

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 989 281**

51 Int. Cl.:

H04L 51/222 (2012.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **18.10.2017 PCT/US2017/057216**

87 Fecha y número de publicación internacional: **17.05.2018 WO18089178**

96 Fecha de presentación y número de la solicitud europea: **18.10.2017 E 17869764 (5)**

97 Fecha y número de publicación de la concesión europea: **07.08.2024 EP 3549037**

54 Título: **Sistema y método para detectar geolocalizaciones en redes sociales**

30 Prioridad:

09.11.2016 US 201662419609 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
25.11.2024

73 Titular/es:

**THOMSON REUTERS ENTERPRISE CENTRE
GMBH (100.0%)
Landis + Gyr-Strasse 3
6300 Zug, CH**

72 Inventor/es:

**NOURBAKSH, ARMINEH y
SHAH, SAMEENA**

74 Agente/Representante:

GONZÁLEZ PECES, Gustavo Adolfo

ES 2 989 281 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Sistema y método para detectar geolocalizaciones en redes sociales

5 **REFERENCIA CRUZADA A SOLICITUDES RELACIONADAS**

La presente solicitud reivindica la prioridad con respecto a la solicitud de patente provisional de Estados Unidos n.º 62/419.609 presentada el 9 de noviembre de 2016, y la solicitud de patente de Estados Unidos n.º 15/787.416, presentada el 18 de octubre de 2017. La presente solicitud también se refiere a la solicitud de patente de Estados Unidos n.º 15/143.730, presentada el 2 de mayo de 2016.

INFORMACIÓN DE ANTECEDENTES

La detección de geolocalización a partir del texto es una tarea difícil. La detección de geolocalizaciones a partir de datos de contenido social se complica aún más por la prominencia de los hashtags, jerga específica de la plataforma, falta de puntuación, capitalización y gramática correcta. Algunos de los principales retos a la hora de identificar con precisión ubicaciones de publicaciones en redes sociales incluyen las siguientes:

1) **Falta de normas o heurísticas adecuadas:** No existen estrategias definitivas para identificar lugares en el texto, ya que pueden expresarse de diversas formas.

2) **Palabras ambiguas:** Hay muchas palabras ambiguas, por ejemplo nombres de lugares que también pueden ser nombres de personas.

3) **Falta de gramática estándar:** Muchos usuarios de redes sociales utilizan un lenguaje informal y algo subestándar en sus mensajes, y muchas redes sociales tienen su propia jerga. Esto significa que los modelos que han sido entrenados en inglés estándar no pueden funcionar bien con datos sociales.

4) **Abundancia de hashtags:** Los hashtags se utilizan en muchas plataformas sociales para indicar metadatos relacionados con un mensaje, por ejemplo, su tema. Con los años de uso en redes sociales, los hashtags han cobrado vida propia, intercalando o sucediendo a un mensaje con tokens ingeniosos o creativos. En muchas ocasiones los usuarios mezclan más de una palabra para hacer un hashtag compuesto o expresan la ubicación de un evento mediante un hashtag de seguimiento. En estos casos, los analizadores sintácticos automáticos no son capaces de descomponer los hashtags correctamente.

5) **Coherencia de las ubicaciones de los usuarios autoidentificados:** A menudo, los usuarios pueden optar por identificar su ubicación en su perfil. En muchas plataformas de redes sociales, esta ubicación no tiene que validarse y puede expresarse como texto libre. Esto ha conducido al inevitable protagonismo de lugares creativos pero inviábiles.

6) **Granularidad de la información:** Algunos equipos de respuesta ante catástrofes, y departamentos de policía y bomberos crearon cuentas oficiales en redes sociales para informar de las emergencias en tiempo real. Los lugares que identifican en sus mensajes suelen ser específicos de su ubicación. Por ejemplo, "Se informa de accidente con heridos en la autopista 183 NB a la altura de la calle Loyola. Retrocedan hasta MLK" incluye una descripción granular de la dirección de un accidente, que puede resultar difícil de analizar. Además, la dirección puede ser difícil de localizar, ya que en muchas ciudades diferentes puede existir una dirección o intersección similar.

7) **Identificar las coordenadas geográficas correctas:** Incluso si las palabras que se refieren a lugares se identifican con precisión, a veces pueden asignarse a varias coordenadas geográficas. Por ejemplo, hay varias ciudades llamadas "Orlando" en Estados Unidos (por ejemplo, en Florida, Oklahoma, Virginia Occidental, Nueva York, Virginia, Kentucky, Carolina del Norte y Arkansas).

8) **Identificar el lugar principal de un acontecimiento:** Por ejemplo, el mensaje "Grupos rebeldes apoyados por Turquía y EE.UU. se enfrentan en Siria con un grupo kurdo respaldado por EE.UU." menciona a tres países. Puede ser importante saber en qué lugar tuvo lugar el suceso (es decir, Siria).

9) **Requisitos de escala de tiempo y sostenibilidad:** Aunque los modelos de aprendizaje automático puedan arrojar buenas cifras de precisión/recuperación, a menudo son demasiado lentos para ser aplicables en tiempo real. Además, como muchos de estos modelos se entrenan con datos de entrenamiento estáticos, requerirán actualizaciones y ajustes periódicos.

La publicación US 2013/325975 A1 se refiere a la estimación de la ubicación de usuarios de redes sociales. En una realización, se recibe una pluralidad de mensajes de redes sociales generados por un usuario determinado. Se extrae una pluralidad de características de localización de los mensajes de las redes sociales. Cada una de las características de localización se procesa con al menos un clasificador de un conjunto de clasificadores. Cada uno de los clasificadores genera una clasificación de ubicación para cada una de los mensajes en redes sociales. Cada

clasificación comprende una ubicación y una ponderación asociada a esa ubicación. A partir de las clasificaciones de ubicación se selecciona una de las ubicaciones como la ubicación del usuario dado basándose en una combinación de las ponderaciones de las clasificaciones de ubicación.

5 El documento US 2014/122604 A1 proporciona una técnica para estimar una ubicación relativa a un usuario que no ha proporcionado información sobre su ubicación en un campo del perfil en redes sociales como un microblog. Un método para estimar la asociación entre un usuario de redes sociales y una ubicación incluye los pasos de adquirir un primer contenido publicado en redes sociales por un primer usuario asociado a una primera ubicación, determinar la localización regional del primer contenido en función de la primera ubicación, adquirir un segundo contenido publicado en redes sociales por un segundo usuario no asociado a una ubicación, determinar el grado de relación entre el primer contenido y el segundo contenido, y asociar la primera ubicación con el segundo usuario sobre la base de la localización y el grado de la relación.

15 El documento US 2010/312769 A1 describe métodos, sistemas y software de análisis de mensajes en microblogs para detectar actividades anómalas de interés. El sistema incluye un agrupador para agrupar mensajes en microblogs recibidos durante un primer periodo de tiempo, un clasificador para puntuar los mensajes agrupados; una base de conocimientos, un generador de reglas para generar reglas de clasificación a partir de la base de conocimientos; y un cotejador para cotejar los mensajes puntuados con las solicitudes de información. Se describen los métodos de funcionamiento del sistema y sus componentes.

20 Por lo tanto, se desea un sistema que aborde todos los retos anteriores y proporcione un modelo validado frente a otros servicios de geolocalización.

25 De acuerdo con la invención, se proporciona un método como se reivindica en la reivindicación 1, un medio de almacenamiento como se reivindica en la reivindicación 15 y un sistema como se reivindica en la reivindicación 16.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

30 Para que puedan comprenderse las características de la presente invención, a continuación se describen varios dibujos. Sin embargo, los dibujos adjuntos ilustran únicamente realizaciones particulares de la invención y, por tanto, no deben considerarse una limitación de su ámbito de aplicación, ya que la invención puede abarcar otras realizaciones igualmente eficaces.

35 La Fig. 1 es un diagrama esquemático que representa una realización de un sistema para detectar geolocalizaciones en sistemas de publicaciones en redes sociales, como microblogs, de acuerdo con una realización de la divulgación.

La Fig. 2 es un diagrama de flujo que representa una realización de un método para detectar geolocalizaciones en publicaciones en sistemas de redes sociales tales como microblogs de acuerdo con una realización de la divulgación.

40 La Fig. 3 es un diagrama de flujo que representa una realización de un ejemplo de arquitectura del sistema para detectar geolocalizaciones en publicaciones en sistemas de redes sociales tales como microblogs de acuerdo con una realización de la divulgación.

La Fig. 4(a)-4(d) muestra ejemplos de publicaciones en redes sociales con geolocalizaciones que pueden detectarse de acuerdo con una realización de la divulgación.

45 La Fig. 5 es un diagrama de flujo que representa una realización de un método de preprocesamiento de acuerdo con una realización de la divulgación.

La Fig. 6 es un diagrama de flujo que representa una realización de un método de identificación de la ubicación de acuerdo con una realización de la divulgación.

50 La Fig. 7 es un diagrama de flujo que representa una realización de un método para identificar ubicaciones a partir del texto de una publicación en redes sociales de acuerdo con una realización de la divulgación.

La Fig. 8 es un diagrama de flujo que representa una realización de un método que implementa un enfoque basado en taxonomía de acuerdo con una realización de la divulgación.

La Fig. 9 es un diagrama de flujo que representa una realización de un método que implementa un enfoque basado en la heurística de acuerdo con una realización de la divulgación.

55 La Fig. 10 es un diagrama de flujo que representa una realización de un método que implementa un enfoque basado en base de conocimientos (KB) de acuerdo con una realización de la divulgación.

La Fig. 11 es un diagrama de flujo que representa una realización de un método para determinar coordenadas geográficas de ubicación de acuerdo con una realización de la divulgación.

60 La Fig. 12(a)-12(b) muestra un ejemplo de salida de un ejemplo de biblioteca de ubicaciones de acuerdo con una realización de la divulgación.

La Fig. 13 es un diagrama de flujo que representa una realización de un método para calificar ubicaciones de acuerdo con una realización de la divulgación.

La Fig. 14 es un diagrama de flujo que representa una realización de un método para calificar ubicaciones de una biblioteca de ubicaciones de acuerdo con una realización de la divulgación.

65 La Fig. 15 es un diagrama de flujo que representa una realización de un método para calificar ubicaciones usando heurística comunitaria de acuerdo con una realización de la divulgación.

La Fig. 16 es un diagrama de flujo que representa una realización de un método para determinar coordenadas geográficas de ubicaciones primarias de acuerdo con una realización de la divulgación.

DESCRIPCIÓN DETALLADA DE REALIZACIONES EJEMPLARES

5 Se divulgan realizaciones de sistemas y métodos para detectar geolocalizaciones en publicaciones en sistemas de redes sociales tales como microblogs. En referencia a las figuras, la Fig. 1 muestra un diagrama esquemático que representa una realización de un sistema 100 para detectar geolocalizaciones en publicaciones en redes sociales. El sistema 100 incluye un sistema de redes sociales 104, un sistema de geolocalización 108, una aplicación 112, y un sistema de usuario 116.

15 El sistema de redes sociales 104 proporciona una plataforma para que sus usuarios envíen publicaciones y/o contenido a una red de otros usuarios que utilizan cuentas del sistema 100. El sistema de redes sociales 104 incluye un sistema servidor 120 de la red social que tiene una interfaz de comunicaciones 124. El sistema servidor 120 de la red social proporciona funcionalidad al sistema de redes sociales 120 a los usuarios y, como se analiza en el presente documento, con la interfaz de comunicaciones 124 que proporciona comunicacionales a través de una o más redes de comunicaciones 128 entre el sistema de redes sociales 120 y otros sistemas. En realizaciones, el sistema de redes sociales 120 puede adoptar varias formas diferentes. En un ejemplo, el sistema de redes sociales 120 puede ser Twitter, en la que los usuarios utilizan sus cuentas para, entre otras cosas, publicar publicaciones y/o contenidos breves, llamados Tweets, en el sistema. En otras realizaciones, el sistema de redes sociales 120 puede ser otro sistema, tales como uno o varios de Facebook, Instagram, Snapchat, Tumblr, Pinterest, Flickr o Reddit, etc.

25 El sistema de geolocalización 108 incluye un módulo identificador 132 de menciones de ubicación, un módulo determinador 136 de ubicaciones y un módulo calificador 140. El sistema de geolocalización 108 tiene una interfaz de comunicaciones 144 que interactúa con el sistema de redes sociales 120 para recuperar publicaciones en redes sociales y enviarlas al módulo identificador 132 de menciones de ubicación para identificar y/o detectar cualquier mención de una ubicación en las publicaciones o conjunto de publicaciones. El módulo identificador 132 de menciones de ubicación también puede extraer ubicaciones especificadas por los usuarios del sistema de redes sociales 120.

30 El módulo identificador 132 de menciones de ubicación tiene una base de datos 148 para almacenar diversas ubicaciones, rangos/puntuaciones de ubicación y coordenadas geográficas. El módulo determinador 136 de la ubicación encuentra la información de las coordenadas geográficas de latitud y longitud asociadas con las ubicaciones detectadas usando una biblioteca de ubicaciones/coordenadas geográficas. En algunas realizaciones, la biblioteca de ubicaciones es una biblioteca de terceros.

35 El módulo calificador 140 utiliza varios métodos para calificar y/o desambiguar las ubicaciones, encontrar las coordenadas geográficas correctas de cada ubicación y clasificar las ubicaciones en función de un nivel de relevancia para un evento. El módulo calificador 140 envía la información de ubicación a una interfaz de comunicaciones 152. En la realización mostrada, cada uno del módulo identificador 132 de menciones de ubicación, módulo determinador 136 de ubicaciones y módulo calificador 140 se pueden comunicar entre sí. En la realización mostrada, la interfaz de comunicaciones 152 del módulo geolocalizador 108 envía información de la ubicación en forma de metadatos en un objeto de publicación en la publicación en redes sociales usando una o más redes de comunicaciones 156 a la aplicación 112 para mostrarla en un sistema de usuario 116. En la realización mostrada, la aplicación 112 puede incluir una interfaz de programación de aplicaciones (API) 160. En la realización mostrada, la aplicación 112 se comunica con el sistema de usuario mediante una o varias redes de comunicaciones 164. Como alternativa, el sistema de geolocalización puede enviar directamente al sistema de usuario 116.

50 El sistema de usuario 116 puede ser cualquier plataforma informática, tales como uno o varios de un ordenador, un ordenador de sobremesa, un ordenador portátil, una tableta, un teléfono inteligente u otros dispositivos estacionarios o móviles, etc., que el usuario utiliza para comunicarse con otros sistemas mediante una o varias redes de comunicaciones 164.

