

Refinamiento de imprecisiones en información geográfica voluntaria

Emmanuel Juárez-Carbajal, Marco Moreno-Ibarra, and Magdalena Saldana-Pérez

Resumen—En este artículo se propone una metodología para el refinamiento de las imprecisiones asociadas con datos geográficos generados bajo el enfoque de Información Geográfica Voluntaria (VGI); la propuesta considera los aspectos descriptivos (texto) y geoespaciales (distancia, topológicos) de los datos. En ese sentido, con el objetivo de contextualizar el dominio geográfico para soportar los análisis a los datos, se considera el uso de datos geográficos de la zona de estudio. El problema que se aborda es de relevancia, ya que promueve la generación de datos geográficos más precisos, lo que favorece el desarrollo de aplicaciones y atender de mejor manera, temas como la prevención de desastres, delincuencia, seguimiento de enfermedades contagiosas, entre otros. En el presente artículo se describe la metodología planteada, siendo este el inicio de una investigación del área geoespacial. En el documento se expresan las características de la metodología, las herramientas que se propone emplear, así como posibles fuentes de información. Como aportación preliminar, se propone el desarrollo de un prototipo, para realizar este refinamiento de manera automatizada, así como la geocodificación, la aplicación de métricas y el resto de los procesos involucrados. Se considera que los resultados de este trabajo pueden tener impacto en la comunidad usuaria de datos geográficos, ya que se favorece la reducción en los tiempos de procesamiento, consumo de recursos y ajustes de las bases de datos para verificar su calidad en cuanto a exactitud.

Palabras clave—Información geográfica voluntaria, geoespacial, geocodificación, exactitud, procesamiento de lenguaje natural.

Refining Inaccuracies in Voluntary Geographic Information

Abstract—In this paper, a methodology for the refinement to the inaccuracies associated with geographic data generated under the Volunteered Geographic Information (VGI) approach is proposed. The proposal considers the descriptive (text) and geospatial (distance, topological) aspects of the data. In this sense, to contextualize the geographic domain to support the data analysis, the use of geographic data from the study area is considered. The problem that is addressed is relevant, since it promotes the generation of more accurate geographic data, which favors the development of applications and better addresses issues such as disaster prevention, crime, monitoring of contagious diseases, among others. This article describes the proposed methodology, which is the beginning of an investigation of the geospatial area. The document expresses the characteristics of the methodology, the tools it proposes to use, as well as possible sources of information. As a preliminary contribution, the development of a prototype is proposed, to carry out this

refinement in an automated way, as well as geocoding, the application of metrics and the rest of the processes involved. It is considered that the results of this work may have an impact on the geographic data user community, since it favors the reduction in processing times, consumption of resources and adjustments of the databases to verify their quality in terms of accuracy.

Index Terms—Volunteered geographic information, geospatial, geocoding, accuracy, natural language processing.

I. INTRODUCCIÓN

La información geográfica voluntaria (VGI, por sus siglas en inglés) ha sido un fenómeno en los últimos años, ya que se ha constituido como un mecanismo alternativo para la adquisición y la recopilación de información geográfica. Esto se logra por la participación voluntaria de ciudadanos que reportan sus observaciones mediante plataformas webs, aplicaciones móviles y medios sociales [3,8].

La VGI es un término relacionado con el Crowdsourcing, que implica la recolección de información por parte de los usuarios involucrados con o sin experiencia en el ámbito geográfico, para el desarrollo de productos cartográficos y realicen análisis para la visualización del problema estudiado. Crowdsourcing es un término que se refiere a la colaboración de personas que contribuyen con otros para realizar y mejorar tareas que a uno solo le llevaría más tiempo y recursos [6].

Los datos proporcionados por usuarios como texto, videos, fotografías e incluso las coordenadas de los sitios extraídos de diferentes medios: redes sociales, blogs de internet, entre otros, son fuentes de VGI [9]. Como tal, este enfoque ofrece ventajas sustanciales ya que favorece la generación de datos sobre el entorno geográfico, sin embargo, por su naturaleza voluntaria, es susceptible a imprecisiones y errores; lo cual, sino es considerado, eventualmente generará inconsistencia en los datos y en los resultados de los análisis a realizar [6].

