

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 752 470**

51 Int. Cl.:

**G06Q 30/02** (2012.01)

**G06Q 50/14** (2012.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **05.06.2008 PCT/EP2008/056963**

87 Fecha y número de publicación internacional: **24.12.2008 WO08155225**

96 Fecha de presentación y número de la solicitud europea: **05.06.2008 E 08760542 (4)**

97 Fecha y número de publicación de la concesión europea: **07.08.2019 EP 2158567**

54 Título: **Sistema y método para integrar y mostrar consejos de viaje obtenidos a partir de una pluralidad de fuentes fiables**

30 Prioridad:

**20.06.2007 US 765633**

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

**06.04.2020**

73 Titular/es:

**AMADEUS S.A.S. (100.0%)  
485 Route du Pin Montard, Sophia Antipolis  
06410 Biot, FR**

72 Inventor/es:

**NATH, GOURABMOY**

74 Agente/Representante:

**SUGRAÑES MOLINÉ, Pedro**

ES 2 752 470 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 integrar y mostrar consejos de viaje obtenidos a partir de una pluralidad de fuentes fiables

### 5 **Campo de la invención**

La presente invención se refiere, en general, a un método y sistema para la obtención, la integración y la entrega de información a los viajeros sobre sus destinos de viaje. Está más específicamente dirigido a proporcionar consejos de viaje obtenidos a partir de una pluralidad de fuentes fiables en línea seleccionadas y entregarlos en la forma de un informe combinado organizado semánticamente cuyo contenido puede filtrarse adicionalmente para su visualización.

### **Antecedentes de la invención**

Para los millones de viajeros de todo el mundo que planean visitar países extranjeros para negocios o placer, a menudo por primera vez, el problema surge al tener que recopilar todo tipo de información de viaje acerca de su destino antes de la salida. Si parte de la información recopilada no puede poner en peligro un viaje, y solo lo hace más o menos agradable y/o costoso si parece ser finalmente incorrecto o inadecuado; sigue siendo crucial esa información sobre las condiciones de seguridad y protección, las preguntas de salud y los requisitos de entrada y no puede pasarse por alto bajo la pena de experimentar problemas graves durante el viaje. Sin embargo, los viajeros a menudo se quedan solos con la tarea de recopilar al menos una parte significativa de esta información clave, a pesar de que pueden recibir ayuda de su aerolínea o agente de viajes si han reservado su viaje a través de una agencia de viajes tradicional. A menudo, hoy en día, los viajes también se reservan en línea sin intervención humana. Los viajeros entonces solo pueden confiar en los consejos y advertencias que hayan sido preparados para su destino de viaje actual por parte de los responsables del sitio del proveedor de viajes en línea, si es que de hecho se proporcionan algunos. Además, los viajeros individuales en general no tienen una idea a priori de qué información es importante conocer sobre su destino y de dónde pueden obtenerla, más allá del hecho de que en general piensan referirse a algunos sitios oficiales de su nacionalidad o país de residencia como los puestos en marcha por la mayoría del departamento de asuntos exteriores del país y otros sitios oficiales gubernamentales o internacionales que los viajeros puedan conocer, como la Organización de Salud Mundial (OMS), el brazo de salud pública de las Naciones Unidas (ONU). Sin embargo, incluso si los viajeros han podido recopilar parte de la información necesaria sobre sus destinos de viaje, no pueden estar seguros de que el conjunto de documentos dispares que han recopilado contengan todo lo que deben conocer antes de partir. Por lo tanto, existe la necesidad de un sistema de viajes capaz de entregar automáticamente, tras una solicitud de un viajero, un conjunto exhaustivo y sintético de consejos y advertencias de viaje actuales recopilado a partir de una pluralidad de fuentes fiables para un destino de viaje determinado.

La técnica anterior incluye un número de sistemas y métodos para generar resúmenes y recopilaciones de múltiples documentos agrupados en grupos. La solicitud de patente estadounidense número de serie 2005/0203970A1 desvela un sistema de generación de documentos para obtener historias de noticias a partir de varias fuentes, agrupando estas historias en grupos de documentos relacionados y a continuación generar un resumen legible de los documentos en el grupo.

### **Sumario de la invención**

El sistema y método de la presente invención satisface la necesidad anterior y describe un sistema de viajes informatizado y un método para proporcionar consejos y advertencias por destino de viaje a los usuarios finales del sistema de viajes. Para este fin, el sistema de la invención tiene un módulo de servicio al cliente destinado a recibir consultas de los usuarios finales y a entregarles informes combinados de consejos y advertencias de viaje para los destinos de viaje consultados. Un módulo de descarga y análisis se usa para analizar documentos sin procesar de viaje, descargados desde una pluralidad de fuentes de viajes fiables en línea, en documentos sin procesar limpiados que comparten una estructura de formato común que incluye secciones, encabezados de sección y párrafos. Un módulo algorítmico integra los documentos sin procesar limpiados en los informes combinados de consejos y advertencias de viaje por destino de viaje. La integración de los documentos sin procesar limpiados en los informes combinados se realiza a partir de un documento base elegido entre todos los documentos sin procesar limpiados relevantes para cada destino de viaje consultado. Las secciones comparables se determinan sobre la base del contenido de los encabezados de sección y los párrafos semánticamente cercanos se colocan uno junto a otro en las secciones correspondientes de los informes combinados.

Por lo tanto, un objeto principal de la invención es proporcionar un informe combinado exhaustivo de consejos y advertencias de viaje a los viajeros mediante la agregación, sin ninguna pérdida de información, de contenidos de documentos de múltiples fuentes de consejos de viaje en línea fiables.

Un objeto adicional de la invención es obtener que la agregación de la información sea no solo una anexión física de todos los documentos individuales obtenidos, sino que sea más bien un informe combinado organizado semánticamente donde los párrafos de sección con altas similitudes semánticas estén cerca unos de otros.

Un objeto de la invención es también permitir el filtrado de los informes combinados y documentos de consejos de

viaje individuales, a imprimir o mostrar, sobre la base de filtros semánticos para ver solamente, por ejemplo, recomendaciones o eventos.

5 Otro objeto de la invención es describir un sistema que puede integrarse en un sistema de viajes más grande o que puede funcionar también en un modo autónomo con un acoplamiento fácil con otros sistemas de viajes desde los que podría invocarse.

10 Otros objetos, características y ventajas de la presente invención serán evidentes para los expertos en la materia tras el examen de la siguiente descripción haciendo referencia a los dibujos adjuntos. Se pretende que cualquier ventaja adicional se incorpore en el presente documento.

### Breve descripción de los dibujos

15 La figura 1 representa un sistema de consejos de viaje de acuerdo con la invención en su entorno.

La figura 2 describe las etapas del método de la invención para producir y mostrar un informe combinado organizado semánticamente de consejos y advertencias de viaje.

20 La figura 3 ilustra a través de un ejemplo cómo se construye un informe combinado de consejos de viaje a partir de documentos individuales descargados de fuentes fiables de información de viajes.

La figura 4 describe los detalles de la etapa de integración del método de la invención.

### Descripción detallada

25 La siguiente descripción detallada de la invención se refiere a los dibujos adjuntos.

30 La invención describe un constructor de consejos de viaje (TAB), destinado a la obtención y a la integración de la información relacionada con los viajes a partir de fuentes fiables. Especialmente, el TAB está destinado a recopilar todo tipo de advertencias y consejos de viajes disponibles a partir de fuentes autorizadas, tales como sitios oficiales del gobierno, oficinas de turismo del país, consulados y embajadas y destinados a viajeros que preparan un viaje o mientras viajan. Los consejos de viaje pueden incluir, pero no se limitan a, la información actualizada sobre el estado de los asuntos locales, las condiciones sanitarias y de salud, los incidentes recientes, los posibles riesgos, las amenazas terroristas, las precauciones pertinentes y las recomendaciones relativas a un destino determinado o un país. Además, la información y los consejos de viaje obtenidos pueden filtrarse con filtros semánticos de tal manera que solo se muestren, por ejemplo, recomendaciones (tal como qué hacer o qué evitar en los destinos considerados) o eventos locales.

