

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 753 524**

51 Int. Cl.:

G16H 10/20 (2008.01)

G16H 10/60 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **07.12.2015** E 15198304 (6)

97 Fecha y número de publicación de la concesión europea: **07.08.2019** EP 3089060

54 Título: **Sistema para calcular similitudes entre fármacos y utilizar las similitudes para extrapolar efectos secundarios**

30 Prioridad:

30.04.2015 GB 201507501

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

13.04.2020

73 Titular/es:

**FUJITSU LIMITED (100.0%)
1-1, Kamikodanaka 4-chome, Nakahara-ku
Kawasaki-shi Kanagawa 211-8588, JP**

72 Inventor/es:

**MUÑOZ JIMÉNEZ, EMIR FERNANDO;
ABDELRAHMAN, AHMED;
NOVACEK, VIT y
VANDENBUSSCHE, PIERRE-YVES**

74 Agente/Representante:

CARVAJAL Y URQUIJO, Isabel

ES 2 753 524 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 para calcular similitudes entre fármacos y utilizar las similitudes para extrapolar efectos secundarios

5 La presente invención se refiere a un método que permite la señalización en etapa temprana de efectos secundarios de fármacos utilizando datos abiertos. El método puede determinar los efectos secundarios potencialmente asociados con los fármacos, mediante el cálculo de la similitud de los fármacos en función de las propiedades intrínsecas. La invención pertenece al campo de la farmacoinformática y tiene aplicaciones en todo el mundo de la medicina.

Antecedentes de la invención

10 Los efectos secundarios son efectos no buscados o no deseados del uso de un fármaco (cualquier sustancia (con la excepción de los alimentos y el agua) que altera la función del cuerpo, para humanos o animales) que ocurren junto con los efectos deseados del fármaco. Ejemplos incluyen agitación, dolores de cabeza, infección de heridas y náuseas. A menudo se utilizan ensayos y estudios previos a la comercialización de fármacos para determinar inicialmente los efectos secundarios asociados con un fármaco. Sin embargo, los ensayos de fármacos previos a la comercialización son, en general, costosos y riesgosos. A pesar de tales estudios, se informan numerosos efectos secundarios desconocidos después de la liberación al mercado de un fármaco (es decir, en un período posterior a la

15 comercialización), que puede, en muchos casos, causar una morbilidad sustancial e incluso mortalidad. Cuando ocurren tales eventos, entre otras acciones, el fármaco se retira del mercado, lo que puede dañar enormemente la imagen y la reputación de una compañía farmacéutica y causar problemas legales. Por lo tanto, sistemas de bajo coste para detectar de manera prematura (es decir, en la etapa temprana de la fabricación de fármacos) los efectos secundarios de los fármacos ayudarían a minimizar los costes futuros, así como a evitar riesgos tanto para los

20 productos farmacéuticos como para el público en general.

Para hacer frente a la falta de conclusión de los ensayos clínicos previos a la comercialización, los efectos secundarios se controlan a nivel mundial después de la liberación de un fármaco por parte de organizaciones como la Food and Drug Administration (FDA). Estas organizaciones recopilan informes sobre los efectos secundarios de los fármacos a partir de múltiples fuentes, como pacientes, farmacéuticos, médicos y fabricantes de productos farmacéuticos.

25 Además, las agencias gubernamentales reguladoras han establecido sistemas de informes que permiten al público en general informar los efectos secundarios no deseados del uso de un fármaco. Recientemente, se han empleado soluciones basadas en software para optimizar el proceso y el almacenamiento de cualquier informe generado por el público.

30 Debido al coste y al tiempo que requieren los ensayos preclínicos, muchos fármacos se liberan con efectos secundarios no descubiertos. Por esta razón, después de que se libera un fármaco, sus efectos secundarios se descubren principalmente por casualidad debido a reacciones adversas del paciente. Más recientemente, se han hecho intentos para diseñar técnicas y métodos que permitan que tales eventos se descubran prematuramente, evitando daños a los pacientes. Esto se ha hecho mediante el uso de datos disponibles, como informes históricos, para tratar de determinar causalidades, inferencias y correlaciones.

35 Los sistemas de software existentes que predicen la aparición de un efecto secundario con el uso de un determinado fármaco(s) se basan principalmente en datos posteriores a la comercialización o se generan después de la comercialización. A este respecto, dichos datos pueden incluir: datos del paciente, informes de eventos adversos del público en general o estadísticas posteriores a la comercialización. Los sistemas requieren la disponibilidad de dichos datos para funcionar. A pesar de esto, los métodos existentes tienen las siguientes características/inconvenientes, que

40 los inventores han clasificado en cuatro grupos, de la siguiente manera:

1. Métodos "*Posteriores a la comercialización*" que requieren la liberación de un fármaco abiertamente al mercado con el fin de recopilar, después de la liberación, comentarios médicos de las partes interesadas (por ejemplo, agencias de informes, pacientes). Estos métodos se basan principalmente en la inferencia estadística que se detecta a través de incidentes de eventos adversos, es decir, después de que el efecto secundario ha sido soportado por los pacientes.

45 Estos métodos también pueden contener un lapso de tiempo significativo entre los efectos secundarios que se experimentan y se informan a las agencias. Además, la recopilación de dichos comentarios de una muestra estadísticamente significativa requiere mucho tiempo y esfuerzo, lo que provoca más demoras y puede poner en peligro al público en general.

50 2. Los métodos basados en "*Datos del paciente*" requieren la recopilación de datos personales y confidenciales de los pacientes. Estos métodos son de naturaleza intrusiva y tienen un alto coste debido a los diferentes pasos legales que se necesitan para adquirir dichos datos. Estos métodos tienden a estar asociados con ensayos clínicos muy caros y monitoreo de pacientes, y generalmente son de alto riesgo en caso de uso indocumentado de un fármaco.

55 3. Métodos de "*Datos pequeños*" donde los efectos secundarios se predicen en base a vectores de dimensiones pequeñas de características extraídas de un solo fármaco. Por ejemplo, estos métodos no tienen en cuenta las características de múltiples fármacos al mismo tiempo y los efectos secundarios asociados con esos fármacos.

4. Métodos de "*Datos completos*" en los que se necesita un conjunto de características (rasgos) requeridas para que el método funcione. En estos métodos, cada fármaco en cuestión debe someterse a investigaciones exhaustivas para determinar las características intrínsecas necesarias. En muchos escenarios, los fabricantes de fármacos y las agencias reguladoras pueden no tener todas estas características y por lo tanto no sería posible predecir los efectos secundarios.
- 5 Willighagen E. L. et al. (2013) "The ChEMBL database as linked open data", Journal of Cheminformatics, vol. 5, no. 1, page 2, (XP021155887) revela desarrollos recientes en un proyecto en curso que convierte datos de la base de datos ChEMBL en triples RDF.
- 10 Willighagen E. L. et al. (2010) "Linking Open Drug Data to Cheminformatics and Proteochemometrics", Proceedings of the Workshop on SWAT4LS, Amsterdam, The Netherlands, Nov 20, 2009, CEUR Workshop Proceedings 559, (XP055267705) se enfoca en hacer que los datos relacionados con el descubrimiento de fármacos estén disponibles como RDF, y aún más en la integración de los enfoques de RDF con el análisis de datos de información molecular en campos de descubrimiento de fármacos como la quiminformática y la proteoquímica.
- 15 QU X. A. et al. (2009) "Inferring novel disease indications for known drugs by semantically linking drug action and disease mechanism relationships", BMC Bioinformatics, vol. 10, no. Suppl 5, page S4, (XP055267676) desarrolla un marco de conocimiento capaz de modelar e interconectar acciones de fármacos y mecanismos de enfermedades en diversos contextos de sistemas biológicos, mediante el diseño de una Ontología de Correlación de Fármacos y Enfermedades (DDCO), formalizada en OWL, que integra múltiples ontologías vocabularios controlados y esquemas de datos y los vincula con diversos conjuntos de datos extraídos de dominios farmacológicos y biológicos.
- 20 Las realizaciones de la presente invención intentan superar los inconvenientes mencionados anteriormente, a través de un método que i) es un método de detección temprana que utiliza datos disponibles antes del lanzamiento al mercado, y/o ii) explota el uso de datos de bajo coste mediante el uso de datos de acceso abierto y no se basa en o no requiere datos de pacientes y/o ensayos experimentales, y/o iii) es capaz de operar con datos incompletos al tiempo que permite la integración incremental del conocimiento adquirido a través del proceso de fabricación de fármacos y/u
- 25 otras partes interesadas.
- De acuerdo con una realización de un aspecto de la invención, se proporciona un sistema configurado para calcular similitudes entre fármacos y configurado para usar las similitudes para extrapolar los efectos secundarios de un fármaco específico, el sistema comprende:
- un aparato de cálculo de similitud que comprende:
- 30 un módulo de integración de datos configurado para obtener datos relacionados con un conjunto de fármacos de una pluralidad de fuentes de datos abiertos e integrar los datos para generar triples RDF;
- un módulo RDF configurado para almacenar un gráfico RDF de los triples RDF; y
- un módulo de cálculo de similitud configurado para recuperar una lista de todos los fármacos presentes en el gráfico RDF y calcular la similitud de cada fármaco con todos los demás fármacos en el gráfico RDF; donde
- 35 el módulo de cálculo de similitud está configurado para encontrar un conjunto de predicados y objetos/sujetos de RDF triples en los cuales, para cada fármaco, el fármaco es un sujeto o un objeto respectivamente y para generar un vector de característica a partir de los predicados y objetos/sujetos que representan el características del fármaco, en donde el módulo de cálculo de similitud está configurado además para calcular la similitud de cada fármaco con cualquier otro fármaco en el gráfico RDF en comparación con los vectores de características de los dos fármacos; donde
- 40 el módulo de cálculo de similitud está configurado para calcular la similitud entre dos fármacos teniendo en cuenta el número de características de intersección de los vectores de características, en las cuales el predicado y el sujeto u objeto es el mismo para ambos fármacos; y en donde
- el módulo de cálculo de similitud está configurado para almacenar las similitudes calculadas como distancias fármaco-fármaco para el conjunto de fármacos en una base de datos de similitud a la que está conectado el aparato de cálculo de similitud; y
- 45 un aparato de determinación de efectos secundarios configurado para extrapolar los efectos secundarios de un fármaco específico a partir de datos abiertos relacionados con el conjunto de fármacos, con:
- una conexión a la base de datos de similitud;
- una interfaz de usuario configurada para permitir que un usuario especifique un fármaco del conjunto de fármacos;
- 50 un módulo de clasificación de vecinos configurado para usar las similitudes de fármacos para obtener vecinos más cercanos del fármaco especificada en términos de similitudes; y

un módulo de propagación de efectos secundarios configurado para recopilar efectos secundarios conocidos de los vecinos más cercanos y para combinar y clasificar los efectos secundarios de acuerdo con su ocurrencia conjunta entre los vecinos.

5 Dado que el método aprende de los efectos secundarios conocidos identificados para fármacos en el pasado, cuantos más fármacos contenga el sistema, mejor. Sin embargo, el conjunto de fármacos puede corresponder a un tratamiento médico específico o subgrupo, omitiendo otros fármacos.

