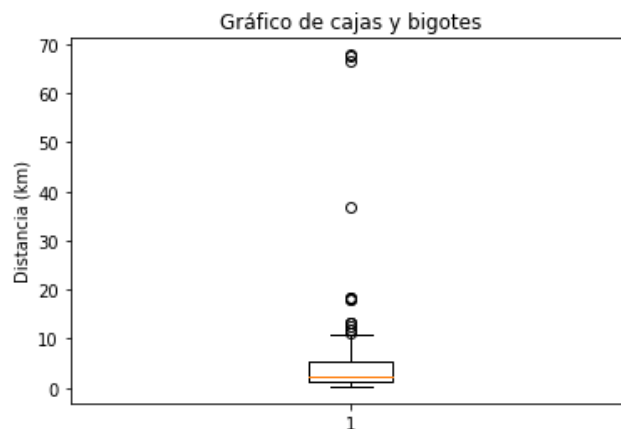
Distancia entre ubicación real y georreferenciada.

	Long 1	Lat 1	Long 2	Lat 2	distancia
0	6.8225	-73.8189	6.443471	-73.3374	54.85217
1	6.914167	-73.6542	6.933413	-73.6463	1.061142
2	4.404444	-75.4358	4.397336	-75.4448	1.014406
3	6.09422	-75.6287	6.091297	-75.6336	0.549719
4	5.509722	-75.175	5.499546	-75.1837	1.01398
5	7.1925	-72.6992	7.197731	-72.7009	0.260189
6	7.582222	-73.2358	7.570444	-73.2352	0.384385
7	3.56	-74.9436	3.556923	-74.9388	0.545868
8	1.827778	-76.1722	1.776927	-76.1748	1.380791
9	2.548889	-76.4667	2.570275	-76.5164	5.562034
10	2.860833	-75.2033	2.883036	-75.2351	3.590711

Al analizar el total de los resultados se encuentra una media de 6.175 ± 12.115 km donde la menor distancia presentada fue 0.261 km y la mayor fue de 67.859 km como lo describe la **Figura 4**.

Figura 4

Gráfico de cajas y bigotes de distancias de todos los datos georreferenciados

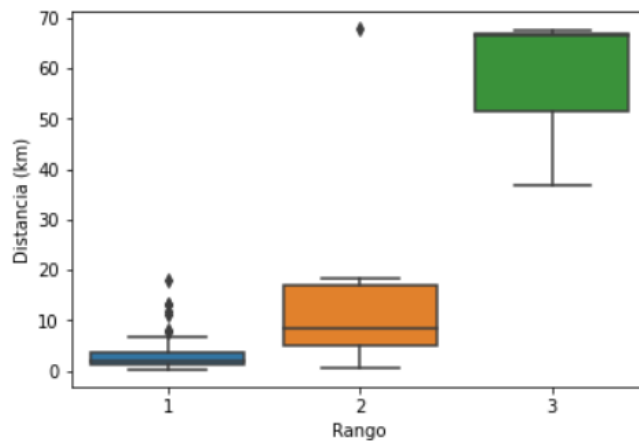


Se observa que al obtener 82 coordenadas de rango 1 la mayoría de los valores de las distancias se agrupan en distancias inferiores a 10 km, al analizar los datos por rango, se encuentra

que los resultados pueden ser más precisos tomando únicamente los de rango 1 ya que cuentan una media de 3.073 ± 3.194 km, si se desea mayor rigurosidad en los análisis geoespaciales, se pueden omitir los valores que no sean de rango 1 para trabajar con datos más fiables, ya que como se muestra en la **Figura 5**, los valores de rango 2 y 3 pueden aumentar la incertidumbre en los análisis; por lo tanto, los rangos sirven como criterio a la hora de realizar filtros en los datos según lo requiera el usuario.