40 La figura 1 representa un sistema (100) que incluye los medios necesarios para realizar la invención. Un sistema de este tipo puede ejecutarse a partir de diversos recursos informáticos independientes (105) actuales de tal manera que los servicios correspondientes se entreguen directamente a los viajeros, por ejemplo, bajo la forma de un sitio web especializado o, indirectamente, en nombre de los proveedores de servicios de viajes tradicionales (es decir, los agentes de viajes) y en línea. Esto puede estar integrado también en un sistema informático más grande como los implementados por unos pocos sistemas de distribución global (GDS), tal como AMADEUS, un proveedor de servicios mundial para la industria de viajes.

Un sistema de acuerdo con la invención (100) está compuesto por cinco módulos de la siguiente manera:

- 50 – Un módulo de servicio de cliente (110) destinado a recibir las solicitudes de entrada de usuario final, es decir, consultas de los clientes (112) que comprenden normalmente el país de destino y la nacionalidad de viajeros. En respuesta a las consultas de los clientes, la salida del módulo de servicio al cliente consiste en diversos informes (114), que incluyen informes combinados e informes específicos de nacionalidad, posiblemente en forma comprimida, elaborados por el sistema de la invención como se describe más detalladamente en la siguiente descripción.
- 55 – Una base de datos relacional (120) para almacenar todos los datos necesarios para operar la invención, que incluye diversos informes para todos los países de destino que son una fuente específica, los informes combinados producidos por el módulo algorítmico (140), los criterios usados para analizar los documentos sin procesar obtenidos a partir de diferentes fuentes de información de viajes, la lista de sitios web que sirven como fuentes de información de viajes y desde donde se descargan los documentos sin procesar de viajes.
- 60 – Un módulo de análisis y descarga (130) destinado a la descarga de los documentos sin procesar de viaje de la lista de sitios web fiables seleccionados almacenada en la tabla correspondiente de la base de datos relacional (120). Los documentos sin procesar descargados se analizan sobre la base de los criterios de análisis específicos de la fuente que también se almacenan en la base de datos relacional. Las salidas de este módulo son los documentos sin procesar limpiados que sirven como entradas para el siguiente módulo algorítmico.
- 65 – Un módulo algorítmico (140) para procesar los documentos sin procesar anteriores y producir los informes combinados a almacenarse en la base de datos relacional, y destinados a los viajeros. Por lo tanto, el módulo

algorítmico es el componente clave de la invención que logra la integración de la información de viaje obtenida a partir de las fuentes fiables seleccionadas. El funcionamiento de este módulo se describe con más detalle en las siguientes figuras.

- Un módulo de administración (150) que, tal como su nombre indica, se usa para configurar, controlar y monitorizar un sistema de acuerdo con la invención que realiza todas las tareas administrativas convencionales, el manejo de errores y la presentación de informes necesarios para operar y para mantenerlo.

En cuanto a su entorno se refiere, el sistema de la invención (100) se fabrica capaz de comunicarse con diversas aplicaciones de software remoto a través de cualquier combinación de redes privadas y públicas (160), incluyendo Internet. Para este fin, usa medios y protocolos convencionales, tal como TCP/IP, la red y el conjunto de protocolos de red y transporte de internet. Por lo tanto, el sistema puede interactuar a través del módulo de servicio al cliente (110) con usuarios finales remotos individuales desde sus ordenadores personales (162) usando en general cualquiera de las aplicaciones de software disponibles (navegadores) que se usan para navegar en la Web, la aplicación ubicua de Internet y de muchas otras redes privadas. El sistema de la invención también puede comunicarse con los sistemas de viaje corporativos (164), tales como los implementados por diversos proveedores de servicios de viaje y GDS (166) ya tratados anteriormente. Y, como ya se ha mencionado también, los documentos de viaje sin procesar se descargan desde sitios web seleccionados que se ejecutan en cualquier parte del mundo desde sus propias plataformas informáticas remotas (170).

La figura 2 es el diagrama de flujo de las etapas de procesamiento realizadas por el sistema de la invención para integrar la información de viaje obtenida con el fin de producir los informes combinados.

En primer lugar se asume que el usuario final introduce un destino de viaje válido, en general, bajo la forma de un nombre de país (210), o de cualquier destino de viaje que el sistema puede reconocer. La verificación del nombre de destino se realiza consultando la tabla de la base de datos correspondiente que almacena la lista de países soportados (posiblemente incluyendo todos los países del mundo) y otros destinos regionales o temáticos que se han definido por los diseñadores del sistema. Opcionalmente, el usuario final puede tener que introducir información adicional, tal como la nacionalidad del o los viajeros, con el fin de que la integración de los documentos de viaje sin procesar pueda realizarse de manera diferente en función de su nacionalidad y fuente. Por ejemplo, los visados se solicitan o no de acuerdo con la nacionalidad de los solicitantes y es posible que se deban tomar precauciones de salud específicas si viaja desde ciertos países.

A continuación, si se ha introducido un destino válido, el sistema verifica si los documentos sin procesar correspondientes de cada una de las fuentes de datos predefinidas individuales ya existen en la base de datos y si no están actualizados. Si este es el caso, no es necesario descargar los documentos correspondientes y puede procederse directamente a la etapa (230) siguiente.

Sin embargo, si no todos los documentos están en la base de datos y/o están desactualizados deben ejecutarse las siguientes dos etapas.

La invención supone que una tabla maestra de enlaces de fuentes de datos fiables por país o destino válido, tal como se ha tratado anteriormente, se ha construido por primera vez en la base de datos relacional (222). Esta tabla se administra a través del módulo administrativo por un administrador del sistema. La tabla almacena los nombres y las direcciones de la página web maestra, es decir, el localizador uniforme de recursos o la URL de cada una de las fuentes de datos seleccionadas. Además, la URL maestra de cada fuente de datos se amplía posiblemente con rutas adicionales que apuntan a las páginas de consejos de viaje para el destino o país correspondiente. Por lo tanto, un recolector de información, al leer y analizar la página maestra, puede aumentar de este modo una URL maestra con las rutas correspondientes a cada país para el que la fuente de datos puede proporcionar consejos de viaje. Por lo tanto, si una URL maestra de fuente de datos es, por ejemplo: [www.data.com](http://www.data.com), es potencialmente la raíz de los consejos de viaje para muchos países y posiblemente para todos los países. En cuyo caso, la URL exacta para un país de destino en particular llamado D puede ser, por ejemplo: [www.data.com/D](http://www.data.com/D) u otras rutas como: [www.data.com/travel\\_advice/103456.html](http://www.data.com/travel_advice/103456.html). Esta información se almacena en la tabla de país de fuente de datos. Cada fila corresponde a un país seguido por la URL relativa del país de destino. La etapa (222) se ejecuta normalmente la primera vez que se ejecuta el sistema y también, posiblemente, a intervalos regulares o cuando se encuentren documentos sin procesar faltantes o desactualizados en la etapa (220) con el fin de rellenar las columnas relevantes de la tabla de fuente de datos de país en la base de datos relacional y, también, para mantener sus contenidos actualizados mientras el TAB está en funcionamiento.

Como resultado de la ejecución de la etapa (222) los enlaces para cada fuente de datos se conocen de este modo y los documentos sin procesar correspondientes pueden descargarse. Las descargas se realizan bajo el control del protocolo subyacente usado por la Web, es decir, HTTP, que significa "protocolo de transferencia de hipertexto". HTTP define cómo se formatean y transmiten los mensajes a través de Internet y otras redes públicas o privadas usando el conjunto de protocolos TCP/IP. Esto se logra desde el sistema TAB a través de las interfaces de programa de aplicación convencionales o las API de tal manera que la información de viaje pueda descargarse de cada fuente de datos seleccionada y de los documentos sin procesar, finalmente almacenados en las tablas de la base de datos sobre la base de su país o destino de viaje correspondiente, estando disponibles para su posterior procesamiento.