El RDF proporciona las conexiones entre fármacos para que el método de realización de la invención funcione. Los datos en otros formatos de datos, como bases de datos relacionales, archivos CSV, archivos JSON, etc., se pueden usar con una asignación correspondiente a RDF.

10 Las realizaciones de la invención usan datos abiertos para proporcionar datos de similitud de fármacos a partir de un gráfico RDF integrado, que luego puede usarse en un aparato de cálculo de similitud (descrito más adelante) para extrapolar los efectos secundarios de un fármaco específico de otros fármacos.

15 RDF (Marco de Descripción de Recursos) es un método general para la descripción conceptual o el modelado de información que es un estándar para las redes semánticas. La estandarización del modelado de información en una red semántica permite la interoperabilidad entre aplicaciones que operan en una red semántica común. RDF mantiene un vocabulario con semántica formal inequívoca, al proporcionar el Esquema RDF (ERDF) como lenguaje para describir vocabularios en RDF.

20 Los triples proporcionan la codificación de los datos del gráfico caracterizando los datos del gráfico como una pluralidad de expresiones sujeto-predicado-objeto. En ese contexto, el sujeto y el objeto son nodos gráficos de los datos gráficos, y como tales son entidades, objetos, instancias o conceptos, y el predicado es una representación de una relación entre el sujeto y el objeto. El predicado afirma algo sobre el sujeto al proporcionar un tipo específico de enlace al objeto. Por ejemplo, el sujeto puede denotar un recurso web (por ejemplo, a través de un Identificador de Recurso Único, URI), el predicado denota un rasgo, característica o aspecto particular del recurso, y el objeto denota una instancia de ese rasgo, característica, o aspecto (quizás otro URI). En otras palabras, una colección de declaraciones triples representa intrínsecamente datos de gráficos direccionales. El estándar RDF proporciona una estructura formalizada para tales triples. Se puede consultar un gráfico RDF utilizando el protocolo SPARQL y el lenguaje de consulta RDF (SPARQL).

30 Preferiblemente, el módulo de integración de datos está configurado para permitir la integración de nuevos datos durante el tiempo de ejecución (es decir, después de una configuración inicial), y el módulo RDF está configurado para agregar los nuevos datos al gráfico RDF. Los nuevos datos pueden ser de fuentes de datos existentes o de nuevas fuentes de datos. Las fuentes de datos no necesariamente deben expresarse en RDF inicialmente; se pueden convertir a RDF usando un lenguaje de mapeo. La fuente puede ser una base de datos relacional, un archivo CSV (valor separado por comas), un archivo JSON (JavaScript ObjectNotation), etc.

35 El módulo de integración de datos puede configurarse para reconocer términos alternativos para el mismo fármaco y clasificar todos los datos asociados con los términos alternativos como pertenecientes al mismo fármaco. Por ejemplo, el módulo puede usar el lenguaje web de ontología (OWL) para procesar automáticamente el idioma descargado, por ejemplo, para definir una relación "igual que".

El módulo de integración de datos puede configurarse para usar nombres alternativos, sinónimos y medidas de distancia de cadena para reconocer términos alternativos para el mismo fármaco.

40 Además, el módulo de integración de datos puede configurarse para vincular diferentes instancias del mismo fármaco, por ejemplo, utilizando un identificador único. Todas estas características juntas pueden permitir la integración exitosa de datos de diferentes fuentes.

45 Por ejemplo, el módulo de integración de datos puede configurarse para integrar datos de una o más de las bases de datos de propiedades de fármacos, efectos secundarios, clasificación de fármacos y enfermedades. La lista de bases de datos se puede ampliar para incluir otras bases de datos.

50 En cuanto al módulo de cálculo de similitud, utiliza la funcionalidad RDF para calcular la similitud. Por ejemplo, se puede configurar para encontrar un conjunto de predicados y objetos/sujetos de RDF triples en los cuales, para cada fármaco, el fármaco es un sujeto o un objeto respectivamente y generar un vector de característica a partir de los predicados y objetos/sujetos que representan las características del fármaco, en donde el módulo de cálculo de similitud está configurado además para calcular la similitud de cada fármaco con cualquier otro fármaco en el gráfico RDF en comparación con los vectores de características de los dos fármacos. En este contexto, cada una de las n combinaciones de predicado y objeto/sujeto se puede ver como una característica y el vector de características se puede ver como un vector n-dimensional.

55 Los efectos secundarios se pueden incluir en el cálculo de similitudes o no, no son un requisito para que el método funcione. Cualquier atributo de un fármaco puede usarse como una característica para calcular la similitud, incluidos los efectos secundarios (si los hay). Si el método se aplica sobre un nuevo fármaco sin ningún efecto secundario previo

conocido, el método informará sobre nuevos efectos secundarios solo en función de su composición química. Si el método se aplica a un fármaco existente con efectos secundarios, el método informará efectos secundarios adicionales. El método está diseñado para trabajar con datos parciales.

5 El módulo de cálculo de similitud puede configurarse para calcular la similitud entre dos fármacos teniendo en cuenta el número de características de intersección de los vectores de características, en donde el predicado y el sujeto u objeto (que constituye el triple junto con el fármaco) es el mismo para ambos fármacos. Por lo tanto, en los triples que se comparan, los fármacos son diferentes, pero el predicado y el tercer elemento (sujeto u objeto) en el triple son los mismos.

10 El módulo de cálculo de similitud puede configurarse para almacenar las similitudes calculadas como distancias de fármaco a fármaco para el conjunto de fármacos en una base de datos de similitud a la que está conectado el aparato de cálculo de similitud.

15 El aparato de determinación de efectos secundarios utiliza las similitudes calculadas para buscar posibles efectos secundarios. El aparato de determinación de efectos secundarios está configurado de este modo para extrapolar los efectos secundarios de un fármaco específico en un conjunto de fármacos a partir de datos abiertos relacionados con el conjunto de fármacos. El aparato comprende: una conexión a una base de datos de similitud, la base de datos de similitud está configurada para almacenar datos de similitud derivados de una pluralidad de fuentes de datos abiertos, los datos de similitud proporcionan similitud de fármaco entre cada fármaco y cada otro fármaco en un conjunto de fármacos; una interfaz de usuario configurada para permitir que un usuario especifique un fármaco del conjunto de fármacos; un módulo de clasificación de vecinos configurado para usar las similitudes de fármacos para obtener vecinos más cercanos del fármaco especificado en términos de similitudes; y un módulo de propagación de efectos secundarios configurado para recopilar efectos secundarios conocidos de los vecinos más cercanos y para combinar y clasificar los efectos secundarios de acuerdo con su concurrencia conjunta entre los vecinos.

20 Por lo tanto, el aparato de determinación de efectos secundarios no se basa en ningún efecto secundario que pueda estar indicado para el fármaco especificado, sino que busca los fármacos más cercanos al fármaco especificado, en función de los datos de similitud, y propaga los efectos secundarios de estos fármacos más similares, utilizando un método de clasificación.

Los datos de similitud pueden proporcionarse como distancias fármaco-fármaco en un modelo de espacio vectorial en un espacio n-dimensional, y el módulo de clasificación de vecinos puede configurarse para seleccionar k vecinos más cercanos en función de las distancias dentro del espacio n-dimensional.

30 El módulo de propagación de efectos secundarios puede configurarse para calcular la frecuencia de las características del fármaco que ocurren con cada efecto secundario encontrado entre los efectos secundarios conocidos para construir una correlación entre cada efecto secundario y cada característica del fármaco.

35 Por ejemplo, las características del fármaco pueden incluir cualquiera de las propiedades del fármaco, la clasificación del fármaco y las enfermedades, que ya están incluidas en los datos de similitud. Las características utilizadas aquí pueden ser las mismas que las utilizadas para calcular los datos de similitud, o un subgrupo de características. El procesamiento de las características es diferente aquí en la función aplicada. Claramente los efectos secundarios son ahora los pivotes.

40 El aparato de determinación de efectos secundarios puede comprender un módulo de notificación configurado para emitir una advertencia, alarma o notificación con una lista de efectos secundarios y/o si se extrapolan ciertos efectos secundarios (por ejemplo, más graves o especificados).

45 La invención puede implementarse en circuitos electrónicos digitales, o en hardware de ordenador, firmware, software o en combinaciones de ellos. La invención puede implementarse como un programa informático o producto de programa informático, es decir, un programa informático incorporado de forma tangible en un soporte de información, por ejemplo, en un dispositivo de almacenamiento legible por máquina o en una señal propagada, para su ejecución o para controlar la operación de uno o más módulos de hardware. Un programa de ordenador puede tener la forma de un programa independiente, una porción de programa de ordenador o más de un programa de ordenador y puede escribirse en cualquier forma de lenguaje de programación, incluidos los lenguajes compilados o interpretados, y puede implementarse en cualquier forma, incluso como un programa independiente o como un módulo, componente, subrutina u otra unidad adecuada para su uso en un entorno de sistema de comunicación. Un programa de ordenador puede implementarse para ejecutarse en un módulo o en múltiples módulos en un sitio o distribuirse en múltiples sitios e interconectarse mediante una red de comunicación.

55 Los pasos del método de la invención pueden ser realizados por uno o más procesadores programables que ejecutan un programa de ordenador para realizar funciones de la invención operando datos de entrada y generando salida. El aparato de la invención puede implementarse como hardware programado o como circuito lógico de propósito especial, que incluye, por ejemplo, un FPGA (matriz de compuerta programable en campo) o un ASIC (circuito integrado específico para la aplicación).

Los procesadores adecuados para la ejecución de un programa informático incluyen, a modo de ejemplo, microprocesadores de propósito general y especial, y uno o más procesadores de cualquier tipo de ordenador digital. Generalmente, un procesador recibirá instrucciones y datos de una memoria de solo lectura o una memoria de acceso aleatorio o ambas. Los elementos esenciales de un ordenador son un procesador para ejecutar instrucciones
5 acopladas a uno o más dispositivos de memoria para almacenar instrucciones y datos.

La invención se describe en términos de realizaciones particulares. Otras realizaciones están dentro del alcance de las siguientes reivindicaciones. Por ejemplo, los pasos de la invención se pueden realizar en un orden diferente y aún lograr resultados deseables.

10 El aparato según las realizaciones preferidas se describe como configurado o dispuesto para llevar a cabo ciertas funciones. Esta configuración o disposición podría ser mediante el uso de hardware o middleware o cualquier otro sistema adecuado. En realizaciones preferidas, la configuración o disposición es por software.

15 Los elementos de la invención se han descrito usando términos tales como "módulo". El lector experto apreciará que estos términos y sus equivalentes pueden referirse a partes del sistema que están separadas espacialmente pero que se combinan para cumplir la función definida. Igualmente, las mismas partes físicas del sistema pueden proporcionar dos o más de las funciones definidas.

Las características y subcaracterísticas de los aspectos del sistema y del aparato pueden aplicarse a los aspectos del método y viceversa.

