

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 616 953**

51 Int. Cl.:

G06F 17/30 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **04.09.2012 PCT/EP2012/067227**

87 Fecha y número de publicación internacional: **14.03.2013 WO2013034553**

96 Fecha de presentación y número de la solicitud europea: **04.09.2012 E 12758822 (6)**

97 Fecha y número de publicación de la concesión europea: **30.11.2016 EP 2754072**

54 Título: **Método y aparato para la generación automática de recomendaciones**

30 Prioridad:

08.09.2011 EP 11180631

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

14.06.2017

73 Titular/es:

**FUNKE DIGITAL TV GUIDE GMBH (100.0%)
Schiffbauerdamm 22
10117 Berlin / DE**

72 Inventor/es:

**KORST, JAN;
BARBIERI, MAURO;
CLOUT, RAMON y
PRONK, SERVERIUS, PETRUS, PAULUS**

74 Agente/Representante:

ROEB DÍAZ-ÁLVAREZ, María

ES 2 616 953 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

Método y aparato para la generación automática de recomendaciones

5 La presente invención se refiere a un motor de recomendaciones para recomendar artículos de contenido a un usuario, un depósito de contenidos, un método para hacer funcionar un motor de recomendaciones para recomendar artículos de contenido a un usuario y un medio de almacenamiento legible por ordenador.

10 Los sistemas de recomendación automáticos se utilizan a menudo para ayudar a los usuarios a seleccionar los artículos que se ajustan a su gusto. De un gran conjunto de artículos que una persona puede elegir, un sistema de recomendaciones hace una selección que se ajusta al gusto de un usuario determinado.

15 Antes de que un sistema de recomendaciones pueda dar recomendaciones verdaderamente personalizadas, primero tendrá que aprender el gusto del usuario. Para ello, el usuario normalmente tiene que clasificar una serie de artículos, por ejemplo, especificar en qué medida le gusta o no le gusta un número de artículos.

20 Los sistemas de recomendación pueden dividirse en dos categorías, a saber, los sistemas de recomendación basados en el contenido y los de recomendación basados en la filtración colaborativa. Para el primer tipo, los artículos tienen que caracterizarse por una serie de características. Por ejemplo, una película puede caracterizarse por el título, el género, el director, el elenco, etc. El historial de calificación de un usuario (la especificación de gustos y no gustos de un número de artículos) puede usarse entonces para estimar la correlación entre los valores-características y la probabilidad de que el usuario quiera un artículo con estos valores-características. Por el contrario, un sistema de recomendaciones que usa el filtrado colaborativo usa las clasificaciones de una gran comunidad de usuarios para extraer a partir de esta una similitud entre los usuarios (porque les gustan o no los mismos artículos) o una similitud entre los artículos (porque les gusta o no a los mismos usuarios). A continuación, esta información se usa o para recomendar los artículos que son similares a los artículos que el usuario ha especificado que desea o para recomendar los artículos que son apreciados por los usuarios que son similares al usuario determinado (y aún no se han visto o comprado por el usuario determinado). Los enfoques de filtrado colaborativo no necesitan una caracterización de los artículos en términos de valores-características.

30 En los últimos años, la popularidad de los servicios de redes sociales, tal como Facebook y LinkedIn, ha aumentado considerablemente. Estos servicios ayudan a un usuario a intercambiar fácilmente ideas, intereses, etc. con amigos, familiares y colegas. Estos servicios también ofrecen a los usuarios la posibilidad de expresar sus intereses enlazando entidades, tales como películas, música, celebridades, organizaciones, productos, etc. Cada una de estas entidades se especifica en una página web que da más detalles de la entidad específica. Por ejemplo, Facebook tiene una extensa colección de estas entidades que pueden buscarse por los usuarios.

35 Supongamos que un usuario quiere expresar sus intereses para una entidad determinada. Si ya existe una página web para esta entidad, entonces el usuario simplemente puede presionar el botón "me gusta" correspondiente y un enlace a esta página web se añadirá al perfil del usuario. Si ninguna página web adecuada expresa la entidad de su interés, el usuario puede crear dicha página web añadiendo adicionalmente una información textual sobre la entidad. Para muchas entidades, esta información se extrae de Wikipedia u otras fuentes, proporcionando información detallada de alta calidad.

45 Para ambas categorías de sistemas de recomendación mencionadas anteriormente, un usuario que es nuevo en el sistema de recomendaciones primero tiene que clasificar una serie de artículos antes de que el recomendador pueda generar unas recomendaciones personalizadas útiles. Esto puede obstaculizar el uso generalizado de un sistema de recomendaciones, ya que los usuarios no siempre están dispuestos a invertir inicialmente tiempo y esfuerzo en "explicar" al sistema sus gustos. Aún, los usuarios esperan recomendaciones inmediatas. Un sistema de recomendaciones podrá aprender el gusto de un usuario a través del tiempo, pero en ese caso las recomendaciones inicialmente no estarán optimizadas para el usuario específico.

50 Una forma de abordar este problema es permitir que el sistema de recomendaciones recomiende inicialmente unos artículos que son apreciados por muchos usuarios. Sin embargo, un usuario crítico puede no apreciar estas recomendaciones como muy valiosas, y él o ella puede dejar de usar el sistema de recomendaciones antes de que haya sido capaz de ajustar sus recomendaciones.

55 Otro enfoque se encuentra en Chumki Basu ET AL: "Technical paper recommendation: A study in combining multiple information sources", Journal of Artificial Intelligence Research 1,1 de enero de 2001 (01-01-2001), páginas 231-252. En este artículo se propone el uso del sistema WHIRL para recuperar los aciertos de múltiples fuentes de información, en las que los investigadores pueden recibir punteros diarios a los periódicos en sus campos de interés.

60 De acuerdo con un primer aspecto de la presente invención, un motor de recomendaciones para recomendar artículos de contenido a un usuario, comprende

65

- una unidad de generación de perfiles que tiene una entrada de pre-perfil, que está configurada para recibir de un base de datos, que es externa al motor de recomendaciones, datos de pre-perfil que comprenden datos de texto de pre-perfil adecuados para identificar las entidades de interés para un usuario determinado, y que tiene una unidad de análisis de pre-perfiles, que está conectada con la entrada de pre-perfil y configurada para extraer de los datos de pre-perfil datos de identificación que identifican las entidades de interés y para generar un conjunto de datos de perfil de usuario inicial para el usuario determinado a partir de los datos de identificación extraídos;
- una unidad de generación de consultas, que está conectada con la unidad de generación de perfiles y configurada para generar, usando los datos de identificación extraídos del conjunto de datos de perfil de usuario inicial, al menos dos consultas semánticamente diferentes una de otra para dirigirse a al menos un depósito de contenidos;
- una unidad de recuperación de contenidos, que está conectada con la unidad de generación de consultas y configurada para emitir las consultas generadas al al menos un depósito de contenidos y que está configurada para recibir desde el al menos un depósito de contenidos, en respuesta a las consultas, unos datos de respuesta relacionados con el contenido que comprenden unas listas de aciertos respectivas que tienen al menos un identificador de localización de contenido respectivo indicativo de una localización de almacenamiento de un artículo de contenido respectivo; y
- una unidad de entrelazado, que está conectada con la unidad de recuperación de contenidos y que está configurada para generar a partir de las diferentes listas de aciertos una única lista de recomendaciones entrelazando entre sí los identificadores de localización de contenido comprendidos en las diferentes listas de aciertos.

La unidad de entrelazado combina las listas de aciertos resultantes de las al menos dos consultas explícitas que son semánticamente distintas entre sí. En otras palabras, se entrelazan al menos dos listas de aciertos para generar las listas de recomendaciones en lugar de usar una sola consulta disyuntiva. La creación explícita de varias consultas semánticamente diferentes que conducen a unas listas de aciertos distintas crea la posibilidad de aplicar unos algoritmos de entrelazado avanzados en las listas separadas, por ejemplo, para lograr un grado suficiente de diversidad en la lista resultante.

El motor de recomendaciones del primer aspecto de la invención se basa en el concepto de permitir la generación de recomendaciones de artículos de contenido a un nuevo usuario creando automáticamente un perfil de usuario inicial sobre la base de los datos relacionados con el usuario, denominados en el presente documento como datos de pre-perfil, que se reciben desde una base de datos externa. Un nuevo usuario de un motor de recomendaciones puede, por ejemplo, proporcionar acceso a datos bajo su cuenta en una base de datos de una red social electrónica, ejemplos de los cuales se conocen bajo las marcas comerciales Facebook o LinkedIn. Existen muchas otras redes sociales electrónicas.

