



## **Exploración sobre un algoritmo de compresión con pérdida aplicado a conjuntos de coordenadas geográficas para reducción del peso computacional en mapeo de rutas**

*Miguel Ángel García Chávarro<sup>1\*</sup>, Santiago Yustre Bautista<sup>1</sup>, Apolonio Hernández Montoya<sup>1</sup>*

<sup>1</sup>Universidad Autónoma de Tamaulipas

\*agcmiguel15@gmail.com

### **RESUMEN**

Los métodos para la reducción de datos conservando la esencia de la información son uno de los objetivos de los algoritmos de compresión con pérdida, donde a un archivo original se le extraerá los datos claves que permitan representar de manera eficiente las partes más importantes, sacrificando entonces un porcentaje de la calidad; en este trabajo se tomó la teoría usada generalmente en imágenes, videos y audios, para aplicarla en la reducción del peso en archivos que contienen coordenadas geográficas usando las herramientas de la documentación de Google Maps plataforma, de manera que se pueda comprobar que en lugar de almacenar latitud y longitud, se pueda representar de una manera más simple y reducir el coste computacional si se usa en otros algoritmos.

**Palabras claves:** Algoritmos, Coordenadas geográficas, Coste computacional, Compresión con pérdida.

### **ABSTRACT**

Data reduction methods preserve the information essence are one of the objectives of lossy compression algorithms, where the key data will be extracted from an original file that allows to represent in a more efficient way the most important parts, then sacrificing a percentage of the quality; in this work, theory generally used in images, videos and audios was taken to apply it the weight reduction in the data that contains geographic coordinates using Google Maps platform documentation tools, so that it can be verified that instead of storing latitude and longitude, it can be represented in a simpler way and reduce the computational cost if it is used in other algorithms.

**Keywords:** Data reduction methods, geographic coordinates, lossy compression algorithms, computational cost.

## INTRODUCCIÓN

Para el 2022 a 16 años del lanzamiento de Google maps qué con 25,463,180 es el servicio de mapas más usado en términos de descargas en Estados Unidos, la cantidad de usos en servicios geográficos han aumentado, siendo de utilidad en servicios de Global Positioning System (GPS), mapeo satelital e incluso como guía en servicios como Booking o Airbnb mostrando a los huéspedes cómo encontrar con exactitud la ubicación del alojamiento (Sharma, 2022). Por otra parte, el sistema de coordenadas usadas en los mapas como Google maps está basada en el sistema geodésico World Geodetic System 84 (WGS 84) que permite expresar un punto en la tierra sin necesidad de otro punto referencial, usando tres valores (x, y, z) dados en unidades de grados ( $^{\circ}$ ), siendo la tierra representada por una forma elipsoide y tomando como referencia el punto inicial sobre el meridiano de Greenwich (Stanaway, 2016).

En los tipos de archivos más utilizados para el uso de coordenadas están Keyhole Markup Language (KML), Keyhole Markup language Zipped (KMZ), basados en un lenguaje de marcas de hipertexto como HTML, que es ampliamente usado por Google maps y Google Earth o GeoJSON que está basado en JavaScript Object Notation (Json) pero con una representación más sencilla para coordenadas (Peterson, 2014); estos archivos son usados en la práctica para ser interpretados por un lenguaje y mostrarlos en mapas a través de rutas, marcas, puntos en mapas o información de un lugar asociado a esas coordenadas haciendo un llamado a las coordenadas contenidas en el archivo y listándolas como información o gráficamente sobre una API de un servicio de mapas como Google Maps, Open Street Maps o Leaflet (Kadyrovich, 2022).

Usar el algoritmo que permite simplificar los datos correspondientes a coordenadas geográficas WGS 84, tomando como base la documentación de Google maps para el uso de polilíneas codificadas (Google Maps API, 2016), de manera que no se tome por completo el valor sino, que se represente con menos datos sin perder el sentido de la información (Sayood, 2018) y por tanto, ofrecer una manera de comprimir, reducir el peso

de los archivos y expresar una manera en que los algoritmos pudieran leer rápidamente cadenas de texto que representan puntos en un mapa que constituyen una polilínea, ejemplo los puntos geográficos que conforman la ruta de un camión de transporte público reemplazando parejas de latitud y longitud por caracteres ASCII en una sola cadena de texto o String.