Los documentos están en HTML, que significa “lenguaje de marcado de hipertexto”, el lenguaje de creación usado para crear documentos en la Web. HTML define la estructura y el diseño de un documento web usando una variedad de etiquetas y atributos.

5 Cualquier documento sin procesar individual que ya estaba disponible (221) en la base de datos o simplemente se ha descargado como resultado de la ejecución de las etapas (222) y (224) que se acaban de describir; la siguiente etapa, es decir, la etapa (230), consiste en analizar los documentos sin procesar. Esto se logra mediante el módulo de descarga y análisis mostrado en la figura 1.

10 El objetivo clave de la etapa de análisis (230) es obtener que todos los documentos individuales que proporcionan consejos de viaje para cada uno de los países de destino tengan finalmente la misma estructura HTML de tal manera que más tarde pueden analizarse de manera más eficaz por el módulo algorítmico anteriormente tratado. Debido a que las diferentes fuentes de datos web seleccionadas probablemente tengan estructuras HTML muy diferentes, ya que se han preparado por organizaciones (por ejemplo, embajadas de países, consulados, etc.) que  
 15 no tienen nada en común, el módulo de análisis se diseña para convertir los contenidos y las diferentes estructuras organizativas de información en una jerarquía simple común que comprende secciones, encabezados y párrafos. Esto se logra simplemente en HTML por los expertos en la materia a través de la selección y el uso de algunas etiquetas apropiadas. Por lo tanto, una vez convertidos, los documentos de viaje descargados individuales tienen estructuras simples idénticas que permiten una extracción directa del texto entre las etiquetas usadas para  
 20 formatearlos. A continuación, cada párrafo puede analizarse directamente en términos de una lista de oraciones.

Para ayudar a lograr este objetivo clave del módulo de análisis y permitir la implementación de una conversión eficaz de los documentos sin procesar descargados, la tabla de países de la base de datos relacional recuerda sus  
 25 estructuras sobre la base de sus fuentes. Por lo tanto, las estructuras HTML, es decir, las estructuras de etiquetas se almacenan (por ejemplo, bajo la forma de plantillas) para cada fuente diferente seleccionada de la información de viaje. Se anticipa que la organización de información, la estructura HTML, será exactamente la misma para cada país para el que una fuente de datos determinada en realidad proporciona la información de viaje. Obtener estas etiquetas de una vez por todas, y después de cualquier cambio de estructura importante de las fuentes web, es suficiente. Ya que el número de fuentes de datos considerado por la invención es mucho menor (generalmente, 4 o  
 30 5) en comparación con el número de países (hay alrededor de 190 países independientes en el mundo), estas etiquetas pueden obtenerse a menudo por inspección humana de unos pocos documentos descargados que usan un visor de estructura de marcado (por ejemplo, un visor de lenguaje de marcado extendido o de XML). Los métodos automatizados también son factibles. Entre ellos, pueden aplicarse técnicas de inducción de envoltura convencionales que usan métodos de aprendizaje inductivo con el fin de obtener automáticamente las etiquetas de  
 35 una fuente de datos determinada. Por lo tanto, una tabla de base de datos relacional de fuente de datos almacena estas etiquetas para cada fuente de datos. Dado un documento a analizar, sus etiquetas relevantes se obtienen a partir de esta tabla de fuente de datos. A continuación, es fácil extraer el contenido entre estas etiquetas para obtener la estructura organizativa del documento descargado. Todas las otras etiquetas HTML extrañas, excepto los enlaces de hipertexto, podrían eliminarse. Al final de esta fase, cada estructura de encabezado de sección de  
 40 documento, sección y párrafo junto con sus contenidos quedan de este modo plenamente establecidos.

La siguiente etapa de procesamiento (250) realizada por el sistema de la invención es la etapa donde se realiza finalmente la integración de los documentos sin procesar descargados individuales (limpiados por el módulo de  
 45 descarga y análisis, como se explica en la etapa anterior) para producir los informes combinados destinados a los viajeros como se trata en la figura 1 a quien se entregan finalmente (260). Esta etapa se describe en detalle en la figura 3 y la figura 4 a continuación.

El fin de la etapa (240) que aparece en la figura 2 es filtrar los contenidos de los diversos documentos individuales y combinados que se manipulan y se producen por el sistema de la invención. La etapa (240), es decir, el filtrado,  
 50 puede invocarse en cualquier etapa del proceso (242) para permitir que los usuarios finales vean solo parte de la información de viaje obtenida en el documento activo actual, es decir: el documento que se está viendo. Las operaciones de filtrado más simples son aquellas que muestran secciones específicas de párrafos de un documento activo para estructurarlo para su visualización. Sin embargo, los filtros semánticos, es decir, los filtros basados en el significado del texto, pueden invocarse con el fin de mostrar solo la información específica contenida en el  
 55 documento activo. Un filtro semántico es, por ejemplo: “Mostrar todas las recomendaciones (cosas que hacer o evitar) de los consejos de viaje actuales”. El siguiente texto, donde X es el nombre de un país e Y es el nombre de una ciudad, es un ejemplo de un texto sin procesar a filtrar:

*Existe un riesgo continuo de actividad terrorista en X por parte de grupos militantes. Le aconsejamos que actúe con  
 60 cautela y supervise los acontecimientos que puedan afectar a su seguridad en X debido al riesgo de actividad terrorista por parte de grupos militantes. Preste mucha atención a su seguridad personal y supervise los medios de información para obtener información sobre posibles nuevos riesgos de seguridad. Los objetivos de ataque podrían ser edificios gubernamentales prominentes, instituciones nacionales, lugares de culto, transporte público, aeropuertos y aviones, lugares religiosos y festivales y sitios turísticos. Se han producido incidentes de violencia de  
 65 la mafia en la ciudad Y desde noviembre de 2003. En octubre de 2005, tres explosiones de bombas golpearon las zonas turísticas de Y, lo que provocó 59 muertes y más de 300 heridos.*

Si aplicamos el filtro semántico anterior se obtiene:

5 *Le aconsejamos que actúe con cautela y supervise los acontecimientos que puedan afectar a su seguridad en X debido al riesgo de actividad terrorista por parte de grupos militantes. Preste mucha atención a su seguridad personal y supervise los medios de comunicación para conseguir información sobre posibles nuevos riesgos de seguridad.*

10 Otro filtro semántico podría ser, por ejemplo: "Mostrar todos los eventos". En cuyo caso lo que se muestra se convierte en:

*Se han producido incidentes de violencia de la mafia en la ciudad Y desde noviembre de 2003. En octubre de 2005, tres explosiones de bombas golpearon las zonas turísticas de Y, lo que provocó 59 muertes y más de 300 heridos.*

15 Obviamente son posibles muchos otros filtros semánticos. Por ejemplo, un filtro legal podría simplemente retener oraciones concernientes a artículos de la ley, sanciones o multas. Con el mismo espíritu, podría utilizarse un filtro de delitos para las fuentes de datos que no tienen una sección de delitos especializada.

20 La implementación del filtrado de acuerdo con la invención se realiza de una manera similar a SQL, el lenguaje convencional de facto usado para crear, actualizar y consultar registros de una base de datos relacional. Al igual que SQL, una orden SELECT permite filtrar lo que se mostrará. Los siguientes ejemplos ilustran cómo puede solicitarse el filtrado desde TAB:

- 25 – SELECT oraciones del documento activo donde la oración contiene una recomendación.
- SELECT párrafos del documento activo donde el párrafo contiene una recomendación.
- SELECT oraciones del documento activo donde la oración contiene un evento.
- SELECT párrafos del documento activo donde el párrafo contiene un evento.