Breve descripción de los dibujos

Se hace referencia, solo a modo de ejemplo, a los dibujos adjuntos en los que:

20 La Figura 1 es un diagrama de bloques del aparato que representa un aparato de cálculo de similitud de acuerdo con una realización general de la invención;

La Figura 2 es un diagrama de bloques del aparato que representa un aparato de determinación de efectos secundarios de acuerdo con una realización general de la invención;

25 La Figura 3 es un diagrama de flujo del método en el aparato de cálculo de similitud de acuerdo con una realización general de la invención;

La Figura 4 es un diagrama de flujo del método en el aparato de determinación de efectos secundarios de acuerdo con una realización general de la invención;

La Figura 5 es un diagrama de bloques/flujo que representa un sistema/arquitectura de ejemplo de acuerdo con una realización de un aspecto de la invención;

30 La Figura 6 es un diagrama de bloques que muestra la interacción entre un módulo de integración de datos y un módulo de gráfico RDF de acuerdo con una realización de un aspecto de la invención;

La Figura 7 es un diagrama esquemático de un gráfico RDF de ejemplo de acuerdo con una realización de un aspecto de la invención;

35 La Figura 8 es un subconjunto de ejemplos de relaciones para dos fármacos de acuerdo con una realización de un aspecto de la invención;

La Figura 9 es un diagrama de bloques de un módulo de Cálculo de Similitud de acuerdo con una realización de un aspecto de la invención;

La Figura 10 es un ejemplo de un módulo de Clasificación de Fármacos Vecinos de acuerdo con una realización de un aspecto de la invención; y

40 La Figura 11 es un ejemplo de un módulo de Propagación de Efectos Secundarios según una realización de un aspecto de la invención.

45 La Figura 1 muestra un diagrama de bloques de un aparato de cálculo de similitud de acuerdo con una realización general de la presente invención. El aparato de cálculo de similitud puede realizarse como un servidor o una PC u otro aparato de cómputo con conexión a través de Internet. Comprende un módulo 105 de integración de datos, un módulo 106 RDF y un módulo 107 de cálculo de similitud. El aparato de cálculo de similitud también tiene acceso a una o más fuentes 100 de datos abiertas (bases de datos o más almacenes de datos simples). El aparato de cálculo de similitud puede estar conectado directa o indirectamente a la (s) fuente 100 (s) de datos abierta (s), a través de una conexión de red adecuada (por ejemplo, a través de Internet). Alternativa o adicionalmente, el aparato de cálculo de similitud puede comprender una o más fuentes de datos locales. La fuente 100 de datos abierta puede comprender uno o más
50 de: una base 101 de datos de perfil de fármaco; y/o una base 102 de datos de efectos secundarios; y/o una base 103 de datos de clasificación de fármacos; y/o una base de datos de enfermedades.

- 5 El aparato de cálculo de similitud utiliza datos abiertos para calcular similitudes entre fármacos. En primer lugar, el módulo 105 de integración de datos obtiene datos, que se relacionan con un conjunto de dos o más fármacos (por ejemplo, fármacos), de la fuente 100 de datos abierta. Los datos relacionados con el conjunto de fármacos se integran luego por el módulo 105 de integración de datos para generar RDF triple para cada uno de los fármacos en el conjunto de fármacos. El módulo 105 de integración de datos también puede, al integrar los datos, filtrar y/o unir los datos. Los RDF se triplican en relación con los datos para el conjunto de fármacos y luego se transfieren al módulo 106 RDF.
- 10 El módulo 106 RDF almacena un gráfico RDF de los triples RDF generados por el módulo 105 de integración de datos. El módulo 107 de cálculo de similitud luego usa el gráfico RDF almacenado en el módulo 106 RDF para recuperar una lista de todos los fármacos que están contenidos en el gráfico RDF. El módulo 107 de cálculo de similitud calcula entonces la similitud de los fármacos contenidas en el conjunto de fármacos con los otros fármacos en el conjunto de fármacos. Por ejemplo, para un conjunto de fármacos que contienen fármacos A, B y C, el módulo 107 de cálculo de similitud puede calcular la similitud entre el fármaco A y el fármaco B, la similitud entre el fármaco A y el fármaco C, y la similitud entre el fármaco B y el fármaco C. Estas similitudes calculadas pueden almacenarse en una base 109 de datos de similitud (no mostrada).
- 15 La Figura 2 muestra un diagrama de bloques de un aparato de determinación de efectos secundarios de acuerdo con una realización general de la presente invención. El aparato de determinación de efectos secundarios puede ser un servidor o un PC u otro aparato informático con conexiones a través de Internet. Comprende una interfaz 112 de usuario, un módulo 110 de clasificación de vecinos y un módulo 111 de propagación de efectos secundarios. El aparato de determinación de efectos secundarios también tiene acceso a una o más bases 109 de datos de similitud. El aparato de determinación de efectos secundarios puede estar conectado directa o indirectamente a las bases 109 de datos de similitud, a través de una conexión de red adecuada. Alternativa o adicionalmente, el aparato de determinación de efectos secundarios puede comprender además una o más bases 109 de datos de similitud locales. La base 109 de datos de similitud es capaz de almacenar datos de similitud relacionados con la similitud entre cada fármaco de un conjunto de fármacos y cualquier otro fármaco en el conjunto.
- 20 El aparato de determinación de efectos secundarios es capaz de determinar los efectos secundarios (o efectos secundarios potenciales/probables) asociados con un medicamento (o fármaco) especificado. La interfaz 112 de usuario permite a un usuario especificar un fármaco particular del conjunto de dos o más fármacos. El fármaco especificado se transfiere luego al módulo 110 de clasificación de vecinos, que utiliza los datos de similitud almacenados en la base 109 de datos de similitud para obtener los fármacos del conjunto de fármacos que son los vecinos más cercanos del fármaco especificado en términos de similitudes.
- 25 El módulo 111 de propagación de efectos secundarios puede recopilar una lista de efectos secundarios conocidos de los vecinos más cercanos seleccionados al fármaco especificado. Los efectos secundarios en la lista se combinan y clasifican por el módulo 111 de propagación de efectos secundarios, de acuerdo con su ocurrencia conjunta entre los fármacos vecinos. Es decir, los efectos secundarios que tienen más probabilidades de estar asociados con el fármaco especificado se clasifican más alto que los que no.
- 30 La Figura 3 es un diagrama de flujo que representa el método en el aparato de cálculo de similitud de acuerdo con una realización general de la presente invención. En primer lugar, en el paso S31, los datos relacionados con un conjunto de dos o más fármacos se obtienen de una o más fuentes de datos abiertas. Los datos adquiridos se integran luego y los RDF se triplican para el conjunto de fármacos en el paso S32. Un gráfico RDF de los triples RDF se almacena en el paso S33, para poder acceder en el paso S34 al calcular la similitud de cada fármaco en el conjunto de fármacos con todos los demás fármacos en el conjunto de fármacos en el paso S35.
- 35 La Figura 4 es un diagrama de flujo que representa el método en el aparato de determinación de efectos secundarios de acuerdo con una realización general de la presente invención. En primer lugar, en la etapa S41 se especifica un fármaco, preferiblemente por un usuario, a partir de un conjunto de dos o más fármacos. Las similitudes de fármacos almacenados en una base de datos de similitud de fármacos se usan luego en el paso S42 para determinar los fármacos del conjunto de fármacos que son los vecinos más cercanos del fármaco especificado en términos de similitudes. A continuación, se recopila una lista de los efectos secundarios conocidos de los fármacos vecinos más cercanos al fármaco especificado en el paso S43. Los efectos secundarios recopilados se combinan y clasifican en el paso S44 según su coexistencia entre los fármacos vecinos.
- 40 El método de las realizaciones de la invención puede verse como doble: 1) una vista centrada en el fármaco en la que todos los datos disponibles para fármacos se utilizan para calcular la similitud entre fármacos; y 2) una vista centrada en los efectos secundarios donde los efectos secundarios concurrentes se usan para determinar las probabilidades de ocurrencia de un efecto secundario en otros fármacos. El primer enfoque se puede usar para identificar los verdaderos efectos secundarios positivos de un fármaco determinado, mientras que el último enfoque se puede usar para detectar falsos negativos en los datos, es decir, los efectos secundarios que no tienen posibilidad de estar asociados con ciertos fármacos. Esta nueva configuración permite el método aquí presentado con la filtración de ruido en los resultados de salida, lo que resulta en mejores resultados de predicción.
- 45
- 50
- 55

Las realizaciones particulares de la presente invención descritas en el presente documento utilizan el conocimiento, en forma de Datos Vinculados (legible por máquina), para descubrir posibles efectos secundarios desconocidos de un fármaco. Las notificaciones, alertas o señales de advertencia pueden ser emitidas o comunicadas como resultado. Las principales iniciativas y gobiernos están publicando datos tales como datos de acceso abierto en la Web. Datos

5 Abiertos Vinculados (LOD) es un paradigma en donde los datos vinculados de una variedad de fuentes se ponen a disposición de las aplicaciones, por lo tanto, a los efectos de este documento, "Datos Vinculados" y "Datos Abiertos Vinculados" o "LOD" pueden usarse de manera más o menos intercambiable.

10 Esto contrasta con los enfoques actuales conocidos en la técnica, que hacen uso de fuentes de datos que son más lentas de recopilar, como, por ejemplo, informes de agencias y/o datos personales intrusivos de pacientes y/o informes de reacciones de pacientes riesgosos. Además, las realizaciones de la presente invención también pueden permitir la alimentación de otras fuentes de datos más nuevas, tales como informes de efectos secundarios y literatura biomédica. Esto puede requerir la aplicación de procesos de unión y filtrado adecuados.

15 El proceso de señalización se puede comparar con un sistema de recomendación, recuperando advertencias ponderadas (recomendaciones) de posibles efectos secundarios (elementos) causados por un fármaco (usuarios). Las realizaciones de la presente invención aprovechan los datos estructurados para calcular la similitud entre los fármacos que se usan y sugerir posibles efectos secundarios para un fármaco.

20 La Figura 5 muestra un diagrama de bloques con flujos de información que representan un sistema/ arquitectura de ejemplo para usar datos vinculados para descubrir efectos secundarios de fármacos de acuerdo con una realización de un aspecto de la invención. En general, el método tiene dos etapas: una etapa (que se muestra en el lado izquierdo de la Figura 5) que, debido al tiempo de procesamiento requerido, se ejecuta preferiblemente fuera de línea en una máquina de servidor adecuada para satisfacer los requisitos de consulta; y una segunda etapa (que se muestra en el lado derecho de la Figura 5) que preferiblemente se ejecuta completa o parcialmente en línea, permitiendo a los usuarios acceder y visualizar señales o asociaciones detectadas entre fármacos y efectos secundarios. En términos de flujo, el proceso se puede separar en los tres bloques principales que se muestran en la Figura 5 y se analizan a

25 continuación. Estos bloques son: (1) construcción de base de conocimiento, (2) cálculo de similitud y (3) señalización de efectos secundarios.

1) Construcción de la Base de Conocimiento

30 La presente invención propone un método novedoso para la detección de señales de los efectos secundarios del fármaco utilizando datos abiertos. Diversas iniciativas conocidas se centran en la transformación y publicación de datos de Ciencias de la Vida utilizando un formato de datos abierto, intercambiable y flexible, como el Marco de Descripción de Recursos (RDF). Linked Data se basa en RDF (utilizado como modelo de datos), HTTP (utilizado para transferir datos a través de Internet) y un identificador uniforme de recursos (UIR) (utilizado para identificar recursos) con el fin de publicar datos de manera que puedan ser leídos automáticamente por ordenadores.