De esta manera, el objetivo general de la investigación, propone comparar un archivo de texto de conjuntos de coordenadas con un archivo tipo GeoJSON para reportar las observaciones obtenidas en el cotejo, para ello se necesita: 1) Explorar la naturaleza del sistema de coordenadas más usado (WGS 84), 2) Proponer y desplegar la metodología para la compresión de las coordenadas en un contexto regional, usando ubicaciones cercanas, 3) Evaluar la fidelidad de la información presentada en la documentación verificando que exista una mejora en las coordenadas codificadas con respecto a otro tipo de archivo para uso de puntos geográficos, 4) Reportar los resultados mediante una comparación que establezca las diferencias en Coste computacional o peso entre los archivos de coordenadas codificadas y archivos de parejas de coordenadas numéricas.

## **METODOLOGÍA**

Los algoritmos de compresión con pérdida buscan reducir los datos y representar la misma información sin perder los puntos más importantes (Lee et al., 2022), en la Ilustración 1 y la Ilustración 2, se plantea el ejemplo de una fotografía que pasó por el proceso de compresión con pérdida, dónde el algoritmo suprime la información de colores como los plateados que pueden expresarse como escalas de grises, reduciendo el nivel de detalle en los pixeles de los bordes pero conservando el orden de la imagen, colores destacados y la posición de la fuente de brillo.

La Ilustración 1 es la fotografía comprimida con pérdida de información con un peso total de 369.01 Kb, mientras la Ilustración 2 es la imagen original con un peso de 3.68 Mb, es decir que para el ejemplo hubo una reducción del 84% del peso del archivo, sin embargo, no se ven grandes diferencias al momento de observarlas.



Ilustración 1 "Fotografía comprimida"; Año: Septiembre,2022; Fuente: Propia.



Ilustración 2 "Fotografía Original"; Año: Septiembre,2022; Fuente: Propia.

**Explorar Naturaleza de las coordenadas.** Para el caso de estudio se usarán coordenadas geográficas basadas en los patrones de Latitud y Longitud en un sistema de referencia de coordenadas angulares esféricas, donde la latitud es la distancia en ángulos entre el centro de la tierra o línea ecuatorial y un punto determinado en la tierra (Chang, 2016); por otra parte, la Longitud es la distancia angular entre un punto determinado en la tierra y un punto 0, para el caso de WSG 84 es el meridiano de Greenwich.

**Proponer la metodología para la compresión de coordenadas.** Para el desarrollo de la metodología es necesario establecer el tipo de coordenadas que se usarán y entender su naturaleza, es decir que, si se tienen los valores de latitud y longitud, son estos numéricos los que se deberán transformar en texto según la documentación de Google Maps plataforma (Google Maps API, 2016), por otra parte, se hará una prueba de factibilidad con un punto fijo (ubicación Instituto Tecnológico Superior de Pánuco) para garantizar que la transformación de numéricos a String se está dando de manera correcta.

El proceso de conversión tratará de tomar un valor de punto fijo decimal a binario y convertirlo al American Standard Code for Information Interchange (ASCII) es decir en forma de carácter y posteriormente concatenar una serie de puntos fijos formando rutas o trazos llamados polilíneas de puntos geográficos, que son ampliamente usados en técnicas de mapeos o direccionamiento.