55 En algunas realizaciones, el sistema 100 para detectar geolocalizaciones en diversas publicaciones en redes sociales puede incluir las características representadas o analizadas en el presente documento con respecto a la Fig. 1.

60 La Fig. 2 muestra un diagrama de flujo que representa una realización de un método 200 para detectar geolocalizaciones en publicaciones en diversos sistemas de redes sociales tales como microblogs de acuerdo con una realización de la divulgación. El método se puede llevar a cabo o puede implicar componentes del sistema 100 de la Fig. 1, tales como el sistema de geolocalización 108. El método empieza en el etapa 204.

65 En la etapa 208, se recuperan una o más publicaciones en redes sociales del sistema servidor 120 de la red social. En las realizaciones mostradas, el módulo de geolocalización 108 se comunica con el sistema servidor 120 de la red social a través de las interfaces de comunicaciones 124, 144 en la red 128 para recuperar una o varias publicaciones en redes sociales de una o varias cuentas de redes sociales.

En la etapa 212, se preprocesan las publicaciones en redes sociales recuperadas. En algunas realizaciones, el preprocesamiento se lleva a cabo en el sistema de geolocalización 108. En otras realizaciones, el preprocesamiento se puede llevar a cabo en un módulo de procesamiento independientes y a continuación las publicaciones en redes sociales preprocesadas se introducen en el sistema de geolocalización 108. El preprocesamiento implica una o más funciones realizadas por el sistema para limpiar y preparar las publicaciones antes de identificar las ubicaciones potenciales de las publicaciones, como se analiza en el presente documento. Por ejemplo, el preprocesamiento puede incluir cualquier combinación de los rasgos de los sistemas y métodos de las Figs. 3-5, o cualquier combinación de cualquier subconjunto y/u ordenación alternativa de las características de dicho sistema o métodos.

En la etapa 216, el sistema determina las menciones de ubicación para las publicaciones en redes sociales recibidas y preprocesadas. En las realizaciones mostradas, el módulo identificador 132 de menciones de ubicación determina las menciones de ubicación realizando una o más de identificar las menciones de ubicación a partir del texto de la publicaciones en redes sociales y/o a partir de la ubicación del usuario de la cuenta de una red social, como se analiza en el presente documento. Por ejemplo, determinar las menciones de ubicación puede incluir cualquier combinación de las características de los sistemas y métodos de las Figs. 6-10, o cualquier combinación de cualquier subconjunto y/u ordenación alternativa de las características de dicho sistema o métodos.

En la etapa 220, el sistema determina las ubicaciones basándose en las menciones de ubicación determinadas. En las realizaciones mostradas, el módulo determinador 136 de ubicaciones recibe las menciones de ubicación determinadas desde el módulo identificador 132 de menciones de ubicación y determina las posibles coordenadas geográficas (es decir, las coordenadas de latitud y longitud) de las ubicaciones geográficas correspondientes a las menciones de ubicación determinadas, como se analiza en el presente documento. Por ejemplo, determinar las ubicaciones basándose en las menciones de ubicación puede incluir cualquier combinación de las características de los sistemas y métodos de las Figs. 11-12, o cualquier combinación de cualquier subconjunto y/u ordenación alternativa de las características de dicho sistema o métodos.

En la etapa 224, el sistema determina una ubicación primaria de cada mención de ubicación basándose en las ubicaciones determinadas. En las realizaciones mostradas, el módulo calificador 140 recibe las ubicaciones geográficas determinadas del módulo determinador 136 de ubicaciones y determina qué ubicación es una ubicación primaria correspondiente a cada mención de ubicación en una publicación en redes sociales, como se analiza en el presente documento. Por ejemplo, determinar las ubicaciones basándose en las menciones de ubicación puede incluir cualquier combinación de las características de los sistemas y métodos de las Figs. 13-16, o cualquier combinación de cualquier subconjunto y/u ordenación alternativa de las características de dicho sistema o métodos.

En la etapa 228, el sistema almacena las coordenadas geográficas para la ubicación primaria determinada en un objeto de publicación en redes sociales. En las realizaciones mostradas, para cada mención de ubicación en una publicación en redes sociales, el sistema agrega las coordenadas geográficas de la ubicación primaria correspondiente a dicha mención de ubicación. En algunas realizaciones, las coordenadas geográficas se pueden agregar como uno o varios campos de metadatos al objeto de la publicación. En algunas realizaciones, las coordenadas geográficas de cada ubicación se pueden comunicar al módulo identificador 132 de menciones de ubicación para su almacenamiento en una base de datos 148 en un medio de almacenamiento legible por máquina no transitorio para futuros procesos de determinación de la ubicación en futuras publicaciones en redes sociales recuperadas.

En la etapa 232, los objetos de publicación en redes sociales que contienen las coordenadas geográficas sus menciones de ubicación pueden enviarse a cualquier aplicación posterior sin interrupciones y en tiempo real. En las realizaciones mostradas, los objetos de publicación se pueden enviar desde el módulo geolocalizador 108 a una API 160 de aplicación 112 mediante la red 156 y a continuación enviarse para su presentación en un sistema de usuario 116 mediante la red 164. Como alternativa, los objetos de publicación se pueden enviar directamente al sistema de usuario 116. El método finaliza en el etapa 236.

En algunas realizaciones, el método 200 para detectar geolocalizaciones en diversas publicaciones en redes sociales puede incluir las características representadas o analizadas en el presente documento con respecto a la Fig. 2.

La Fig. 3 muestra un diagrama de flujo que representa un ejemplo de realización del sistema de geolocalización 108 para detectar geolocalizaciones en publicaciones de diversos sistemas de redes sociales tales como microblogs más detalladamente de acuerdo con una realización de la divulgación. En la realización mostrada, el sistema 108 recupera una o varias publicaciones en redes sociales desde un flujo de datos 304 de contenido social. En algunas realizaciones, el flujo de datos 304 de contenido social se puede emitir desde el sistema servidor 120 de la red social. En la realización mostrada, el sistema puede recuperar publicaciones en redes sociales en un modo 308 de publicación única y/o en un modo 312 de publicación múltiple. En el modo 308 de publicación única, el sistema determina la información de geolocalización basándose en publicaciones individuales desde cuentas de usuario, como se analiza en el presente documento. En el modo 312 de publicación múltiple, el sistema determina información de geolocalización basándose en múltiples publicaciones recibidas desde múltiples cuentas de usuario. En cada modo, cada publicación en redes sociales puede incluir una publicación 316 en texto y/o una ubicación 320 del usuario. En

la realización mostrada, las publicaciones de texto 316 de las redes sociales son entradas del módulo identificador 132 de menciones de ubicación.

En la realización mostrada, las menciones de ubicación procedentes del texto de la publicación 316 de las publicaciones en redes sociales y las ubicaciones 320 del usuario de las publicaciones en redes sociales son entradas del módulo determinador 136 de ubicaciones. En la realización mostrada, las posibles coordenadas geográficas de ubicación de cada mención de ubicación son entradas del módulo calificador 140. En la realización mostrada, las coordenadas geográficas de cada ubicación primaria para cada ubicación mencionada en el texto de la publicación 316 y/o la ubicación 320 del usuario son salidas del módulo calificador 140 y se agregan al objeto de publicación en redes sociales recibida desde el flujo de datos 304 de contenido social.

En algunas realizaciones, el sistema 300 para detectar geolocalizaciones en diversas publicaciones en redes sociales puede incluir las características representadas o analizadas en el presente documento con respecto a la Fig. 3.

Las Figs. 4(a)-4(d) muestran ejemplos de publicaciones en redes sociales 400 con geolocalizaciones que pueden detectarse de acuerdo con una realización de la divulgación. En las realizaciones mostradas, se representan varios ejemplos de publicaciones en redes sociales que se pueden recibir desde el sistema servidor 120 de la red social como parte del flujo de datos 304 de contenido social. En cada publicación en redes sociales 400 mostrada, existe una geolocalización 404, 408, 412, 416 mencionada en el texto de la publicación. Cuando cada publicación en redes sociales 400 se analiza mediante el sistema divulgado, el objeto de la publicación se puede actualizar con las coordenadas geográficas de la ubicación primaria para cada mención de ubicación.

En algunas realizaciones, los ejemplos de publicaciones en redes sociales 400 con geolocalizaciones que se pueden detectar pueden incluir las características representadas o analizadas en el presente documento con respecto a las Figs. 4(a)-4(d).

La Fig. 5 muestra un diagrama de flujo que representa una realización de un método de preprocesamiento 500 de acuerdo con una realización de la divulgación. En algunas realizaciones, el método de preprocesamiento 500 se lleva a cabo como la etapa 212 del método 200. Antes de procesar la una o varias publicaciones en redes sociales para identificar posibles ubicaciones, el sistema puede realizar algunas etapas de preprocesamiento para limpiar y preparar la una o varias publicaciones. El método 500 empieza en la etapa 504. En la etapa 508, el sistema elimina los truncamientos del texto de la publicación. En plataformas de redes sociales como Twitter, las publicaciones truncadas son habituales en las aplicaciones automatizadas para compartir publicaciones. Por ejemplo, cuando una publicación supera el límite de 140 caracteres de Twitter, las aplicaciones de terceros eliminan automáticamente la parte final de la publicación. Esto puede causar problemas potenciales en el sistema. Por ejemplo, si se considera un tweet truncado que dice "Celebremos Nueva York...", no está claro si la palabra final se refiere a una ubicación tal como la "ciudad de Nueva York" o si lo que se pretendía era decir, "la revista New Yorker" Por lo tanto, en la etapa 508, el sistema ignora y/o elimina todos los símbolos de truncamiento y las palabras y frases truncadas del texto de la publicación.

Sin embargo, en algunas realizaciones, eliminar los truncamientos no es tan fácil como solo identificar las publicaciones que tengan un símbolo de truncamiento (es decir, una elipsis representada por "...") al final de la publicación. O, en muchas ocasiones, las aplicaciones automatizadas pueden añadir hashtags, menciones o URL adicionales al final de la publicación (por ejemplo, "Celebremos Nueva York... con @Awesome App"). Esto suele ir en detrimento de la longitud de la publicación original. Además, no todas las publicaciones que terminan en elipsis están truncadas; a veces, los usuarios utilizan el símbolo simplemente como modo de expresión. Para abordar estas dificultades, el sistema, en la etapa 508, puede utilizar dos pistas principales para determinar si un tweet está truncado: 1) si la longitud de la publicación está cerca del límite de caracteres, y, 2) si la publicación finaliza con uno cualquiera de los dos símbolos de truncamiento principales (es decir "...") y el carácter Unicode de la elipsis horizontal) potencialmente seguido por una expresión estándar de truncamiento por ejemplo, "mediante @handle" o "mediante #hashtag"). En realizaciones, si una publicación cumple los criterios anteriores, la última palabra o token antes del símbolo de truncamiento se elimina, así como el resto del final de la publicación.

En la etapa 512, el sistema divide los hashtags que aparecen en el texto de la publicación de la publicación en redes sociales. En las publicaciones en redes sociales, los hashtags pueden desempeñar un papel importante en la identificación de ubicaciones, especialmente cuando no se dispone de otras pistas. Muchas cuentas oficiales fiables (por ejemplo, de equipos de respuesta a catástrofes, canales meteorológicos, monitores de tráfico, etc.) utilizan frecuentemente hashtags para transmitir información sobre su ubicación (por ejemplo, "#BuelahHillFire"). En muchos casos, estas cuentas oficiales suelen tener cuidado de utilizar diferentes tipos de letra en sus publicaciones para denotar las ubicaciones dentro de los hashtags. Para utilizar la información del hashtag, el sistema inspecciona cada publicación para determinar si está escrita TODA EN MAYÚSCULA (es decir, escrita usando todas las letras en mayúsculas). Si no se usa TODO EN MAYÚSCULAS, el sistema divide cada hashtag basándose en la ubicación de sus letras mayúsculas. Por ejemplo, el hashtag #BuelahHillFire se puede descomponer en el texto "#Buelah Hill Fire". Si aparecen varias letras mayúsculas una junto a otra, el sistema vuelve a unir las letras que hayan quedado sueltas. Por ejemplo, el hashtag "#LAFlood" de descompondrá en "#L A Flood," y, a continuación, las letras sueltas "L" y "A"

se volverán a unir para leer "#LA Flood."

En algunas realizaciones, el sistema conserva el símbolo del hashtag para distinguir entre ubicaciones basadas en el hashtag y otras ubicaciones determinadas a partir del texto de la publicación. Esto ayuda a delimitar el inicio de las ubicaciones basadas en hashtags del resto del texto de la publicación. Por ejemplo, se dispone de la publicación "Hurricane Matthew moving towards Florida #HaitiDisaster". Una vez que el hashtag se descompone siguiendo los procesos descritos, la publicación se leerá "Hurricane Matthew moving towards Florida #Haiti Disaster". Si se elimina el símbolo del hashtag, el sistema identificará incorrectamente "Florida Haiti" como una única ubicación debido al uso consistente del uso de mayúsculas. Por lo tanto, retener el símbolo del hashtag garantiza que el sistema determina "Florida" y "Haiti" como dos ubicaciones diferentes.

En la etapa 516, el sistema elimina los caracteres especiales que aparecen en el texto de la publicación de la publicación en redes sociales. Los caracteres especiales pueden incluir caracteres no alfanuméricos que se encuentren en el texto de la publicación. En realizaciones, el sistema elimina todos los caracteres especiales excepto un conjunto seleccionado de caracteres especiales conservados por su utilidad para identificar ubicaciones. Por ejemplo, el sistema puede eliminar todos los caracteres especiales, tales como los que incluyen, por ejemplo, paréntesis, asteriscos, signos de porcentaje, barra diagonal inversa, salvo los siguientes:

Guion: Algunas ubicaciones incluyen guiones (por ejemplo, "Al-Hasakah").

Apóstrofe o comilla simple: Estos símbolos se utilizan ocasionalmente para denotar formas posesivas, (por ejemplo, "Austin's PD reports a three-alarm fire downtown"). Como se analiza en el presente documento, el sistema puede utilizar nombres de los organismos públicos (por ejemplo, departamentos de policía, parques de bomberos, etc.) para encontrar ubicaciones expresadas en forma posesiva.

Hashtag: Como se ha analizado anteriormente, este símbolo se conserva para distinguir las ubicaciones basadas en hashtags.

Barra diagonal. A veces, este símbolo se utiliza para conectar varios lugares (por ejemplo, "Aviso de huracán para los condados de Kings/Queens"). Conservar este símbolo puede ayudar a identificar estos casos para que el "condado" pueda alternar entre "King" y "Queen".