Cuando la VGI se realiza por medio de la descripción textual de eventos o entidades, pudiera tener algún tipo de argot, abreviaturas, acrónimos, además de que frecuentemente es limitada a un pequeño conjunto de caracteres [4]. Esto pudiera generar algunos errores en procesos como la geocodificación que consiste en la asignación de una ubicación geográfica a una descripción textual. En este sentido, Küçük y Avdan [1] describen que, debido a las incoherencias en los formatos de las direcciones, los errores ortográficos, el uso de abreviaturas, existe una baja precisión al momento de geocodificar la VGI, por lo que para atender estos casos se aplican correcciones al texto aplicando el procesamiento de lenguaje natural y métodos como el algoritmo de distancia Levenshtein.

The authors are with the Instituto Politécnico Nacional (IPN), Centro de Investigación en Computación (CIC), México {ejuarezc2023, marcomoreno, amagdasaldana}@cic.ipn.mx).

Manuscript received 01/05/2023, accepted for publication on 12/08/2024.

Es importante destacar que, en el contexto geográfico, se encuentran implícitos los aspectos geométricos y topológicos, ya que estos van de la mano de la ubicación, la forma y la manera en que se relacionan las entidades en el espacio geográfico. Como se ha comentado, el proceso de transformación de la VGI representada en formato de texto en datos geográficos se conoce como geocodificación [4].

Este proceso permite obtener explícitamente la ubicación de los eventos y los objetos geográficos, lo cual permite que se generen datos que eventualmente pueden utilizarse para la prevención de desastres, delincuencia, seguimiento de enfermedades contagiosas, y otras temáticas relacionadas al entorno geográfico. Las imprecisiones que causan discrepancias en la exactitud al momento de geocodificar, han sido comentados por Küçük y Avdan [1] y Escamilla [4], en donde se utilizó la limpieza a los datos aplicando algunas técnicas de procesamiento de lenguaje natural, minería de datos, algunos algoritmos como la distancia Levenshtein, métricas de evaluación de datos integrados entre otros.

Este artículo aborda el tema de las imprecisiones de la información VGI, y se propone de manera preliminar una alternativa para refinarla. Para lo cual, se presenta una metodología y la propuesta de un prototipo para automatizar los procesos relacionados. El resto del documento se compone de 4 secciones distribuidas de la siguiente manera; en estado del arte, se presenta al lector lo relacionado a trabajos previos para el desarrollo del trabajo, por otra parte, metodología, muestra los pasos a seguir para lograr lo cometido, mientras que en resultados preliminares, se describe una propuesta de implementación y por último, en conclusiones, se describe el aporte que tendrá este desarrollo, las ventajas, algunas dificultades que pueda presentar y lo que sigue a futuro para esta propuesta.

II. ESTADO DEL ARTE

Algunos trabajos relacionados con el problema de imprecisiones se describen a continuación. Se debe destacar que la valoración de imprecisiones en la VGI ha suscitado un gran debate en la comunidad geográfica, por lo que se han desarrollado diversos estudios. Como lo plantean Küçük y Avdan [1] el problema de las imprecisiones de la información se puede abordar mediante un proceso de normalización.

Para ello emplearon un método que descompone la información utilizada como datos de entrada en la Geocodificación, identificando las faltas de ortografías y las abreviaturas y reorganiza la información mediante el Proceso de Lenguaje Natural (NLP).

Por su parte, en Jackson et. al [2], describe la evaluación de la calidad y precisión de las contribuciones de la Información Geográfica Voluntaria (VGI), la propuesta consistió en un método para cuantificar la exhaustividad y precisión de un subconjunto de datos geográficos asociados a puntos de infraestructuras aportadas voluntariamente. Sus resultados mostraron las ventajas de incluir el control de la calidad en el proceso de recolección de datos, identificando que los errores en los datos están distribuidos en forma heterogénea.

En Saldaña et al. [3] se describe como la VGI representada por tuits (ahora X) son una fuente de datos geográficos para el análisis espacial. En dicho trabajo se comenta que este tipo de

enfoque presenta imprecisiones en los mismos, debido a su naturaleza, asociada con su forma de generarlos por parte de los usuarios de manera voluntaria y colaborativa. Por lo cual, se tiene la necesidad de realizar un preprocesamiento de minería de datos, procesamiento de lenguaje natural, clasificación, identificación de patrones y elementos entre otros [10], para que puedan ser bien aprovechados.