30 Cualquier filtro semántico, cualquiera que sea un evento o una recomendación como en los ejemplos anteriores, u otra cosa, se define por una firma. La firma para un filtro de recomendación debe ser parte de un conjunto de palabras clave. Por ejemplo, algunos de los términos en el conjunto de palabras clave de firmas para un filtro de recomendación son, por ejemplo, los siguientes:

35 {Diga, obtenga, guarde, use, no, verifique, empaque, debería evitar, debe, viajeros, visitantes, tenga precaución, preste atención, tenga especial cuidado, se recuerda, se advierte, su responsabilidad, se recomienda fuertemente, se aconseja fuertemente}

40 Antes, un conjunto de palabras clave se definía o bien manualmente o con medios automáticos, la operación de filtrado consiste simplemente en verificar si alguna de las palabras clave de una sola palabra o secuencias de múltiples palabras está efectivamente presente en un elemento de texto, es decir: una oración o párrafo. Debido a que los documentos de consejos de viaje están bien escritos y son textos validados, vale la pena señalar que las letras mayúsculas y minúsculas pueden tenerse en cuenta para el filtrado. Por lo tanto, las letras mayúsculas en la palabra clave anterior significan que coinciden con el comienzo de las oraciones.

45 Un filtro de evento por el contrario se define por una firma que se caracteriza por una fecha. Las fechas aparecen en oraciones bajo diferentes formas, por ejemplo: [Ene 99, Enero 99, Enero de 1999, 3 de enero de 1999, etc.]. Tan pronto como una de las posibles formas gramaticales de una fecha se detecte en el elemento de texto (párrafo u oración), puede clasificarse como un evento.

50 Una ventaja del dominio de consejos de viaje es que en los documentos sin procesar suele haber muchas repeticiones de firmas para un filtro determinado en los diferentes países. Esto permite formular rápidamente un conjunto de documentos de texto de firma basado en el conocimiento para un filtro semántico determinado que sea eficaz en un amplio intervalo de países de destino. Además, la longitud máxima de las secuencias de múltiples palabras en el conjunto de palabras clave es, en la práctica, muy limitada para el dominio de aplicación considerado, evitando de este modo que el TAB tenga que examinar largas secuencias de palabras. Las secuencias de una, dos y tres palabras son en general suficientes como se ilustra en el ejemplo mostrado anteriormente de un conjunto de palabras clave de filtrado. Finalmente, la formulación de un patrón gramatical que implemente mejor el filtro semántico a aplicar con el fin de obtener un filtrado determinado, puede facilitarse mediante observaciones heurísticas de apariciones posicionales de ciertas secuencias de múltiples palabras, por ejemplo: que se producen solo cerca del principio o al final de las oraciones.

60 La figura 3 describe a través de un ejemplo la integración de documentos sin procesar individuales en un documento combinado de acuerdo con la invención.

65

La integración de los contenidos de documentos sin procesar para la síntesis del nuevo documento se realiza con respecto a uno del documento sin procesar designado como el documento base. DocA (310), DocB (320) y DocC (330) son los documentos de consejos de viaje analizados y limpiados después de que se haya aplicado la etapa (230) descrita en la figura 2. Los documentos, provenientes de diferentes fuentes de datos, se refieren a un mismo país o destino. En este ejemplo, se elige el DocC (330) para desempeñar el papel de documento base. Como se ha explicado anteriormente, los documentos analizados y limpiados tienen una estructura HTML simple realizada solo de secciones, encabezados de sección y párrafos. Ejemplos de los cuales se designan respectivamente por las siguientes referencias: (350), (352) y (354). Idealmente, el objetivo de la etapa de integración sería producir una verdadera unión semántica de los documentos sin procesar. Sin embargo, una verdadera unión semántica, que puede definirse como una sin información perdida en cada documento y que no queda ninguna redundancia, todavía es posible solo cuando la realiza un editor humano experto. Lo que un mecanismo automático puede producir razonablemente, sin requerir una cantidad inaceptable de recursos informáticos, solo puede tender hacia una unión semántica. En otras palabras, en un extremo, una unión física sería una simple anexión de todos los documentos individuales, uno después de otro, que obviamente es fácil de implementar y requiere muy pocos recursos computacionales. En el otro extremo se tendría una verdadera unión semántica que, de ser factible, requeriría de todos modos hoy en día demasiados recursos computacionales para aplicaciones prácticas. Por lo tanto, la etapa de integración de la invención se realiza de la siguiente manera para producir un documento combinado, organizado semánticamente, sin ninguna pérdida de información:

- En el documento combinado (340), los párrafos semánticamente cercanos se agrupan en la proximidad física adyacente. Se muestra un ejemplo de esto en la referencia (360) donde un párrafo del DocA (5555...), que se encuentra semánticamente cercano al primero (18181818...) de la segunda sección del documento base (330), se realiza físicamente cerca entre sí en el documento combinado.
- La estructura organizativa de uno de los documentos individuales se elige para que actúe como una plantilla para formar la unión. Este es el documento base (330) ya mencionado que se elige en este ejemplo. Para el dominio de los consejos de viaje, el documento base (DocC) puede ser de cualquier fuente de datos que prefiera el usuario. Por ejemplo, si el viajero es ciudadano de los Estados Unidos, el formato base podría ser el suministrado por la embajada de los Estados Unidos para el destino seleccionado y dentro del que se insertan los clips de otros documentos. Por lo tanto, los encabezados de sección del DocC (342, 344) se mantienen como los encabezados de sección del documento combinado (340).
- No se hace ningún intento para eliminar redundancias si existe alguna entre los párrafos semánticamente cercanos. Las secciones no relacionadas se ponen al final (346).

Por lo tanto, lo que se obtiene es una unión física lógicamente organizada de los documentos A, B y C sobre la base del DocC y se denomina como LOPU<sub>AB/C</sub> (340). Una LOPU no está diseñada para mostrarse como está. Como ya se ha tratado en la figura 2, y más específicamente en la etapa (240), el filtrado se aplica en general sobre los documentos para mostrar solo una parte de sus contenidos. Las formas de mostrar las secciones relacionadas para mejorar la legibilidad dependen de la aplicación y pueden depender de si se ven en monitores de pantalla o se imprimen. Otros métodos de visualización pueden considerarse como todo tipo de métodos jerárquicos y de vista de árbol y diversos métodos de ratón sobre punta de herramienta.

Los beneficios de la reorganización de los documentos en una LOPU son muchos. La integración con respecto a un documento base es a menudo muy eficaz, específicamente en el dominio de consejos de viaje, ya que es más probable que los viajeros se interesen principalmente en los consejos de su fuente de datos preferida (por ejemplo, consejos de su propio gobierno) y les gustaría ver los párrafos relacionados de otros documentos a la luz de su documento preferido. Al colocar juntos párrafos semánticamente cercanos, el usuario es capaz de elegir rápidamente las diferencias entre los párrafos similares para obtener una idea más holística detrás del tema de los párrafos. Además, la comparación de párrafos completos que tienen redundancias semánticas (mientras que las expresiones de lenguaje y redacción usadas pueden ser muy diferentes) puede enfatizar la importancia de ciertos aspectos, debido a las repeticiones semánticas, de tal manera que es más probable que los lectores se den cuenta y tengan cuidado con los mismos. Además, los párrafos estrechamente relacionados pueden considerarse convenientemente para un proceso de resumen, si es necesario. Puede extraerse convenientemente una o dos oraciones de cada grupo de párrafos semánticamente cercanos para contribuir al resumen general. Finalmente, solo reformulando el problema de esta forma conveniente (es decir, una LOPU), se disminuye enormemente la complejidad de producir una integración global completa de documentos en los que tendrían que compararse todos los pares de párrafos posibles en todos los documentos. La comparación puede limitarse en este caso para comparar solo párrafos semánticamente cercanos.

La figura 4 describe adicionalmente las subetapas de la etapa de integración general, es decir: la etapa (250) tratada en la figura 2.