El procedimiento para transformar los puntos geográficos consiste en:

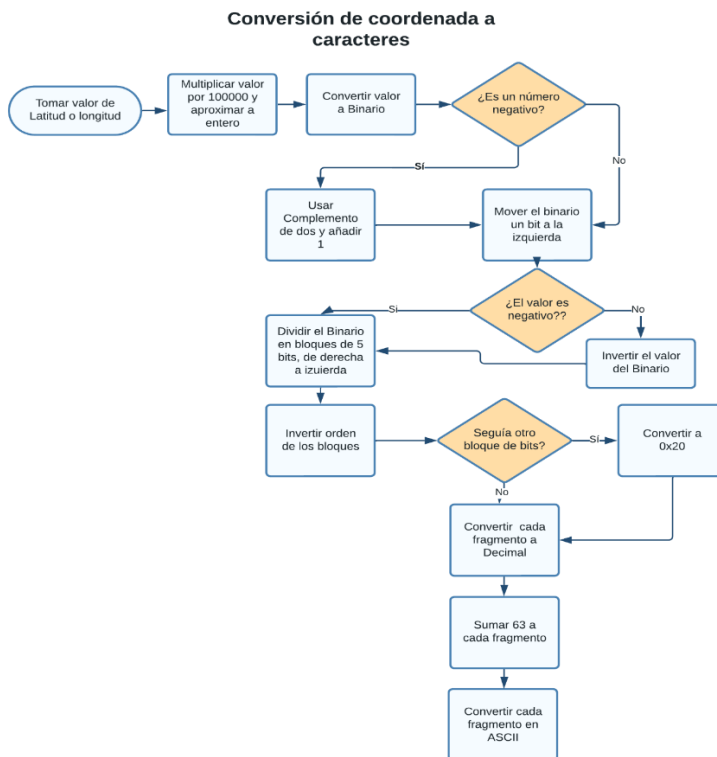


Ilustración 3 "Diagrama de flujo para conversión de coordenadas"; Año: Septiembre,2022; Fuente: propia.

**Evaluar la fidelidad de la información comprimida como un factor de mejor en términos de coste en cómputo.** Se imputarán los valores ASCII que representarán un punto Latitud, Longitud o una cadena compuesta por varios puntos mediante la función de JavaScript PathDecode en la API de Google maps, de manera que el patrón decodificado pueda ser usado en aplicaciones de geolocalización de la misma manera que se usaría una coordenada dada.

**Reportar los resultados mediante una comparación experimental.** Constituye la comparación del peso de los archivos y la cuantificación de mejora de desempeño en términos de porcentaje para proponer algunos usos del sistema de archivos ASCII como alternativa de los formatos más utilizados como GeoJSON.

**Prueba de hipótesis.** Para dar solución a la hipótesis de mejora en la metodología propuesta, se estudiará un nuevo caso para tener un diseño de tres modelos: uno con pocos puntos (9), otro con cantidad media de puntos (16) y el tercer caso será el de cantidad más grande de puntos sobre un mismo mapa (36).

Finalmente se mostrará en porcentajes de mejora y en una tabla los resultados obtenidos, de los cuales quedará demostrada la eficiencia del método propuesto en comparación de otras maneras de contener las coordenadas, como los archivos GeoJson.

## RESULTADOS

Dado el sistema de coordenadas, se tiene que para el Instituto Tecnológico Superior de Pánuco las coordenadas son 22.03941 y -98.19103, para esto, cualquier servicio de mapas que utilice el mismo sistema de coordenadas podrá entregar esta información, en este caso la ilustración 4 mediante la herramienta de Google Maps

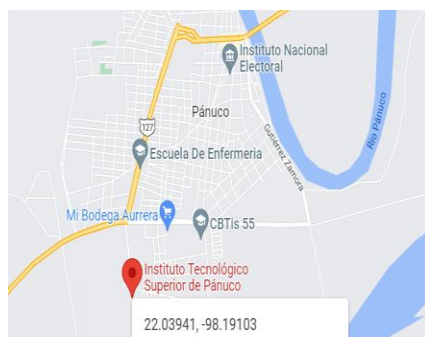


Ilustración 4 "Mapa y coordenadas Instituto Tecnológico Superior de Pánuco"; Año: Septiembre,2022; Fuente: Google Maps.

La ilustración 5 muestra el desarrollo de una polilínea decodificada en un mapa alrededor del Instituto Tecnológico Superior de Pánuco, los puntos azules proporcionan una latitud y longitud que se codificó mediante el algoritmo de compresión con pérdida y posteriormente se decodificó para representar la ruta compuesta por la unión de los siete puntos.