35 En una realización preferida de la invención, RDF se usa como el formato de datos para representar un gráfico que representa las fuentes de conocimiento de fondo subyacentes relacionadas con el fármaco. A pesar de que algunas fuentes de datos no se publican originalmente como RDF o son transformaciones incompletas, en este trabajo se llevó a cabo un proceso de extracción, transformación e integración de datos no trivial. Por ejemplo, el mismo fármaco puede estar contenido en diferentes fuentes de datos con un nombre u ortografía diferente, lo que hace que sea imposible realizar una coincidencia directa de cadenas. Este problema se aborda combinando los nombres de los

40 fármacos y los efectos secundarios, utilizando una coincidencia más compleja que se basa en nombres y sinónimos alternativos. Por mencionar, dos ejemplos:

45 1) Los inventores generaron un volcado RDF más completo de Side Effect Resources Sider (www.sideeffects.embl.de/) extendiéndolo con la Construcción Automática de Taxonomía, taxonomía ATC para generar un esquema jerárquico, términos preferidos en un Diccionario médico para Actividades Reguladoras (MedDRA) y la frecuencia de los efectos secundarios extraídos de su sitio web público.

2) Los inventores integraron los nombres de los fármacos del recurso DrugBank (www.drugbank.ca/), y los nombres de los efectos secundarios de Diseasesome (www.diseasome.eu/), utilizando nombres alternativos (relación owl:sameAs), sinónimos y medidas de distancia de cuerda.

50 Como se muestra en la Figura 5, se pueden considerar varias fuentes de datos, tales como, por ejemplo, perfiles (101) de fármacos, base (102) de datos de efectos secundarios, base (103) de datos de clasificación de fármacos y/o base (104) de datos de enfermedades. Las fuentes de datos están integradas por el módulo (105) de Integración de Datos, para luego ponerlas a disposición a través de un Gráfico (106) RDF para ser consultado. Las fuentes de datos utilizadas no se limitan a los ejemplos enumerados anteriormente y las diferentes fuentes públicas, como los informes posteriores a la comercialización, también se puede usar en el presente método para construir una base de conocimiento de gráficos RDF. Debido a la flexibilidad del RDF, también se pueden utilizar otras fuentes no

55 estructuradas que contienen las propiedades de los fármacos (por ejemplo, información química, biológica, fenotípica), efectos secundarios, clasificación de fármacos y enfermedades. Las fuentes mencionadas inicialmente están disponibles para los usuarios, lo que minimiza los costes de recopilación de datos y el método general.

La Figura 6 muestra una vista ampliada de la interacción entre el módulo (105) de integración de datos y el módulo (106) RDF Gráfico. El módulo (105) de integración de datos es responsable de procesar una fuente (201) de datos y de realizar la integración de datos con filtrado para generar triples RDF. Los triples RDF se almacenan e indexan en un almacén (204) RDF para consultas futuras. Las fuentes de datos generalmente requieren un proceso (203) de enlace para hacer accesibles las mismas entidades utilizando un identificador único. Es decir, si un fármaco contiene nombres alternativos, el sistema debe asegurarse de que cualquier nombre alternativo que se use, el identificador de fármaco recuperado sea el mismo. Esto evita problemas de referencia y mantiene un nivel de integridad aceptable en los datos.

2) Cálculo de Similitud

El módulo (107) de Cálculo de Similitud es el módulo responsable de calcular la similitud de los fármacos. Primero, el módulo (107) de Cálculo de Similitud recupera una lista de todos los fármacos presentes en el gráfico RDF (106). Un gráfico RDF es un gráfico dirigido en donde cada borde es un triple RDF con una estructura (*sujeto, predicado, objeto*) y donde el borde comienza en el nodo *sujeto* y termina en el nodo *objeto*. Luego, para cada fármaco X en la lista, el módulo (107) de Cálculo de Similitud realiza una consulta contra el módulo Gráfico RDF (106) preguntando por el conjunto de predicados (relaciones) donde X está en una posición de sujeto o de objeto. Formalmente, usando una consulta por patrones podríamos representar estas consultas como (X,?,?) y (?,?, X) para sujeto y objeto, respectivamente. A partir de esos predicados se genera un conjunto de características para el fármaco X, que se conoce como un vector de características.

El gráfico RDF es un gráfico dirigido, es decir, los bordes tienen una dirección y existen bordes entrantes y salientes. Por lo tanto, la consulta (X, ?, ?) se usa para los bordes salientes, y la consulta (?, ?, X) se usa para los bordes entrantes, respectivamente.

Luego, cada fármaco se asocia a un vector de características extraído del gráfico RDF subyacente en 106. Estas características interactúan con las características de otros fármacos para calcular una similitud entre cualquier par de fármacos. En una realización preferida, los vectores de características y similitudes para un par de fármacos pueden generarse como se describe a continuación y con referencia a la Figura 7.

La Figura 7 muestra un gráfico RDF para dos fármacos, conocidos como recursos en el dominio del gráfico RDF. Se debe calcular la similitud de los dos fármacos del gráfico RDF que se muestra en la Figura 7, **a** y **b**. En este ejemplo, se utiliza una métrica de similitud basada en el trabajo descrito en R. Meymandpour and J. G. Davis, "Recommendations Using Linked Data," in PIKMT Maui, Hawaii, 2012. Sin embargo, se puede adaptar cualquier otra métrica adecuada para calcular similitud entre recursos en un gráfico RDF.

Sea A el conjunto de características para el recurso **a**, donde A contiene dos conjuntos: uno para las relaciones de salida y otro para las relaciones de borde entrantes, es decir, $A = \{\{\text{saliente}\}, \{\text{entrante}\}\}$. El recurso **a** en la Figura 7 tiene relaciones salientes con los recursos **c**, **e** y **f** con los predicados ℓ_1 , ℓ_3 y ℓ_4 , respectivamente. Además, tiene relaciones entrantes desde el recurso **d** con el predicado ℓ_2 . Entonces, A se define como se muestra en la Ecuación (1) a continuación.

$$A = \text{Rasgos}_{LD}(a) = \{ \{(\ell_1, c), (\ell_3, e), (\ell_4, f)\}, \{(\ell_2, d)\} \} \quad (1)$$

De manera similar, B se define como el conjunto de características para el recurso **b**, como se muestra en la Ecuación (2).

$$B = \text{Rasgos}_{LD}(b) = \{ \{(\ell_4, e), (\ell_4, f), (\ell_5, g)\}, \{(\ell_2, d)\} \} \quad (2)$$

Intuitivamente, cuantos más elementos **a** y **b** tienen en común, más similares son. Esto es consistente con los principios de las métricas -basadas en características, donde un mayor nivel de características compartidas implica un mayor valor de similitud. Por lo tanto, para calcular la similitud entre un par de recursos en el gráfico RDF, usamos la intersección entre conjuntos de características. Por ejemplo, la intersección entre A y B viene dada por la Ecuación (3).

$$A \cap B = \{ \{(\ell_4, f)\}, \{(\ell_2, d)\} \} = C \quad (3)$$

En la técnica anterior asociada, "Recomendaciones Usando Datos Vinculados", la similitud se define usando métodos de Teoría de la Información para identificar la importancia de las características en función de sus probabilidades.

$$\text{Similitud}(A,B) = IC_{LD}(C) \quad (4)$$

El Contenido de Información (IC) de un conjunto de características en Datos Vinculados (LD) se define como la suma del IC de cada componente en el vector.

$$IC_{LD}(A) = \sum_{\forall a_i \in A} IC(a_i) \quad (5)$$

Para un componente en el vector, su IC se calcula como una puntuación de frecuencia normalizada.

$$IC_{\forall a_i \in A}(a_i) = -\log\left(\frac{\phi(a_i)}{n}\right) \quad (6)$$

5 Donde $\phi(a_i)$ es la frecuencia de la característica a_i en el gráfico RDF y n es el número máximo de ocurrencias de una característica.

10 Se muestra en la Figura 8 y se analiza a continuación un ejemplo detallado de dos fármacos: fenacetina (identificada por dBank: DB03783) y acetaminofén (identificada por dBank: DB00316), con un subconjunto de sus relaciones que incluyen: etiqueta, marca, forma de dosificación y proteínas diana. Como se puede ver en la Figura 8, los fármacos dBank: DB03783 (Fenacetina) y dBank: DB00316 (Acetaminofén) comparten algunas relaciones con los siguientes literales y recursos: "Humans and other mammals"@en, dBank:capsuleOral, dBank:liquidOral, dBank:analgesics,Non-narcotic, dBank:target-20.

La lista de cinco relaciones compartidas por los dos fármacos es:

- 15
- dBank:Organismo afectado
 - dBank:Forma de dosificación (dos veces)
 - dBank:Categoría de fármaco
 - dBank:diana

20 Sobre la base de estas relaciones, se pueden calcular los vectores de características correspondientes, y se puede determinar una puntuación de similitud entre dBank:DB03783 (Fenatina) y dBank:DB00316 (Acetaminofén).

$$A = Rasgos_{LD}(dBank:DB03783) = \left\{ \begin{array}{l} (rdfs: etiqueta, "Fenacetina" @en), \\ (adverso: suturaid, "CID0000123620" @en), \\ (dBank: Organismoafectado, "Humanos y otros mamíferos" @en), \\ (dBank: Nombredemarca, "Disprol" @en), \\ (dBank: Formadedosis, dBank:cápsulaOral), \\ (dBank: Formadedosis, dBank:líquidoOral), \\ (dBank: Formadedosis, dBank:suspensiónOral), \\ (dBank: Categoriadefármaco, dBank: analgésicos, No narcóticos), \\ (dBank: Categoriadefármaco, dBank: antipiréticos), \\ (dBank: diana, dBank: diana - 20), \\ (dBank: diana, dBank: diana - 290) \end{array} \right\}, \emptyset \quad (7)$$

$$\begin{aligned}
 B &= \text{Rasgos}_{LD}(\text{dBank:DB00316}) \\
 &= \left\{ \left\{ \begin{array}{l} (\text{rdfs: etiqueta, "Acetaminofen"@en}), \\ (\text{adverso: suturaid, "CID0005481350"@en}), \\ (\text{dBank: Organismoafectado, "Humanos y otros mamíferos" @en}), \\ (\text{dBank: Nombredemarca, "Acodicina" @en}), \\ (\text{dBank: Formadedosis, dBank:cápsulaOral}), \\ (\text{dBank: Formadedosis, dBank:líquidoOral}), \\ (\text{dBank: Categoriadefármaco, dBank: analgésicos, No narcóticos}), \\ (\text{dBank: diana, dBank: diana - 20}), \\ (\text{owl: igual que, dbpedia: Fenacetina}), \\ (\text{dBank: Nombregenérico, "Fenacetina"@en}) \end{array} \right\}, \emptyset \right\} \quad (8)
 \end{aligned}$$

Por lo tanto, para obtener el puntaje de similitud entre ambos fármacos, la intersección entre A y B debe calcularse, como se muestra en la Ecuación 9.

$$\begin{aligned}
 A \cap B &= \left\{ \left\{ \begin{array}{l} (\text{dBank: Organismoafectado, "Humanos y otros mamíferos" @en}), \\ (\text{dBank: Formadedosis, dBank:cápsulaOral}), \\ (\text{dBank: Formadedosis, dBank:líquidoOral}), \\ (\text{dBank: Categoriadefármaco, dBank: analgésicos, No narcóticos}), \\ (\text{dBank: diana, dBank: diana - 20}), \end{array} \right\}, \emptyset \right\} \quad (9) \\
 &= C
 \end{aligned}$$

- 5 La intersección contiene el conjunto de características compartidas. Nótese que ni A ni B contienen bordes entrantes, por lo que el conjunto de entidades del lado derecho está vacío. Para cada característica, su contenido de información se calcula para obtener una puntuación final de similitud.