La subetapa (410) está dirigida a determinar si los contenidos de dos secciones son comparables sobre la base de sus encabezados de sección. Si los encabezados tienen contenidos de información similares, se asume que la sección es comparable. La comparación de los encabezados de sección se realiza entre todos los encabezados de sección de los documentos individuales (excepto el designado como que es el documento base) y los del documento base hasta que se agotan las comparaciones (412). Para detectar si dos encabezados de sección tienen contenidos

de información similares, se usa el siguiente procedimiento:

- Identificar las palabras clave de contenido relevante en los dos encabezados:
  - Identificar los sustantivos propios y comunes, los verbos y los adjetivos para formar dos listas de palabras clave.
  - Eliminar las palabras de “ruido” (sustantivos comunes, por ejemplo: yo, nosotros; verbos, por ejemplo: tener, ser; así como las palabras de enlace). Las palabras de ruido son palabras que se usan comúnmente en cualquier texto independientemente del tema y no ayudan o ayudan poco a distinguir entre el contenido semántico principal y el tema semántico de las dos partes del texto.

Este también es, en general, el caso de la siguiente descripción de la invención, a menos que se especifique lo contrario. A continuación, la subetapa (410) procede de la siguiente manera:

- Convertir en testigos P1 y P2 para formar dos conjuntos múltiples de palabras clave (es decir, conjuntos en los que pueden repetirse elementos). Llamarlos  $w(P1)$  y  $w(P2)$ .
- Calcular el conjunto de intersección de los conjuntos múltiples y llamarlo  $w(Pi) \cap w(P2)$ . El número de elementos en el conjunto es:  $n(w(Pi) \cap w(P2))$ .
- Calcular la unión de los conjuntos múltiples y llamarlo  $w(Pi) \cup w(P2)$ . El número de elementos en el conjunto es:  $n(w(Pi) \cup w(P2))$

En el cálculo de la intersección y la unión anteriores de conjuntos múltiples, se consideran las equivalencias sintácticas y semánticas. Se supone que dos palabras clave de encabezado seccionales son semánticamente equivalentes si están presentes en un diccionario de datos de equivalencia almacenado en las tablas de la base de datos relacional del sistema. Además, tal como ya se ha mencionado, las palabras clave pueden ser palabras clave simples, dobles y, en general, de  $n$  palabras como se define por el diccionario de datos.

- Entonces, la similitud ( $P1, P2$ ) es:  $n(w(Pi) \cap w(P2)) / n(w(P1) \cup w(P2))$ . Por lo tanto, la similitud es el número de palabras clave comunes dividido por el número total de palabras clave de los encabezados que se comparan. Por lo tanto, el valor de similitud es 0 si no tienen nada en común y 1 si comparten todas las palabras clave. Si el valor de similitud supera un umbral predefinido, se dice que las dos secciones son comparables basándose en sus encabezados.

Vale la pena señalar en este caso que se ha descubierto que los encabezados de sección en el dominio de consejos de viaje son muy cortos, en general, se componen de solo unas pocas palabras y se caracterizan por el uso de un mismo conjunto de palabras clave o de palabras clave sinónimas. Además, se ha descubierto que el conjunto de unión completo de encabezados de sección para todas las fuentes de datos posibles es extremadamente pequeño y tiene un alto grado de cosas en común. Por lo tanto, dicha detección se basa en el conocimiento y tiene lugar tanto a nivel sintáctico como semántico, basándose la comparación semántica en un diccionario de datos de equivalencia. En el nivel sintáctico, el deletreo exacto se usa para identificar palabras comunes. En el nivel semántico, el diccionario de datos se usa para asignar la equivalencia entre los mismos.

La subetapa (410) que ha determinado qué secciones en los documentos son comparables con la siguiente subetapa de la etapa de integración general, es decir: la subetapa (420), se dirige a la comparación de  $n$  párrafos de cada sección del documento base con cada uno de los  $m$  párrafos en las secciones comparables de los otros documentos hasta que se agoten las comparaciones de párrafos (426). El párrafo del documento base, y el párrafo de destino en otro documento con el que se está comparando, se denominan, en lo sucesivo en el presente documento, respectivamente el párrafo base y el párrafo comparativo. En general, en los documentos de consejos de viaje, el número de párrafos  $n$  y  $m$  es pequeño. Y, debido a que también están muy bien editados, se anticipa que el número promedio de palabras en un párrafo también será pequeño (por lo general, 150 palabras o menos).

Mientras que los párrafos se están comparando, la subetapa (420) identifica en primer lugar las palabras clave de contenido relevante en el párrafo base (421). Se extraen las palabras que transmiten la mayor parte de la información semántica en oraciones de texto, es decir, verbos, sustantivos (sustantivos propios y sustantivos comunes) y adjetivos. Si bien esto puede lograrse aplicando un etiquetador gramatical o de parte del habla convencional (es decir, un software para identificar palabras como sustantivos, verbos, adjetivos, etc.), la invención sigue una manera más simple y sencilla de eliminar todas las frases de enlace de una sola palabra y de varias palabras a través de un diccionario de datos. El conjunto de palabras eliminado incluirá artículos (el, la, uno, una), pre y posposiciones (para, desde, en, adentro, hacia, a, siguiente, aparte, fuera de, lejos, por lo tanto, no obstante, en, arriba, abajo, a través de, etc.), conjunciones y adverbios (una regla simple es palabras que terminan con “ly”, por ejemplo: (luckily) suerte, happily (felizmente), etc.). Por lo tanto, las palabras restantes son verbos, sustantivos y adjetivos. Además, la invención también usa cualquier algoritmo convencional de reconocimiento de entidades nominadas (NER) para distinguir los sustantivos propios de los sustantivos comunes. A continuación, todas las palabras clave, excepto aquellas que se identifican como entidades nominadas, se reducen a su raíz morfológica usando un algoritmo de derivación convencional como el algoritmo de derivación de Levin.

Lo que sigue es un ejemplo de que la subetapa (421) puede lograr identificar palabras clave de contenido relevante en el siguiente párrafo base sobre consejos de viaje a Francia:

5 *En París, en el aeropuerto Charles de Gaulle o en la estación de tren Gare du Nord, debería proteger su equipaje contra el robo y tener cuidado con los carteristas. Los ladrones y los carteristas también operan en las líneas de metro y RER (suburbano) de París, especialmente en la línea B de RER, que da servicio a los aeropuertos Charles de Gaulle y Orly y a la estación Gare du Nord. También se han producido robos en el aeropuerto de Niza, especialmente en los aparcamientos de coches de alquiler donde se guardan las maletas, ya que los conductores cargan el equipaje en los automóviles de alquiler. Los peatones deben tener cuidado con los ladrones de bolsos que operan tanto a pie como en motocicletas.*

El resultado antes de la derivación es el siguiente. Los números entre paréntesis indican la frecuencia de uso de las palabras clave correspondientes en el texto anterior.

15 *{París (2), Charles De Gaulle (2), aeropuerto (3), Gare du Nord (2), ferrocarril, estación, protección, equipaje, robo (2), cuidado (2), carterista (2), operar (2), metro, RER (2), suburbano, línea (2), servir, Orly, ocurrir, Niza, automóvil (3), alquilar, estacionar, bolso (2), tirón (2), conductor, carga, equipaje, alquiler, peatón, pie, motocicleta, B}*

20 Después de derivar algunas de las palabras en el conjunto anterior se sustituyen por sus raíces morfológicas (por ejemplo, operado y operar ambas se cambiarán a oper). La siguiente subetapa (422) hace lo mismo sobre el texto de un párrafo a comparar y proveniente de otro documento.

25 *Existe una alta incidencia de delitos menores, especialmente el tirón de bolsos y el carterismo en toda Francia, especialmente en las calles de ciudades más grandes tales como París, Marsella y Niza. Los aeropuertos, el transporte público, las áreas turísticas y las playas son objetivos principales para los ladrones que trabajan frecuentemente en pandillas, y usan una variedad de métodos para distraer a las posibles víctimas. Hay informes frecuentes de delitos, incluidos robos y atracos, en el sistema de trenes que da servicio a los aeropuertos de Roissy, Charles-de-Gaulle y Orly. Cada vez más, tales crímenes están siendo acompañados por acciones o amenazas de violencia. Existe una creciente incidencia de ataques violentos contra turistas por parte de grupos de jóvenes, que en general se producen a altas horas de la noche en torno a las principales atracciones turísticas tales como los Campos Elíseos, el distrito de Les Halles y el Barrio Latino en el centro de París.*

35 Las palabras clave de contenido relevante se extraen de una manera similar. El resultado es el siguiente. En este ejemplo, se notará que “violen” es el resultado de la derivación de “violencia”.