Figura 5

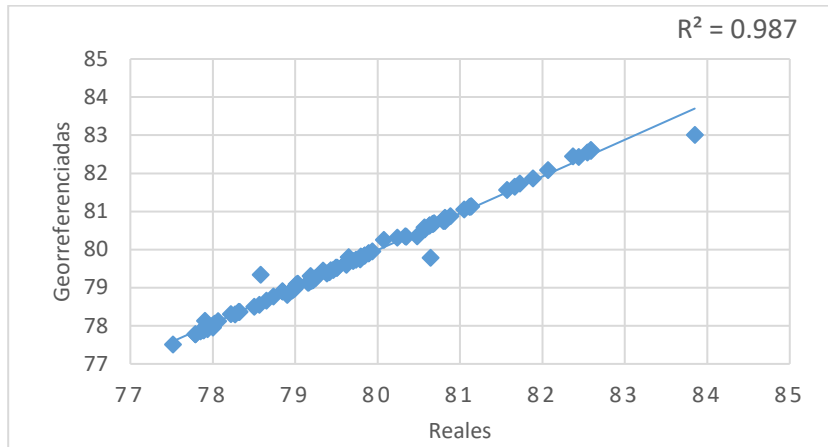
Gráfico de cajas y bigotes de distancias por rango de coincidencia.



Finalmente, para comprobar la correlación entre los datos de las coordenadas reales contra las coordenadas georreferenciadas se realiza la regresión lineal siguiendo las indicaciones de la sección **3.3 Aplicación, corrección y validación.**, obteniendo la **Figura 6**, donde se observa que los datos se correlacionan linealmente con coeficiente de determinación R^2 de 0.987. Por lo tanto, incluso incluyendo los datos más atípicos correspondientes a los valores con coordenadas de rango 2 y 3, se encuentra que, sin ser rigurosos en la exactitud, las coordenadas georreferenciadas utilizando la herramienta, se acercan a las coordenadas de las ubicaciones reales.

Figura 6

Coordenadas reales contra georreferenciadas.

**4.4 Georreferenciación con Google Maps.**

Para evaluar el programa, se implementó el servicio de georreferenciación que ofrece Google Maps, a comparación de la georreferenciación utilizando bases de datos previamente georreferenciadas este método no tiene límites en el tipo localizaciones que se desean buscar, lo que permite resultados más precisos, sobre todo en zonas urbanas donde Google tiene registro de calles, edificios, barrios, etc. Por otro lado, en ocasiones no encuentra ubicaciones o la georreferencia incorrectamente al buscar algunas ubicaciones rurales que no se encuentren en los registros de Google. Además, utilizar este servicio tiene costos económicos y de eficiencia elevados si se utiliza de forma masiva; por lo tanto, con fines ilustrativos se utilizó una clave de acceso abierta para comparar los resultados encontrados con la georreferenciación de Google y la georreferenciación con bases de datos previamente georreferenciadas.

Se georreferenció la base de datos de prueba utilizando Google y se procedió a medir las distancias entre los puntos originales y los obtenidos a través de Google. La **Figura 7** muestra los resultados obtenidos.

Figura 7

Gráfico de cajas y bigotes de distancias de datos georreferenciados con Google

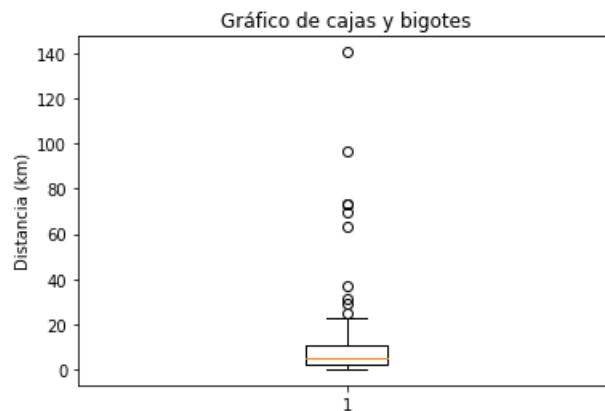


Como se observa en la **Figura 7**, hay datos que se alejan demasiado del dato original, como se mencionó anteriormente, es muy común encontrar datos incorrectos georreferenciando con Google, esto se debe principalmente a que encuentra ubicaciones más relevantes en otras regiones con el mismo nombre de la ubicación buscada.

Manualmente se eliminaron los datos más alejados correspondientes a 4 coordenadas de ubicaciones fuera de Colombia para obtener una mejor visualización de los resultados, como se muestra en la **Figura 8**. Donde aún se observan muchos datos alejados con una distancia promedio de 12.171 ± 21.651 km.