Una red social electrónica contiene datos específicos del usuario sobre entidades, tales como personas, artistas, grupos de personas, ciudades, países, clubes, partidos políticos, empresas, ideas, teorías, ciencias, cosas de todo tipo, juegos, obras de arte, por ejemplo, piezas de música, películas, obras de teatro, artículos, libros, fotos, grabados, pinturas, estilos de arte, eventos, actividades, deportes, etc. El término entidad se usa en la presente solicitud para referirse a cualquiera de tales intereses de un usuario determinado.

En consecuencia, el motor de recomendaciones de la presente invención acelera fuertemente el proceso de aprendizaje del gusto del usuario permitiendo recomendaciones sobre la base de la actividad documentada del usuario en un entorno externo al motor de recomendaciones, tal como una red social electrónica. Al mismo tiempo, la invención reconoce y supera otro problema importante del uso temprano de un motor de recomendaciones por parte de un usuario mediante el entrelazado de recomendaciones de diferentes listas de contenido, denominadas en el presente documento también como listas de aciertos, recuperadas por el motor de recomendaciones. Este entrelazado mejora la diversidad de las recomendaciones, que es un requisito de funcionalidad importante de un motor de recomendaciones, en particular para la región superior de la lista de recomendaciones, tal como puede medirse por la frecuencia de uso general de las recomendaciones generadas. Por lo tanto, el motor de recomendaciones de la presente invención está capacitado para presentar al usuario una lista única de recomendaciones no solo de lo que el usuario ya conoce, sino, entrelazando las recomendaciones que resultan de dos consultas semánticamente diferentes, un conjunto diverso de recomendaciones.

Por lo tanto, al combinar la evaluación automatizada de los datos de pre-perfil importados y un entrelazado de recomendaciones de diferentes listas de aciertos de contenido para un usuario determinado, el motor de recomendaciones de la presente invención consigue una estrecha adhesión de las recomendaciones a los intereses y a las expectativas reales del usuario de un sistema de recomendaciones desde el principio, sin requerir una entrada de redacción o una interacción inicial con el usuario con respecto a su gusto. Ambas medidas, por lo tanto, aceleran de manera colaborativa el proceso de aprendizaje específico del usuario del motor de recomendaciones, inmediatamente después de la primera interacción con el usuario determinado. Para un usuario interesado que recibe recomendaciones cercanas a sus intereses desde el principio, interactúa más frecuentemente y, por lo tanto, proporciona más rápidamente la información necesaria para refinar el conjunto de datos de perfil de usuario inicial. Esto a su vez mejora la percepción general de la calidad y la adecuación de las recomendaciones y proporciona una mejor experiencia del producto al usuario.

A continuación, se describirán las realizaciones del motor de recomendaciones del primer aspecto de la invención. Las características adicionales de las diferentes realizaciones pueden combinarse entre sí para formar otras realizaciones, a menos que se excluyan explícitamente en la presente memoria descriptiva.

5 Preferentemente, el motor de recomendaciones ordena las recomendaciones de acuerdo con la relevancia. Con este fin, la unidad de recuperación de contenidos está configurada en una realización adicional para recuperar los datos de texto de contenido asociados con los identificadores de localización de contenido respectivos en respuesta a las consultas. De esta manera, la relación textual entre los datos de pre-perfil y los datos de texto de contenido puede usarse para evaluar automáticamente la relevancia. Con este fin, una realización comprende una unidad de
 10 clasificación, que está conectada con la unidad de recuperación de contenidos y que está configurada
 - para asignar a los identificadores de localización de contenido, que se han recibido en respuesta a cada una de las al menos dos consultas, unos indicadores de relevancia basándose en un criterio de similitud que evalúa una relación textual entre los datos de texto de pre-perfil y los datos de texto de contenido,
 - para ordenar las listas de aciertos de acuerdo con la relevancia que se expresa por los indicadores de relevancia, y
 15 - para proporcionar las listas de aciertos ordenadas a la unidad de entrelazado.

Por lo tanto, la unidad de clasificación difiere de otros enfoques de clasificación en que la clasificación no se basa en el perfil de usuario (como en la técnica anterior) sino en los datos de pre-perfil, por ejemplo, sobre la base de los datos relacionados con el usuario que pueden recuperarse por ejemplo, de las redes sociales.

20 En otra realización, la unidad de generación de perfiles comprende una base de datos de clasificación, que asigna una clase respectiva de entidades de acuerdo con al menos un criterio de clasificación de entidades a un conjunto respectivo de al menos una palabra clave a incluir en una consulta.
 - en el que la unidad de generación de perfiles está configurada para asignar al menos una clase a los datos de
 25 identificación extraídos que identifican una entidad de interés respectiva de acuerdo con la base de datos de clasificación, y
 - en el que la unidad de generación de consultas está configurada para generar las consultas usando los datos de identificación respectivos y al menos una de las palabras clave asignadas a la clase respectiva de los datos de
 30 identificación de acuerdo con la base de datos de clasificación.

Las ventajas de esta realización se ilustran mediante el siguiente ejemplo de aplicación: sabiendo que una entidad identificada en los datos de pre-perfil se refiere a una persona, puede resultar - en virtud del motor de recomendaciones de la presente realización - en una consulta, en la que el nombre de la persona se combina con la palabra clave "entrevista" o "biografía". Estas consultas, emitidas en un depósito de contenidos de un proveedor de
 35 contenidos, como YouTube o Wikipedia, pueden dar como resultado unas recomendaciones de videos que contienen una entrevista con la persona o un sitio web que contiene material biográfico sobre la persona determinada.

En una realización adicional, la unidad de análisis de pre-perfiles está configurada además para clasificar los datos de texto de pre-perfil por su lenguaje y para proporcionar en su salida al menos un identificador de lenguaje indicativo de un lenguaje respectivo usado en los datos de texto de pre-perfil. La base de datos de clasificación del motor de recomendaciones de esta realización comprende las palabras clave en diferentes lenguajes. La unidad de generación de consultas está configurada para generar las consultas usando las palabras clave en el lenguaje correspondiente al identificador de lenguaje. Por ejemplo, si la entidad se refiere a un autor francés, entonces el nombre del autor se combina con "écrit par" en lugar de "escrito por" en una consulta para encontrar libros que él o ella ha escrito. En una variante de esta realización, la unidad de generación de consultas está configurada adicionalmente para generar las consultas usando las palabras clave en un lenguaje predeterminado si no están presentes palabras clave en otro lenguaje especificado en la base de datos de clasificación para la clase respectiva.

40 Para mejorar adicionalmente la diversidad de la lista de recomendaciones generada, la unidad de generación de consultas está configurada en una realización para incluir los datos de identificación extraídos de al menos dos entidades de interés en las consultas.

En otra realización, la unidad de generación de perfiles está configurada para detectar en los datos de pre-perfil un enlace de recursos a un recurso de contenido en una localización de red de una red de datos de área amplia, para acceder al recurso de contenido y para añadir los datos de texto disponibles desde el recurso de contenido a los datos de texto de pre-perfil. Tales enlaces se refieren normalmente a entidades que están relacionadas semánticamente con una entidad "me gusta" de origen. Por ejemplo, si una entidad "me gusta" de origen es un director de cine, entonces los enlaces normalmente especifican los nombres de las películas que él o ella ha dirigido.
 55 Para un autor, pueden especificar los nombres de los libros que él o ella ha escrito. Los enlaces también pueden referirse a artistas relacionados o el tipo de movimiento de arte con el que la entidad "me gusta" suele estar asociada. Por lo tanto, generando consultas específicas que combinan estas partes con posiblemente diferentes palabras clave de dirección adicionales como "escrito por", "influenciado por", etc., y emitiendo estas consultas a los depósitos de contenido específico, se obtiene una amplia gama de resultados devueltos usando los diversos enlaces
 60 potenciales que pueden encontrarse en la descripción textual.
 65

En una variante de esta realización, la unidad de generación de perfiles está configurada como alternativa o adicionalmente para explorar los datos de pre-perfil para los segmentos resaltados por tipos de etiquetas de marcado distintas de las que identifican un enlace, tal como una apariencia en negrita, etc., y para incluir tales segmentos en las consultas como se ha descrito anteriormente.