Coma: Este símbolo se utiliza habitualmente para asociar dos ubicaciones (por ejemplo, "Orlando, Florida").

Punto, signo de exclamación, signo de interrogación, dos puntos, punto y coma. Estos símbolos se utilizan a menudo para identificar el final de las frases o el final de las oraciones, algo esencial para identificar correctamente las ubicaciones. Se dispone de la publicación "Big celebration in New York City. Tonight at 8pm." Si se eliminaran los puntos de la frase, el sistema podría leer erróneamente que "New York City Tonight" es el nombre de un sitio. En la etapa 520, una vez que la una o más publicaciones se ha procesado, finaliza el método 500.

En algunas realizaciones, el método de preprocesamiento 500 puede incluir las características representadas o analizadas en el presente documento con respecto a la Fig. 5.

La Fig. 6 muestra un diagrama de flujo que representa una realización de un método 600 para identificar la ubicación de acuerdo con una realización de la divulgación. En algunas realizaciones, el método 600 para identificar la ubicación se lleva a cabo en la etapa 216 del método 200 por el módulo identificador 132 de menciones de ubicación. En la realización mostrada, el método 600 empieza en la etapa 604. Como se ha analizado anteriormente, el sistema puede funcionar en dos modos: modo de publicación simple y múltiple. En ambos casos, en la etapa 608, el sistema identifica las ubicaciones mencionadas en el texto de cada publicación que recupera. En la etapa 612, si se han identificado una o varias ubicaciones a partir del texto de la publicación, el método 620 finaliza. En algunos casos, no hay información suficiente en una publicación (o conjunto de publicaciones) para identificar ninguna ubicación. En la etapa 612, si no se ha identificado ninguna ubicación a partir del texto de la publicación, el método 600 pasa a la etapa 616 para identificar ubicaciones a partir de los perfiles de usuario. En la etapa 612 se pueden utilizar diferentes umbrales de identificación de ubicaciones para determinar si se han identificado suficientes ubicaciones. Por ejemplo, en la etapa 616 si el sistema encuentra menos de dos menciones de ubicación en el mensaje, también extrae las ubicaciones de los usuarios y las agrega al conjunto de ubicaciones potenciales. En algunas realizaciones, las ubicaciones de los usuarios se pueden identificar aunque las ubicaciones se identifiquen a partir del texto. Por ejemplo, la etapa 612 se puede eliminar del método 600 y las ubicaciones se pueden identificarse tanto a partir del texto como a partir de las ubicaciones de los usuarios.

En algunas realizaciones, el método 600 para identificar la ubicación puede incluir las características representadas o analizadas en el presente documento con respecto a la Fig. 6.

Volviendo a la Fig. 3, la realización representada del módulo identificador 132 de menciones de ubicación incluye un módulo clasificador 324 basado en taxonomía, un módulo clasificador 328 basado en heurística y un módulo clasificador 332 basado en una base de conocimientos (*knowledge-base*, KB). En la realización mostrada, el módulo

identificador 132 de menciones de ubicación también incluye una lista y/o la tabla taxonómica de ubicaciones 336 y una base de datos KB 340 de alias aunque uno o varios de estos elementos se puedan encontrar fuera del módulo identificador 132 de menciones de ubicación. El módulo clasificador 324 basado en taxonomía está configurado para realizar un enfoque basado en taxonomía para detectar las menciones de ubicación a partir del texto de una publicación en redes sociales, el módulo clasificador basado en heurística está configurado para realizar un enfoque basado en heurística y el módulo clasificador 332 basado en KB está configurado para realizar un enfoque basado en KB. En la realización mostrada, el módulo clasificador 324 basado en taxonomía se puede comunicar con una lista y/o tabla taxonómica de ubicaciones 336 y el módulo clasificador 332 basado en KB se puede comunicar con la base de datos KB 340 de alias. En la realización mostrada, el módulo clasificador 324 basado en taxonomía, cada uno del módulo clasificador 328 basado en heurística y el módulo clasificador 332 basado en la KB reciben una o varias publicaciones en redes sociales, tales como desde la fuente 304 de publicaciones sociales mediante la interfaz de comunicaciones 144 (omitida en esta figura por claridad de la ilustración). Los procesos de recepción de las publicaciones en redes sociales y/o la implementación de los procesos de detección de ubicaciones pueden ejecutarse en paralelo y/o en serie.

La Fig. 7 muestra un diagrama de flujo que representa una realización de un método 700 para identificar ubicaciones a partir del texto de una publicación en redes sociales de acuerdo con una realización de la divulgación. En la realización mostrada, el método 700 se lleva a cabo por el módulo identificador 132 de menciones de ubicación y se ejecuta en serie. En realizaciones, el método 700 para identificar la ubicación se puede llevar a cabo como la etapa 216 del método 200 y/o como la etapa 608 del método 600. En la realización mostrada, el método 700 empieza en la etapa 704. En la etapa 708, se detectan las menciones de ubicación de una o varias publicaciones en redes sociales usando un enfoque basado en taxonomía. En la etapa 712, se detectan las menciones de ubicación de una o varias publicaciones en redes sociales usando un enfoque basado en heurística. En la etapa 716, se detectan las menciones de ubicación de una o varias publicaciones en redes sociales usando un enfoque basado en KB. El método 700 finaliza en la etapa 720.

En algunas realizaciones, el método 700 para detectar ubicaciones a partir del texto de una publicación en redes sociales puede incluir las características representadas o analizadas en el presente documento con respecto a la Fig. 7.

La Fig. 8 muestra un diagrama de flujo que representa una realización de un método 800 que implementa un enfoque basado en taxonomía de acuerdo con una realización de la divulgación. En realizaciones, el método 800 se implementa en la etapa 708 del método 700 por el módulo clasificador 324 basado en taxonomía. El método 800 empieza en el etapa 804 y utiliza taxonomía convencional para detectar ubicaciones comunes en un texto de la publicación. Los nombres de geolocalizaciones grandes y/o bien conocidas, tales como países, continentes, estados y provincias pueden cogerse y/o recopilarse para generar una lista taxonómica 336 de nombres de ubicaciones. En algunas realizaciones, la lista taxonómica 336 puede incluir miles de nombres. En la etapa 808, los nombres de ubicaciones que componen la lista taxonómica 336 se recuperan por el módulo clasificador 324 basado en taxonomía y se vuelven a comparar con el texto de cada publicación en redes sociales recibida para ver si algún nombre de ubicación de la lista taxonómica 336 aparece en algún lugar del texto de la publicación.

En la etapa 812, el sistema determina si existe una coincidencia exacta entre una mención de ubicación en el texto y una ubicación incluida en la lista taxonómica 336. Si no hay una coincidencia exacta, el método 800 avanza a la etapa 816. Si hay una coincidencia exacta, el método 800 avanza a la etapa 820. En la etapa 816, cuando no se encuentre una coincidencia exacta, el sistema determina si existe una coincidencia aproximada. En esta etapa, el sistema analiza el texto en busca de palabras que puedan estar relacionadas con una ubicación concreta pero que no coincidan exactamente con una ubicación de la lista taxonómica 336. Por ejemplo, las nacionalidades que aparecen en el texto se pueden asignar a sus correspondientes países de la lista taxonómica 336 (por ejemplo, "canadiense" se puede emparejar con "Canadá"). Las formas posesivas también se pueden asignar a sus ubicaciones correspondientes de la lista taxonómica 336 (por ejemplo, "Louisiana's fire departments" [Cuerpo de bomberos de Luisiana] se puede emparejar con "Luisiana"). En la etapa 820, cuando una palabra del texto de la publicación coincide (tanto de manera exacta como aproximada) con un término de la taxonomía 336, se elimina y se sustituye por una máscara de ubicación (por ejemplo, "<loc>") en una versión procesada de la publicación en redes sociales. Eliminando las menciones de ubicación cuando coinciden, el sistema puede evitar procesar varias veces la misma mención de ubicación y preservar los recursos del sistema. Por ejemplo, la publicación mostrada en la Fig. 4(c) contiene el término "Bruselas". Si "Bruselas" está incluida como una ubicación en la lista taxonómica 336, el término "Bruselas" se elimina y se sustituye por una máscara de ubicación para evitar un procesamiento innecesario. El método 800 finaliza en la etapa 824. En algunas realizaciones, aunque aumente la sobrecarga de procesamiento, el método 800 puede determinar coincidencias próximas incluso cuando también se determinan coincidencias exactas.

En algunas realizaciones, el método 800 para implementar un enfoque basado en taxonomía puede incluir las características representadas o analizadas en el presente documento con respecto a la Fig. 8.

La Fig. 9 muestra un diagrama de flujo que representa una realización de un método 900 que implementa un enfoque basado en heurística de acuerdo con una realización de la divulgación. En la realización mostrada, el método 900 se implementa en la etapa 712 del método 700 por el módulo clasificador 328 basado en heurística. El método

900 empieza en la etapa 904. Además de la taxonomía convencional, el sistema utiliza un conjunto de heurística convencional para identificar palabras o frases de la publicación en redes sociales que probablemente se refieran a ubicaciones. En la etapa 908, el sistema genera una lista de reglas que correlacionan palabras o frases concretas encontradas en determinadas posiciones del texto con posibles ubicaciones. En la etapa 912, el sistema comprueba la publicación en busca de texto que coincida con las reglas de la lista de reglas. La Tabla 1 siguiente enumera algunas de estas palabras/frases como "pistas" junto con algunos ejemplos de posibles ubicaciones correlacionadas.

Tabla 1. Lista de pistas utilizadas para identificar heurísticamente las menciones de ubicación.

Pista	Posición w.r.t la ubicación	Ejemplos	Ejemplos de excepciones
Direcciones cardinales	Prefijo	Norte de Alepo <u>Suroeste</u> de Alepo Norte de Alepo <u>Sureste</u> de Alepo	<u>Norte</u> de la ciudad <u>Banco</u> del Noroeste
Identificadores de puntos de referencia	Prefijo/Sufijo	Ciudad de Alepo Alepo <u>Ciudad</u> Golfo de <u>México</u> <u>Suburbio</u> de Nueva Jersey	<u>City</u> Bank Departamento de <u>Estado</u> <u>Estado Islámico</u>
Indicadores de distancia	Prefijo	a 3 millas de Nueva York Cinco <u>kilómetros</u> de NJ	Cinco <u>millas</u> de campos de trigo
Indicadores de puntos de referencia urbanos	Sufijo	Instituto <u>St. Johns</u> <u>Aeropuerto</u> JFK <u>Ayuntamiento</u> de Manila	<u>El instituto</u> de mi sobrina
Indicadores de puntos de referencia naturales	Prefijo/Sufijo	<u>Río</u> Rojo <u>Monte</u> Rushmore <u>Lago</u> Verde	<u>Río</u> Profundo Cuenca <u>fluvial</u>
Agencias	Sufijo	<u>Policía</u> de Nueva York <u>Cuerpo</u> de bomberos de Los Ángeles	Mayor presencia <u>policial</u>
Área urbana	Prefijo	<u>Centro</u> de Los Ángeles	Chica de <u>clase alta</u>
Movimiento indicadores	Prefijo	El huracán <u>avanza</u> hacia NJ El tifón <u>arrasa</u> Nueva York El festival <u>llega</u> a Los Ángeles	<u>Llegando</u> a casa de mi amigo <u>Estar cerca</u> de un acuerdo
Indicadores de frontera	Prefijo/Sufijo	<u>Frontera</u> entre NY y NJ <u>Intersección</u> de la avenida principal y la calle 17 <u>Frontera</u> NY/NJ	<u>Médicos sin fronteras</u> <u>Límite</u> de la locura
Pronombres	Prefijo	<u>En</u> Nueva York <u>En</u> el JFK	<u>En</u> las conversaciones <u>En</u> casa

Por ejemplo, el sistema puede detectar el término "Norte" en el texto de la publicación, y determinar que este término se correlaciona con una regla de la lista de reglas, concretamente, que el término "Norte" es una dirección cardinal y se utiliza como prefijo. La columna central derecha de la tabla enumera ejemplos en los que la regla detecta correctamente una mención de ubicación en el texto. Sin embargo, la columna más a la derecha de la tabla 1 enumera ejemplos en los que las reglas coinciden con palabras del texto pero la palabra que las precede o sucede no es una ubicación. Para reconocer estos casos, el sistema implementa la etapa 916 para determinar una ubicación basada en el uso de mayúsculas. Si la palabra del texto de la publicación parece tener un uso de mayúsculas correcto (por ejemplo, si la palabra no está escrita en MAYÚSCULAS o en forma de título (es decir, usando solo letras mayúsculas en la primera letra de cada palabra)), entonces, el sistema se basa en el uso de mayúsculas para determinar si las palabras adyacentes a las palabras que coinciden con una regla se refieren a una ubicación. Es menos probable que las palabras que no empiezan con mayúscula sean ubicaciones. Como se muestra en la Tabla. 1, cuando se determina que el término "Norte" está cerca de la palabra "Alepo" en mayúsculas, se determina que "Alepo" es una mención de ubicación. Sin embargo, cuando se determina que el término "Norte" se encuentra cerca de la palabra "ciudad" sin mayúsculas no se determina que "ciudad" sea una mención de ubicación concreta.

Como se puede observar, la etapa 916 puede seguir haciendo coincidir palabras con ubicaciones que no son ubicaciones (es decir, "Northwest Bank" satisface la regla del uso de mayúsculas, pero no es una mención de ubicación. Para reducir estas falsas coincidencias, el sistema implementa la etapa 920 para eliminar y/o ignorar determinados términos de una lista de términos prohibidos. En algunas realizaciones, cada regla tiene una lista de términos prohibidos asociada con la misma. Por ejemplo, la regla de identificadores de puntos de referencia en la Tabla 1, un término de la lista de términos prohibidos es "El Estado Islámico" ya que se trata de un término de uso frecuente que no corresponde a una mención de ubicación para el identificador general del punto de referencia "estado". El método finaliza en la etapa 924.

En algunas realizaciones, el método 900 para implementar un enfoque basado en heurística puede incluir las características representadas o analizadas en el presente documento con respecto a la Fig. 9.