Figura 8

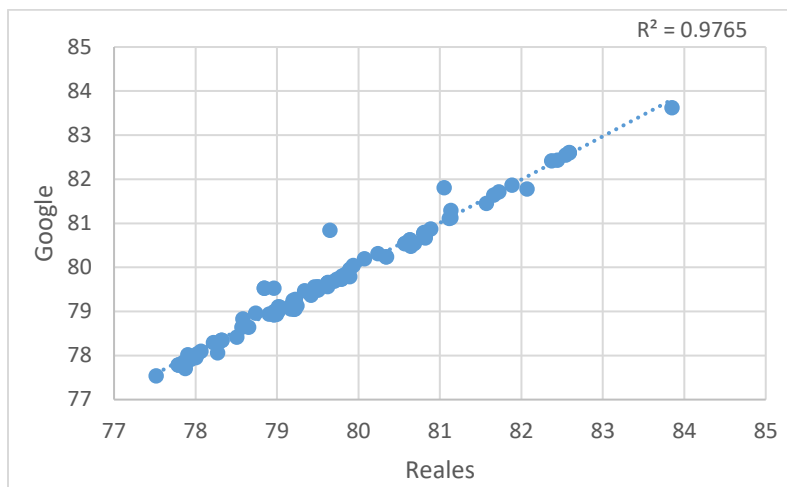
Gráfico de cajas y bigotes de distancias de datos georreferenciados con Google corregido.



Con los datos corregidos, se graficaron las coordenadas reales contra las coordenadas georreferenciadas con Google, obteniendo la **Figura 9**, donde se observa que, aunque los datos se correlacionen linealmente con un R^2 de 0.97655, se encuentran más datos atípicos y una menor correlación entre los datos que en la **Figura 6**, esto teniendo en cuenta que se eliminaron los datos más con margen de mayor error.

Figura 9

Coordenadas reales contra georreferenciadas con Google.



Comparando la **Figura 4** con la **Figura 8** y la **Figura 6** con la **Figura 9**, se observa que se obtiene distancias mucho menores georreferenciando con bases de datos previamente georreferenciadas que georreferenciando con Google, por lo tanto, se puede decir que el primer método es más preciso al georreferenciar zonas rurales como veredas y corregimientos y también mucho más rápido y fiable, ya que al no encontrar una coincidencia exacta como una vereda o corregimiento el programa utiliza la coordenada del municipio o departamento por lo tanto las coordenadas del dato georreferenciado no se alejarán mucho de las reales, además, los rangos de coincidencia permiten dar una idea de que tanta incertidumbre puede tener un dato.

4.5 Interfaz gráfica.

Para fácil uso de esta herramienta, se elaboró una página web con Flask (Flask, 2022), donde el usuario puede subir los archivos que desea georreferenciar y descargar los resultados de manera

sencilla. **Figura 10** muestra la página de inicio con instrucciones de que realiza la herramienta y de cómo se deben subir los archivos que se desean georreferenciar.

Figura 10

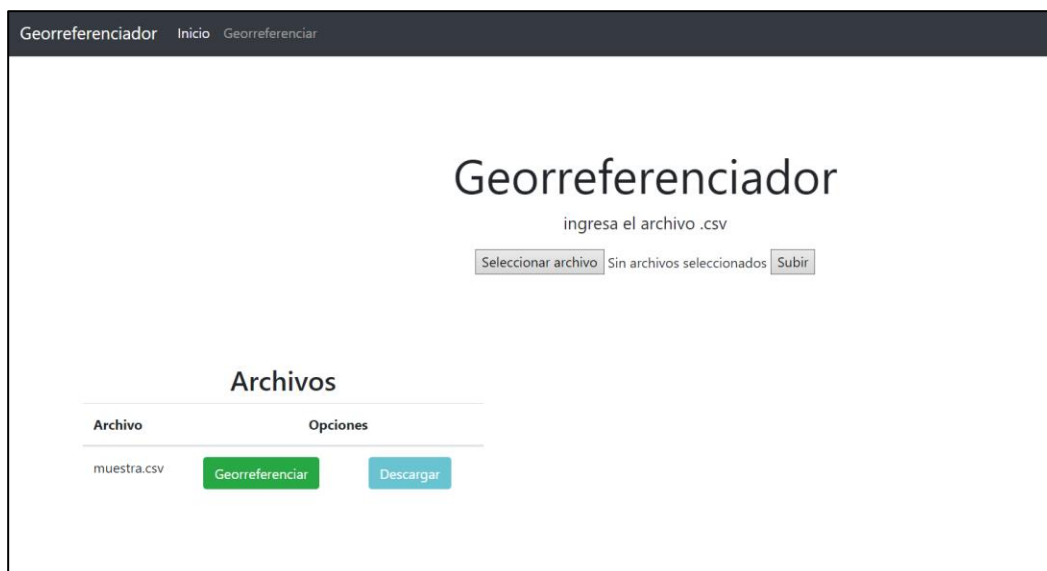
Interfaz gráfica, página de inicio.