40 *{alta, incidencia (2), pequeña, bolso, tirón, carterista, Francia, calle, grande, ciudad, París, Marsella, Niza, aeropuerto (2), público, transporte, turista, área, playa, primer, objetivo, ladrones, trabajo, pandilla, variedad, método, distracción, potencial, víctima, frecuente, informe, crimen (3), robo, atraco, tren, sistema, servicio, Roissy, Charles-de-Gaulle, Orly, acompañar, actuar, violen (2), amenaza, violencia, ataque, turista (2), grupo, joven, gente, se produce, tarde, noche, mayor, atracción, Campos Elíseos, Les Halles, distrito, Barrio Latino, centro, París}*

45 A continuación, la próxima subetapa (423) consiste en la identificación de espacios vectoriales para la evaluación de similitudes entre el párrafo base y el párrafo comparativo. Una vez que se han identificado las palabras clave de contenido relevante en ambos párrafos, el proceso de comparación puede comenzar. El principio de comparación es el siguiente.

- 50 – Cada palabra en el párrafo base representa un eje de dimensión. Si, por ejemplo, se tienen 150 palabras en la representación de palabras clave de contenido relevante del párrafo base, implica que se tienen 150 dimensiones. El texto completo es una combinación lineal de estas dimensiones, los pesos o coeficientes lineales de cada eje corresponden a las frecuencias de las palabras. Por lo tanto, todo el texto es un vector en este espacio.
- 55 – El párrafo de comparación tiene un conjunto diferente de dimensiones determinado por el número de palabras únicas, por ejemplo: 80; con algunas dimensiones en común, por ejemplo: 7 palabras en común exactas. Del mismo modo, es por lo tanto un vector en este espacio de 80 dimensiones.
- 60 – Para calcular la similitud, se considera la unión de las dimensiones de los vectores del párrafo base y del párrafo comparativo. En este ejemplo, la unión de dimensiones proporciona:  $(150 - 7) + 7 + (80 - 7) = 223$ . Por lo tanto, la unión es un espacio de 223 dimensiones. En este espacio, una práctica convencional de minería de textos para estimar similitudes consiste en calcular el ángulo del coseno entre el vector de párrafo base y el vector de párrafo de comparación. Si los dos vectores son exactamente iguales, el ángulo entre ellos es 0 y el valor del coseno es 1 (comparten exactamente las mismas palabras clave). Si los vectores son completamente diferentes, son ortogonales en el espacio y el valor del coseno es 0 (no tienen ninguna palabra clave en común).

65 Sin embargo, antes de la aplicación del principio de comparación anterior, la invención intenta reducir las dimensiones del espacio para ser computacionalmente más eficaz en el trabajo en un espacio de comparación de

5 contenido relevante semántico que es un subespacio de los espacios de párrafo anteriores. Para este fin, en primer lugar, se eliminan las dimensiones raras. Lo que es raro o común depende de si una palabra clave específica se usa por los otros documentos de consejos de viaje seleccionados para el mismo país. Para hacer esto, la invención tabula la frecuencia de las palabras clave en el contenido de cada fuente de datos para el mismo país. Aquellas de las palabras clave que tienen una frecuencia de 1 pueden considerarse raras en el espacio de los consejos de viaje del país en cuestión y se eliminan. A continuación, para reducir aún más las dimensiones de espacio, las palabras clave léxicamente cercanas se combinan en una única dimensión. Esto se logra mediante un mapa de equivalencias léxicas que se construye para el dominio de consejos de viaje por un ser humano. Por lo tanto, las siguientes palabras clave, que no están cubiertas por el algoritmo de derivación, se combinan sin embargo en una única dimensión, por ejemplo:

- bolsa frente a equipaje;
- robo contra robar, ladrón

15 El mapa de equivalencia léxica y un mapa de asociación semántica se agregan y se colocan en un archivo de texto que se verifica para reducir las dimensiones de espacio. El mapa de asociación semántica permite orientar el espacio de comparación hacia conceptos semánticos pertenecientes al dominio de consejos de viaje con el objetivo de hacer que los valores de similitud sean lo más significativos posibles. El ajuste fino de las evaluaciones de similitud es posible entonces si diferentes viajeros expresan sus demandas de diferentes maneras. Por ejemplo, dos entradas en este mapa de asociación semántica serían:

- tren-metro, estación, tren, ferrocarril,
- ataque-violencia

25 El mapa semántico permite que dos conceptos semánticos con significados altamente correlacionados se colapsen en la misma dimensión. Esto significa que si bien la derivación reduce sintácticamente las palabras “operar” y “operado” a “oper”, de tal manera que se vuelvan iguales (la dimensión se reduce de 2 a 1), las entradas de mapa semántico se reducen adicionalmente, por ejemplo, las dos dimensiones “Ataque” y “Violencia” a una palabra equivalente, resultando de este modo una dimensión que se convierte en 1.

30 Después de que se hayan creado los espacios vectoriales y las dimensiones se hayan reducido adicionalmente, se establece una medida de la similitud entre los dos párrafos a comparar en la subetapa (424). Al calcular el coseno de sus vectores espaciales, que se ha explicado anteriormente, se obtiene una puntuación de similitud.

35 Finalmente, en la subetapa (430), todos los párrafos de comparación de las secciones de documentos comparables actuales se insertan de manera exhaustiva (432) junto al párrafo base que devuelve la mejor puntuación de similitud, es decir, el valor más alto del coseno. Si se encuentran múltiples valores de similitud idénticos, puede elegirse uno cualquiera de los párrafos base correspondiente para insertar el siguiente párrafo de comparación, por ejemplo, el primero. Los ejemplos de inserciones de párrafos se han mostrado y se han tratado ya en la figura 3.

40

**REIVINDICACIONES**

1. Un sistema de viajes informatizado (100) para proporcionar consejos y advertencias por destino de viaje a los usuarios finales del sistema de viajes, que comprende:

- un módulo de servicio al cliente (110) para recibir consultas (112) de los usuarios finales y entregar informes combinados (114) de consejos y advertencias de viaje para los destinos de viaje consultados;
- un módulo de descarga y análisis (130) para analizar documentos sin procesar de viaje que tienen una estructura de lenguaje de marcado de hipertexto (HTML) descargados desde una pluralidad de fuentes de viaje fiables en línea (170) en documentos sin procesar limpiados que comparten una estructura de formato común que incluye secciones (350), encabezados de sección (352) y párrafos (354);
- un módulo algorítmico (140) para integrar los documentos sin procesar limpiados en los informes combinados de consejos y advertencias de viaje por destino de viaje;
- una base de datos relacional (120) para almacenar y recuperar toda la información de viaje y los controles para operar el sistema de viajes que incluye los documentos sin procesar limpiados y los informes combinados;
- un módulo administrativo (150) para administrar y controlar la creación y la entrega de los informes combinados a los usuarios finales del sistema de viajes,

en el que el sistema de viajes informatizado (100) está configurado para:

- realizar la integración de los documentos sin procesar limpiados con los informes combinados (340) a partir de un documento base (330) elegido entre todos los documentos sin procesar relevantes (310, 320 y 330) para cada destino de viaje consultado, en el que los encabezados de sección del documento base se conservan como encabezados de sección del informe combinado,
- determinar, para cada sección del documento base, las secciones comparables entre la sección del documento base y las secciones de todos los otros documentos sobre la base del contenido de sus encabezados de sección (352),
- determinar, para cada párrafo de las secciones comparables, una puntuación de similitud semántica con los párrafos de la sección correspondiente del documento base, en el que las similitudes semánticas de los párrafos de los documentos sin procesar se evalúan sobre la base de un conjunto reducido de palabras clave de contenido relevante por párrafo y en el que se calculan las puntuaciones de similitud semántica frente a los párrafos base correspondientes (420),
- insertar cada párrafo de la sección comparable junto al párrafo de la sección del documento base que devuelva la mejor puntuación de similitud semántica en el informe combinado.