La Fig. 10 muestra un diagrama de flujo que representa una realización de un método 1000 que implementa un enfoque basado en una base de conocimientos (KB) de acuerdo con una realización de la divulgación. En realizaciones, el método 1000 se implementa en la etapa 716 del método 700 por el módulo clasificador 332 basado en una KB. Un problema habitual para identificar menciones de ubicación es la abundancia de términos tales como "NYC" "LA," "PDX," "Big Apple," y otros alias comunes de ubicaciones. Algunos de estos alias son conocidos y pueden integrarse manualmente utilizando un glosario de términos. Sin embargo, a veces, los medios de comunicación locales popularizan un término que no se utiliza en medios de comunicación globales. Por ejemplo, en medios de comunicación indios se utiliza habitualmente "J&K" para referirse a "Jammu y Kashmir", pero no en otros lugares. Adicionalmente, algunos eventos generan hashtags de noticias vinculados a un lugar concreto (por ejemplo, "#LAFlood"). Adicionalmente, algunas cuentas oficiales de servicios meteorológicos utilizan un formato estándar para indicar un lugar. Por ejemplo, "#nywx" se refiere a las previsiones meteorológicas para Nueva York. Este tipo de alias no son fáciles de detectar o aplicar usando los glosarios de términos tradicionales. Por lo tanto, puede ser importante detectar dinámicamente estos términos y asociarlos a su ubicación correspondiente, en especial si aparecen frecuentemente. El sistema utiliza una base de conocimientos (*Knowledge-Base*, KB) autoajustable dinámicamente para lograr esta funcionalidad. En algunas realizaciones, la base de conocimientos (KB) autoajustable dinámicamente se almacena en la base de datos 340 mostrada en la Fig. 3.

En la realización mostrada, el método 1000 empieza en la etapa 1004. En la etapa 1008, el sistema construye la KB utilizando información de aparición simultánea para determinar asociaciones entre ubicaciones y sus alias. Por ejemplo, si "Ciudad de Nueva York" aparece frecuentemente junto con "NYC" en publicaciones en redes sociales, el sistema actualiza la KB para emparejar ambos términos.

Usando la KB, el sistema puede eliminar emparejamiento incorrectos. En la etapa 1012, el sistema determina si dos términos de la KB satisfacen un umbral mínimo de aparición simultánea. Por ejemplo, las expresiones "Ciudad de Nueva York" y "NYC" deben aparecer simultáneamente más de 1000 veces (es decir, el umbral mínimo) para que se tengan en cuenta. Si el sistema determina que los términos no alcanzan el umbral mínimo, los términos se eliminan de la KB.

En la etapa 1016, el sistema determina un orden de reciprocidad de los términos, basándose en la información de aparición simultánea. En algunas realizaciones, el sistema ordena en primer lugar cada término según la frecuencia de su aparición simultánea con otro término. Por ejemplo, si "NYC" es el segundo término más frecuente que aparece simultáneamente con "Ciudad de Nueva York", entonces, su rango con respecto a "Ciudad de Nueva York" será "2." Después de calcular todos los rangos, el sistema comprueba si "NYC" y "Ciudad de Nueva York" tienen el mismo rango. Por ejemplo, si "NYC" es el término más frecuente que aparece simultáneamente con "Ciudad de Nueva York", (es decir, con rango "1"), el sistema determina si "Ciudad de Nueva York" también es el término más frecuente (es decir, con rango "1") que aparece simultáneamente con "NYC". Si los rangos coinciden, se determina que los dos términos satisfacen la reciprocidad de rango y se reconocen como pares de alias en la KB. En realizaciones, la KB se puede actualizar dinámicamente de forma periódica (por ejemplo, cada semana) para ampliar su colección de emparejamientos de alias.

En la etapa 1020, el sistema usa la KB para detectar los alias que aparecen en la publicación y determina una ubicación correspondiente a cada alias. Durante el procesamiento de una publicación para determinar ubicaciones, si aparece un alias como posible ubicación, tanto el alias como su correspondiente coincidencia pueden agregarse como ubicación. Por ejemplo, si "NYC" aparece en la publicación (tal como el ejemplo mostrado en la Fig. 4(d)), tanto "NYC" como "Ciudad de Nueva York" se marcan como posibles menciones de ubicación. El método finaliza en la etapa 1024.

El método 1000 para implementar un enfoque basado en una KB incluye las características representadas o analizadas en el presente documento con respecto a la Fig. 10.

Volviendo a la Fig. 3, la realización representada del módulo determinador 136 de ubicaciones incluye un georreferenciador 344, una biblioteca y/o servicio 348 de georreferenciación y un validador 352 del tipo de usuario. En otras realizaciones, el validador 352 del tipo de usuario puede estar situado fuera del módulo determinador 136 de ubicaciones. El georreferenciador 344 está configurado para recibir los resultados de los tres enfoques de detección

de ubicación (por ejemplo, basado en taxonomía, basado en heurística y basado en KB) analizados anteriormente en forma de una lista de ubicaciones potenciales. En la realización mostrada, el georreferenciador 344 también está configurado para recibir una o varias ubicaciones de usuario validadas desde el validador 352 del tipo de usuario. En la realización mostrada, el validador 352 del tipo de usuario puede recibir una publicación 316 y una ubicación 320 del usuario y determinar si la cuenta de publicación representa un tipo concreto de cuenta tal como un cuerpo de bomberos o de policía, noticiero local u otra cuenta de servicio público. En la realización mostrada, el georreferenciador 344 se puede comunicar con una biblioteca y/o servicio 348 de georreferenciación para determinar las coordenadas geográficas de cada una de las posibles menciones de ubicación recibidas desde el módulo identificador 132 de menciones de ubicación. La biblioteca y/o servicio 348 de georreferenciación puede ser una biblioteca de terceros que comprende una base de datos que correlaciona ubicaciones concretas con sus coordenadas geográficas.

La Fig. 11 muestra un diagrama de flujo que representa una realización de un método 1100 para determinar coordenadas geográficas de ubicación de acuerdo con una realización de la divulgación. En realizaciones, el método 1100 se implementa en la etapa 220 del método 200 por el módulo determinador 136 de ubicaciones mostrado en la Fig. 3. En la realización mostrada, el método 1100 empieza en la etapa 1104. En la etapa 1108, el sistema busca cada una de las posibles ubicaciones recibidas del módulo determinador 132 de ubicaciones incluida en la biblioteca y/o servicio 348 de georreferenciación. En algunas realizaciones, la biblioteca 348 puede ser una biblioteca de ubicaciones de terceros, tales como, por ejemplo, Nominatim, aunque se pueden usar otras bibliotecas de ubicaciones de terceros. Esta biblioteca en particular recibe datos del proyecto OpenStreetMap (OSM), que periódicamente proporciona un volcado de todas las geolocalizaciones del planeta que se puede descargar a una base de datos. La biblioteca de ubicaciones proporciona un mecanismo para acceder y consultar fácilmente la base de datos y proporciona un servicio REST y una GUI para facilitar la navegación. En la etapa 1112, el sistema recibe resultados de geolocalización desde la biblioteca de ubicaciones para cada posible ubicación. Si se busca una palabra en la biblioteca de ubicaciones, devuelve un conjunto de resultados de geolocalización que pueden corresponder a dicha palabra. En la etapa 1116, cada uno de los resultados de la geolocalización puede tener una puntuación, tal como una puntuación de importancia, que representa la intensidad de la asociación entre la posible ubicación y una geolocalización en particular. En realizaciones, la puntuación se puede recibir desde la biblioteca 348. En otras realizaciones, la puntuación se puede calcular por separado. En la etapa 1120, finaliza el método.

En algunas realizaciones, el método 1100 para detectar la determinación de las coordenadas geográficas de una ubicación puede incluir las características representadas o analizadas en la presente con respecto a la Fig. 11.

Las Figs. 12(a)-12(b) muestran ejemplos de salidas 1200 de un ejemplo de biblioteca de ubicaciones de acuerdo con una realización de la divulgación. En la Fig. 12(a), la consulta "orlando" 1204 se buscó en una biblioteca de ubicaciones que devolvió una lista de resultados 1208. En la realización mostrada, la lista de resultados 1208 se puede complementar y/o sustituir por una selección de botón 1212. En la Fig. 12(b), cada resultado 1208 incluye datos 1216 que incluyen un par de coordenadas geográficas de latitud/longitud, un polígono, algunos metadatos que indican un contexto más amplio del resultado (por ejemplo, provincia, país), y una puntuación de importancia que muestra un grado de coincidencia de la asociación entre la consulta 1204 y el resultado 1208. En otras palabras, la puntuación de importancia representa el grado de correlación entre la ubicación consultada y cada resultado de geolocalización. Por ejemplo, la puntuación de importancia para Orlando, Florida es de aproximadamente 0,71, mientras que la puntuación de importancia para Orlando, Arkansas es 0,37. El sistema puede utilizar esta información para encontrar las coordenadas geográficas más probables de cada ubicación potencial.

En algunas realizaciones, el ejemplo de salida 1200 de un ejemplo de biblioteca de ubicaciones puede incluir las características representadas o analizadas con respecto a las Figs. 12(a)-12(b).

Volviendo a la Fig. 3, la realización representada del módulo calificador 140 incluye un motor de desambiguación 356 de ubicaciones, un validador 360 de la ubicación del usuario y un motor de ordenación 364 de ubicaciones. En la realización mostrada, el módulo calificador 140 está configurado para recibir los resultados procedentes de la biblioteca de ubicaciones anteriormente analizados en forma de una lista o conjunto de coordenadas geográficas y puntuaciones de importancia de posibles ubicaciones. En la realización mostrada, el módulo calificador 140 está configurado para calificar y validar la lista de posibles ubicaciones y determinar una ubicación primaria y sus correspondientes coordenadas geográficas.

La Fig. 13 muestra un diagrama de flujo que representa una realización de un método 1300 para calificar ubicaciones de acuerdo con una realización de la divulgación. En realizaciones, el método 1300 se implementa en la etapa 224 del método 200 por el módulo calificador 140 de ubicaciones mostrado en la Fig. 3. En la realización mostrada, el método 1300 empieza en la etapa 1304. En la etapa 1308, el sistema califica los resultados de ubicación recibidos desde la biblioteca de ubicaciones usando las puntuaciones de importancia. En la etapa 1312, el sistema califica las ubicaciones de los usuarios utilizando heurística comunitaria basada en los tipos de cuentas de usuario y las ubicaciones de los usuarios. En algunas realizaciones, estas etapas pueden incluir procesos que identifican y eliminan las palabras que no son ubicaciones. Por ejemplo, se trata de palabras que se han identificado como ubicaciones por error o ubicaciones identificadas por los propios usuarios como no viables, como "Planeta Tierra" En la etapa 1316, el sistema encuentra y georreferencia una primera ubicación determinada a partir de la lista de posibles ubicaciones calificadas restantes. Por ejemplo, una vez que se han eliminado algunas ubicaciones, para el resto de

las ubicaciones, el sistema puede determinar las coordenadas geográficas "mejores" o más probables de la lista proporcionada por la biblioteca de ubicaciones. En la etapa 1320, finaliza el método.

En algunas realizaciones, el método 1300 para calificar ubicaciones puede incluir las características representadas o analizadas en el presente documento con respecto a la Fig. 13.

La Fig. 14 muestra un diagrama de flujo que representa una realización de un método 1400 para calificar ubicaciones de una biblioteca de ubicaciones de acuerdo con una realización de la divulgación. En realizaciones, el método 1400 se implementa en la etapa 1308 del método 1300 por el motor de desambiguación 356 de ubicaciones. El método 1400 empieza en la etapa 1404. En la etapa 1408, el sistema analiza los resultados de la biblioteca de ubicaciones, tal como comparando la puntuación de importancia de cada resultado con un umbral de puntuación de importancia. En la etapa 1412, el sistema determina si una puntuación de importancia de un resultado en particular es menor que el umbral. Si la puntuación de importancia es menor que el umbral, el sistema descarta el resultado de la ubicación en la etapa 1416. El sistema realiza este análisis para cada resultado de ubicación devuelto por la biblioteca de ubicaciones. Si la puntuación de importancia del primer resultado devuelto por la biblioteca de ubicaciones está por debajo del umbral, probablemente significa que la ubicación es demasiado incierta para ser fiable. Por ejemplo, "La Vía Láctea" podría corresponder con el nombre de un bar en Nueva Jersey, pero el sistema no reconocerá (es decir, descartará) esta ubicación si su puntuación de importancia es inferior al umbral.

Si se determina que la puntuación de importancia de un resultado es mayor o igual que el umbral de puntuación, el proceso pasa a la etapa 1420 y analiza la varianza de la puntuación de importancia de los resultados de la lista. En la etapa 1412, el sistema determina si la varianza en las puntuaciones de importancia en la totalidad de la lista de resultados es menor que un umbral de varianza. Si la varianza es inferior al umbral de varianza, el sistema descarta el resultado de la ubicación en la etapa 1416. Si la varianza en la puntuación de importancia de los resultados está por debajo de un umbral de varianza, probablemente significa que el sistema no tiene suficiente confianza para dar una geolocalización definitiva de una posible ubicación. Por ejemplo, "La Vía Láctea" podría ser el nombre tanto de un bar en Nueva Jersey como de un café en Washington, D.C. Sin embargo, puesto que ambos puntos de referencia son relativamente inciertos, la biblioteca de ubicaciones asignará puntuaciones de importancia análogamente bajas a estos resultados y posteriormente se eliminarán del conjunto de posibles ubicaciones. Si la varianza es superior al umbral de varianza, el sistema mantiene el resultado de la ubicación en la etapa 1428. En la etapa 1432, finaliza el método 1400.

En algunas realizaciones, el método 1400 para calificar ubicaciones de una biblioteca de ubicaciones puede incluir las características representadas o analizadas en el presente documento con respecto a la Fig. 14.

La Fig. 15 muestra un diagrama de flujo que representa una realización de un método 1500 para calificar ubicaciones de los usuarios usando heurística comunitaria de acuerdo con una realización de la divulgación. En realizaciones, el método 1500 se implementa en la etapa 1312 del método 1300 por el validador 360 de la ubicación del usuario. Las ubicaciones de los usuarios no siempre están relacionadas con la ubicación de un acontecimiento concreto. Esto puede ser especialmente cierto en el caso de los ataques selectivos, como atentados terroristas, que a menudo pueden producirse en ubicaciones concurridas y propensas al tránsito, como aeropuertos y atracciones turísticas. Por ejemplo, en referencia a la Fig. 4(b), una publicación en redes sociales que incluya la referencia de un testigo de un ataque terrorista en Niza (mención de ubicación 408) puede proceder de un usuario de Mónaco. Por lo tanto, el sistema puede calificar las ubicaciones de los usuarios para determinar una geolocalización fiable.