La **Figura 11** muestra la página para georreferenciar, donde se deben seleccionar los archivos que se desean georreferenciar en formato csv y subirlos, los archivos subidos aparecerán en la sección de archivos en la parte inferior izquierda.

Figura 11

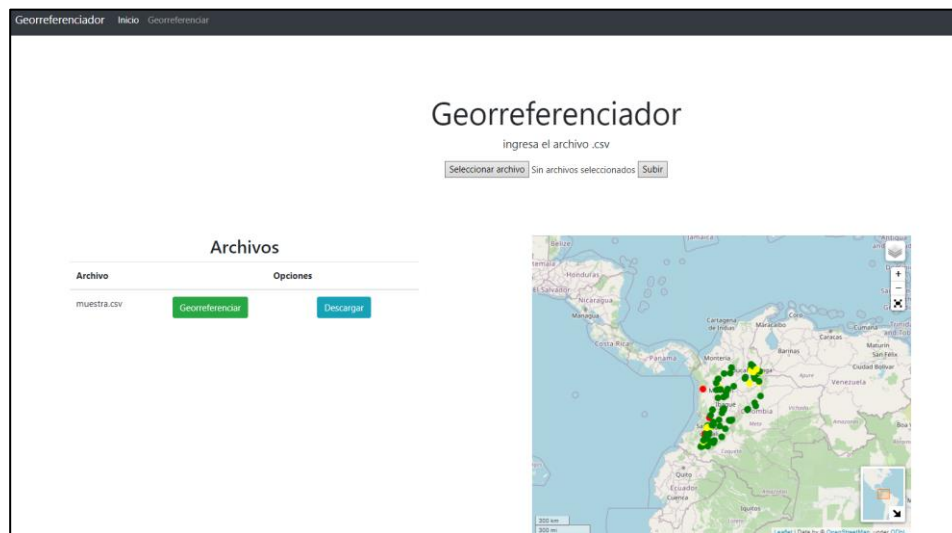
Interfaz gráfica, página georreferenciar.



Finalmente, una vez se oprime el botón de georreferenciar, la página ejecuta el proceso de georreferenciación como se describe en la sección **4.2 Desarrollo del código.**, se habilita la opción de descargar el archivo con los datos georreferenciados y se muestra una visualización de las coordenadas obtenidas como se ilustra en la Figura 12.

Figura 12

Interfaz gráfica, página georreferenciar con los datos procesados.



5 Conclusiones

La herramienta digital: georreferenciador de veredas y centros poblados de Colombia para bases de datos e inventarios surge por la necesidad de añadir utilidad a grandes cantidades de datos espaciales sin georreferenciar para realizar análisis espaciales o gestiones territoriales. Esta herramienta puede georreferenciar grandes cantidades de datos correspondientes a veredas y centros poblados utilizando bases de datos previamente georreferenciadas, dando un rango que permite estimar la precisión de los resultados obtenidos. Este proceso se puede realizar fácilmente gracias a una interfaz simple que permite subir y procesar los archivos fácilmente y visualizar los resultados en un mapa.

Para los resultados obtenidos georreferenciando veredas y centros poblados, esta herramienta demuestra un mejor desempeño en tiempo y precisión que georreferenciar estas ubicaciones utilizando herramientas robustas como lo es Google Maps. Esto se debe a que el buscador de Google Maps prioriza ubicaciones más relevantes que poseen el mismo nombre de la ubicación buscada mientras que el programa realizado se asegura que la ubicación coincida, en lo posible, en el departamento, municipio y vereda de la ubicación buscada.