En resumen, la Figura 9 muestra un gráfico de bloques de grano fino para el módulo de cálculo de similitud. El Método (503) de Similitud obtiene un conjunto de vectores de características (502) como entrada, que se generaron previamente a partir del gráfico RDF subyacente con Información de Fármacos (501). Las Similitudes de Fármacos se almacenan en la Base (504) de Datos de Similitudes de Fármacos para ser consumidas durante el procesamiento en línea a pedido del usuario.

3) Señalización de efectos secundarios

- 15 Nótese que durante el proceso de cálculo de similitud (107 en la Figura 5) una serie de distancias de fármaco a fármaco se calculan y almacenan en la base de datos de similitudes de fármacos (109 en la Figura 5 y 504 en la Figura 9), que luego se usa durante la parte en línea del sistema.

20 Conceptualmente, la base de datos de similitudes de fármacos representa un modelo de espacio vectorial donde cada fármaco se representa como un punto en un espacio n -dimensional (donde n es la cardinalidad de los vectores de características), y se satisfacen ciertas propiedades. La Figura 10 muestra una pequeña muestra de este espacio de fármacos donde el punto negro denota un fármaco seleccionado, y los cinco puntos no llenos conectados son los 5 fármacos más similares, también llamados vecinos más cercanos. Este conjunto de cinco más uno de fármacos conforma un vecindario donde los fármacos están altamente asociados en función de sus propiedades (por ejemplo, información química, biológica o fenotípica).

- 25 Para determinar un vecindario, se utiliza el conocido algoritmo k -Vecinos más Cercanos (o k -NN) sobre un fármaco para obtener los k vecinos más cercanos. Aquí el valor de $k + 1$ representa el tamaño del vecindario.

30 La Figura 11 muestra el módulo (111) de Propagación de Efectos Secundarios basado en un enfoque basado en efectos secundarios. Para un fármaco determinado, el sistema recopila los efectos secundarios ya conocidos entre todos sus k vecinos (representados por 701). Los efectos secundarios se combinan y clasifican de acuerdo con su ocurrencia conjunta entre los fármacos vecinos para determinar el peso de la señal. Los efectos secundarios también pueden clasificarse utilizando diferentes métodos de clasificación, no limitados al descrito aquí.

5 El filtro (702) de concurrencia calcula la frecuencia de las características del fármaco que ocurren con cada efecto secundario. Estas características se utilizan para construir un histograma normalizado de las características del fármaco para un efecto secundario dado. Por ejemplo, el dolor abdominal de efecto secundario ocurre 271 veces con fármacos conocidos, y solo el 5% de los fármacos que causan dolor abdominal curan la fibrilación ventricular; o el 98% de los fármacos que causan dolor abdominal contienen una subestructura química 12. Todas estas correlaciones pueden expresarse en una matriz de correlación entre todos los efectos secundarios y las características del fármaco. Esta matriz de correlación luego se usa como un perfil de fármacos para clasificar los efectos secundarios propuestos por los fármacos vecinos. Por lo tanto, el método aquí descrito es capaz de determinar cuándo es poco probable que ocurra un efecto secundario con el uso de un fármaco dado. Por ejemplo, usando la matriz de perfil podríamos decir que un fármaco fabricado para curar la fibrilación ventricular que no contiene la subestructura química 12 y que se dirige a la proteína Q9NY47 es poco probable que cause dolor abdominal.

10 Por lo tanto, el filtro (702) de concurrencia selecciona los efectos secundarios como potencialmente significativos para un fármaco determinado, y se emite una señal para cada uno de los posibles efectos (703) secundarios para realizar un análisis adicional.

15 Como ejemplo, considerando el vecindario extraído como se discutió anteriormente y determinando los efectos secundarios presentes en los 5 fármacos vecinos principales, el método puede analizar las coincidencias entre los 5 fármacos, para determinar los pesos de las señales. Una lista de los efectos secundarios se muestra en la Tabla 1 a continuación.

Tabla

Anafilaxia aguda	Diarrea	Prurito
Angiitis cutánea alérgica	Mareo	Erupción
Dolor de pecho	Dispepsia	Insuficiencia renal
Desorientación	Disuria	Somnolencia
Disnea	Fatiga	Estomatitis
Calambre muscular	Aumento de la temperatura corporal	Transpiración
Síndrome nefrótico	Fascitis	Trombocitopenia
Palpitaciones	Flatulencia	Tinnitus
Síncope	Enrojecimiento	Urticaria
Fascitis necrotizante	Gastritis	Vértigo
Personas con discapacidad auditiva	Hemorragia gastrointestinal	Vómitos
Agranulocitosis	Alucinación	Función hepática anormal
Choque anafiláctico	Dolor de cabeza	Perforación gastrointestinal
Anemia	Hematuria	Insuficiencia renal
Anemia hemolítica	Hemorragia	Visión borrosa
Angioedema	Hepatitis	Fotosensibilidad
Anorexia	Hipersensibilidad	Síndrome hipersensible
Astenia	Ictericia	Disturbio visual
Broncoespasmo	Náusea	Dolor gastrointestinal
Colestasis	Nefritis	Insomnio
Estado de confusión	Nefritis intersticial	
Estreñimiento	Nerviosismo	
Dermatitis	Edema	
Dermatitis exfoliativa	Proteinuria	

Se destacan los efectos secundarios mejor clasificados, los que tienen mayores probabilidades de ocurrir con el uso de fenacetina.

Tabla 2

1) Desorientación	9) Mareos
2) Choque anafiláctico	1) Desorientación
3) Angioedema	2) Choque anafiláctico
4) Anorexia	5) Estado de confusión
5) Estado de confusión	6) Estreñimiento
6) Estreñimiento	3) Angioedema
7) Dermatitis	4) Anorexia
8) Diarrea	7) Dermatitis
9) Mareos	8) Diarrea
10) Fatiga	10) Fatiga
11) Hemorragia gastrointestinal	20) Somnolencia
12) Hemorragia	11) Hemorragia gastrointestinal
13) Hipersensibilidad	12) Hemorragia
14) Ictericia	13) Hipersensibilidad
15) Náuseas	14) Ictericia
16) Nerviosismo	15) Náuseas
17) Edema	
18) Prurito	17) Edema
19) Erupción	18) Prurito
20) Somnolencia	
21) Tinnitus	21) Tinnitus
22) Vómitos	22) Vómitos
(a)	(b)

- 5 Una vez que se clasifican los efectos secundarios, el filtro de coocurrencia ayuda a volver a clasificar los efectos secundarios para dar más importancia a algunos efectos secundarios en comparación con otros. Además, el mismo filtro ayuda a descartar ciertos efectos secundarios que es poco probable que ocurran con el consumo de Fenacetina. La lista filtrada y ponderada de efectos secundarios está lista para ser visualizada para un usuario que lo solicite utilizando una interfaz de usuario adecuada, que muestre el fármaco con las señales ponderadas.
- 10 La precisión de las realizaciones del método de la invención se puede medir de las siguientes maneras:
1. Descartar los efectos secundarios ya conocidos (presentes en la base de conocimiento) para un fármaco; realizando el método descrito para extraer señales de efectos secundarios y comparando ambas listas para medir la proporción de efectos secundarios recuperados.
 2. Tener en cuenta las etiquetas disponibles en los efectos secundarios, que indican cuándo se descubrió un efecto secundario en una etapa posterior a la comercialización o cuando se descubrió en una etapa previa a la comercialización; descartando los efectos secundarios posteriores a la comercialización, la precisión se puede medir sobre la base de la detección de los datos posteriores a la comercialización.
- 20 El método de bajo-coste para la señalización en etapa temprana de los efectos secundarios de los fármacos descrito en la presente invención permite a diferentes usuarios evaluar las hipótesis con respecto a la posible asociación de los efectos secundarios a los fármacos.

Como se indicó anteriormente, las realizaciones de la presente invención intentan superar los inconvenientes de la técnica anterior a través de un método que cumple uno o más de los siguientes criterios.

5 A. Es un método de detección temprana que utiliza los datos disponibles antes de su lanzamiento al mercado, haciendo uso de datos disponibles públicamente de fuentes confiables como, por ejemplo, DrugBank, Sider, y Diseasome a un coste menor.

10 B. Explota el uso de datos de bajo coste al utilizar datos de acceso abierto y no depender o requerir datos de pacientes y/o ensayos experimentales. Los datos abiertos permiten que el sistema funcione y ejecute la detección de efectos adversos antes del lanzamiento al mercado y sin restricciones de disponibilidad de datos del paciente.

15 C. Es capaz de operar con datos incompletos mientras permite la integración incremental del conocimiento adquirido a través del proceso de fabricación de fármacos y/u otras partes interesadas. Los métodos no requieren datos completos para ejecutar las predicciones y, gracias al formato RDF, los datos nuevos se pueden integrar sin ningún coste adicional.