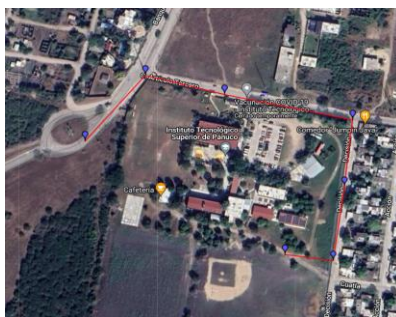


Ilustración 5 "Ruta de polilínea compuesta por puntos geográficos"; Año: Septiembre,2022; Fuente: Propia a través de Google maps.

En la Tabla 1 se muestran en números los 7 puntos mostrados en la ilustración 5, siendo el punto 1 el que se muestra más a la izquierda del mapa, los valores de las coordenadas no son tan diferentes entre ellos porque los puntos están cerca en relación con otras coordenadas en el mundo.

Tabla 1 "Coordenadas asociadas a cada punto de la ruta"; Año: Septiembre,2022; Fuente: Propia a través de Google maps.

Punto	Latitud	Longitud
1	22.03975	-98.19240
2	22.04040	-98.19180
3	22.04021	-98.19101
4	22.03996	-98.18974
5	22.03931	-98.18982
6	22.03858	-98.18993
7	22.03864	-98.19041

Siguiendo el algoritmo, el punto 1 con latitud 22.03975 y longitud -98.19240 se convierten en: Latitud 22.03975 = msoeC?

Longitud -98.19240 = ?neivQ

Y la coordenada completa se expresa como Latitud + Longitud eliminando el carácter genérico "?", Coordenada = msoeCneivQ y continúa punto por punto hasta obtener una sola cadena de texto como se muestra a en la Tabla 2, los trayectos son los trazos de la polilínea que se va formando, los puntos son los marcadores de coordenadas que están dentro de la polilínea y la codificación es la suma de coordenadas formando la ruta, pero dadas como cadena de texto:

Tabla 2 "Coordenadas asociadas a cada punto de la ruta"; Año: Septiembre,2022; Fuente: Propia.

Trayecto	Puntos	Codificación
1	1	msoeCneivQ
2	1-2	msoeCneivQaCwB
3	1-2-3	msoeCneivQaCwBd@}C
4	1-2-3-4	msoeCneivQaCwBd@}Cp@}F
5	1-2-3-4-5	msoeCneivQaCwBd@}Cp@}F`CN
6	1-2-3-4-5-6	msoeCneivQaCwBd@}Cp@}F`CNpCT
7	1-2-3-4-5-6-7	msoeCneivQaCwBd@}Cp@}F`CNpCTK-A

En la ilustración 6 se tiene el caso de la Laguna de Pueblo Viejo en Veracruz y se tomaron 36 puntos en coordenadas del contorno de la laguna, luego la unión de todos los puntos constituye una ruta que se asemeja al contorno de la laguna.

En lugar de comparar los puntos y caracteres como en el ejemplo anterior, se dará un ejemplo de un archivo para coordenadas geográficas que se usa convencionalmente y una cadena de texto que represente la ruta de contorno de la laguna

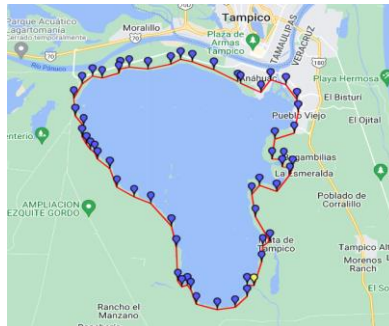


Ilustración 6 "Ruta de polilínea Laguna Pueblo viejo"; Año: Septiembre,2022; Fuente: Propia a través de Google maps.