Troncoso y Fernández [12] aplicaron una geocodificación a grandes bases de datos utilizando minería de texto, describiendo que las direcciones son escritas generalmente por usuarios, provocando problemas de precisión en el registro, como errores ortográficos, datos innecesarios o falta de datos mínimos. Para resolver estos problemas aplican una metodología para la limpieza y corrección de las direcciones optimizando el proceso de geocodificación, para esto aplican el proceso Knowledge Discovery in Text (KDT), Procesamiento de Lenguaje Natural (NLP) y el algoritmo de Levenshtein. Demostrando en su trabajo un aumento de geocodificaciones.

Lee et al. [13] describen el mejoramiento de la geocodificación basada en calles utilizando técnicas de Machine Learning. En dicho trabajo, reportan los problemas encontrados en la información como abreviaturas, faltas de ortografía, jergas entre otros. Por lo cual, se prueban tres métodos diferentes de machine learning, con el fin de encontrar el método más preciso, encontrando que, el refuerzo de gradiente fue el que obtuvo la mayor precisión, concluyendo que el mejor método para ser aplicado para la geocodificación es el refuerzo de gradiente.

Con respecto al ajuste de datos geoespacial, en Montejano et al. [5] describen una estandarización de las áreas geoestadísticas básicas urbanas, para resolver un problema de ajuste geoespacial en el aspecto topológico, donde encontraron múltiples inconsistencias en la cartografía censal utilizada, apreciando una rotación, desplazamiento, y en algunos casos reducción en el tamaño de los polígonos urbanos.

Además, se describen algunos métodos para el ajuste espacial como transformación afín, transformación de similitud, transformación proyectiva, media cuadrática y residual, transformaciones a ciegas [7]. Estos métodos presentados se verificarán para que posteriormente, se integren como herramientas de apoyo para el proceso de ajuste geoespacial con base a lo que describió el autor. Dicho esto, en el siguiente apartado se describen de forma general cada método utilizado.

- Transformaciones a fin: puede escalar los datos diferencialmente, segarlos, rotarlos y traducirlos.
- Transformación de similitud: Escala rota y traduce los datos.
- Transformación proyectiva: se basa en una fórmula más compleja que requiere un mínimo de cuatro vínculos de desplazamiento.
- Media cuadrática y residual: los parámetros de transformación son un mejor ajuste entre los puntos de control de origen y destino.
- Transformación a ciegas: Es posible configurar la transformación creando vínculos de desplazamiento con el extremo de origen en ubicaciones.

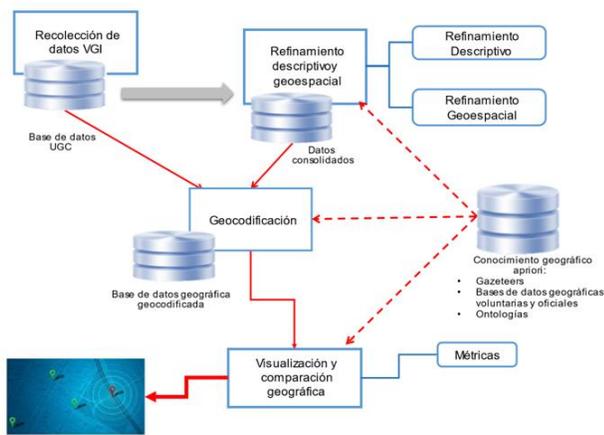


Fig. 1. Etapas de la metodología propuesta

Algunos trabajos relacionados con esta investigación, describieron los problemas que conlleva el uso de información proveniente de usuarios sin conocimiento alguno del espacio geográfico, planteando diferentes procesos para ajustar este tipo de información.

Con base en lo revisado, en el siguiente apartado se describirá la propuesta metodológica para lograr solventar estos problemas de imprecisiones con la VGI.

III. METODOLOGÍA PROPUESTA

La metodología propuesta, a comparación de los trabajos relacionados a esta investigación, tienen como objetivo de obtener información de buena calidad para realizar geocodificaciones con mayor exactitud y así, obtener información viable para ser representada en un producto final. Sin embargo, los trabajos mencionados solo describen el proceso de refinamiento descriptivo donde aplican técnicas de PLN, Machine learning, algoritmos de distancia como Levenshtein, minería de datos entre otros, sin tomar en cuenta el refinamiento geoespacial donde se aplican las relaciones topológicas para categorizar los datos en “error”, “imprecisión” y “vaguedad”.