En la realización mostrada, el método 1500 empieza en la etapa 1504. En la etapa 1508, el sistema determina si las publicaciones en redes sociales se han recuperado en modo de publicación única o en modo de publicación múltiple, tal como en función de la configuración o entrada de un usuario u otra aplicación. Si se ha determinado un modo de publicación única, el método 1500 avanza a la etapa 1512. En realizaciones, la etapa 1512 puede ser realizada por el validador 352 del tipo de usuario. En el modo de publicación única, el sistema solo puede aceptar la ubicación del usuario si la cuenta es de un tipo fiable. A menudo, las cuentas fiables son cuentas oficiales de las agencias de noticias locales, equipos de respuesta a catástrofes, cuerpos de policía o bomberos. Para determinar si una cuenta de usuario es fiable, el sistema puede utilizar un conjunto estándar de palabras (por ejemplo, "fuego," "policía", etc.) y cotejarlas con la descripción y el nombre de la cuenta. Para identificar el conjunto estándar de palabras, el sistema puede confeccionar una lista de cuentas de redes sociales que pertenezcan a agencias de noticias locales, agencias gubernamentales locales y cuerpos de bomberos y policía. El sistema también puede recopilar las descripciones de estas cuentas de redes sociales, tokenizarlas, eliminar las palabras o significativas y determinar los términos más frecuentes que aparecen en las descripciones de las cuentas. En algunas realizaciones, un ejemplo de conjunto estándar de palabras y/o lista taxonómica puede incluir palabras tales como "ciudad", "estado", "condado", "local," "departamento", "dept.," "policía", "emergencia", "emergencias", "incendios", "911" [112] "última hora", "noticias", "servicio", "tiempo", "estación", "tráfico" y "canal" La lista de cuentas también puede incluir cuentas que tengan entradas que empiecen o terminen por "PD" o "FD" (en mayúsculas) y cuantas cuyas descripciones incluyan la palabra "oficial". Adicionalmente, la lista de cuentas puede incluir todas las cuentas que tengan una URL que termine en ".gov." En la etapa 1512, el sistema compara el nombre de la cuenta de publicación con el conjunto de palabras determinado. En la etapa 1516, se determina si el nombre de la cuenta de publicación coincide con una entrada del conjunto de palabras determinado. Si el nombre de la cuenta coincide con una entrada del conjunto de palabras

determinado, la ubicación o ubicaciones del posible usuario se mantienen en la etapa 1520. Si el nombre de la cuenta no coincide con ninguna entrada del conjunto de palabras determinado, se descartan la posible o posibles ubicaciones del usuario (por ejemplo, se ignoran) en la etapa 1524. Por ejemplo, en referencia a la Fig. 4(a), una publicación en redes sociales puede contener el término "policía" pero puede haber sido publicada por un usuario situado en una ubicación diferente a la del evento descrito (por ejemplo, Miami en lugar de Orlando). El sistema puede comparar el nombre de la cuenta de la publicación en redes sociales con el conjunto de palabras determinado y, si el nombre de la cuenta no coincide con ninguna entrada del conjunto de palabras determinado, se ignora la posible ubicación del usuario.

En la realización mostrada, si se ha determinado un modo de publicación múltiple, el método 1500 avanza a la etapa 1532. En el modo de publicación múltiple, el sistema solamente puede aceptar la ubicación del usuario si al menos un determinado porcentaje predeterminado de usuarios (por ejemplo, el 75 %) tienen la misma ubicación de usuario. Si el sistema determina que un determinado porcentaje de usuarios coincide con la ubicación del usuario que realiza la publicación, la ubicación se mantiene en la etapa 1520. Si el sistema determina que un determinado porcentaje de usuarios no coincide con la ubicación del usuario que realiza la publicación, se descarta la ubicación (por ejemplo, se ignora) en la etapa 1524. Por ejemplo, si se introducen tres publicaciones en el sistema, al menos dos usuarios deberán tener ubicaciones coincidentes para que se mantengan sus mensajes. Las ubicaciones coincidentes pueden identificarse como lugares "cercaños" (es decir, ubicaciones situadas a una distancia o radio predeterminado de la ubicación del usuario). Por ejemplo, un usuario podría identificar una ubicación de usuario como "Norte de Londres" y otro usuario podría identificar una ubicación de usuario como "Sur de Londres". La biblioteca de ubicaciones puede devolver una lista de resultados/aciertos para cada ubicación. Si, entre la lista de aciertos hay al menos un par de usuarios a una distancia o radio predeterminado (por ejemplo, 20 millas [32 km] entre sí, los usuarios pueden considerarse "cercaños" El método 1500 finaliza en la etapa 1536.

En algunas realizaciones, el método 1500 para calificar ubicaciones de usuarios usando heurística comunitaria puede incluir las características representadas o analizadas en el presente documento con respecto a la Fig. 15.

La Fig. 16 muestra un diagrama de flujo que representa una realización de un método 1600 para determinar coordenadas geográficas de una ubicación primaria de acuerdo con una realización de la divulgación. En la realización mostrada, el método 1600 se implementa en la etapa 1316 del método 1300 por el motor de ordenación 364 de ubicaciones. En el método 1600, el sistema ordena todas las ubicaciones restantes después de finalizar los procesos anteriores analizados en lo que antecede. En las realizaciones mostradas, la ordenación se basa en un nivel de confianza de la relevancia de la ubicación para un evento que es el objeto de la publicación en redes sociales. En la etapa 1608, el sistema ordena cada ubicación restante basándose en el conjunto de ubicaciones de usuario calificadas anteriormente descritas. Esta ordenación representa el nivel de confianza más alto de que una ubicación se corresponde con el evento. En la etapa 1612, el sistema ordena cada ubicación restante basándose en la lista taxonómica 336 de ubicaciones analizada anteriormente. Esta ordenación representa un nivel de confianza menor de que una ubicación se corresponde con el evento, en comparación con una ubicación ordenada usando la ubicación de usuario calificada. En la etapa 1616, el sistema ordena cada ubicación restante basándose en la heurística usada por el módulo clasificador 328 basado en heurística y el módulo clasificador 332 basado en la KB analizados anteriormente. Esta ordenación representa el nivel de confianza más bajo de que una ubicación se corresponda con el evento, comparada tanto con una ubicación ordenada usando una ubicación de usuario calificada o una ubicación ordenada usando la lista taxonómica 336 de ubicaciones.

En el método de ordenación 1600, cada fuente (por ejemplo, taxonomía, basada en reglas y KB) recibe una ponderación de confianza independiente, tal y como se ha analizado anteriormente en las etapas 1608, 1612 y 1616. Por ejemplo, las ubicaciones basadas en taxonomía pueden tener una puntuación de confianza más alta que las basadas en heurística o en la KB. Cada ubicación también puede recibir una puntuación de ordenación basada en los siguientes factores: 1) su posición dentro de la publicación (por ejemplo, ubicación más a la izquierda, ubicación más a la derecha, segundo lugar a la izquierda, etc.); y 2) su inclusión en otras ubicaciones dentro de la publicación (por ejemplo, "Inundaciones en París, Francia" tendrá etiquetados como ubicaciones tanto "París" como "Francia", pero como París se encuentra en Francia, tendrá una puntuación más alta en la ordenación). La puntuación de ordenación final puede ser una interpolación lineal de las ponderaciones de confianza de las fuentes y las puntuaciones de ordenación. El comportamiento de ordenación se puede evaluar dentro de la evaluación integral del sistema, y las ponderaciones se pueden ajustar para proporcionar al sistema el máximo rendimiento. En algunas realizaciones, cada una de las etapas 1608, 1612 y 1616 se lleva a cabo en cada ubicación restante. En otras realizaciones, la etapa 1612 solamente se lleva a cabo en una ubicación restante que no se ha podido ordenar en la etapa 1608 y la etapa 1616 solamente se lleva a cabo en una ubicación restante que no se ha podido ordenar en ninguna de las etapas 1608 o 1612.

En la etapa 1620, el sistema identifica las coordenadas geográficas que más se aproximan a cada una de las ubicaciones como coordenadas geográficas de ubicación primaria. Después de haber ordenado las ubicaciones según los criterios anteriores, el sistema puede utilizar un proceso de distancia mínima de emparejamiento para determinar los resultados/aciertos que generan las distancias más cortas. Esta etapa se puede ilustrar con referencia a los siguientes escenarios. En el escenario A, se recibe la siguiente publicación social: "Dos alarmas de incendio en el 30

de Main St". Para esta publicación, una ubicación de usuario calificada es "Fort Lee, NJ." En el escenario B, se recibe la siguiente publicación social: "Alerta de huracán en París, Texas". Para esta publicación, no se ha podido determinar una ubicación de usuario calificada y, por tanto, no se utiliza la ubicación del usuario.

5 En el escenario A, el sistema determina una mención de ubicación en "30 de Main St" y una ubicación de usuario calificada de "Fort Lee, NJ" después de implementar los métodos divulgados en el presente documento. El sistema recibe resultados de una biblioteca de ubicaciones que incluye tres aciertos: "30 de Main St., Brooklyn, NY," "30 de Main St., Flushing, NY," y "30 de Main St., Fort Lee, NJ." Porque la publicación incluye una ubicación de usuario calificada, el sistema implementa la etapa 1608 y determina un rango de las ubicaciones usando la ubicación de usuario calificada "Fort Lee, NJ." Basándose en esta ordenación, el resultado de ubicación "30 Main St., Fort Lee, NJ" se mantiene, y las otras dos ubicaciones se descartan. En este escenario, las etapas 1612 y 1616 no se llevan a cabo porque una ordenación de ubicaciones usando una ubicación de usuario calificada representa el mayor nivel de confianza, por tanto, el procesamiento adicional es innecesario. A continuación, el sistema implementa la etapa 1620 para encontrar un par de coordenadas geográficas que correspondan a "30 de Main St., Fort Lee, NJ," y designa estas coordenadas geográficas como correspondientes a la ubicación primaria del evento descrito en la publicación.

10 En el escenario B, el sistema determina una mención de ubicación de "París, Texas" después de implementar los métodos divulgados en el presente documento. El sistema recibe resultados de una biblioteca de ubicaciones que incluye tres aciertos: "París, Francia", "París, Texas", y "París, Illinois". Sin embargo, puesto que el sistema no puede calificar la ubicación del usuario, se basa exclusivamente en las ubicaciones mencionadas en el texto de la publicación para determinar la ubicación primaria. El sistema omite la etapa 1608 porque no existe una ubicación de usuario calificada e implementa la etapa 1612 para determinar un rango de las ubicaciones basado en la lista taxonómica. Como se determina que "París" y "Texas" son los nombres de una ciudad y un estado, respectivamente, de la lista taxonómica, "París, Texas" se ordena como la ubicación más probable en la lista de aciertos. A continuación, el sistema implementa la etapa 1620 para encontrar un par de coordenadas geográficas que correspondan a "París, Texas", y designa estas coordenadas geográficas como correspondientes a la ubicación primaria del evento descrito en la publicación.

15 En las realizaciones mostradas, el método 1600 finaliza en la etapa 1624. Cuando se han determinado las coordenadas geográficas de las ubicaciones primarias, el sistema puede enriquecer el flujo 304 de publicaciones sociales entrante agregando las coordenadas geográficas como campos de metadatos en el uno o varios objetos de publicación. El flujo puede ser aprovechado por cualquier aplicación posterior sin interrupciones y en tiempo real. Las aplicaciones posteriores pueden incluir, por ejemplo, detección de noticias, detección de catástrofes, creación de perfiles de usuarios, etc. Por ejemplo, en algunas realizaciones, los objetos de publicación modificados se pueden enviar a otra aplicación mediante la comunicación con un sistema de usuario a través de una o varias redes de comunicaciones. En algunas realizaciones, los objetos de publicación modificados se pueden enviar directamente a un usuario a través de una o varias redes de comunicaciones. En algunas realizaciones, el aplicación puede ejecutarse en dos modos: mensaje único y mensajes múltiples. Cada modo puede utilizarse en modo "*streaming*" (distribución continua) o "*pull*" (distribución puntual). Por ejemplo, los grupos de publicaciones geoetiquetadas en un modo de mensajes múltiples se pueden aprovechar mediante llamadas a la API o a través de un servicio de *streaming* que enriquece una UI u otra aplicación. Puede aprovecharse de manera similar una única publicación geoetiquetada en modo de mensaje único.

20 En algunas realizaciones, el método 1600 para determinar las coordenadas geográficas de una ubicación primaria puede incluir las características representadas o analizadas con respecto a la Fig. 16.

25 Se puede apreciar que las funciones anteriormente descritas se pueden ejecutar en muchos tipos diferentes de aplicaciones de programas informáticos, como aplicaciones web o aplicaciones para dispositivos móviles. Si se implementa en firmware y/o software, las funciones descritas anteriormente se pueden almacenar en forma de una o varias instrucciones o código en un medio legible por ordenador no transitorio. Los ejemplos incluyen medios legibles por ordenador no transitorios codificados con una estructura de datos y medios legibles por ordenador no transitorios codificados con un programa informático. Los medios legibles por ordenador no transitorios incluyen medios de almacenamiento de equipos físicos. Un medio de almacenamiento físico puede ser cualquier medio disponible al que pueda acceder un ordenador. Como ejemplo y no como limitación, dichos medios legibles por ordenador no transitorios pueden comprender RAM, ROM, EEPROM, CD-ROM u otro almacenamiento en disco óptico, almacenamiento en disco magnético u otros dispositivos de almacenamiento magnético, o cualquier otro medio físico que pueda utilizarse para almacenar el código de programa deseado en forma de instrucciones o estructuras de datos y al que pueda acceder un ordenador. Disco y discos incluyen discos compactos (CD), discos láser, discos ópticos, discos versátiles digitales (DVD), disquetes y discos Blu-ray. Por lo general, los discos reproducen los datos magnética y ópticamente. También se incluyen combinaciones de los anteriores en el ámbito de los medios legibles por ordenador no transitorios. Además, las funciones descritas anteriormente se pueden conseguir usando dispositivos dedicados en lugar de programas informáticos, como un circuito de hardware compuesto por circuitos VLSI personalizados o matrices de puertas, semiconductores comerciales, como chips lógicos, transistores u otros componentes discretos, todos ellos no transitorios. Otros ejemplos son los dispositivos de hardware programables, como las matrices de puertas programables en campo, lógica de matriz programable, dispositivos lógicos programables o similares, todos ellos no transitorios. Otros ejemplos incluyen circuitos integrados de aplicación

específica (ASIC) o circuitos integrados a muy gran escala (VLSI). De hecho, los expertos en la técnica pueden utilizar cualquier cantidad de estructuras adecuadas capaces de ejecutar operaciones lógicas de acuerdo con las realizaciones descritas.