El archivo GeoJSON que describe la ruta de la Ilustración 6, tiene una estructura como la que se muestra en la ilustración 7, los valores de latitud y longitud mostrados son de 3 puntos en la ruta, sin embargo, el archivo original se extiende con las otras 33 coordenadas.

```
{
  "type": "FeatureCollection",
  "features": [
    {
      "type": "Feature",
      "properties": {},
      "geometry": {
        "type": "LineString",
        "coordinates": [
          [
            -97.86518096923828,
            22.199007896720307
          ],
          [
            -97.87702560424805,
            22.205365246479296
          ],
          [
            -97.88801193237305,
            22.209497369425055
          ],
          [
            -97.86518096923828,
            22.199007896720307
          ]
        ]
      }
    }
  ]
}
```

Ilustración 7 "Fragmento archivo GeoJSON"; Año: Septiembre,2022; Fuente: Propia.

Mientras la decodificación mediante el algoritmo de compresión con pérdida resume toda esta ruta de coordenadas en la cadena de caracteres ASCII a continuación:

```
yvnfCjhitQwf@`iAyXrcA}Fja@zN`^|MzfA_Ezw@~C|ZzMhD|Nnr@_E~[xYb^tm@xYry@i
Cx^uWvc@dAzU{KzQyJ|G{KxVqGx^se@r{@mSx]st@zUor@njAm`ApeAaOj|BgBzS}L_
EcO|NwJtn@eP|Og}@}F{w@yb@mc@yc@?~@oTolAqVuu@kD}MmTooA`l@oiAvJy]q
VzRut@ut@uf@yXsH_Ab^{TgB_Anc@qaAsG{Ra{@oeAyJwg@fBuo@ha@{Odn@vh@
~[yf}@
```

el archivo que GeoJSON tiene un peso en disco de 5340 bytes mientras el archivo texto que contiene la cadena de texto tiene un peso en disco de 271 bytes.

De igual manera, repitiendo la comparación, el archivo de GeoJSON de los puntos cercanos al Instituto Tecnológico Superior de Pánuco tiene un peso de 728 bytes mientras el archivo texto tiene un peso de 31 bytes.

El diseño de pruebas presenta 3 casos de toma de coordenadas, todos los casos son en la Laguna de Pueblo Viejo, pero con cantidades diferentes de puntos.

Las Ilustraciones 8,9,19 son el mapa de la Laguna de Pueblo Viejo, con trazos de líneas con 9, 16, 36 puntos de coordenadas cada una, que se toman como muestras para el diseño experimental de pesos en relación con el formato en que se presentan.

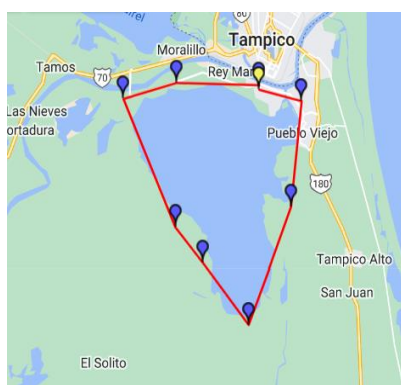


Ilustración 8 "Polilínea con 9 puntos"; Año: Octubre,2022; Fuente: Propia a través de Google Maps.

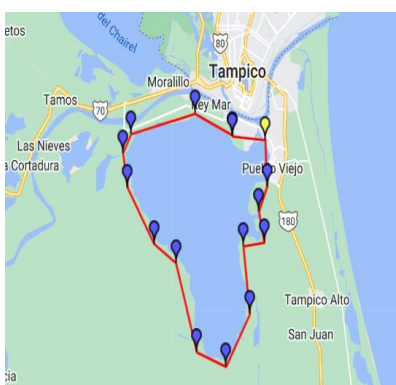


Ilustración 9 "polilínea con 16 puntos"; Año: Octubre,2022; Fuente: Propia a través de Google Maps.

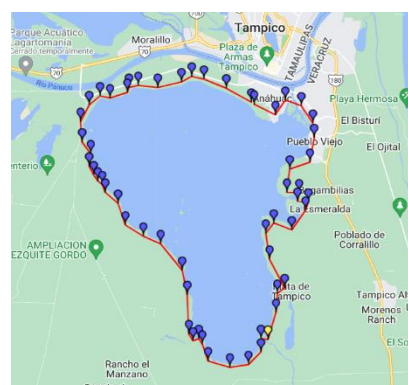


Ilustración 10 "polilínea con 36 puntos"; Año: Octubre,2022; Fuente: Propia a través de Google Maps.