Esta propuesta metodológica plantea el refinamiento geoespacial a la VGI, sin dejar fuera la descriptiva para tener mejor exactitud al momento de geocodificar. A continuación, se describe la metodología propuesta más íntegramente para lograr entender la finalidad de este trabajo.

La metodología considera la integración de diversas fuentes de VGI, las cuales son almacenadas en una base de datos denominada BD de UGC, posteriormente sobre estos datos se realiza un refinamiento, que consiste en identificar y corregir las imprecisiones descriptivas y espaciales, para lo cual se hace uso de conocimiento a priori sobre el dominio geográfico, que incluye una base de datos geográfica, gazetter y ontologías.

Los datos refinados se almacenan en una base de datos consolidada. Una vez terminado esto, los datos de BS UGC y BD Consolidada son geocodificados, para posteriormente compararlos con base en métricas y cuantificar el efecto de la refinación. Por lo anterior, la metodología se compone de las siguientes etapas: recolección de datos VGI, refinamiento descriptivo y geoespacial, geocodificación y, por último,

visualización y comparación geográfica (ver Fig. 1). A continuación, se describen las etapas de la metodología.

Recolección de datos VGI (Base de datos UGC): En esta etapa se recolectarán los datos proporcionados por usuarios de fuentes de internet de situación delictivas, violencia de género, asaltos, etc. Generando la Base de datos de Contenido Generado por Usuarios (BD-UGD). Durante esta etapa se seleccionan las fuentes de datos, en el caso de X, son elegidas algunas cuentas que pueden proporcionar información relevante, ya sea por ser organizaciones dedicadas a la temática de interés o bien usuarios con buena reputación, posteriormente al seleccionar las fuentes confiables se almacenan en diccionarios para un buen manejo al momento refinar los textos [3].

Refinamiento descriptivo y geoespacial. En esta etapa se aplicará el refinamiento a la base de datos llamada BD-UGC, esto se realizará tanto a nivel descriptivo como geoespacial. Para realizarlo, se requiere la definición de una clasificación de imperfecciones, que nos permita guiar esta etapa.

En el refinamiento Descriptivo, los textos se clasifican en un catálogo de imprecisiones como: faltas de ortografía, abreviaturas, tipos de argot, acrónimos, hashtag y algún enriquecimiento en el texto. Para que posteriormente sirvan como normalizadores de las fuentes de información recolectadas. Con base en este catálogo, las imprecisiones en el texto se identifican aplicando procesamiento de lenguaje natural (PLN), permitiendo así la estandarización cambiando el texto original, por el que está en el catálogo. Por ejemplo, “Av. Montebideo” se sustituye por “Avenida Montevideo” [4].

Para lograr este proceso se hace uso de la biblioteca Natural Language Toolkit (NLTK), ya que mediante sus herramientas para el procesamiento de texto como el manejo de n-gramas, bolsas de palabras, entre otras permite procesar el texto original a texto normalizado para su uso posterior. El objetivo principal del uso del PLN en el procesamiento de los textos se deriva de la necesidad de contar con estructuras textuales claras que permitan la identificación de palabras relacionadas a lugares geográficos específicos para la posterior geolocalización. Como primera instancia se propone el uso del algoritmo de Naïve-Bayes, para identificar términos de interés como son nombres de vialidades, calles y puntos de interés, posteriormente se propone probar el conjunto de datos con un algoritmo de máquina de soporte vectorial para comprobar su desempeño.

Otro caso de refinamiento, es el Geoespacial donde se representan entidades geográficas, de manera que se deben verificar las asociaciones con los nombres de calles, negocios, entre otros. Para este caso de refinamiento geoespacial, se clasifican las imprecisiones como, “error”: cuando un dato proporcionado no existe en el entorno geoespacial, “imprecisión”: cuando un dato proporcionado si existe en el espacio geográfico, pero con una referencia decadente y “vaguedad”: cuando un dato tiene falta de claridad para proporcionar una ubicación.