5 Con el fin de mejorar aún más la generación de recomendaciones, la unidad de generación de perfiles de una realización del motor de recomendaciones está configurada para extraer de los datos de pre-perfil diferentes subconjuntos de los datos de texto de pre-perfil, que se distinguen entre sí por su fecha de generación que se asigna por la base de datos externa, y para extraer de los subconjuntos las fechas respectivas de generación de los subconjuntos. Esto permite el filtrado de los subconjuntos de acuerdo con su fecha de generación. La unidad de clasificación de esta realización está a su vez configurada preferentemente para aplicar una ponderación al indicador de relevancia, aumentando más la ponderación del indicador de relevancia cuanto más reciente es la fecha de generación de uno respectivo de los subconjuntos relacionados con una localización de contenido determinada perteneciente a una entidad de interés determinada.

15 En una realización adicional, la unidad de análisis de pre-perfiles está configurada para detectar a partir de los datos de texto de pre-perfil la presencia de palabras clave que indican que el usuario determinado o alguna otra entidad (normalmente una persona) relacionada con el usuario determinado de acuerdo con la base de datos externa le gusta una entidad/artículo y para asignar un indicador de "me gusta" a la entidad/artículo respectivo en el conjunto de datos de perfil de usuario inicial. La unidad de clasificación de esta realización está configurada preferentemente para aplicar una ponderación al indicador de relevancia, aumentando la ponderación del indicador de relevancia si una entidad/artículo de interés determinado tiene un indicador de "me gusta" asociado.

20 En una realización adicional del motor de recomendaciones, la unidad de clasificación está configurada para evaluar un valor de relación textual entre los datos de texto de pre-perfil y los datos de texto de contenido. Preferentemente, la relación textual se evalúa usando un cálculo de un peso de frecuencia de documento inversa de término de frecuencia, en lo sucesivo en el presente documento un peso tf-idf. Una implementación de esta realización puede basar la evaluación en un conjunto de palabras comprendidas en los datos de texto de pre-perfil y en los datos de texto de contenido. Por ejemplo, las palabras que superan un peso tf-idf predefinido pueden identificarse y compararse con los datos de texto de pre-perfil en un lado y con los datos de texto de contenido en el otro lado. Adicionalmente o como alternativa, puede usarse un criterio de similitud predefinido matemáticamente de los pesos td-idf para palabras que se producen en ambos tipos de datos de texto para evaluar la relación textual.

25 En una realización se implementa un enfoque adicional o alternativo para ponderar la relevancia de los aciertos, en la que la unidad de clasificación está configurada para ponderar los indicadores de relevancia basándose en un criterio de similitud que evalúa una relación textual entre los datos de texto de contenido de los diferentes artículos de contenido recuperados en las consultas. En esta realización, la ponderación disminuye el indicador de relevancia si un artículo de contenido a evaluar para la relevancia tiene una relación textual con un artículo de contenido evaluado inmediatamente antes de que la relación textual supere un valor predeterminado. Esta realización mejora además la diversidad de la lista de recomendaciones generada en que se evalúan los resultados de la consulta que tienen una alta similitud recíproca de acuerdo con su relación textual como no igualmente relevantes, favoreciendo por lo tanto automáticamente a solamente uno de los resultados de consulta recíprocamente similares para la única lista de recomendaciones a generar.

30 En una realización adicional, el motor de recomendaciones comprende además una unidad de autenticación, que está configurada para recibir a través de una interfaz de entrada de usuario los datos de autenticación de usuario adecuados para acceder a la base de datos externa. La unidad de generación de perfiles está configurada para acceder a la base de datos externa para recuperar los datos de pre-perfil.

35 Con el fin de continuar la adaptación del perfil de usuario, otra realización del motor de recomendaciones comprende además una unidad de mantenimiento de perfilases además de la unidad de generación de perfiles. La unidad de mantenimiento de perfilases está configurada para extraer datos de identificación adicionales para una entidad adicional a partir de los datos de texto de contenido recuperados. La unidad de mantenimiento de perfilases está configurada preferentemente de manera adicional para añadir los datos de identificación adicionales extraídos al conjunto de datos de perfil de usuario al detectar un valor de relación textual entre los datos de texto de pre-perfil y los datos de texto de contenido que superan un valor umbral predeterminado. Por lo tanto, en esta realización, el perfil de usuario está adaptado además, mediante un mantenimiento adicional del perfil de usuario generado inicialmente.

40 Un segundo aspecto de la presente invención está formado por un depósito de contenidos, que comprende
 - una base de datos de contenidos que comprende unos artículos de contenido en la forma de archivos de datos almacenados en localizaciones de contenido;
 - un motor de recomendaciones de acuerdo con el primer aspecto de la presente invención o de acuerdo con una de sus realizaciones desvelada en la presente memoria descriptiva que incluye las reivindicaciones.

45

En el depósito de contenidos del segundo aspecto de la segunda invención, la unidad de recuperación de contenidos del motor de recomendaciones está configurada para emitir las consultas generadas a la base de datos de contenidos. Esto no es para excluir principalmente la emisión de las consultas generadas a otras bases de datos de contenido externas no comprendidas por el depósito de contenidos del presente aspecto de la invención. Sin embargo, en una realización, la emisión de las consultas está de hecho restringida a la base de datos de contenidos del propio depósito de contenidos.

Las realizaciones del depósito de contenidos comprenden al menos una de las realizaciones del motor de recomendaciones de acuerdo con el primer aspecto de la invención. Las ventajas del depósito de contenidos del segundo aspecto de la invención y sus realizaciones corresponden por lo tanto a las descritas anteriormente en el contexto respectivo del primer aspecto de la invención y no se repetirán en el presente contexto.

De acuerdo con un tercer aspecto de la invención, un método para hacer funcionar un motor de recomendaciones para recomendar artículos de contenido a un usuario, comprende

- recibir de una base de datos, que es externa al motor de recomendaciones, los datos de pre-perfil que comprenden los datos de texto de pre-perfil adecuados para identificar las entidades de interés para un usuario determinado,
- extraer de los datos de pre-perfil los datos de identificación que identifican las entidades de interés
- generar un conjunto de datos de perfil de usuario inicial para el usuario determinado a partir de los datos de identificación extraídos,
- generar, usando los datos de identificación extraídos a partir del conjunto de datos de perfil de usuario inicial, al menos dos consultas semánticamente diferentes una de otra para dirigirse a al menos un depósito de contenidos;
- emitir las consultas generadas al al menos un depósito de contenidos;
- recibir del al menos un depósito de contenidos, en respuesta a la consulta, los datos de respuesta relacionados con el contenido que comprenden las listas de aciertos respectivas que tienen al menos un identificador de localización de contenido respectivo indicativo de una localización de almacenamiento de un artículo de contenido respectivo;
- generar a partir de las diferentes listas de aciertos recuperadas una única lista de recomendaciones entrelazando entre sí los identificadores de localización de contenido comprendidos en las diferentes listas de aciertos.

El método del tercer aspecto corresponde estrechamente a la funcionalidad del motor de recomendaciones del primer aspecto de la invención. Por lo tanto, para una descripción de sus ventajas y realizaciones, se hace referencia a la descripción del motor de recomendaciones del primer aspecto de la invención y sus diversas realizaciones en la presente memoria descriptiva y en las reivindicaciones.

Un cuarto aspecto de la presente invención está formado por un medio de almacenamiento legible por ordenador que almacena un código de programa ejecutable, implementado el código de programa un método para controlar el funcionamiento de un motor de recomendaciones para recomendar artículos de contenido a un usuario de acuerdo con el tercer aspecto de la invención o una de sus realizaciones.

Las realizaciones preferidas de la invención también se definen en las reivindicaciones dependientes. Los aspectos anteriores y otros aspectos de la invención serán evidentes a partir de y se aclararán haciendo referencia a las realizaciones descritas en lo sucesivo en el presente documento. En los siguientes dibujos

la figura 1: muestra un diagrama de bloques de un motor de recomendaciones, y de un depósito de contenidos de acuerdo con una realización;

la figura 2: muestra un diagrama de flujo de un método para hacer funcionar un motor de recomendaciones de acuerdo con una realización adicional.

La figura 1 muestra un diagrama de bloques de un motor de recomendaciones de acuerdo con una realización. La figura 1 también se usará más adelante para explicar una realización de un depósito de contenidos.