La Tabla 3, muestra los resultados obtenidos en cada caso, siendo el primer caso el de los 9 puntos y el tercer caso el más extenso con 36 puntos, también se tabuló el peso del archivo GeoJson y el peso la polilínea codificada, en ambos casos están dados en bytes, finalmente la columna final muestra el porcentaje de reducción en peso entre ambos datos.

Tabla 3 "Relación hipótesis de puntos, pesos y reducción"; Año: octubre ,2022; Fuente: Propia.

Caso	Cantidad de puntos	Peso geojson	Peso codificado en txt	Reducción en peso después de codificado
1	9	785 bytes	51 bytes	93.5 %
2	16	1218 bytes	94 bytes	92.28%
3	36	5340 bytes	271 bytes	94.93%

En la Ilustración 11 hay una gráfica comparativa entre el caso 1, 2 ,3 en relación con el peso en Bytes, dónde el azul representa el GeoJson y el naranja los puntos codificados.

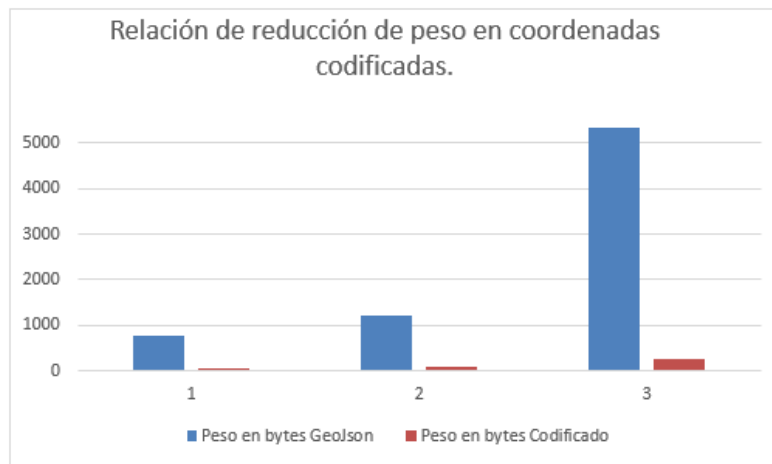


Ilustración 11 "Comparación gráfica de pesos por cada caso "; Año: Octubre,2022; Fuente: Propia

## DISCUSIÓN.

Para efectos prácticos se establecieron los parámetros de investigación, mostrando que como en otros servidores de mapas, Google también hace uso de coordenadas basadas en las distancias angulares entre la línea ecuatorial y el meridiano de Greenwich, en la Ilustración 4 se muestra que un punto ubicado en las proximidades del Instituto Tecnológico Superior de Pánuco se tiene coordenadas 22.03941 y -98.19103, por tanto es objeto de estudio de la presente investigación, continuando con la Ilustración 5 que muestra que la unión de varios puntos genera una ruta, que técnicamente es mostrada mediante polilíneas y por lo tanto calificable para el uso del algoritmo de compresión con pérdida mediante la codificación de decimales a ASCII propuesto por Google maps en su documentación, para ese ejemplo los puntos se mostraron en la tabla 1 en su forma natural y luego en la tabla 2 como los puntos descritos en caracteres de texto, sin embargo, aún no se establecía ninguna comparación sino, el proceso de transformación.

En la Ilustración 6 se hizo una ruta con 36 puntos alrededor de la Laguna de Pueblo Viejo , en este caso se planteó el caso de múltiples puntos de coordenadas que necesitan ser tratados a través de un archivo como GeoJSON, en la ilustración 7 se muestra únicamente un fragmento del archivo que compone toda la ruta, mostrando solo 3 coordenadas para evidenciar que el almacenar múltiples coordenadas requiere de mayor

esfuerzo que almacenar un archivo de texto, sin embargo, se plantea la comparación del peso de ambos archivos, primero la ruta marcada cerca del tecnológico de Pánuco y luego un a comparación del archivo con los 36 puntos,

En el caso del mapa del Instituto Tecnológico de Pánuco, que está compuesto por 7 puntos, la comparación arroja una reducción del 95.71%, siendo 728 bytes el peso del archivo GeoJSON y 31 bytes el archivo de texto.