REIVINDICACIONES

1. Un sistema configurado para calcular similitudes entre fármacos y configurado para usar las similitudes para extrapolar los efectos secundarios de un fármaco específico, el sistema comprende:
- 5 un aparato de cálculo de similitud configurado para usar datos abiertos para calcular similitudes entre fármacos, comprendiendo el sistema:
- un módulo de integración de datos configurado para obtener datos relacionados con un conjunto de fármacos de una pluralidad de fuentes de datos abiertos e integrar los datos para generar triples RDF;
- un módulo RDF configurado para almacenar un gráfico RDF de los triples RDF; y
- 10 un módulo de cálculo de similitud configurado para recuperar una lista de todos los fármacos presentes en el gráfico RDF y calcular la similitud de cada fármaco con todos los demás fármacos en el gráfico RDF; donde
- el módulo de cálculo de similitud está configurado para encontrar un conjunto de predicados y objetos/sujetos de RDF triples en los cuales, para cada fármaco, el fármaco es un sujeto o un objeto respectivamente y para generar un vector de característica a partir de los predicados y objetos/sujetos que representan las características del fármaco, en donde
- 15 el módulo de cálculo de similitud está configurado además para calcular la similitud de cada fármaco con cualquier otro fármaco en el gráfico RDF en comparación con los vectores de características de los dos fármacos; donde
- el módulo de cálculo de similitud está configurado para calcular la similitud entre dos fármacos teniendo en cuenta el número de características de intersección de los vectores de características, en las cuales el predicado y el sujeto u objeto es el mismo para ambos fármacos; y en donde
- 20 el módulo de cálculo de similitud está configurado para almacenar las similitudes calculadas como distancias fármaco-fármaco para el conjunto de fármacos en una base de datos de similitud a la que está conectado el aparato de cálculo de similitud; y
- un aparato de determinación de efectos secundarios configurado para extrapolar los efectos secundarios de un fármaco específico a partir de datos abiertos relacionados con el conjunto de fármacos, con:
- una conexión a la base de datos de similitud;
- 25 una interfaz de usuario configurada para permitir que un usuario especifique un fármaco del conjunto de fármacos;
- un módulo de clasificación de vecinos configurado para usar las similitudes de fármacos para obtener vecinos más cercanos del fármaco especificada en términos de similitudes; y
- un módulo de propagación de efectos secundarios configurado para recopilar efectos secundarios conocidos de los vecinos más cercanos y para combinar y clasificar los efectos secundarios de acuerdo con su ocurrencia conjunta entre los vecinos.
- 30
2. Un sistema según la reivindicación 1, en donde el módulo de integración de datos está configurado para permitir la integración de nuevos datos durante el tiempo de ejecución, y el módulo RDF está configurado para agregar los nuevos datos al gráfico RDF.
- 35
3. Un sistema de acuerdo con la reivindicación 1 o 2, en donde el módulo de integración de datos está configurado para reconocer términos alternativos para el mismo fármaco y clasificar todos los datos asociados con los términos alternativos juntos como pertenecientes al mismo fármaco y/o en donde el módulo de integración de datos está configurado para usar nombres alternativos, sinónimos y medidas de distancia de cadena para reconocer términos alternativos para el mismo fármaco.
- 40
4. Un sistema de acuerdo con cualquiera de las reivindicaciones anteriores, en donde el módulo de integración de datos está configurado para enlazar diferentes instancias del mismo fármaco usando un identificador único.
5. Un sistema de acuerdo con cualquiera de las reivindicaciones anteriores, en donde el módulo de integración de datos está configurado para integrar datos de una o más de: propiedades de fármacos, efectos secundarios, clasificación de fármacos y bases de datos de enfermedades.