El motor de recomendaciones 100 de la figura 1 sirve en general para recomendar artículos de contenido a un usuario. El usuario hace funcionar un dispositivo de visualización de contenidos 102. El motor de recomendaciones 100 de la presente realización se hace funcionar como un dispositivo que está físicamente separado del dispositivo de visualización 102. En otras palabras, en la presente realización el dispositivo de visualización es externo al motor de recomendaciones 100. En otra realización útil, el motor de recomendaciones está integrado con el dispositivo de visualización de contenidos 102. En cualquiera de estas realizaciones, el motor de recomendaciones 100 y el dispositivo de visualización de contenidos 102 están comunicativamente conectados entre sí.

También puede establecerse una conexión comunicativa entre el motor de recomendaciones 100 y un dispositivo de terminal de usuario 104, que normalmente es alguna forma de ordenador, tal como un ordenador de sobremesa, un ordenador móvil como un ordenador portátil o un teléfono inteligente. El dispositivo de terminal 104 y el dispositivo de visualización de contenidos 102 están en muchos casos de aplicación integrados en un solo dispositivo. Sin embargo, por razones de claridad con respecto a la funcionalidad en su interacción con el motor de recomendaciones 100, la figura 1 los muestra como bloques separados, también para señalar que de hecho pueden implementarse en dispositivos físicamente separados.

El dispositivo de terminal 104 está configurado para intercambiar información con una base de datos 106 de una red social electrónica. Como es bien conocido, un usuario puede emplear el dispositivo de terminal 104 para introducir y por lo tanto transmitir a la red social, es decir, para los fines de la presente descripción: a la base de datos 106, su información personal, comentarios, artículos favoritos (me gusta), enlaces a sitios web, etc., y recibir entradas similares de otros usuarios (sus “amigos”) de la red social.

El dispositivo de terminal 104 también puede usarse para interactuar con el motor de recomendaciones 100, como se describirá más adelante.

El motor de recomendaciones 100 comprende una unidad de generación de perfiles 108. La unidad de generación de perfiles 108 está dividida en una unidad de entrada de pre-perfiles 110, que puede conectarse comunicativamente con la base de datos 106 o (no mostrado en la figura 1) el dispositivo de terminal 104, o ambos. La entrada de pre-perfil está configurada para recibir datos de pre-perfil que comprenden datos de texto de pre-perfil adecuados para identificar las entidades de interés para un usuario determinado. Por ejemplo, los datos de pre-perfil pueden comprender una página personal “me gusta”, es decir, un código, normalmente en un lenguaje de marcas como html o xml, que especifica las entidades de interés para el usuario, por medio de publicaciones, comentarios, otros tipos de texto o fragmentos de texto, enlaces para mostrar usando un software de navegador web y que se mantiene por el usuario a través de sus entradas a la base de datos 106, usando también una interfaz de navegador web presentada al usuario por un proveedor de la red social.

La unidad de generación de perfiles 108 comprende además una unidad de análisis de pre-perfiles 112, que está conectada con la unidad de entrada de pre-perfiles 110 y está configurada para extraer de los datos de pre-perfil los datos de identificación que identifican las entidades de interés y para generar un conjunto de datos de perfil de usuario inicial para el usuario determinado a partir de los datos de identificación extraídos. Por ejemplo, una página web que especifica una entidad que le “gusta” a un usuario determinado, contiene normalmente un número de características que están casi siempre presentes, tales como una categoría, una imagen de la entidad y un número de usuarios a los que ha “gustado” la entidad. El texto que especifica la categoría es normalmente un texto libre, es decir, puede ser cualquier fragmento de texto, pero se pueden reconocer las categorías más frecuentes, tales como por ejemplo “músico/banda” o “figura pública”. Además, se extraen las partes de la página “me gusta” que contienen grandes fracciones de texto. Estos se denominan normalmente como “descripción”, “sobre”, etc. Estos fragmentos de texto se analizan para obtener etiquetas de marcado específicas, que indican enlaces, resaltados, negritas, etc. Estas partes se refieren normalmente a entidades que están relacionadas semánticamente con la entidad “me gusta” de origen. Por ejemplo, si la entidad “me gusta” de origen es un director de cine, entonces estas partes pueden especificar los nombres de las películas que él o ella ha dirigido. Para un autor, pueden especificar los nombres de los libros que él o ella ha escrito. Sin embargo, estas partes específicas pueden referirse también a artistas relacionados o el tipo de movimiento de arte con el que la entidad “me gusta” está normalmente asociada. Además, muchos de los medios sociales, tales como Facebook, YouTube y Twitter, permiten a los usuarios “publicar” comentarios e información sobre los medios sociales, también relacionados con los artículos, para actualizar a los amigos acerca de la actividad o intereses de uno. La información contenida en una publicación o comentario y el “me gusta” (no necesariamente perteneciente al usuario que publica el comentario, sino a un amigo) a la que están asociados estas publicaciones y comentarios, también pueden usarse como datos de pre-perfil por la unidad de análisis de pre-perfiles en el proceso de generar un conjunto de datos de perfil de usuario inicial a usar para generar las primeras recomendaciones al usuario.

Para identificar si el comentario de un amigo en una entidad “me gusta” debería interpretarse como positivo o negativo, el sentimiento del comentario puede analizarse en una variante de la presente realización, por ejemplo, por la unidad de análisis de pre-perfiles 112 que busca la aparición de palabras que se asocian normalmente con el sentimiento positivo y las palabras que se asocian normalmente con el sentimiento negativo.

Una opción para la adquisición de datos de pre-perfil es equipar el motor de recomendaciones con una unidad de autenticación, que está configurada para recibir a través de una interfaz de entrada de usuario datos de autenticación de usuario adecuados para acceder a la base de datos 106. Por lo tanto, un usuario desvela su información de autenticación de usuario necesaria para acceder a la base de datos 106 al motor de recomendaciones 100 a través de la unidad de autenticación 114. En este caso, la unidad de generación de perfiles 108 está configurada para acceder a la base de datos externa 106 para recuperar los datos de pre-perfil empleando los datos de autenticación de usuario.

Sin embargo, el suministro de la unidad de autenticación 112 es una característica no necesaria del motor de recomendaciones 100. Los datos de pre-perfil pueden proporcionarse por el propio usuario a través de su dispositivo de terminal 104. Por ejemplo, el usuario puede descargar primero una copia de los datos personales de una red social respectiva y a continuación proporcionar la copia como datos de pre-perfil al motor de recomendaciones a través de la unidad de entrada de pre-perfiles 110.

La unidad de análisis de perfiles 112 de la unidad de generación de perfiles 108 del motor de recomendaciones 100 está configurada adicionalmente para asignar al menos una clase a los datos de identificación extraídos que identifican una entidad de interés respectiva de acuerdo con una base de datos de clasificación 116. Las clases de

entidades pueden diferenciar, por ejemplo, las entidades asignando un atributo tal como persona, artista, grupos de personas, ciudad, estado, país, organización, club, partido político, compañía, idea, teoría, ciencia, cosa (de cualquier clase), juego, obra de arte, por ejemplo, una pieza de música, película, juego, artículo, libro, foto, impresión, pintura, estilo de arte, evento, actividad, deporte, etc. La base de datos de clasificación 116 asigna adicionalmente una clase respectiva de entidades a un conjunto respectivo de al menos una palabra clave a incluir en una consulta.

La base de datos de clasificación 116 comprende ventajosamente las palabras clave en diferentes lenguajes. Esto permite servir al usuario con las recomendaciones en su lenguaje preferido seleccionado a partir de un número de lenguajes disponibles, como se explicará más adelante. En el lado de la unidad de generación de perfiles, esta característica se soporta por la unidad de análisis de pre-perfiles 112 que está configurada además para clasificar los datos de texto de pre-perfil por su lenguaje y para proporcionar en su salida al menos un identificador de lenguaje indicativo de un lenguaje respectivo usado en los datos de texto de pre-perfil.

El conjunto de datos de perfil de usuario inicial generado por la unidad de generación de perfiles se almacena en una base de datos de perfiles de usuario 118.