2. El sistema de la reivindicación 1, en el que la inserción de los párrafos en secciones de los informes combinados se realiza como una función de sus puntuaciones de similitud.

3. El sistema de la reivindicación 1, en el que el módulo algorítmico (140) filtra adicionalmente el contenido de cualquiera de los documentos de consejos de viaje (242) con el fin de imprimir o mostrar solo una parte de los consejos y advertencias de viaje.

4. El sistema de la reivindicación 3, en el que el filtrado (240) se realiza sobre la base de filtros semánticos que incluyen el filtrado de recomendaciones y eventos.

5. Un método implementado por ordenador de un sistema de viajes informatizado para integrar consejos y advertencias a partir de una pluralidad de documentos de consejos de viaje individuales, que comprende:

- recibir consultas de consejos y advertencias para destinos de viaje (210);
- obtener documentos de consejos de viaje individuales (220) que tengan una estructura de lenguaje de marcado de hipertexto (HTML) a partir de una base de datos del sistema de viajes;
- analizar los documentos de consejos de viaje individuales en documentos sin procesar limpiados (230) que comparten una estructura de formato común que incluye secciones, encabezados de sección y párrafos;
- integrar los documentos sin procesar limpiados en informes combinados de consejos y advertencias por destino de viaje (250), incluyendo la etapa de integración la etapa adicional de:

- elegir, entre todos los documentos sin procesar limpiados relevantes, un documento base a partir del que se realiza la integración;
- realizar la integración de los documentos sin procesar limpiados en los informes combinados (340) a partir del documento base (330), en el que los encabezados de sección del documento base se conservan como encabezados de sección del informe combinado;
- determinar, para cada sección del documento base, las secciones comparables entre la sección del documento base y las secciones de todos los otros documentos sobre la base del contenido de sus encabezados de sección (352);
- determinar, para cada párrafo de las secciones comparables, una puntuación de similitud semántica con los párrafos de la sección correspondiente del documento base, en el que las similitudes semánticas de los

- párrafos de documentos sin procesar limpiados se evalúan sobre la base de un conjunto reducido de palabras clave de contenido relevante por párrafo y en el que se calculan las puntuaciones de similitud semántica frente a los párrafos base correspondientes (420),
- 5 – insertar cada párrafo de la sección comparable junto al párrafo de la sección del documento base que devuelva la mejor puntuación de similitud semántica en el informe combinado;
- entregar los informes combinados de consejos y advertencias de viaje para los destinos de viaje consultados a los usuarios finales del sistema de viajes (260).
- 10 6. El método de la reivindicación 5, en el que la etapa de obtención de documentos de consejos de viaje individuales (220) a partir de la base de datos se reemplaza por las etapas de:
- 15 si los documentos no están disponibles o no están actualizados en la base de datos,  
crear o actualizar una tabla maestra de fuentes de datos fiables de consejos y advertencias de viaje (222);  
descargar los documentos sin procesar de las fuentes de datos fiables enumeradas en la tabla maestra (224).
7. El método de la reivindicación 5, en el que la etapa de integración incluye las etapas adicionales de:
- 20 con respecto al documento base usado para producir el informe combinado,  
comparar el contenido de cada párrafo dentro de las secciones comparables (420), comprende además:
- 25 identificar las palabras clave de contenido relevante en el párrafo actual del documento base (421);  
identificar las palabras clave de contenido relevante en el párrafo comparado del documento sin procesar limpiado (422);  
identificar espacios vectoriales para evaluar similitudes de párrafos (423);  
medir las similitudes de párrafos para obtener una puntuación de similitud (424);  
insertar el párrafo comparado en el informe combinado en función de su puntuación de similitud.
- 30 8. El método, en la reivindicación 5, en el que el contenido de cualquiera de los documentos de consejos de viaje (242) se filtra adicionalmente (240) con el fin de imprimir o mostrar solo una parte de los consejos y advertencias de viaje.
- 35 9. El método de la reivindicación 8, en el que la etapa de filtrado (240) incluye filtrar sobre la base de filtros semánticos que incluyen el filtrado de recomendaciones y eventos.
10. Un producto de programa informático almacenado en un medio de almacenamiento legible por ordenador (105), que comprende unos medios de código legibles por ordenador para hacer que al menos un ordenador opere el método para integrar consejos y advertencias de viaje de acuerdo con una cualquiera de las reivindicaciones 5 a 9.