5 Por ejemplo, las realizaciones del sistema de redes sociales 104, el sistema de geolocalización 108,
aplicación 112 y el sistema de usuario 116, y/o cualquier componentes individual, subconjunto de componentes o
todos los componentes de los mismos, se pueden implementar como hardware, software o una mezcla de hardware
y software. Por ejemplo, cada uno de del sistema de redes sociales 104, el sistema de geolocalización 108, aplicación
10 112 y el sistema de usuario 116, y/o cualquier componentes individual, subconjunto de componentes o todos los
componentes de los mismos, se puede implementar utilizando un procesador y un medio de almacenamiento no
transitorio, donde el medio de almacenamiento legible por máquina no transitorio incluye instrucciones de programa
que al ejecutarse en el procesador realizan realizaciones de las funciones de dichos componentes analizadas en el
presente documento. En realizaciones, cada uno de del sistema de redes sociales 104, el sistema de geolocalización
108, aplicación 112 y el sistema de usuario 116, y/o cualquier componente individual, subconjunto de componentes o
15 todos los componentes de los mismos, se puede implementar utilizando uno o más sistemas informáticos tales como,
por ejemplo, un ordenador de sobremesa, ordenador portátil, dispositivo informático móvil, dispositivo de red, servidor,
servidor de Internet, servidor en nube, etc.

20 La memoria descriptiva y los ejemplos anteriores proporcionan una descripción completa de la estructura y
el uso de las realizaciones ilustrativas. De este modo, no se pretende que las varias realizaciones ilustrativas de los
métodos, dispositivos y sistemas divulgados estén limitadas a las formas concretas divulgadas. Por ejemplo, los
componentes se pueden combinar como una estructura unitaria y/o se pueden sustituir las conexiones.

25 Son posibles realizaciones adicionales del sistema de redes sociales 104, el sistema de geolocalización 108,
la aplicación 112 y el sistema de usuario 116, y los métodos asociados, como se analizan en el presente documento.

REIVINDICACIONES

1. Un método para determinar ubicaciones destinado a publicaciones en redes sociales, el método comprendiendo:

5 recuperar (208), mediante la comunicación con al menos una interfaz de programación de aplicaciones (API) de un sistema de redes sociales (104) en una o varias primeras redes de comunicaciones (128), al menos una publicación en redes sociales;
 10 determinar (216) al menos una mención de ubicación en la al menos una publicación en redes sociales;
 detectar si aparece al menos un alias en una base de conocimientos, KB, en la al menos una publicación en redes sociales, la KB incluyendo asociaciones entre ubicaciones y alias, una asociación entre una ubicación en particular y un alias correspondiente en la KB estando basada en la frecuencia con la que la ubicación en particular y el alias correspondiente aparecen simultáneamente en publicaciones en redes sociales, la asociación entre una ubicación en particular y un correspondiente alias estando en la KB cuando la ubicación en particular y el alias correspondiente tienen una ordenación de aparición simultánea recíproca entre sí;
 15 agregar a la al menos una mención de ubicación una ubicación correspondiente al alias detectado en la KB;
 determinar (220) al menos una ubicación basándose en la al menos una mención de ubicación;
 determinar (224) una ubicación primaria a partir de la al menos una ubicación;
 almacenar (228), en al menos una base de datos en un medio de almacenamiento legible por máquina no transitorio, al menos un conjunto de coordenadas geográficas para la ubicación primaria en al menos un objeto de publicación para la al menos una publicación en redes sociales; y
 20 enviar (232), mediante la comunicación con un sistema de usuario (116) usando una o varias redes de comunicaciones secundarias (156, 164), la al menos una publicación en redes sociales con al menos un conjunto de coordenadas geográficas almacenado para su presentación en el sistema de usuario (116).

25 2. El método de la reivindicación 1, en donde el envío (232) incluye transmitir una señal que representa la al menos una publicación en redes sociales con el al menos un conjunto de coordenadas geográficas almacenado a un sistema servidor que ejecuta una aplicación (112), y proporcionar mediante la aplicación (112) la señal o una segunda señal basada en la señal al sistema de usuario (116), en donde la señal se transmite a una API (160) de la aplicación (112);
 30 y en donde la recuperación (208) incluye comunicar una solicitud a través de la una o varias de las primeras redes de comunicaciones (128) a la API del sistema de redes sociales (104), y recibir una respuesta usando la una o varias primeras redes de comunicaciones (128) desde la API del sistema de redes sociales (104).

3. El método de la reivindicación 1, que comprende además preprocesar (212) la al menos una publicación en redes sociales y en donde el preprocesamiento (212) comprende al menos uno de:
 35 eliminar truncamientos (508) de la al menos una publicación en redes sociales;
 dividir los hashtags (512) de la al menos una publicación en redes sociales; o
 eliminar caracteres especiales de la al menos una publicación en redes sociales.

4. El método de la reivindicación 1, en donde determinar (220) la al menos una mención de ubicación comprende al menos uno de:
 40 identificar (608) al menos una ubicación a partir del texto de la al menos una publicación en redes sociales; o
 identificar (616) al menos una ubicación a partir de al menos una ubicación del usuario de la al menos una publicación en redes sociales.
 45

5. El método de la reivindicación 1, que comprende además:
 50 determinar al menos una ubicación del usuario de la al menos una publicación en redes sociales;
 comparar términos en el nombre de una cuenta o en la descripción de una cuenta de la al menos una publicación en redes sociales con una lista taxonómica, validar la al menos una ubicación del usuario cuando al menos uno de los términos coincide con la lista taxonómica, y descartar la al menos una ubicación del usuario cuando ninguno de los términos coincide con la lista taxonómica.

6. El método de la reivindicación 1, en donde determinar (220) la al menos una mención de ubicación incluye identificar (608) al menos una ubicación a partir del texto de al menos una publicación en redes sociales implementando un método de identificación basado en taxonomía (708);
 55 en donde el método de identificación basado en taxonomía (708) comprende:

60 comparar (812) términos de la al menos una publicación en redes sociales con una lista de nombres de ubicaciones;
 retirar (820) al menos un nombre coincidente cuando un nombre de la al menos una publicación en redes sociales coincide con un nombre de la lista de nombres; y
 65 sustituir (820) el al menos un nombre coincidente por una máscara de ubicación.

7. El método de la reivindicación 1, en donde determinar (220) la al menos una mención de ubicación incluye identificar

(608) al menos una ubicación a partir del texto de al menos una publicación en redes sociales implementando un método de identificación basado en heurística (712);
 en donde el método de identificación basado en heurística (712) comprende al menos uno de:

5 determinar al menos una ubicación basándose en comparar la al menos una publicación en redes sociales con al
 menos una regla basada en la al menos una ubicación de posición;
 determinar (916) al menos una mención de ubicación basándose en el uso de mayúsculas en la al menos una
 publicación en redes sociales; o
 retirar (920) al menos un término de la lista de términos prohibidos de la al menos una publicación en redes
 10 sociales.

8. El método de la reivindicación 1, el método comprendiendo además:

15 construir (1008) la KB que tiene las asociaciones entre ubicaciones y alias usando información de aparición
 simultánea;
 y
 actualizar periódicamente la KB.

9. El método de la reivindicación 1, en donde determinar (220) al menos una ubicación comprende:

20 buscar (1108) al menos una posible ubicación en una biblioteca de ubicaciones;
 recibir (1112) uno o más resultados de geolocalización correspondientes con al menos una posible ubicación; y
 determinar una puntuación para cada uno de los uno o varios resultados de la geolocalización,
 en donde la puntuación es una puntuación de importancia que representa un grado de correlación entre la al menos
 25 una posible ubicación y cada uno de los uno o varios resultados de la geolocalización.

10. El método de la reivindicación 9, en donde determinar la al menos una ubicación comprende al menos uno de:

30 calificar el uno o varios resultados de la geolocalización a partir de la biblioteca de ubicaciones (1308); o
 calificar el uno o varios resultados de la geolocalización usando heurística comunitaria (1316),
 en donde determinar al menos una ubicación comprende además determinar al menos una geolocalización
 primaria a partir del uno o varios resultados de la geolocalización calificados.

35 11. El método de la reivindicación 10, en donde calificar el uno o varios resultados de la geolocalización a partir de la
 biblioteca de ubicaciones comprende:

40 comparar (1408) la puntuación de importancia de cada uno del uno o varios resultados de la geolocalización con
 un umbral de puntuación de importancia;
 descartar (1416) el uno o varios resultados de la geolocalización cuando la puntuación de importancia sea menor
 que el umbral de puntuación de importancia;
 analizar (1420) una varianza de la puntuación de importancia cuando la puntuación de importancia sea mayor que
 el umbral de puntuación de importancia;
 mantener (1428) el uno o varios resultados de la geolocalización cuando la varianza de la puntuación de
 importancia sea mayor que un umbral de varianza; y
 45 descartar (1416) el uno o varios resultados de la geolocalización cuando la varianza de la puntuación de
 importancia sea menor que el umbral de puntuación de importancia.

50 12. El método de la reivindicación 10, en donde calificar el uno o varios resultados de la geolocalización usando
 heurística comunitaria comprende determinar si la al menos una publicación en redes sociales se recibe en modo de
 publicación única en el que el método determina la información de geolocalización basándose en publicaciones únicas
 procedentes de cuentas de usuarios, o en un modo de publicación múltiple en el que el método determina la
 información de geolocalización basándose en múltiples publicaciones procedentes de cuentas de usuarios,
 en donde, en el modo de publicación única, el método comprende además:

55 comparar (1512) el nombre de una cuenta de la al menos una publicación en redes sociales con una lista
 taxonómica;
 mantener (1520) el uno o varios resultados de la geolocalización cuando el nombre de la cuenta coincide con la
 lista taxonómica; y
 descartar (1524) el uno o varios resultados de la geolocalización cuando las palabras no coinciden con la lista
 60 taxonómica,

en donde, en el modo de publicación múltiple, el método además comprende:

65 determinar (1532) si el uno o varios resultados de la geolocalización coinciden con una ubicación del usuario para
 al menos un porcentaje predeterminado de usuarios;
 mantener (1520) el uno o varios resultados de la geolocalización cuando coincida con la ubicación de usuario de

al menos el porcentaje predeterminado de usuarios; y
 descartar (1524) el uno o varios resultados de la geolocalización cuando no coincida con la ubicación de usuario
 de al menos el porcentaje predeterminado de usuarios.

5 13. El método de la reivindicación 10, en donde determinar al menos una geolocalización primaria a partir del uno o
 varios resultados de la geolocalización comprende al menos uno de:

ordenar (1608) el uno o varios resultados de la geolocalización basándose en al menos una ubicación de usuario
 calificada;

10 ordenar (1612) el uno o varios resultados de la geolocalización basándose en los resultados de un método de
 identificación basado en taxonomía; o

ordenar (1616) el uno o varios resultados de la geolocalización basándose en los resultados de al menos uno de:
 un método de identificación basado en heurística y un método de identificación basado en una base de
 conocimientos (KB).

15 14. El método de la reivindicación 13, que comprende además identificar (1620) las coordenadas geográficas de la al
 menos una ubicación primaria basándose en la distancia más corta entre una ubicación del usuario y el uno o varios
 resultados de la geolocalización.

20 15. Un medio de almacenamiento legible por máquina no transitorio que incluye instrucciones de programación, que
 cuando son ejecutadas por el al menos un procesador realizan un método para determinar ubicaciones de
 publicaciones en redes sociales, el método comprendiendo:

25 recuperar (208), mediante la comunicación con al menos una interfaz de programación de aplicaciones (API) de
 un sistema de redes sociales (104) en una o varias primeras redes de comunicaciones (128), al menos una
 publicación en redes sociales;

determinar (216) al menos una mención de ubicación en la al menos una publicación en redes sociales;

30 detectar si aparece al menos un alias en una base de conocimientos, KB, en la al menos una publicación en redes
 sociales, la KB incluyendo asociaciones entre ubicaciones y alias, una asociación entre una ubicación en particular
 y un alias correspondiente en la KB estando basada en una frecuencia con la que la ubicación en particular y el
 alias correspondiente aparecen simultáneamente en publicaciones en redes sociales, la asociación entre la
 ubicación en particular y el alias correspondiente estando en la KB cuando la ubicación en particular y el alias
 correspondiente tienen una ordenación de aparición simultánea recíproca entre sí;

35 agregar a la al menos una mención de ubicación una ubicación correspondiente al alias detectado en la KB;

determinar (220) al menos una ubicación basándose en la al menos una mención de ubicación;

determinar (224) una ubicación primaria a partir de la al menos una ubicación;

40 almacenar (228), en al menos una base de datos en el medio de almacenamiento legible por máquina no transitorio,
 al menos un conjunto de coordenadas geográficas para la ubicación primaria en al menos un objeto de publicación
 para la al menos una publicación en redes sociales; y

45 enviar (232), mediante la comunicación con un sistema de usuario (116) usando una o varias redes de
 comunicaciones secundarias (156, 164), la al menos una publicación en redes sociales con al menos un conjunto
 de coordenadas geográficas almacenado para su presentación en el sistema de usuario (116).

16. Un sistema (100) para determinar ubicaciones para publicaciones en redes sociales, el sistema comprendiendo:

45 al menos un procesador; y

un medio de almacenamiento legible por máquina no transitorio que incluye instrucciones de programación que,
 cuando son ejecutadas por el al menos un procesador, realizan un método para determinar ubicaciones para
 publicaciones en redes sociales, el método comprendiendo:

50 recuperar (208), mediante la comunicación con al menos una interfaz de programación de aplicaciones (API)
 de un sistema de redes sociales (104), al menos una publicación en redes sociales;

determinar (216) al menos una mención de ubicación en la al menos una publicación en redes sociales;

55 detectar si aparece al menos un alias en una base de conocimientos, KB, en la al menos una publicación en
 redes sociales, la KB incluyendo asociaciones entre ubicaciones y alias, una asociación entre una ubicación
 en particular y un alias correspondiente en la KB estando basada en la frecuencia con la que la ubicación en
 particular y el alias correspondiente aparecen simultáneamente en publicaciones en redes sociales, la
 asociación entre la ubicación en particular y el alias correspondiente estando en la KB cuando la ubicación en
 particular y el alias correspondiente tienen una ordenación de aparición simultánea recíproca entre sí;

60 agregar a la al menos una mención de ubicación una ubicación correspondiente al alias detectado en la KB;

determinar (220) al menos una ubicación basándose en la al menos una mención de ubicación;

determinar (224) una ubicación primaria a partir de la al menos una ubicación;

65 almacenar (228), en al menos una base de datos en el medio de almacenamiento legible por máquina no
 transitorio, al menos un conjunto de coordenadas geográficas para la ubicación primaria en al menos un objeto
 de publicación para la al menos una publicación en redes sociales; y

enviar (232), mediante la comunicación con un sistema de usuario (116) usando una o varias redes de

comunicaciones secundarias (156, 164), la al menos una publicación en redes sociales con al menos un conjunto de coordenadas geográficas almacenado para su presentación en el sistema de usuario (116).