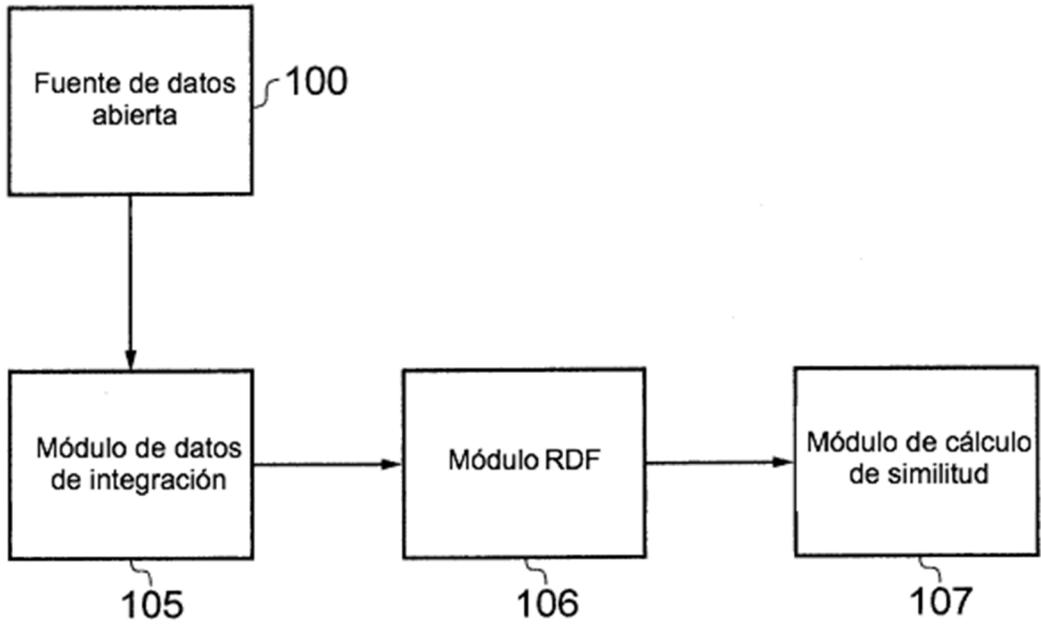


FIG. 1

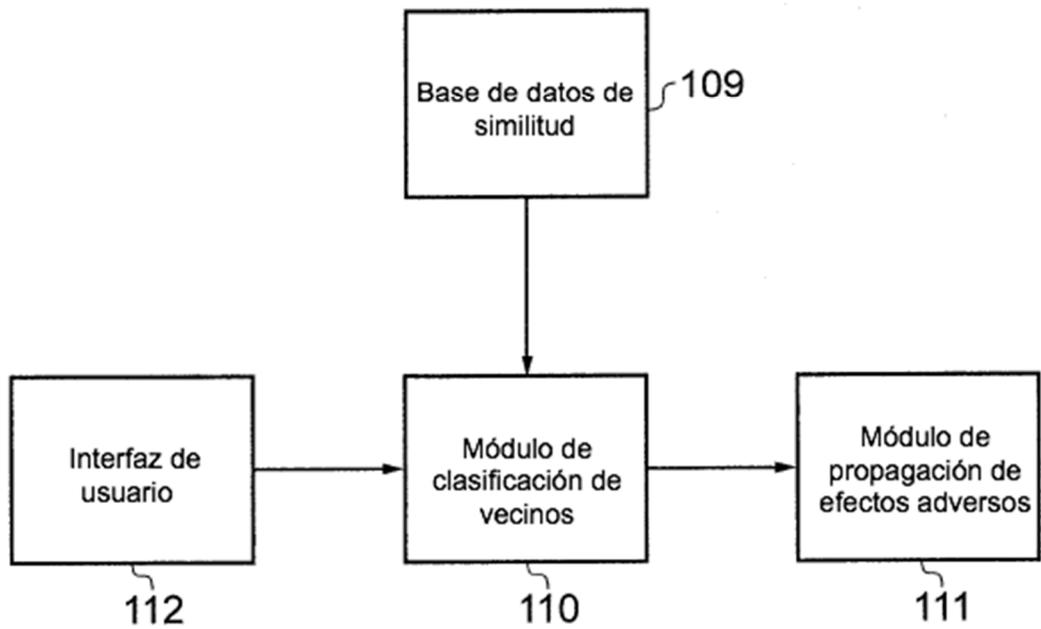


FIG. 2

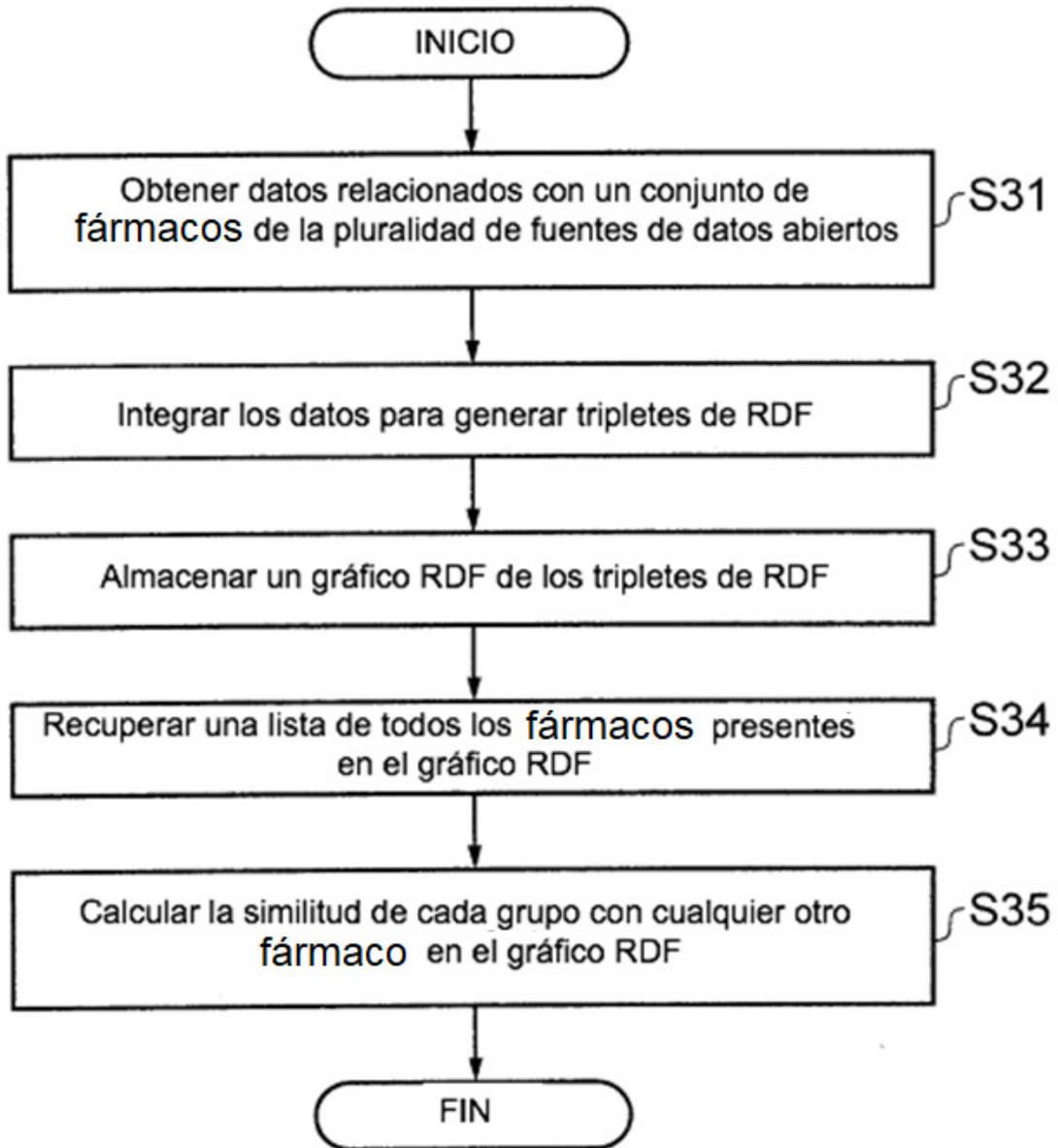


FIG. 3

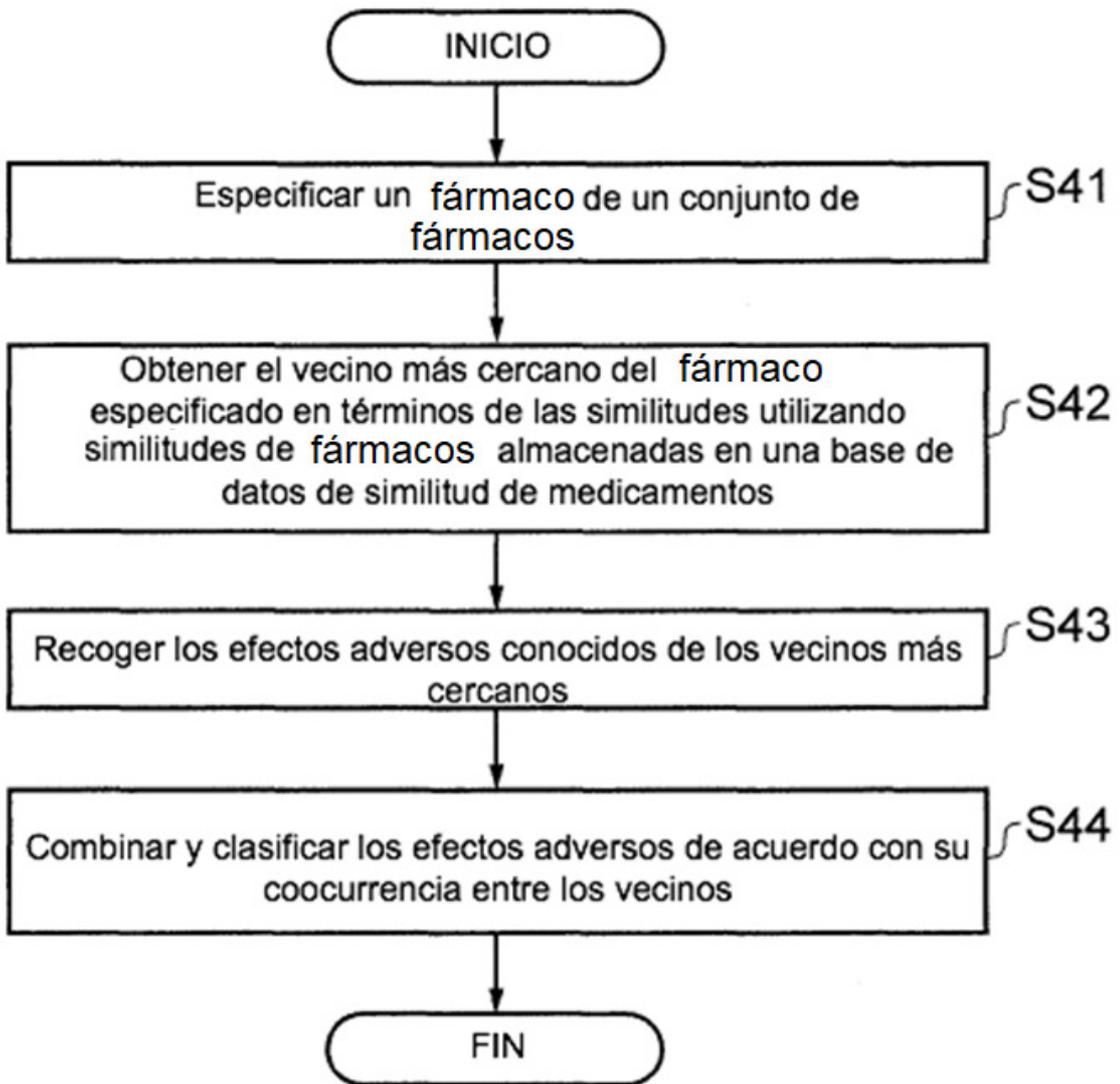


FIG. 4

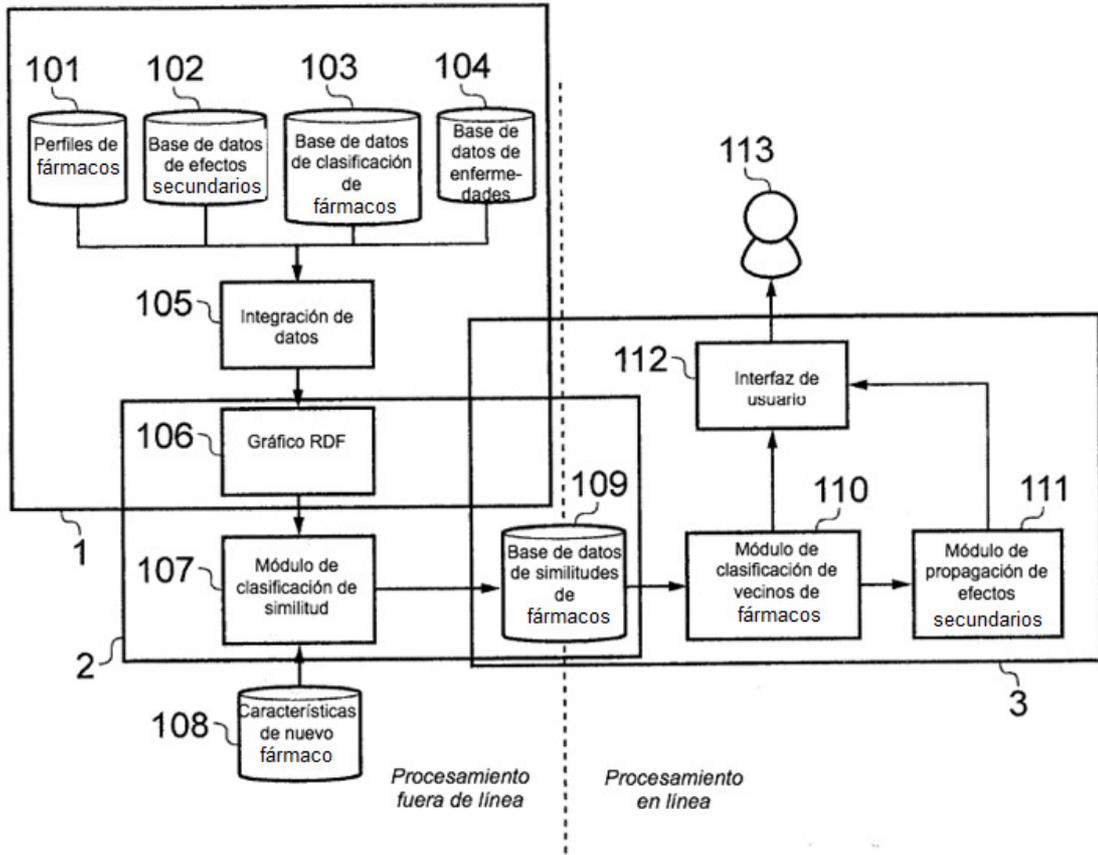


FIG. 5

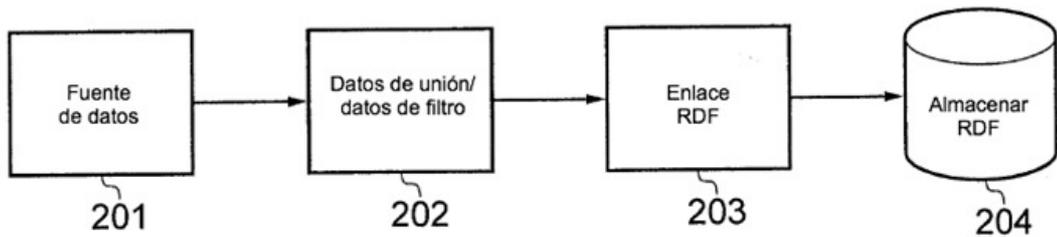