Para identificar las imprecisiones se aplicarán algunas propiedades espaciales representadas en la VGI, por ejemplo, se comparan las relaciones topológicas entre la información, teniendo en cuenta lo siguiente; la adyacencia (entre polígonos), contigüidad (entre línea y polígonos), pertenencia (arcos a polígonos), conectividad (entre arcos, en redes), inclusión (punto en polígono, líneas en polígono y polígono entre polígono). Con base a la clasificación que tenga el texto se

estaría comparando las relaciones topológicas para realizar el refinamiento apoyándose del conocimiento geográfico a priori, base de datos geográfica, gazetters y ontologías.

Geocodificación: En esta etapa se geocodifican las dos bases de datos (BD UGD y BD-Consolidados), utilizando servicios para geocodificación como: QGIS, GeoPy, Google [1,9]. Posteriormente se realiza una comparación con el conocimiento geográfico a priori para realizar comparaciones y determinar la precisión de cada base de datos geocodificados y evaluar el efecto del refinamiento sobre los datos.

Visualización y comparación geográfica: En esta etapa se realiza la visualización en mapas para comparar las dos geocodificaciones y determinar la precisión que se obtiene. Así mismo se aplican las métricas que permiten evaluar diversos aspectos de los datos como la exactitud y la distancia.

Para evaluar la exactitud de las dos bases de datos se aplica el método NMAS, (National Map Accuracy Standard) se trata de un método de control exactitud posicional que establece de una manera muy sencilla una regla de aceptación/rechazo, el NMAS se compone por varios parámetros, de los que toma en cuenta para evaluar, que son: método de comparación, componente posicional, clases de elementos, estándar de exactitud, procedimiento y la fuente [11].

En este apartado se describió la propuesta metodológica planteando algunas formas de realizar el refinamiento descriptivo y geoespacial tomando en cuenta lo revisado del estado de arte, así mismo los problemas que llegan encontrar en la VGI y los métodos que emplean para corregirla. En el siguiente apartado se presentan algunos preliminares que tiene este trabajo de investigación, presentando una estructuración de la metodología en una herramienta denominada “prototipo”.

IV. PRELIMINARES

Con base a la descripción de los autores mencionados anteriormente y la metodología propuesta, se propone un prototipo, para automatizar las tareas relacionadas al refinamiento descriptivo y geoespacial proveniente de la Información Geográfica Voluntaria. El prototipo permitirá la refinación de un texto proveniente de la VGI, como ejemplo: el tweet de un usuario, delincuente desahogado a una usuaria en Av. “Montebideo”, Esq. “Uancayo” un equipo de teléfono móvil, se puede observar en el tweet, contiene faltas de ortografía y abreviaturas, la mal escritura pudo influir el evento ocurrido, generando adrenalina y una baja capacidad de concentración para escribir, para realizar la prueba del prototipo se ingresan los datos para proceder el refinamiento descriptivo y con esto obtener la escritura refinada “Avenida Montevideo” “Esquina Huancayo”.

Otro caso relacionado con el refinamiento geoespacial, cuando usuarios proporcionen información que no existe en el espacio geográfico o que compartan datos ambiguos (con doble referencia), por ejemplo: en esquina Montevideo con Huancayo enfrente del edificio gris no. 502 se suscitó un percance automovilístico.

En este sentido, el tweet se vuelve ambiguo debido a que existen dos referencias, aquí se aplicaría un refinamiento geoespacial caracterizando una relación topológica, que se cumple entre la esquina de Montevideo y Huancayo,

categorizando la parte final del texto por tener más posibilidad de error de ubicación.

Otro ejemplo de corrección, son los mensajes con alto grado de vaguedad, como ejemplo la ubicación de un lugar, la escuela primaria José Martí se encuentra en el estado de Guerrero, expresarlo así es no es muy exacto, sin embargo, expresarla como, la escuela primaria José Martí se encuentra en Chilpancingo de los Bravo, es más preciso, aunque si se utiliza “está en la colonia Dirección Nacional de Caminos”, lo vuelve más preciso para describir la ubicación.

Estos son tipos de imprecisiones, que también pueden generar errores al expresarse, como ejemplo: la escuela José Martí se encuentra en el estado de Puebla, si se trata de ubicar, se estaría generando un error de ubicación al comparar con el conocimiento geográfico a priori.