5

10

15

20

25

30

35

40

45

50

55

60

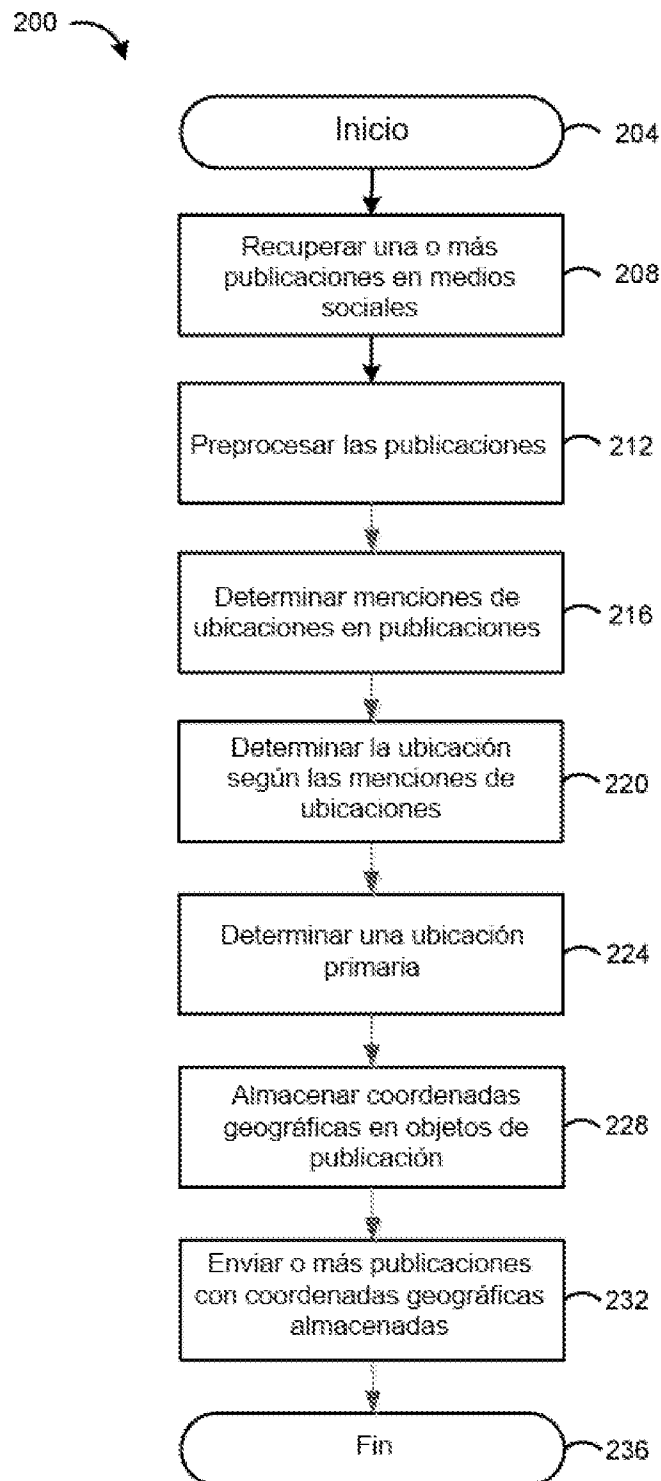


FIG. 2

400

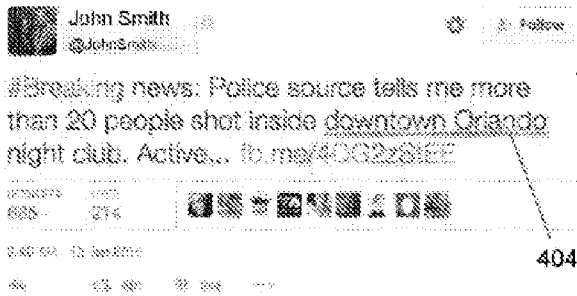


FIG. 4(a)

FIG. 4(b)

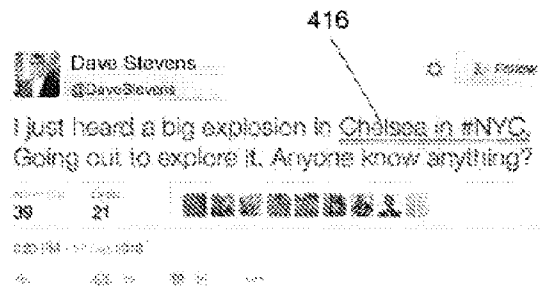
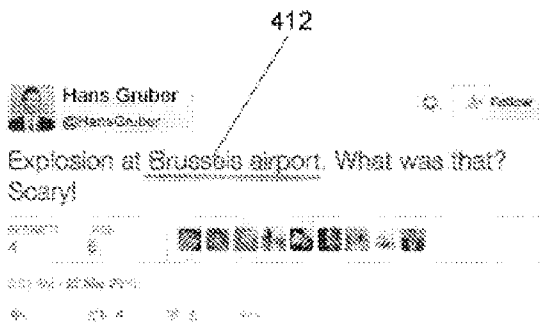


FIG. 4(c)

FIG. 4(d)