El motor de recomendaciones 100 comprende además una unidad de generación de consultas 120. La unidad de generación de consultas 120 está conectada con la unidad de generación de perfiles 108 y configurada para generar, usando los datos de identificación extraídos del conjunto de datos de perfil de usuario inicial, al menos dos consultas semánticamente diferentes una de otra para dirigirse a al menos un depósito de contenidos. Se representan dos depósitos de contenidos diferentes en la figura 1 bajo las etiquetas de referencia 124 y 126, y se resumen para facilitar la representación gráfica bajo la etiqueta de referencia 122. Con el fin de la descripción de la presente realización, se supone que los depósitos de contenido pueden accederse por el motor de recomendaciones a través de una red de comunicación pública tal como Internet. Ejemplos de este tipo de depósitos de contenido son, por ejemplo, las fuentes de contenido de libre acceso tales como YouTube o Wikipedia, pero también pueden ser fuentes de contenido de acceso restringido tales como las bases de datos de películas comerciales.

Al generar consultas específicas que combinan partes identificadas como una entidad de interés por la unidad de generación de perfiles 108 con diferentes palabras clave de dirección adicionales tales como “escrito por”, “influido por”, etc., y al emitir éstas a los depósitos de contenidos específicos 122, se obtiene una amplia gama de resultados devueltos usando las diversas pistas potenciales que se han encontrado en la descripción textual por la unidad de análisis de pre-perfiles. Además, ya que se identifica el lenguaje(s) usado en las descripciones textuales, como se ha mencionado anteriormente, esto se usa de manera ventajosa para generar las consultas usando el mismo lenguaje de las palabras clave que el usado en aquellas partes que identifican una entidad de interés. Por ejemplo, si la entidad se refiere a un autor francés y a una página “me gusta” correspondiente, entonces tiene sentido combinar el nombre del autor con “écrit par” en lugar de “escrito por” para tratar de encontrar los libros que él o ella ha escrito. La unidad de generación de consultas 120 se configura para este extremo para generar las consultas usando las palabras clave en el lenguaje correspondiente al identificador de lenguaje o, si no hay palabras clave presentes en este lenguaje en la base de datos de clasificación de la clase respectiva, en un lenguaje predeterminado.

Una unidad de recuperación de contenidos 128, que está conectada con la unidad de generación de consultas 120, está configurada para emitir las consultas generadas para el al menos un depósito de contenidos 128 y recibir desde el al menos un depósito de contenidos 128, en respuesta a la consultas, los datos de respuesta relacionados con el contenido que comprenden las listas de aciertos respectivas que tienen al menos un identificador de localización de contenido respectivo indicativo de una localización de almacenamiento de un artículo de contenido respectivo.

Obsérvese que la extensión de la descripción de la entidad “me gusta” como recuperada por la unidad de generación de perfiles 108 a partir de los datos de texto de pre-perfil influirá en la diversidad de las consultas emitidas y los resultados devueltos. Por ejemplo, suponiendo que un usuario de Facebook ha especificado que le “gusta” el compositor Mozart. La página de Facebook de Mozart contiene múltiples enlaces, por ejemplo, a ciudades como “Salzburgo” y “Viena”, pero también a “Requiem” y “Constanza”. La diversidad de estos términos, que pueden combinarse en las consultas con “Mozart”, proporciona resultados muy diversos. En particular, en este caso, las consultas que contienen las palabras clave “Viena” o “Salzburgo” pueden devolver resultados que no tienen nada que ver con Mozart. Por lo tanto una etapa adicional es ventajosa para filtrar los resultados irrelevantes que no deberían recomendarse. En consecuencia, el motor de recomendaciones de la presente realización comprende una unidad de clasificación 130, que está conectada con la unidad de recuperación de contenidos y que está configurada para asignar los identificadores de localización de contenido, que se han recibido en respuesta a cada una de las al menos dos consultas, a los indicadores de relevancia basándose en un criterio de similitud que evalúa una relación textual entre los datos de texto de pre-perfil y los datos de texto de contenido, para ordenar las listas de aciertos de acuerdo con la relevancia de acuerdo con lo expresado por los indicadores de relevancia, y para proporcionar en su salida las listas de aciertos ordenadas.

El motor de recomendaciones hace funcionar la unidad de clasificación 130 sobre la base de los resultados obtenidos en varias consultas, posiblemente para diferentes depósitos de contenido. El objetivo de la clasificación es

que la parte superior de la lista final de recomendaciones debería proporcionar un conjunto de resultados relevantes. La relevancia de un resultado se establece determinando la relación textual de un resultado con la entidad "me gusta" de origen. Hay un número de enfoques conocidos en la técnica para identificar la relación textual. El enfoque conocido como tf-idf, que ya se ha mencionado anteriormente en la presente memoria, se usa a menudo en la práctica. También es común el uso de un modelo de espacio vectorial, en el que cada texto (o documento) se representa por un vector multidimensional, donde cada dimensión corresponde a una palabra que se produce en el texto. La entrada de un vector-palabra de este tipo, se calcula multiplicando la frecuencia en término relativo (tf), es decir, el número de veces que la palabra determinada aparece en el texto, dividido por el número total de palabras en el texto, por la frecuencia de documento inversa (idf), que expresa la frecuencia con la que aparece una palabra en un documento de un corpus determinado de documentos. Obsérvese que un corpus es un lenguaje específico. Como ha descrito anteriormente, el lenguaje de un texto, a partir del que se generan las consultas, se identifica y puede usarse para seleccionar el corpus adecuado para calcular la frecuencia de documento inversa. De esta manera, tanto el texto de la entidad "me gusta" de origen y el texto que corresponde a un resultado determinado pueden representarse como vectores en un espacio dimensional alto y el coseno entre estos vectores puede usarse como medida de su relación textual.

Para implementar la clasificación, puede hacerse lo siguiente. Dejar que q_1, \dots, q_n sea la lista de consultas que se han enviado a los depósitos de contenido y, para cada q_i , dejar que $R(q_i)$ indique la lista de resultados devueltos o emitidos q_i . A continuación, para cada q_i , pueden ordenarse los resultados en $R(q_i)$ en orden decreciente de la relación textual a la entidad "me gusta" de origen.

La clasificación puede en algunas variantes implicar también la aplicación de un peso al indicador de relevancia evaluado. Sobre la base de las fechas extraídas de generación de diferentes entradas realizadas por un usuario a la base de datos 106, la unidad de clasificación 130 puede aplicar una ponderación al indicador de relevancia de acuerdo con la edad de la entrada de usuario respectiva. En otras palabras, aumentando más la ponderación del indicador de relevancia cuanto más reciente es la fecha de generación de uno respectivo de los subconjuntos relacionados con una localización de contenido determinada perteneciente a una entidad de interés determinada. Por ejemplo, las entidades "me gusta" pueden contener canales de noticias, que informan sobre las noticias recientes de la entidad correspondiente. También allí, puede proporcionarse una preferencia por la ponderación adecuada a las más recientes adiciones a la entidad "me gusta", para reclasificar los resultados, o incluso para preseleccionar los fragmentos de texto más recientes para la generación de consultas. En otra variante, la unidad de análisis de pre-perfiles está configurada para detectar a partir de los datos de texto de pre-perfil la presencia de palabras clave que indican que el usuario determinado o alguien relacionado con el usuario determinado le gusta una entidad. Por lo tanto, puede asignarse un indicador "me gusta" a la entidad respectiva en el conjunto de datos de perfil de usuario inicial, y la unidad de clasificación puede aplicar una ponderación al indicador de relevancia, aumentando la ponderación el indicador de relevancia si una entidad de interés determinada tiene un indicador "me gusta" asociado.

El suministro de una unidad de clasificación no es obligatorio para el motor de recomendaciones. La clasificación mediante el motor de recomendaciones puede omitirse, por ejemplo, en el proceso de generación de una única lista de recomendaciones, si los depósitos de contenido ya aplican una clasificación a sus listas de aciertos sobre la base de los términos usados para la consulta. Además, la clasificación puede implementarse como una opción que puede activarse o desactivarse por el usuario.

Con el fin de obtener una única lista de recomendaciones que proporcione diversidad en sus entradas, el motor de recomendaciones 100 contiene una unidad de entrelazado 132. La unidad de entrelazado 132 está conectada con la unidad de recuperación de contenidos 128 (en la presente realización a través de la unidad de clasificación 130), y está configurada para generar a partir de las diferentes listas de aciertos una única lista de recomendaciones entrelazando los identificadores de localización de contenido comprendidos en las diferentes listas de aciertos entre sí. El entrelazado de las listas ordenadas de los resultados de las diferentes consultas puede realizarse, por ejemplo, usando un enfoque de todos contra todos o usando un enfoque de programación más avanzado basado en créditos.