En este apartado se describió el proceso que se seguirá al ejecutar el prototipo mostrando algunos ejemplos, de lo que corregirá al momento de ser ejecutado. En el siguiente apartado se describe de manera general como está estructurada la interfaz gráfica para que el usuario pueda ejecutarla, estructurándose de la siguiente manera: conector de base datos, tipo de refinamiento descriptivo, refinamiento geoespacial (corrección topológica), geocodificar, visualización en el mapa base, aplicar métricas. Cabe recalcar que el prototipo presentado es solo una propuesta a futuro de lo que podría resultar al integrarse la metodología.

A. Estructura de la interfaz gráfica

Conector de base de datos: El usuario en este apartado tendrá la opción de ingresar su base de datos o lo podrá hacer desde un gestor de datos como PostGIS, para comenzar el proceso de refinación tendrá que asegurar que su información venga de fuentes confiables para no tener algún error al momento de procesarla.

Refinamiento descriptivo (corrección de texto): En este apartado tendrá la opción de categorizar su información con base al catálogo previamente formado y comenzar el proceso de refinación, cabe mencionar que si no se categoriza primero no tendrá ningún valor el proceso ya que no podrá normalizar los datos. Dentro de la interfaz se pondrá un mensaje de advertencia antes de realizar el proceso.

Refinamiento geoespacial (corrección topológica): En este apartado se realizará el refinamiento geoespacial aplicando las relaciones topológicas, con base a lo siguiente: la adyacencia (entre polígonos), contigüidad (entre línea y polígonos), pertenencia (arcos a polígonos), conectividad (entre arcos, en redes), inclusión (punto en polígono, líneas en polígono y polígono entre polígono) y a la categorización del catálogo de imperfecciones, como error, imprecisión y vaguedad.

Geocodificar: En este apartado el usuario aplicará la geocodificación de manera general a la base de datos y no por lotes o uno por uno, debido a esto el proceso se automatiza para reducir el tiempo de geocodificar.

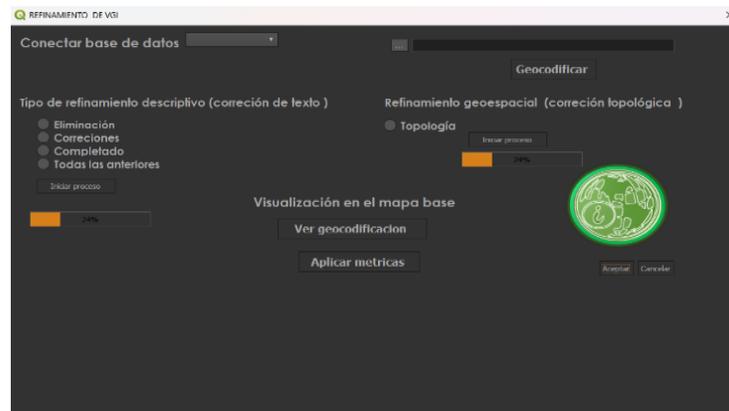


Fig. 2. Interfaz gráfica del prototipo, automatización para el refinamiento de información geográfica voluntaria

Visualización en el mapa base: En este apartado el usuario podrá visualizar las BS UGC y BD Consolidada en conjunto con el conocimiento geográfico a priori (ver fig. 2).

Con base en lo descrito, la interfaz gráfica contiene un ambiente amigable de forma que el usuario pueda hacer uso de ella sin ningún problema, disponiendo de todas las herramientas usadas para el refinamiento de la VGI. En el siguiente apartado se presentan las conclusiones, donde se comentan algunos pros y contras que podría tener esta propuesta, así mismo algún mejoramiento a futuro y algunas ventajas y desventajas que tiene con herramientas comerciales dedicadas al proceso de limpieza de datos.

V. CONCLUSIONES

La calidad de los datos siempre ha sido un tema recurrente, debido al factor de impacto que tiene en las investigaciones, donde realizan procesos de geocodificación para el análisis de un evento que afecta en la sociedad, por lo regular, la información se obtiene de fuentes de internet utilizando el contenido generado por usuarios o VGI, al ser una generación por personas que no tienen conocimiento geográfico se llegan a encontrar imprecisiones, optando por aplicar algunas técnicas para mejorar la precisión al momento de geocodificar.

Con la propuesta metodológica se pretende solventar algunos tipos de problemas que llegan afectar la precisión al momento de asignar coordenadas a los datos, teniendo en cuenta el refinamiento descriptivo y geoespacial.

La propuesta esperamos tendrá un impacto en la comunidad relacionada con el desarrollo de aplicaciones geoespaciales, ciudades inteligentes y en general usuarios de información geográfica, contará con información más precisa.