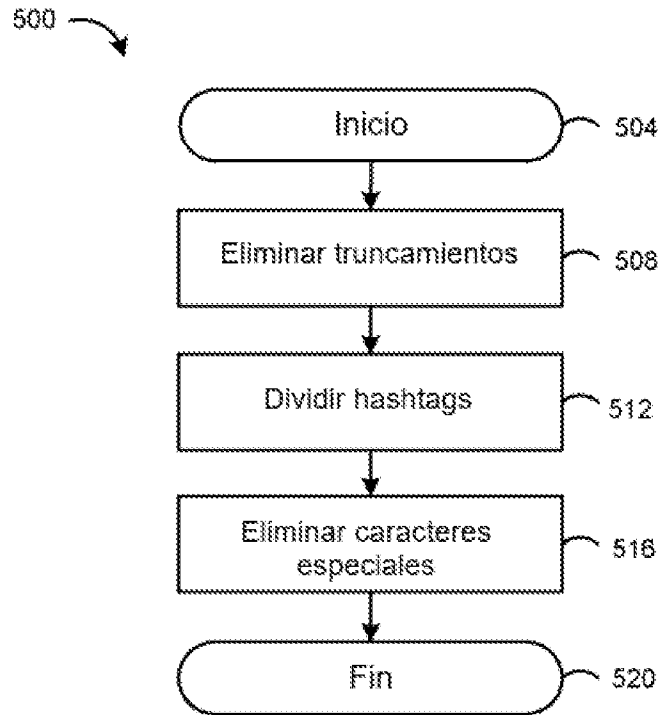


FIG. 5

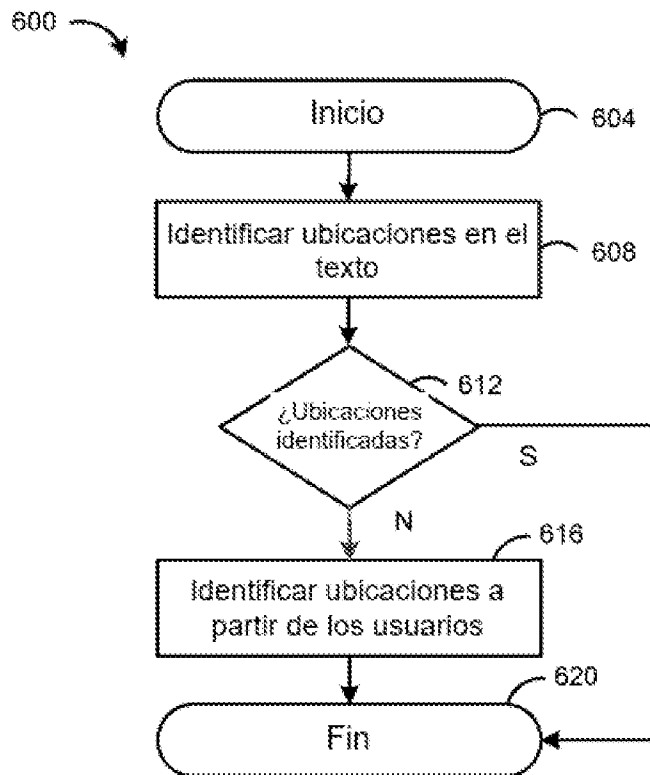


FIG. 6

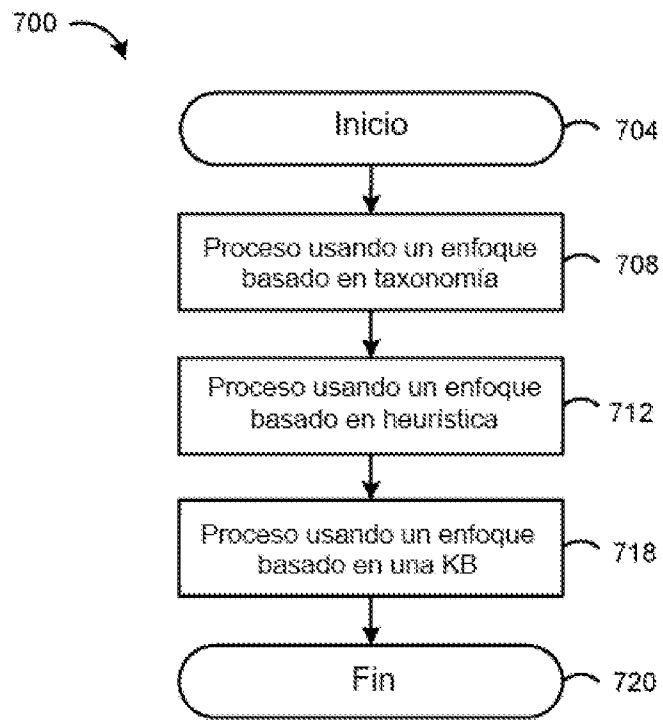


FIG. 7

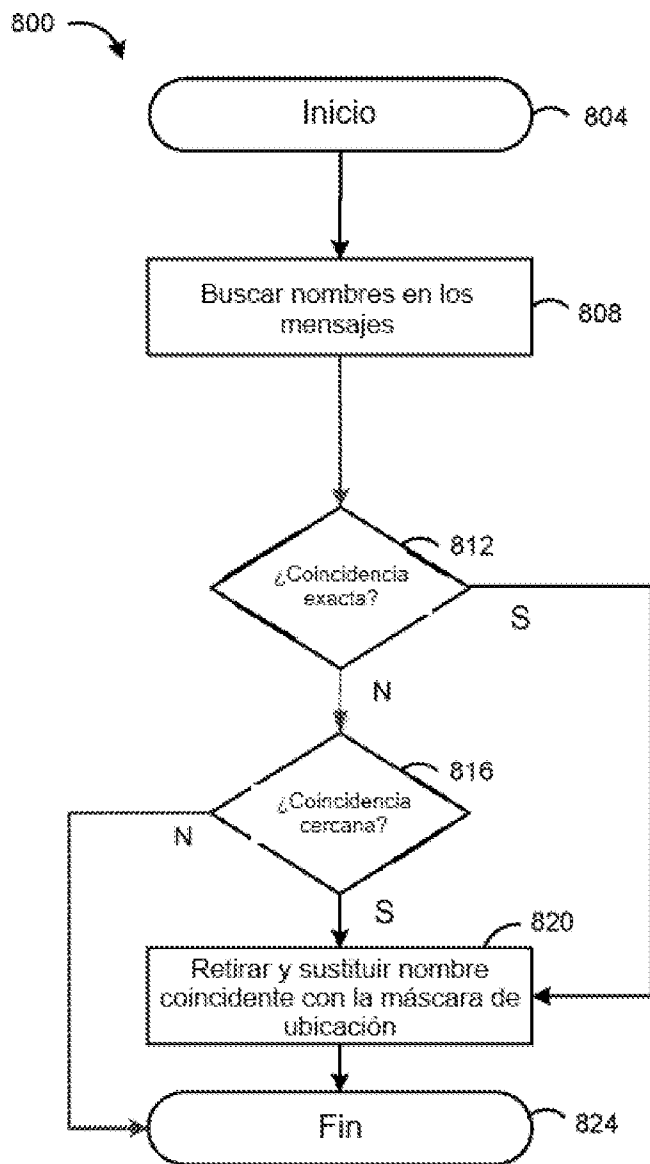


FIG. 8

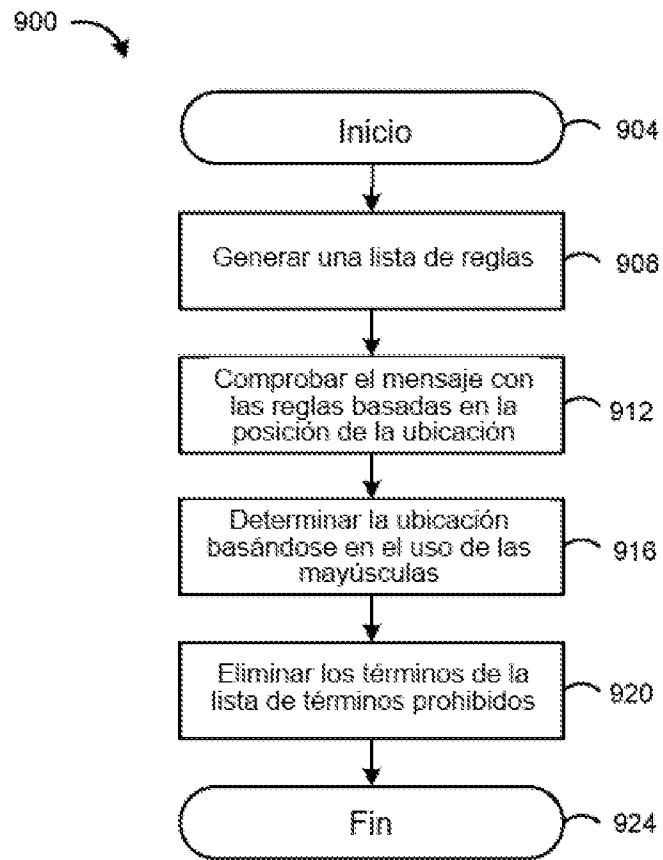


FIG. 9

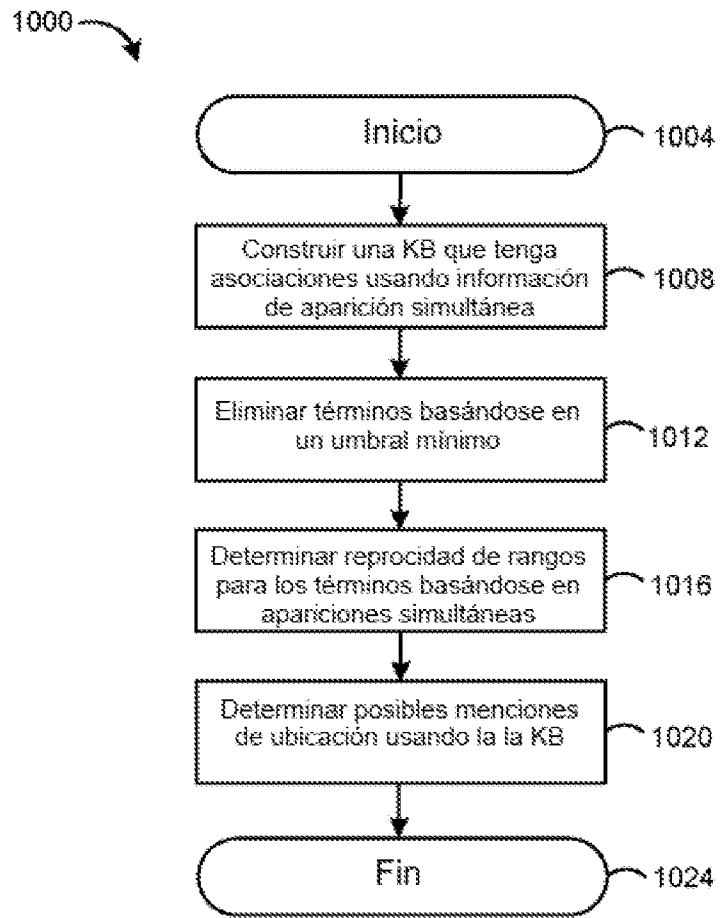


FIG. 10

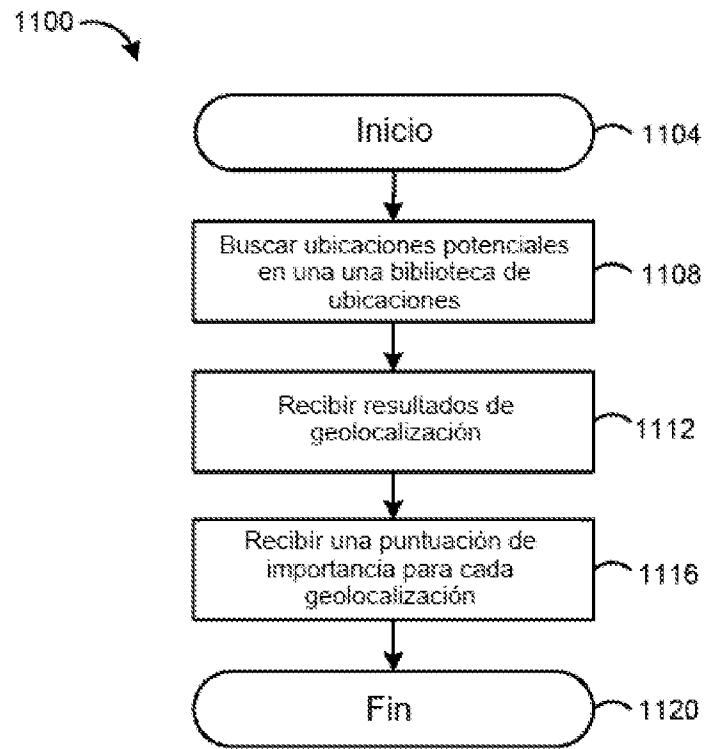


FIG. 11

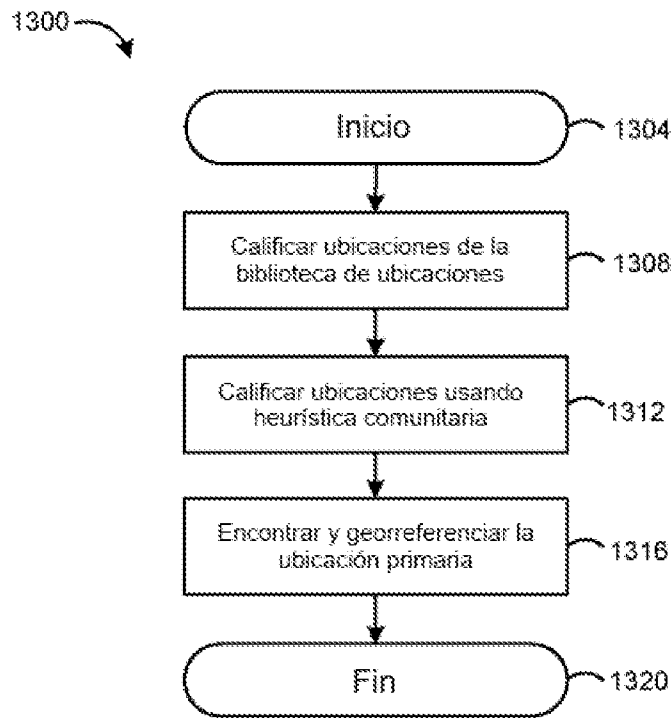


FIG. 13

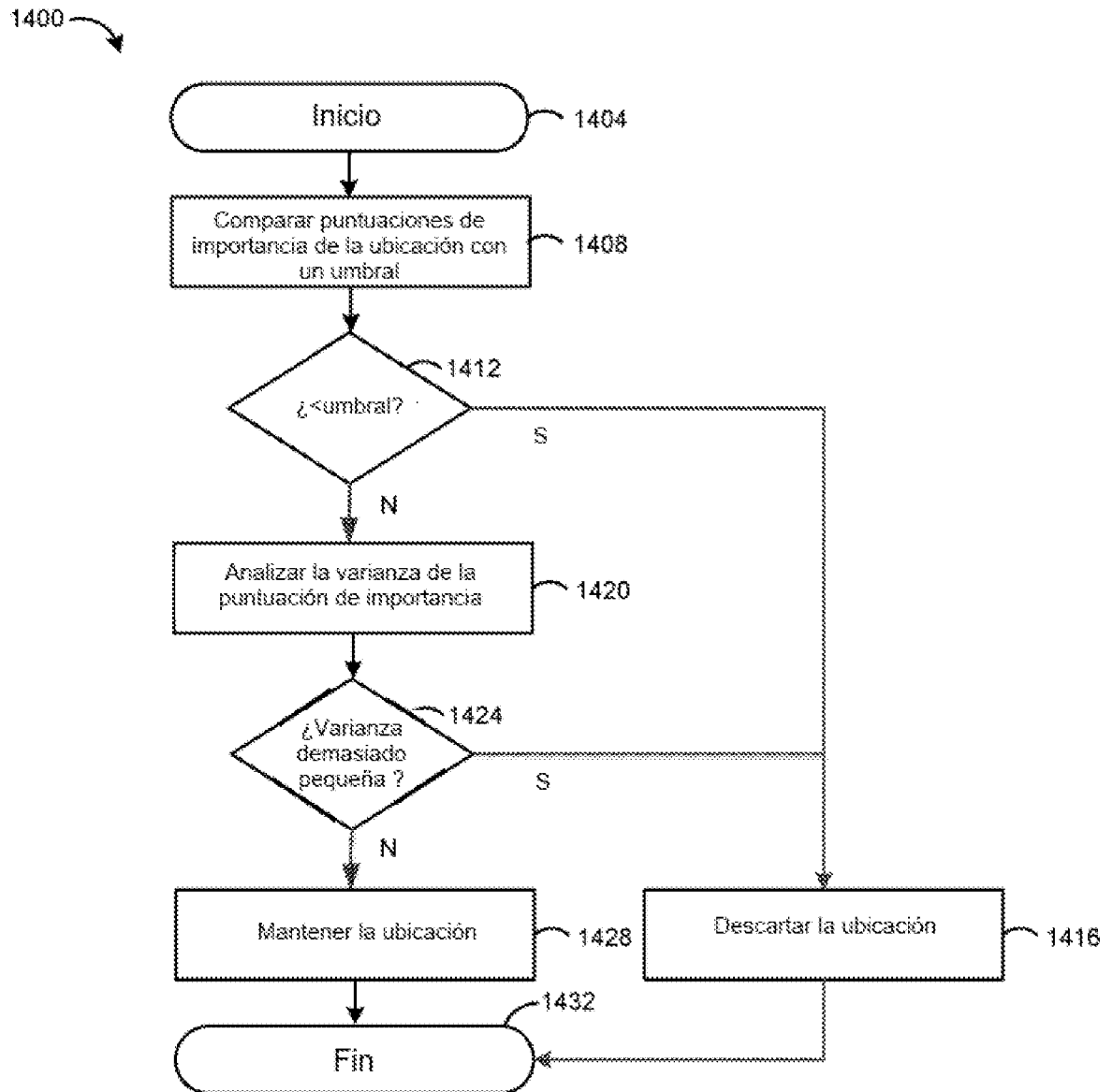


FIG. 14

1500

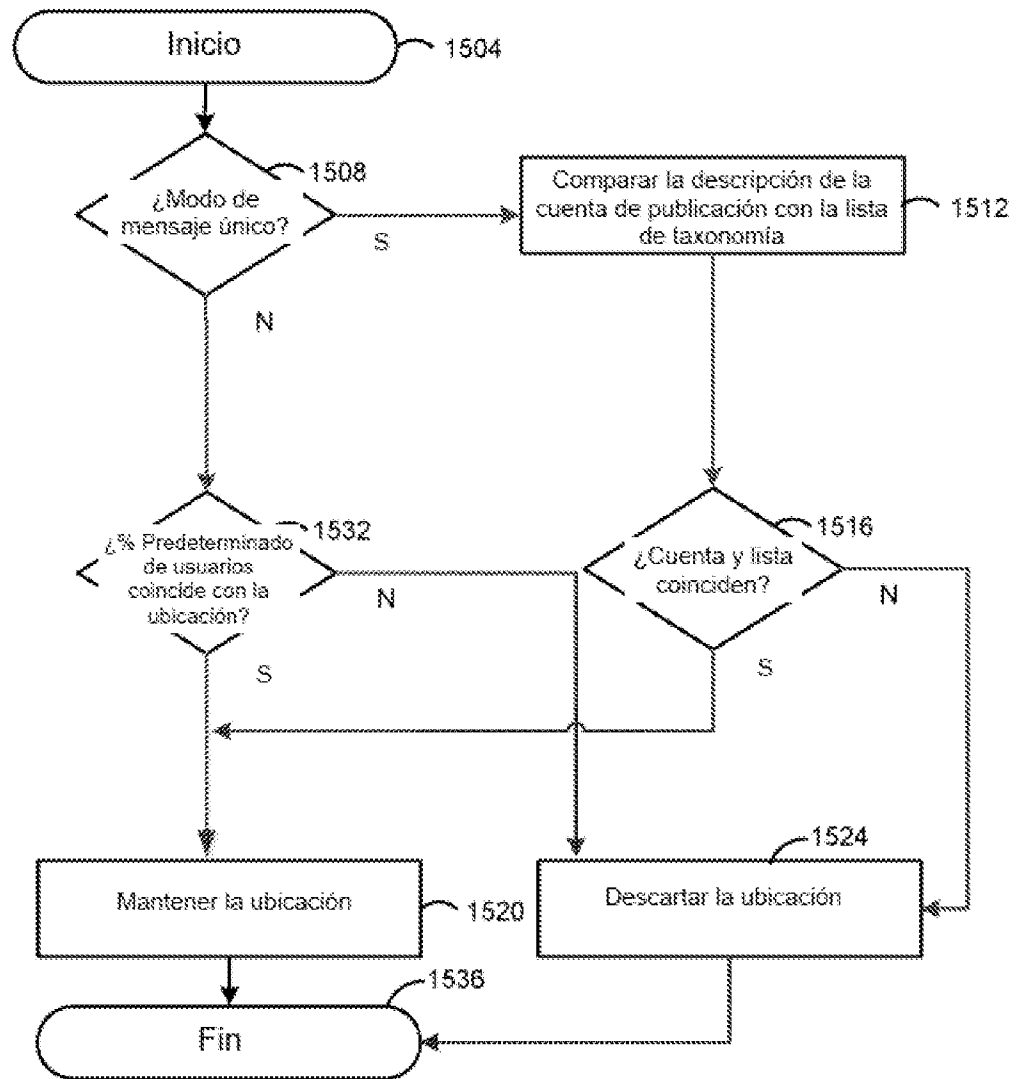


FIG. 15

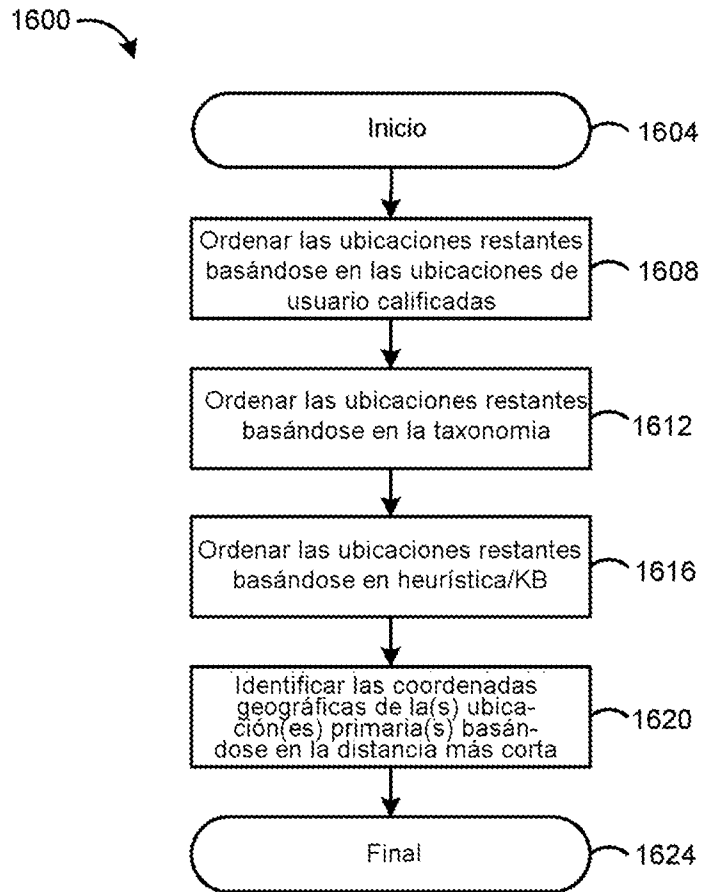


FIG. 16