El entrelazado es una realización basada en una evaluación de la relación textual de los resultados obtenidos de diferentes consultas. La unidad de entrelazado está configurada para ponderar los indicadores de relevancia basándose en un criterio de similitud que evalúa una relación textual entre los datos de texto de contenido de diferentes artículos de contenido recuperados en las consultas, en el que la ponderación disminuye el indicador de relevancia si un artículo de contenido a evaluar para la relevancia tiene una relación textual con un artículo de contenido evaluado inmediatamente antes de que la relación textual supere un valor predeterminado. Además mediante esta implementación, se garantiza que al menos la parte superior de la lista final de recomendaciones contiene suficiente diversidad.

Además de ayudar a superar el período inicial donde el motor de recomendaciones 100 todavía tiene que aprender los gustos de un usuario, puede lograrse una aceleración del proceso de aprendizaje del sistema de recomendaciones de la siguiente manera. Los resultados que se identifican como suficientemente relevantes - por ejemplo, debido a que la similitud de la descripción del coseno con la de la entidad "me gusta" es suficientemente alta - puede interpretarse por el recomendador como un me gusta. Esto puede resultar directamente en un conjunto

relativamente grande de gustos. Para un motor de recomendaciones que usa un enfoque de filtrado colaborativo, éstos podrían añadirse directamente como unos me gustan en una tabla de artículos de usuario. Para un sistema de recomendaciones basado en contenidos, estos “me gusta” pueden traducirse en cambios de un igual grado de pares de característica-valor, siempre que estas características puedan identificarse a partir de los resultados. Esto puede realizarse si el resultado se reconoce como una entrada en una base de datos que almacena información de los pares característica-valor sobre los artículos de contenido. Para permitir tal proceso de aprendizaje mejorado, el motor de recomendaciones 100 de la presente realización comprende también una unidad de mantenimiento de perfiles 134. La unidad de mantenimiento de perfiles está configurada para adaptar el conjunto de datos de perfil de usuario inicial almacenado en la base datos de perfiles de usuario 118 en el ulterior proceso de uso. Con este fin, la unidad de mantenimiento de perfiles está configurada para extraer los datos de identificación adicionales para una entidad adicional a partir de los datos de texto de contenido recuperados, y para añadir los datos de identificación adicionales extraídos al conjunto de datos de perfil de usuario tras detectar un valor de la relación textual entre los datos de texto pre-perfil y los datos de texto de contenido que supera un valor umbral predeterminado.

Un mantenimiento adicional puede basarse en unos algoritmos conocidos que emplean las interacciones de usuario detectadas. Para permitir tal interacción de usuario, se proporciona una interfaz 136 para entregar al dispositivo de visualización 102 la única lista de recomendaciones generada proporcionada por la unidad de entrelazado 132, normalmente en forma de datos que permite una representación gráfica de la lista de recomendaciones para permitir la interacción intuitiva del usuario para la selección, la no selección u otro tipo de valoración de los artículos recomendados por el usuario. Tal entrada de usuario se evalúa por la unidad de mantenimiento de perfiles 134.

El motor de recomendaciones puede implementarse en hardware, usando una circuitería dedicada para las diferentes unidades funcionales del motor de recomendaciones. En otra realización, el motor de recomendaciones se implementa en un hardware de ordenador, usando uno o más procesadores controlados por programas para implementar las diferentes unidades funcionales del motor.

La figura 1 también es adecuada para ilustrar una realización de un depósito de contenidos 200 de acuerdo con la presente invención. En el depósito de contenidos, se hacen funcionar el motor de recomendaciones y al menos una base de datos de contenidos 122 por el mismo proveedor. Un acceso a las bases de datos de contenidos más allá de la base de datos de contenidos interna 122 del depósito de contenidos 200 de acuerdo con la descripción anterior puede o no puede implementarse, de acuerdo con el modelo de negocio del proveedor. Por lo tanto, el depósito de contenidos 200 puede restringir las recomendaciones a los artículos de contenido que figuran en la base de datos de contenidos del proveedor. Al emplear el motor de recomendaciones junto con la base de datos de contenidos, se crea una forma atractiva de publicitar los artículos de contenido a los usuarios de un depósito de contenidos.

La figura 2 muestra un diagrama de flujo de un método para hacer funcionar un motor de recomendaciones para recomendar unos artículos de contenido a un usuario de acuerdo con una realización de la presente invención. La realización comprende las siguientes etapas:

- etapa S1: recibir de una base de datos, que es externa al motor de recomendaciones, los datos de pre-perfil que comprenden los datos de texto de pre-perfil adecuados para identificar las entidades de interés para un usuario determinado;
- etapa S2: extraer de los datos de pre-perfil los datos de identificación que identifican las entidades de interés;
- etapa S3: generar un conjunto de datos de perfil de usuario inicial para el usuario determinado a partir de los datos de identificación extraídos;
- etapa S4: generar, usando los datos de identificación extraídos a partir del conjunto de datos de perfil de usuario inicial, al menos dos consultas semánticamente diferentes una de otra para dirigirse a al menos un depósito de contenidos;
- etapa S5: emitir las consultas generadas al al menos un depósito de contenidos;
- etapa S6: recibir del al menos un depósito de contenidos, en respuesta a la consulta, los datos de respuesta relacionados con el contenido que comprenden las listas de aciertos respectivas que tienen al menos un identificador de localización de contenido respectivo indicativo de una localización de almacenamiento de un artículo de contenido respectivo; y
- etapa S7: generar a partir de las diferentes listas de aciertos recuperadas una única lista de recomendaciones entrelazando entre sí los identificadores de localización de contenido comprendidos en las diferentes listas de aciertos.

Otra realización de un proceso de generar recomendaciones sobre la base de una página “me gusta” determinada, tal como una página web que especifica las entidades de interés, comprende las siguientes etapas.

1. Se analiza el texto de la página web, mediante la extracción de una categoría de una entidad y mediante la extracción de fragmentos de texto específicos a partir del texto.
2. Esta etapa se realiza siguiendo o 2a o 2b o ambos:

2a. Las consultas se generan usando la categoría de entidad me gusta y usando los fragmentos de texto extraídos. Estas consultas se envían a los depósitos de contenido en línea, tal como, por ejemplo, YouTube y Amazon, usando una API que ofrecen estos depósitos para este tipo de uso.

5 2b. En lugar de enviar las consultas a un depósito de contenidos posiblemente externo, se puede hacer coincidir como alternativa, los fragmentos de textos extraídos y la categoría de la entidad me gusta con las entradas en un depósito de contenidos interno, donde se tiene un control total sobre cómo se implementa la coincidencia.

10 3. Los resultados que se devuelven por los depósitos en línea se analizan para determinar, para cada uno de estos resultados, la relación semántica con la entidad “me gusta” de origen. Además, como una opción, puede analizarse la similitud recíproca entre cada par de los resultados. Ambos análisis se basan en la comparación de su similitud textual.

15 4. Posteriormente, los resultados se combinan en una única lista de recomendaciones, con el objetivo de tener en la parte superior de la lista los resultados que se relacionan altamente de manera semántica con el artículo “me gusta” de origen, pero que son suficientemente diferentes entre sí.

20 5. Como una opción, los resultados con una suficientemente alta similitud textual con la entidad “me gusta” de origen pueden incluirse directamente como un “me gusta” en el sistema de recomendaciones, de tal manera que se acelera el proceso de aprendizaje del perfil del usuario.

25 En resumen, la invención permite la generación de recomendaciones personales y diversas, especialmente para los nuevos usuarios que empiezan a usar un sistema de recomendaciones. Por lo tanto, se supera el periodo en el que el recomendador aún no ha aprendido el gusto del usuario, debido a la falta de clasificaciones por el nuevo usuario. Un usuario puede iniciar sesión en el sistema de recomendaciones usando su cuenta de servicio de red social y dando permiso al sistema de recomendaciones para usar los “me gusta” que el usuario ha especificado anteriormente. De esta manera, se proporciona al recomendador tiempo suficiente para aprender el gusto del usuario mientras que la lista de clasificaciones aumenta.

30 Al usar de manera apropiada las API de los diferentes depósitos de contenido, puede usarse la invención para recomendar diversos tipos de contenidos, incluyendo películas, programas de televisión, libros, artículos, documentos digitales, etc., por lo tanto, muy diferentes entidades, que incluyen personas e intereses generales.