Es importante comentar que una de las ventajas que puede tener el desarrollo, es que será integrado en un ambiente Open Source y tendrá la capacidad de ajustar de manera geoespacial los datos para tener una mayor certeza que lo que se está geocodificando es en verdad información verídica, otra ventaja es que trabajará con VGI a comparación con las comerciales, estas solo trabajan con direcciones postales en un formato.

Por ser una herramienta desarrollada internamente se podrá adaptar a más problemas en específico en un futuro, siendo el objetivo para abordar más temas relacionados con las imprecisiones. Se podrían presentar algunas dificultades con el tipo de información que sea agregada para ajustar, debido a que al momento de capturar una base de datos de actos delictivos y

una base de datos de tráfico vehicular, la forma de los datos es poco diferente y no sería la misma aplicación, este podría ser un problema que se pretende resolver a futuro para que el prototipo reciba cualquier tipo de información.

REFERENCIAS

- [1]. D. Küçük, U. Avdan, "Address standardization using the natural language process for improving geocoding results," *Computers, Environment and Urban System*, Vol. 70, pp. 1–8 (2018). DOI: 10.1016/j.compenvurbsys.2018.01.009.
- [2]. S. P. Jackson, W. Mullen, P. Agouris, A. Crooks, Coritoru, A. Stefanidis, "Assessing Completeness and Spatial Error of Features in Volunteered Geographic information," *ISPRS Int. J. Geo-Inf.*, Vol. 2, pp. 507–530 (2013). DOI: 10.3390/ijgi2020507.
- [3]. M. Saldaña, C. Cavalière, M. Torres, M. Moreno, "When Twitter Becomes a Data Source for Geospatial Analysis," *Research in Computing Science*, Vol. 148, No. 10, pp. 357–374 (2019). DOI: 10.13053/rics-148-10-30.
- [4]. I. Escamilla, M. Moreno, M. Torres, R. Quintero, "Geocoding Tweets Base on Semantic Web and Ontologies," in *International Journal on Semantic Web and Information Systems*, Vol. 12, No. 1, pp. 44–61 (2016). DOI: 10.4018/978-1-5225-5042-6.ch014.
- [5]. J. Montejano, G. Ávila, C. Caudillo, "Estandarización de las áreas geoestadísticas básicas urbanas para 100 ciudades del sistema Urbano Nacional," *Realidad datos y espacio revista internacional de estadística y geografía*, Vol. 11, No. 3, pp. 62–89 (2020).
- [6]. M. Goodchild, L. Li, "Assuring the quality of volunteered geographic information," *Spatial statistics*, Vol. 1, pp. 110–120 (2012).
- [7]. ArcGIS Desktop, <https://desktop.arcgis.com/es/arcmap/latest/manage-data/editing-existing-features/about-spatial-adjustment.htm>, (2023).
- [8]. B. Haworth, "Emergency management perspectives on volunteered geographic information: Opportunities, challenges and change," *Computers, Environment and Urban Systems*, Vol. 57, pp. 189–198 (2016). DOI: 10.1016/j.compenvurbsys.2016.02.009.
- [9]. A.M.M. Saldaña, "Análisis de la dinámica urbana con información geográfica voluntaria utilizando técnicas espacio-temporales". Repositorio Dspace (2018).
- [10]. B. Suh, L. Hong, P. Pirolli, E.H. Chi, "Want to be retweeted? large scale analytics on factors impacting retweet in twitter network," in *Social computing, 2010 IEEE second international conference on*, pp. 177–184, (2010). DOI: 10.1109/SocialCom.2010.33.
- [11]. F. Ariza, J. Rodríguez, J. García, C. Roblado, "Guía para la evaluación de la exactitud posicional de datos espaciales",

Instituto Panamericano de Geografía e Historia, pp. 2–60, (2019).

- [12].F. Troncoso, N. Fernández, “Limpieza, corrección y geocodificación de grandes bases de direcciones utilizando minería de texto,” *Universidad, Ciencia y tecnología*, Vol. 25, pp. 80–87 (2021).
- [13].K. Lee, A. Claridades, J. Lee, “Improving a Street-Based Geocoding Algorithm Using Machine Learning Techniques,” *Applied sciences*, Vol. 10, pp. 1–21 (2020).