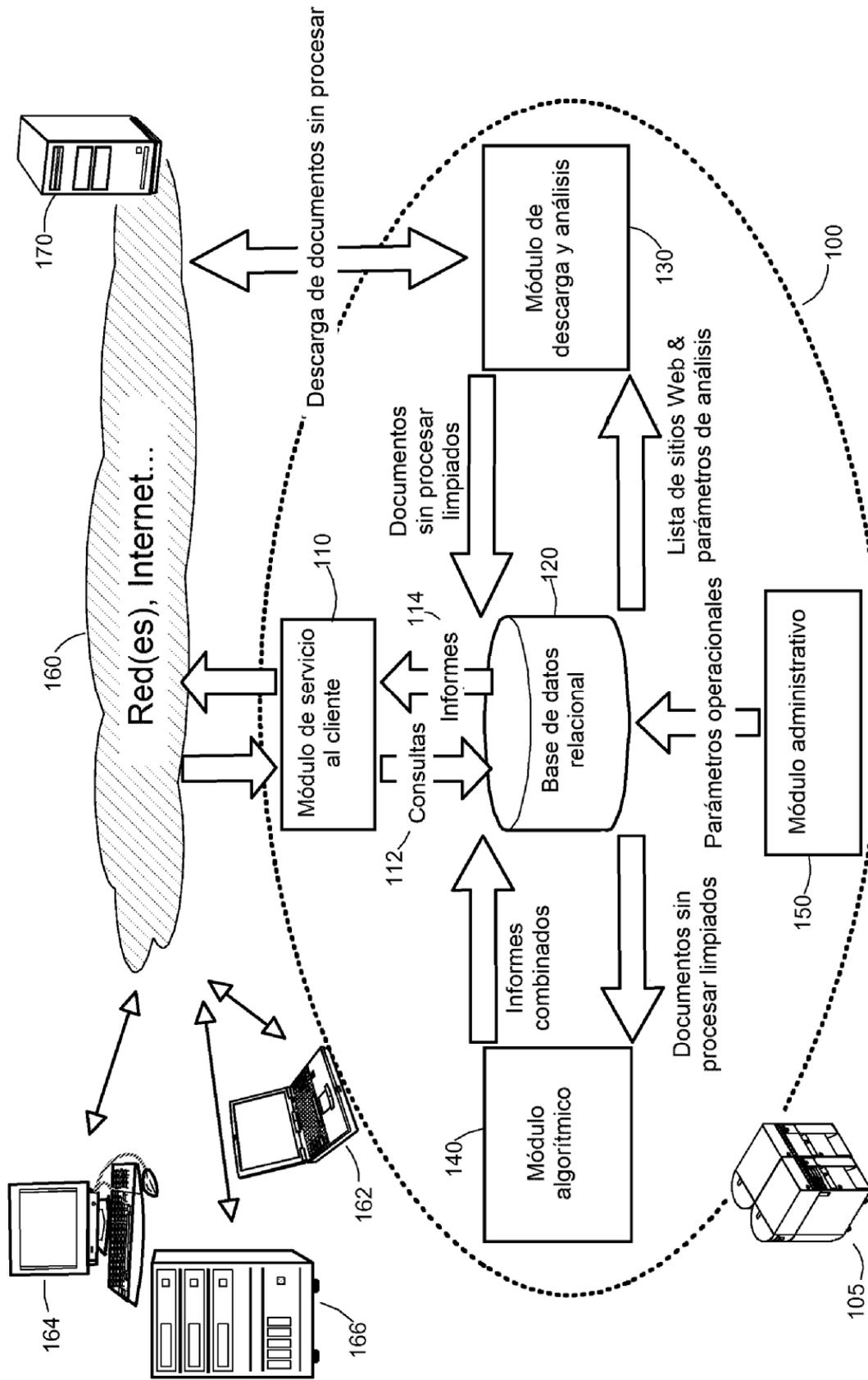


Figura 1

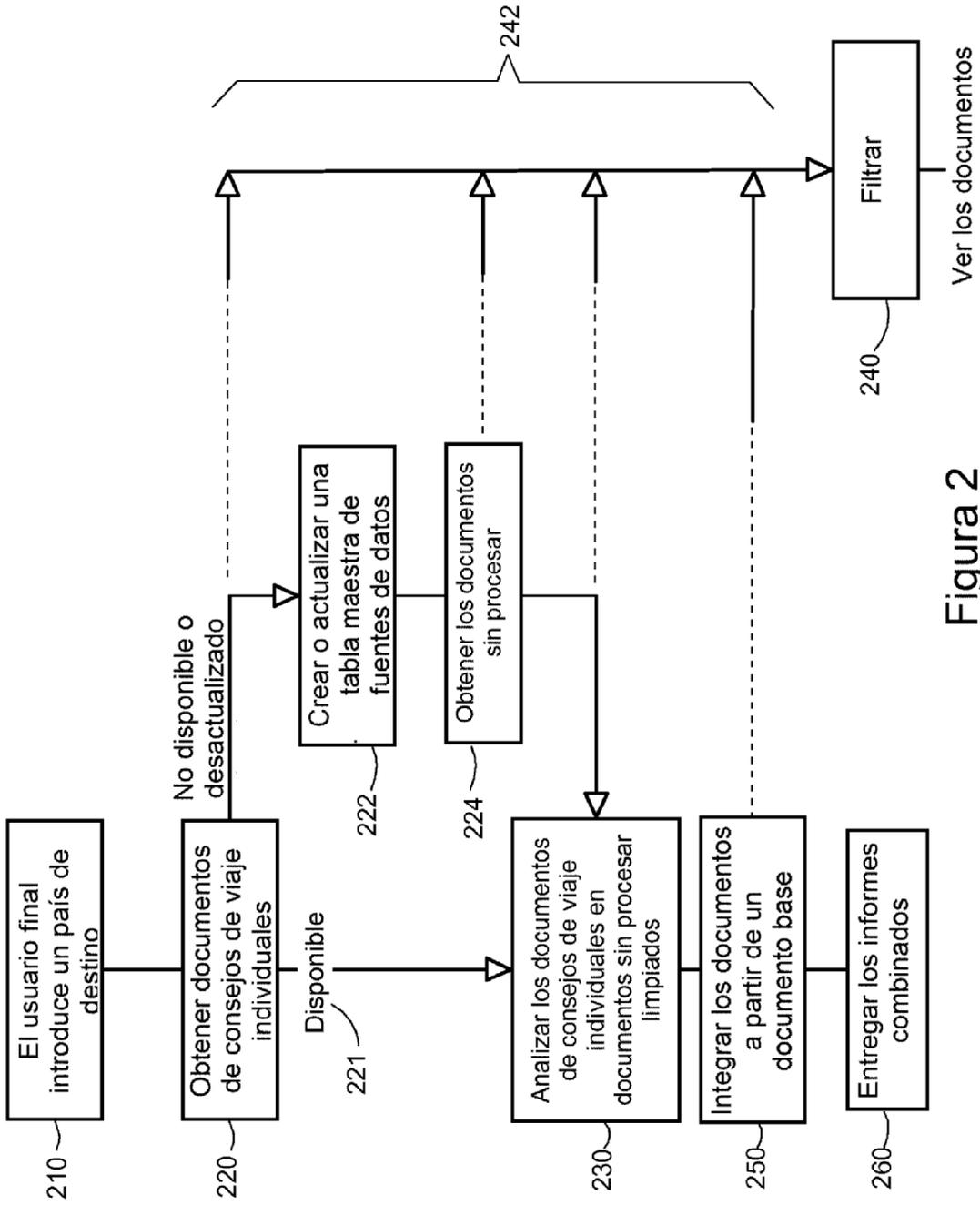


Figura 2



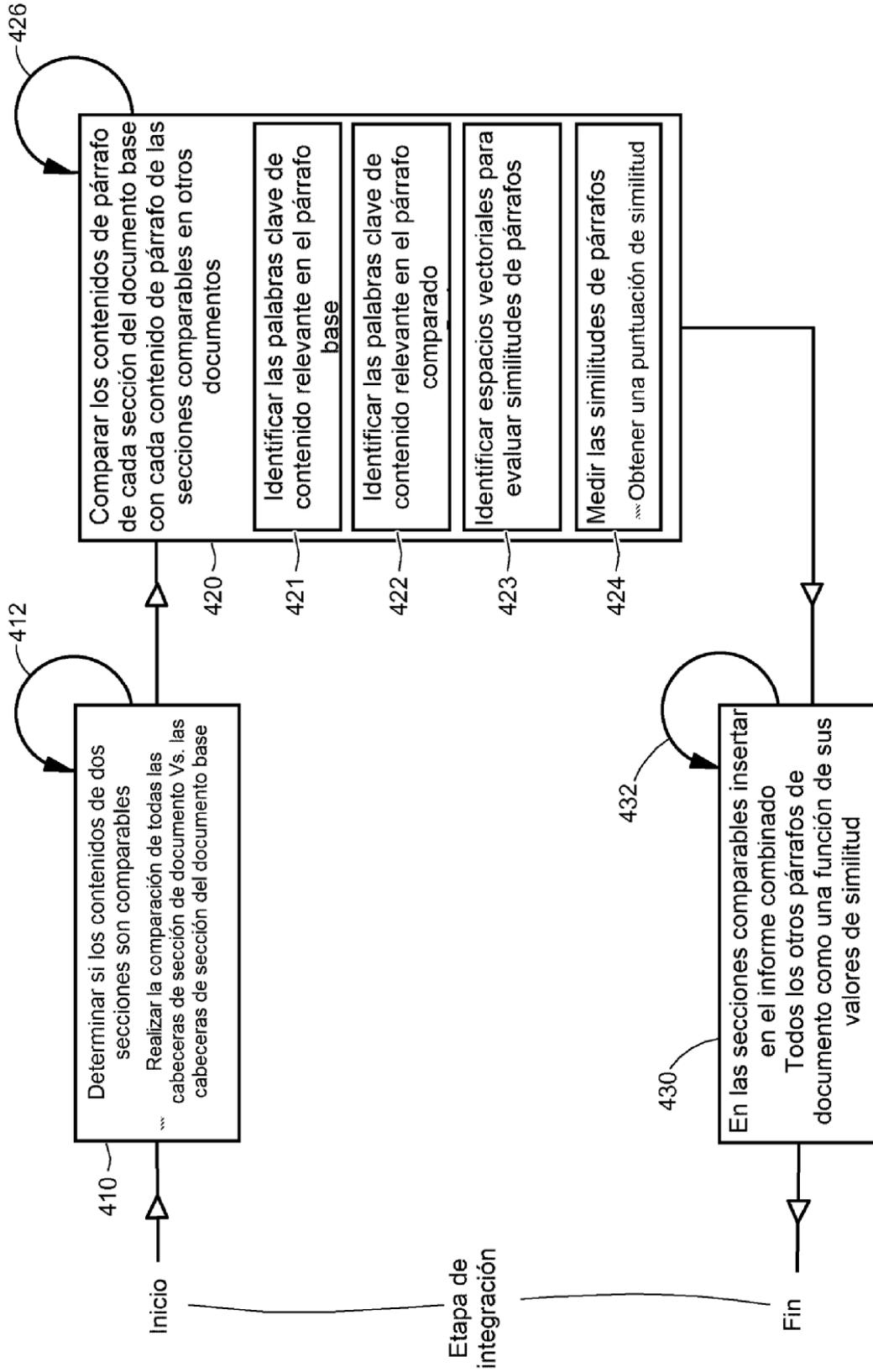


Figura 4