35 Aunque la invención se ha ilustrado y descrito en detalle en los dibujos y la descripción anterior, tal ilustración y descripción han de considerarse ilustrativas o a modo de ejemplo y no restrictivas; la invención no se limita a las realizaciones desveladas.

40 Otras variaciones a las realizaciones desveladas pueden entenderse y efectuarse por los expertos en la materia en la práctica de la invención reivindicada, a partir de un estudio de los dibujos, la divulgación y las reivindicaciones adjuntas.

45 En las reivindicaciones, la palabra “comprende” no excluye otros elementos o etapas, y el artículo indefinido “un” o “una” no excluye una pluralidad. Una sola unidad puede cumplir las funciones de varios artículos citados en las reivindicaciones. El mero hecho de que ciertas medidas se citen en las reivindicaciones dependientes recíprocamente diferentes no indica que una combinación de estas medidas no pueda usarse ventajosamente.

50 Un programa informático puede almacenarse/distribuirse en un medio adecuado, tal como un medio de almacenamiento óptico o un medio de estado sólido suministrado junto con o como parte de otro hardware, pero también puede distribuirse de otras formas, tales como a través de Internet o de otros sistemas de telecomunicaciones por cable o inalámbricos.

Cualquier signo de referencia en las reivindicaciones no debería interpretarse como que limita el alcance.

REIVINDICACIONES

1. Un motor de recomendaciones (100) para recomendar artículos de contenido a un usuario, que comprende

- 5 - una unidad de generación de perfiles que tiene una unidad de entrada de pre-perfiles (110), que está configurada para recibir datos de pre-perfil que comprenden datos de texto de pre-perfil adecuados para identificar las entidades de interés para un usuario determinado, y que tiene una unidad de análisis de pre-perfiles (112), que está conectada con la unidad de entrada de pre-perfiles (110) y configurada para extraer de los datos de pre-perfil datos de identificación que identifican las entidades de interés y para generar un conjunto de datos de perfil de usuario inicial para el usuario determinado a partir de los datos de identificación extraídos;
- 10 - una unidad de generación de consultas (120), que está conectada con la unidad de generación de perfiles y configurada para generar, usando los datos de identificación extraídos del conjunto de datos de perfil de usuario inicial, al menos dos consultas semánticamente diferentes una de otra para dirigirse a al menos un depósito de contenidos;
- 15 - una unidad de recuperación de contenidos (128), que está conectada con la unidad de generación de consultas (120) y configurada para emitir las consultas generadas al al menos un depósito de contenidos (122) y que está configurada para recibir desde el al menos un depósito de contenidos, en respuesta a las consultas, unos datos de respuesta relacionados con el contenido que comprenden unas listas de aciertos respectivas que tienen al menos un identificador de localización de contenido respectivo indicativo de una localización de almacenamiento de un artículo de contenido respectivo; y
- 20 - una unidad de entrelazado (132), que está conectada con la unidad de recuperación de contenidos y que está configurada para generar a partir de las diferentes listas de aciertos una única lista de recomendaciones entrelazando entre sí los identificadores de localización de contenido comprendidos en las diferentes listas de aciertos.

2. El motor de recomendaciones de la reivindicación 1,

- 30 - en el que la unidad de recuperación de contenidos (128) está configurada además para recuperar los datos de texto de contenido asociados con los identificadores de localización de contenido respectivos en respuesta a las consultas; comprendiendo además el motor de recomendaciones (100)
- una unidad de clasificación (130), que está conectada con la unidad de recuperación de contenidos (128) y que está configurada
- 35 - para asignar a los identificadores de localización de contenido, que se han recibido en respuesta a cada una de las al menos dos consultas, unos indicadores de relevancia basándose en un criterio de similitud que evalúa una relación textual entre los datos de texto de pre-perfil y los datos de texto de contenido,
- para ordenar las listas de aciertos de acuerdo con la relevancia que se expresa por los indicadores de relevancia, y
- para proporcionar las listas de aciertos ordenadas a la unidad de entrelazado.

3. El motor de recomendaciones de la reivindicación 1 o 2,

- 40 - en el que la unidad de generación de perfiles (108) comprende una base de datos de clasificación (116), que asigna una clase respectiva de las entidades de acuerdo con al menos un criterio de clasificación de entidad a un conjunto respectivo de al menos una palabra clave a incluir en una consulta,
- 45 - en el que la unidad de generación de perfiles (108) está configurada para asignar al menos una clase a los datos de identificación extraídos que identifican una entidad de interés respectiva de acuerdo con la base de datos de clasificación y
- en el que la unidad de generación de consultas (120) está configurada para generar las consultas usando los datos de identificación respectivos y al menos una de las palabras clave asignadas a la clase respectiva de los datos de identificación de acuerdo con la base de datos de clasificación.

4. El motor de recomendaciones de la reivindicación 3,

- 55 - en el que la unidad de análisis de pre-perfiles (112) está configurada además para clasificar los datos de texto de pre-perfil por su lenguaje y proporcionar en su salida al menos un identificador de lenguaje indicativo de un lenguaje respectivo usado en los datos de texto de pre-perfil;
- en el que la base de datos de clasificación (116) comprende las palabras clave en diferentes lenguajes; y en el que la unidad de generación de consultas (120) está configurada para generar las consultas usando las palabras clave en el lenguaje correspondiente al identificador de lenguaje o, si no hay palabras clave en este lenguaje que estén presentes en la base de datos de clasificación de la clase respectiva, en un lenguaje predeterminado.

5. El motor de recomendaciones de una de las reivindicaciones anteriores, en el que la unidad de generación de consultas (120) está configurada para incluir los datos de identificación extraídos de al menos dos entidades de interés en las consultas.

65

6. El motor de recomendaciones de una de las reivindicaciones anteriores, en el que la unidad de generación de perfiles (108) está configurada para detectar en los datos de pre-perfil un enlace de recursos a un recurso de contenido en una localización de red de una red de datos de área amplia, para acceder al recurso de contenido y añadir los datos de texto disponibles del recurso de contenido a los datos de texto de pre-perfil.
7. El motor de recomendaciones de una de las reivindicaciones anteriores,
- en el que la unidad de generación de perfiles (108) está configurada para extraer de los datos de pre-perfil diferentes subconjuntos de los datos de texto de pre-perfil, que se distinguen entre sí por su fecha de generación asignada por la base de datos externa, y para extraer de los subconjuntos las fechas respectivas de generación de los subconjuntos, y
 - en el que la unidad de clasificación (130) está configurada para aplicar una ponderación al indicador de relevancia, aumentando más la ponderación del indicador de relevancia cuanto más reciente es la fecha de generación de uno respectivo de los subconjuntos relacionados con una localización de contenido determinada perteneciente a una entidad de interés determinada.
8. El motor de recomendaciones de una de las reivindicaciones anteriores, en el que la unidad de análisis de pre-perfiles (112) está configurada para detectar a partir de los datos de texto de pre-perfil la presencia de palabras clave que indican que al usuario determinado o alguien relacionado con el usuario determinado le gusta una entidad y para asignar un indicador de "me gusta" a la entidad respectiva en el conjunto de datos de perfil de usuario inicial, y en el que la unidad de clasificación (130) está configurada para aplicar una ponderación al indicador de relevancia, aumentando la ponderación del indicador de relevancia si una entidad de interés determinada tiene un indicador de "me gusta" asociado.
9. El motor de recomendaciones de una de las reivindicaciones anteriores, en el que la unidad de clasificación (130) o la unidad de entrelazado (132) está configurada para evaluar un valor de relación textual entre los datos de texto de pre-perfil y los datos de texto de contenido.
10. El motor de recomendaciones de una de las reivindicaciones 2 a 9, en el que
- la unidad de entrelazado (132) está configurada para ponderar los indicadores de relevancia basándose en un criterio de similitud que evalúa una relación textual entre los datos de texto de contenido de diferentes artículos de contenido recuperados en las consultas, en el que la ponderación disminuye el indicador de relevancia si un artículo de contenido a evaluar para la relevancia tiene una relación textual con un artículo de contenido evaluado inmediatamente antes, relación textual que supera un valor predeterminado.
11. El motor de recomendaciones de una de las reivindicaciones anteriores,
- que comprende además una unidad de autenticación (114), que está configurada para recibir a través de una interfaz de entrada de usuario unos datos de autenticación de usuario adecuados para acceder a la base de datos externa,
 - en el que la unidad de generación de perfiles (108) está configurada para acceder a la base de datos externa para recuperar los datos de pre-perfil.
12. El motor de recomendaciones de la reivindicación 9, que comprende además una unidad de mantenimiento de perfiles (134), que está configurada para extraer los datos de identificación adicionales para una entidad adicional de los datos de texto de contenido, y que está configurada para añadir los datos de identificación adicionales extraídos al conjunto de datos de perfil de usuario tras detectar un valor de relación textual entre los datos de texto de pre-perfil y los datos de texto de contenido que supera un valor umbral predeterminado.
13. Un depósito de contenidos (200), que comprende
- una base de datos de contenidos que comprende los artículos de contenido en la forma de archivos de datos almacenados en las localizaciones de contenido;
 - un motor de recomendaciones de acuerdo con una de las reivindicaciones anteriores,
 - en el que la unidad de recuperación de contenidos del motor de recomendaciones está configurada para emitir las consultas generadas a la base de datos de contenidos.
14. Un método para hacer funcionar un motor de recomendaciones para recomendar unos artículos de contenido a un usuario, que comprende
- recibir (S1) de una base de datos, que es externa al motor de recomendaciones, los datos de pre-perfil que comprenden los datos de texto de pre-perfil adecuados para identificar las entidades de interés para un usuario determinado;
 - extraer (S2) de los datos de pre-perfil los datos de identificación que identifican las entidades de interés;
 - generar (S3) un conjunto de datos de perfil de usuario inicial para el usuario determinado a partir de los datos