FIG. 6

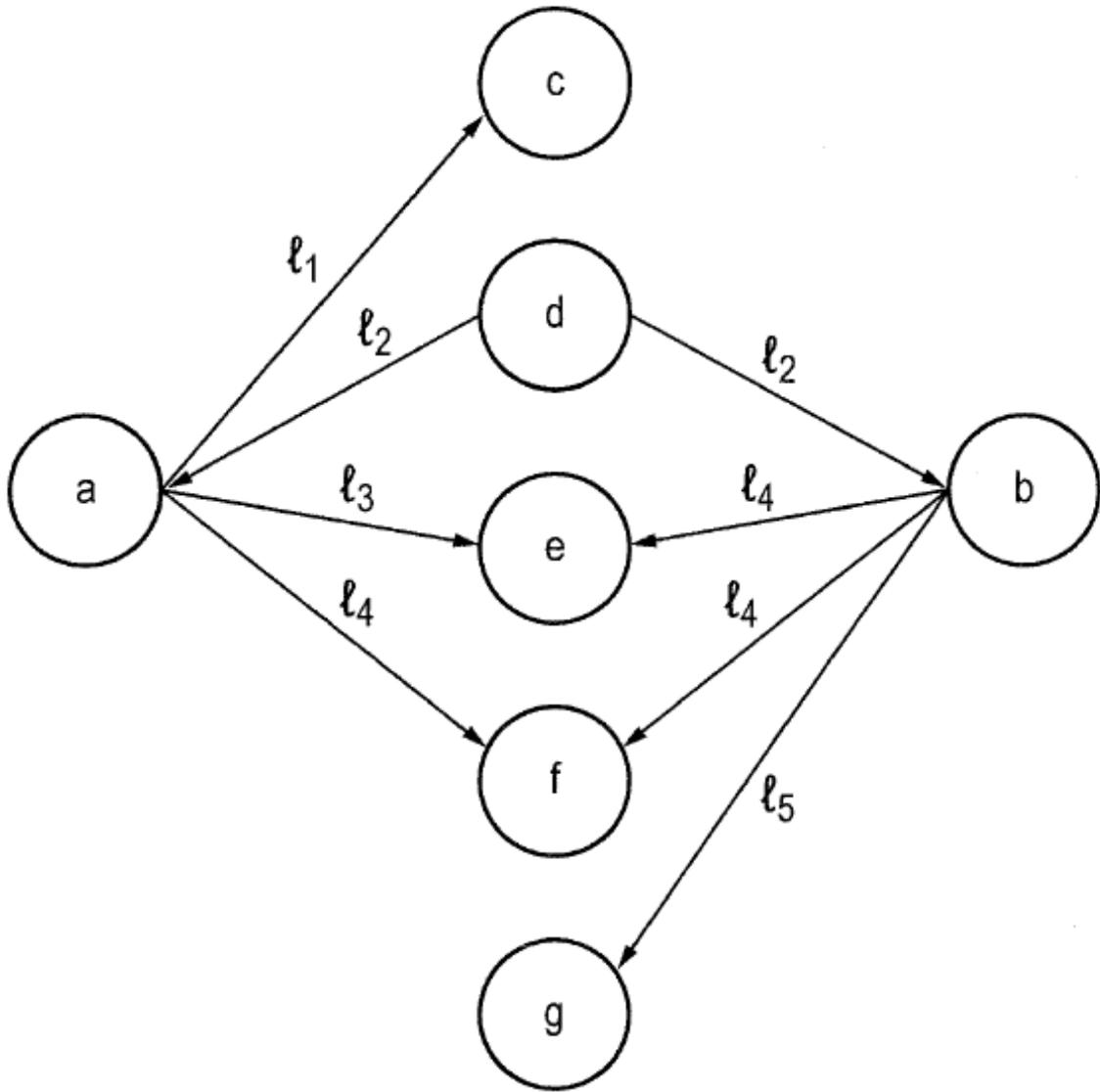


FIG. 7

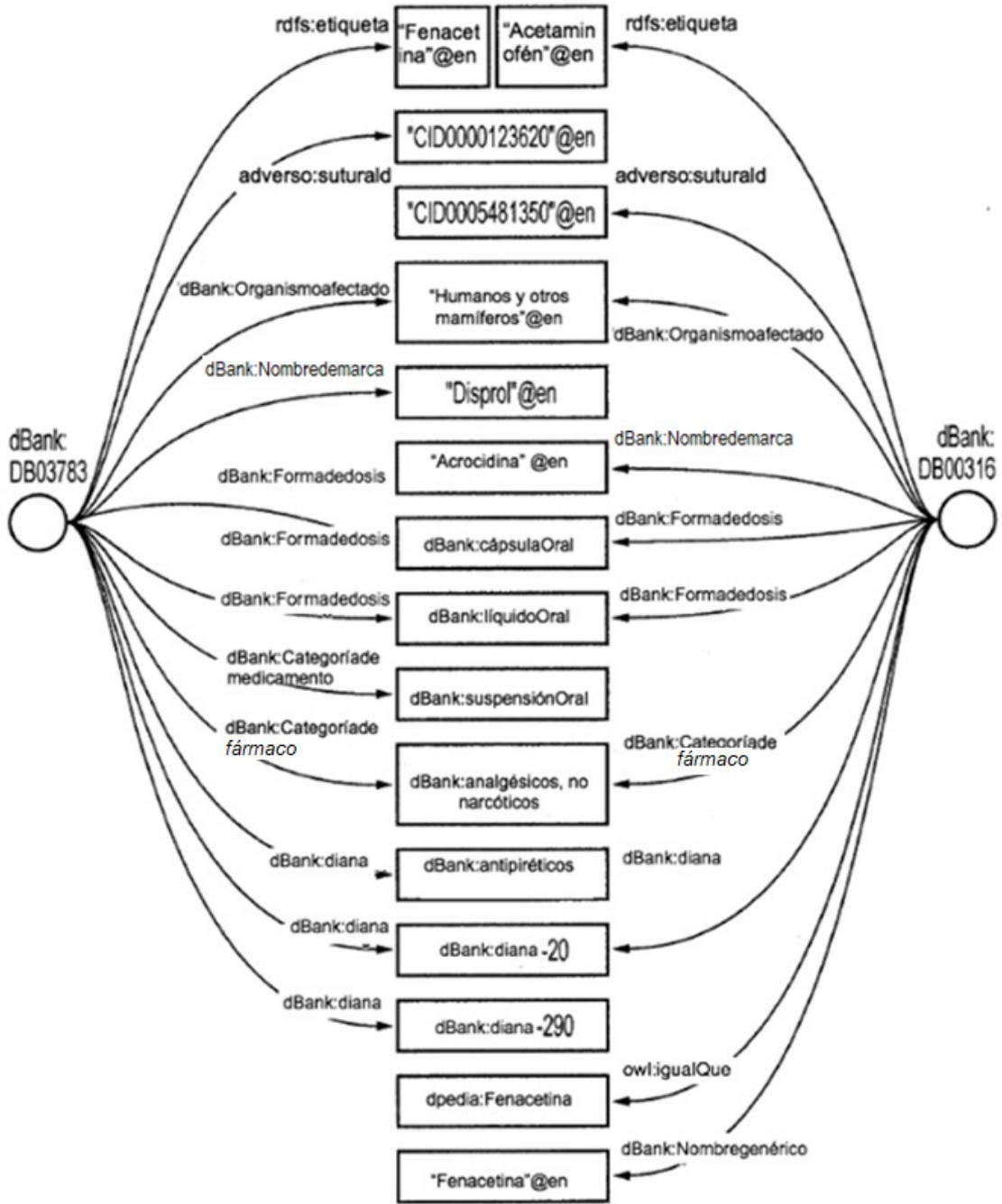


FIG. 8

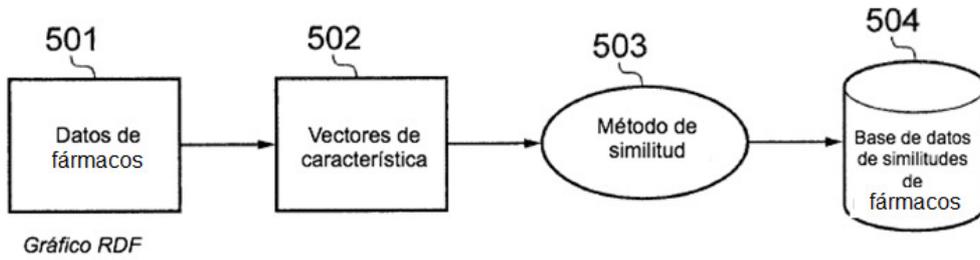


FIG. 9

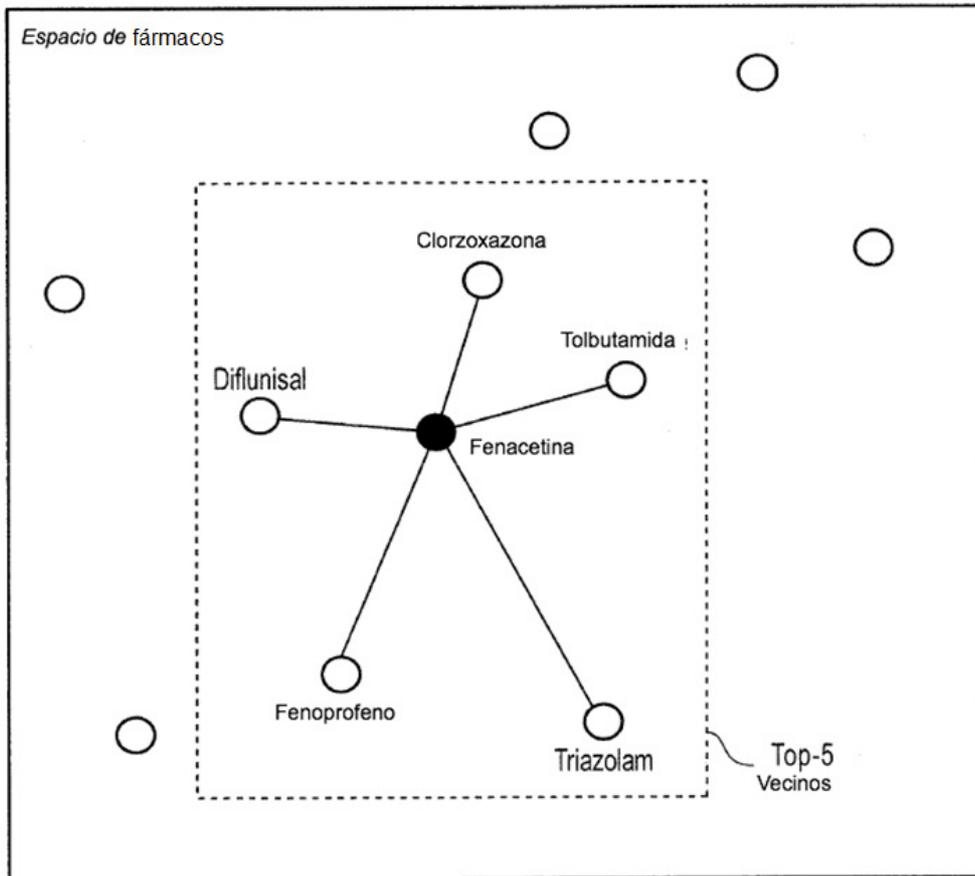


FIG. 10

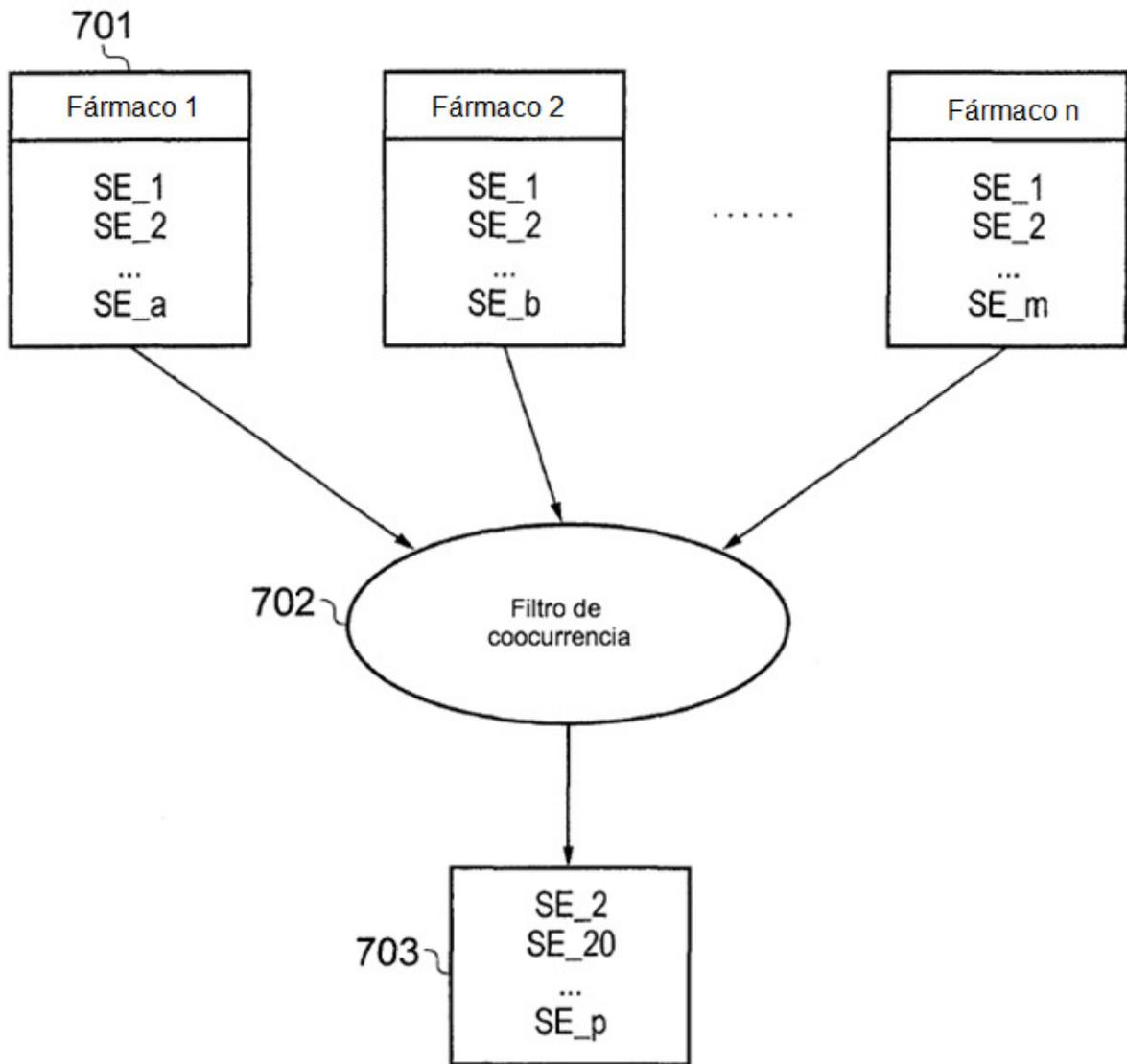


FIG. 11