Mientras en el caso de la Laguna de Pueblo Viejo con los 36 puntos, la reducción fue del 94.93%, siendo 5340 bytes el archivo GeoJSON y 271 el archivo de texto.

Para comprobar la hipótesis de reducción de pesos, se diseñó una comparación usando el mismo fragmento de mapa, pero con tres polilíneas diferentes, donde cada uno de ellos tuvo más parejas de coordenadas que el anterior; en la Tabla 3 se muestran los detalles de cada uno de los casos y los pesos, finalmente un gráfico de barras en la Ilustración 11 para mostrar la notable diferencia de pesos entre el archivo GeoJson que es usado normalmente y el archivo codificado a caracteres, que podría ser usado en casos donde se desea optimizar los recursos.

Además de la reducción en peso, también se da por comprobado el funcionamiento de la utilidad de polilíneas y el algoritmo como codificador de coordenadas bajo compresión con pérdida, pues se representa la misma información con una menor cantidad de datos en extensión.

## **CONCLUSIONES**

Como conclusión se pretende comprobar que el algoritmo de compresión por pérdidas aplicado a coordenadas geográficas funciona y es útil en casos como mapeo de rutas, localización de vehículos, rastreo, pero sobre todo una alternativa para sistemas con limitados recursos como los microcontroladores, dónde el almacenamiento representa uno de los grandes desafíos dada la capacidad de estos.

Sin embargo, también puede ser de utilidad en casos donde se trabaja con grandes cantidades de puntos y los datos se pudieran almacenar en una base de datos, en lugar de utilizar archivos externos como GeoJSON e implementar una infraestructura de

comunicación, se puedan almacenar directamente en cualquier motor de base de datos como si se tratara de una cadena de textos cualquiera pero que contiene información reducida de rutas o polilíneas en un mapa.

## LITERATURA CITADA

- Chang, K. (2016). *Introduction to Geographic Information Systems* (McGraw-Hill, Ed.; 9na ed.). McGraw-Hill.
- Google Maps API. (2016). Encoded Polyline Algorithm Format | Google Maps APIs | Google Developers. <https://developers.google.com/maps/documentation/utilities/polylinealgorithm>
- Kadyrovich, K. E. (2022). HISTORY OF GEOGRAPHICAL INFORMATION SYSTEMS AND ITS IMPORTANCE TODAY. 98–101. <http://papers.euroasiaconference.com/index.php/eac/article/view/617>
- Lee, J., Gong, Q., Choi, J., Banerjee, T., Klasky, S., Ranka, S., & Rangarajan, A. (2022). Error-Bounded Learned Scientific Data Compression with Preservation of Derived Quantities. *Applied Sciences* (Switzerland), 12(13). <https://doi.org/10.3390/app12136718>
- Peterson, M. P. (2014). Mapping GeoRSS Feeds and the Shift from KML to GeoJSON. *Cartogis*, 141–150.
- Sayood, K. (2018). Introduction to Data Compression. In *Introduction to Data Compression* (pp. 1–10). The Morgan Kaufmann Series in Multimedia Information and Systems. <https://doi.org/10.1016/b978-0-12-809474-7.00001-x>
- Sharma, M. (2022). HERE Maintains the Location Platform Leadership, Ahead of Google, and TomTom in 2021 - Counterpoint Research. In Counterpoint. <https://www.counterpointresearch.com/maintains-location-platform-leadership-ahead-google-tomtom/>
- Stanaway, R. (2016). GDA94, ITRF, WGS84: WHAT'S THE DIFFERENCE? WORKING WITH DYNAMIC DATUMS. *Packaging News*, 2016-Sept(September), 53–54.