de identificación extraídos;

- generar (S4), usando los datos de identificación extraídos a partir del conjunto de datos de perfil de usuario inicial, al menos dos consultas semánticamente diferentes una de otra para dirigirse a al menos un depósito de contenidos;

5 - emitir (S5) las consultas generadas al al menos un depósito de contenidos;

- recibir (S6) del al menos un depósito de contenidos, en respuesta a la consulta, los datos de respuesta relacionados con el contenido que comprenden las listas de aciertos respectivas que tienen al menos un identificador de localización de contenido respectivo indicativo de una localización de almacenamiento de un artículo de contenido respectivo; y

10 - generar (S7) a partir de las diferentes listas de aciertos recuperadas una única lista de recomendaciones entrelazando entre sí los identificadores de localización de contenido comprendidos en las diferentes listas de aciertos.

15. Un medio de almacenamiento legible por ordenador que almacena un código de programa ejecutable, implementando el código de programa el método de la reivindicación 14.

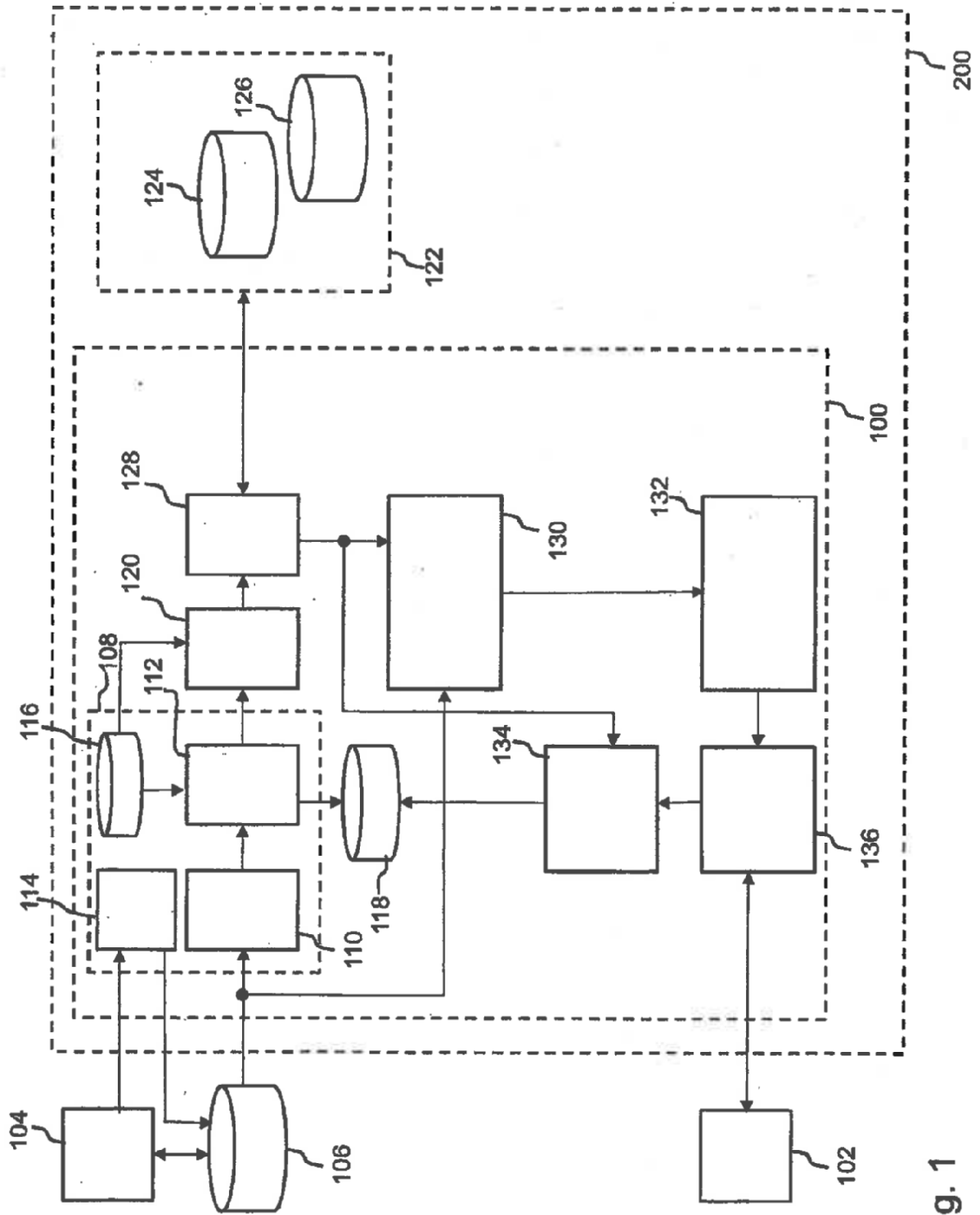


Fig. 1

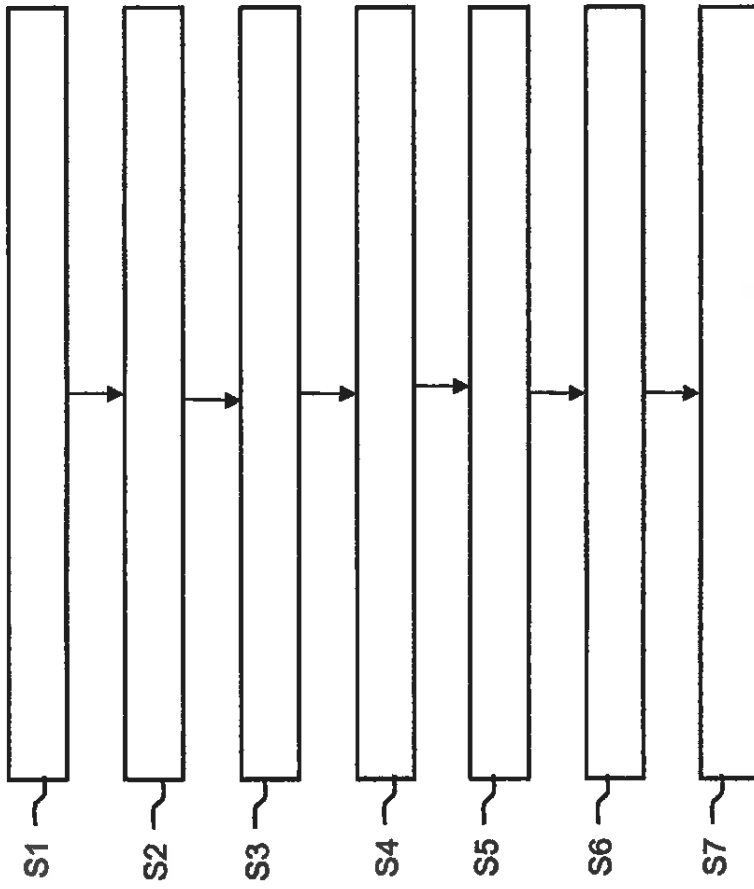


Fig. 2