Una de las principales limitaciones que tiene esta herramienta es el tipo de ubicaciones que puede georreferenciar, ya que se restringe a veredas, centros poblados o corregimientos y municipios. A pesar de que el georreferenciador realiza su tarea de forma eficiente georreferenciando estas ubicaciones, aún tiene un amplio margen de mejora; como la implementación de nuevas bases de referencia que permitan georreferenciar infraestructura importante como puentes, viaductos, túneles; y zonas urbanas como barrios, edificios públicos, direcciones, etc.

Referencias

- Casquero Martín, Miguel Ángel (2022). *Redes de narices electrónicas desplegadas en catástrofes*. E.T.S.I. de Minas y Energía (UPM). Recuperado de <https://oa.upm.es/70473/>
- Chamberlain, R. G. (1996). Geographic Information Systems FAQ. Recuperado de: <http://www.faqs.org/faqs/geography/infosystems-faq/>
- Chapman, A. D., & Wieczorek, J. R. (s/f). Guía de buenas prácticas de georreferenciación. Gbif.org. Recuperado de <https://docs.gbif.org/georeferencing-best-practices/1.0/es/>
- DANE. (2021). Descarga del Marco Geostatístico Nacional (MGN). Disponible en: <http://bit.ly/3Wgkrt2>
- DesInventar. (2018). Disaster Information Management System. (Colombia Profile). Recuperado de <https://bit.ly/3BE7dhT>
- ESRI (2022). Límites administrativos a nivel veredal de Colombia, con sus respectivos atributos de identificación. Disponible en: <https://bit.ly/3HD3ogB>
- Fernández Lobo, Allison Esther (2021). *Detección de patrones en datos espaciales para aplicaciones con geolocalización en smart cities*. Tesis (Master), E.T.S. de Ingenieros Informáticos (UPM). Recuperado de: <https://oa.upm.es/67907/>
- Flask. (2022). Desarrollo web una gota a la vez. Disponible en: <https://flask.palletsprojects.com/en/2.2.x/>
- Folium. (2022). Visualización de datos en mapa interactivo con Python. Disponible en: <https://python-visualization.github.io/folium/>
- GIS Geography. (2021). Latitude, Longitude and Coordinate System Grids. Disponible en: <https://bit.ly/3UYdpYX>
- Gómez Quesada J. C., & Afanador Franco, F. (2009). Modelo geodatabase en la estructuración y almacenamiento de información espacial para la administración y control del litoral Caribe colombiano. *Boletín Científico CIOH*, (27), 135–144. <https://doi.org/10.26640/22159045.206>
- Granados, J. s. (2021). Aplicación de los sistemas de Información Geográfica en los Inventarios de Drenaje Vial. Recuperado de: <http://hdl.handle.net/10654/40412>.
- Lara, E. L., Simeón, C. P., & Navarro, J. G. M. (2006). Los sistemas de información geográfica. *Geoenseñanza*, 11, 16.
- Montero O., J., Beltrán M., L., & Cortés D., R. (1988). Inventario de deslizamientos en la red vial colombiana. *Ingeniería e Investigación*, 17, 16–27. <https://doi.org/10.15446/ing.investig.n17.19652>
- Muñoz, D. A. (2008). Metodología para la georreferenciación de elementos emisores y su implementación a través de un SIG. *Tiempo y espacio*, (21), 24-46.
- NumPy. (2022). El paquete fundamental para computación científica con Python. Disponible en: <https://numpy.org/>
- Pandas. (2022). Manipulación y análisis de datos. Disponible en: <https://pandas.pydata.org/>

Rojas Ugaz, H. O. (2012). Ejecución del inventario vial georreferenciado de Piura y propuesta de optimización de metodologías existentes.

SIMMA, SGC. (2018). Sistema de Información de Movimientos en Masa (SIMMA). Recuperado de <https://simma.sgc.gov.co/>

Q. Wang, X. Wang, H. Tang, J. Li, H. Ma and T. Cheng, "Spatial and Dynamic Analysis of Regional Sustainable Development Using Geographic Information System and Relative Carrying Capacity of Resources," *Natural Resources*, Vol. 2 No. 1, 2011, pp. 28-34. doi: 10.4236/nr.2011.21005.

Anexos

Código de georreferenciador disponible en Google Colab: <https://bit.ly/3FxTOJt>

Código Georreferenciador y página web disponible en repositorio GitHub:

<https://github.com/MiguelAMolinaP/Georreferenciador>