

SISTEMA ECLÉCTICO DE FILTRADO DE INFORMACIÓN BASADO EN INTELIGENCIA COMPUTACIONAL PARA RECOMENDACIÓN DE ATRACTIVOS TURÍSTICOS DEL CARIBE COLOMBIANO

ECLECTIC SYSTEM OF INFORMATION FILTERING BASED ON COMPUTING INTELLIGENCE FOR RECOMMENDATION OF COLOMBIAN CARIBBEAN TOURIST ATTRACTIONS

Luis C. Tovar Garrido¹
Josías D. Montoya Campo²
Raúl J. Martelo Gómez³

Palabras clave: Sistema ecléctico, Sistema de recomendación, Inteligencia computacional, Atractivo turístico, Turismo.

RESUMEN

Los sistemas de recomendación pueden ayudar a las personas a encontrar los ítems que mejor se adapten a sus gustos y necesidades cuando son abrumados con una sobrecarga de información. Estos sistemas se han utilizado en diferentes dominios como la música y las películas, pero también han probado ser útil en el ámbito de atracciones turísticas. En este artículo, se presenta un sistema ecléctico de filtrado de información para atractivos turísticos del caribe colombiano. La solución propuesta sugiere atracciones turísticas de intereses a un usuario con base en su perfil de ratings, preferencias y gustos personales. El sistema adopta un enfoque ecléctico, combinando tres diferentes técnicas de recomendación con el propósito de encontrar el algoritmo más óptimo para este complejo dominio. Todo esto fue orientado a una aplicación móvil y evaluado por un grupo de usuarios de prueba. Los resultados demostraron la utilidad de los sistemas de recomendación para el dominio del turismo permitiendo generar visitas turísticas adaptadas al perfil del usuario, simplificando la tarea de planificar un viaje.

ABSTRACT

Recommendation systems can help people find items that best suit their tastes and needs when overwhelmed with information overload. These systems have been used in different domains such as music and movies, but have also proven to be useful in the field of tourist attractions. This article presents an eclectic information filtering system for tourist attractions in the caribbean region of Colombia. The proposed solution suggests tourist attractions of interest to a user based on their rating profile, preferences and personal tastes. The system adopts a eclectic approach, combining three different recommenders in order to find the most optimal algorithm for this complex domain. All this was oriented to a mobile application and evaluated by a group of test users. The results showed the usefulness of recommendation systems for the tourism domain allowing to generate tourist visits adapted to the profile of the user, simplifying the task of planning a trip.

Keywords: Eclectic system, Recommender system, Computational intelligence, Tourist attraction, Tourism.

¹ Universidad de Cartagena, ltovarg@unicartagena.edu.co

² Universidad de Cartagena, jmontoyac1@unicartagena.edu.co

³ Universidad de Cartagena, rmartelog1@unicartagena.edu.co

INTRODUCCIÓN

Internet ha provocado un importante cambio en la realización de muchas tareas y en la forma de trabajar de la sociedad. Actualmente, antes de realizar un viaje, los turistas utilizan recursos web de todo tipo para consultar servicios y actividades, que podrán realizar al llegar al destino. El volumen de información que se puede encontrar de un determinado destino turístico es habitualmente demasiado grande, por lo que se debe dedicar gran cantidad de tiempo a buscar y encontrar lo que se necesita y realmente interesa.

Los destinos deben tener una visión amplia y dinámica, proporcionando al turista herramientas que ofrezcan información personalizada, tratando de integrar los diferentes productos turísticos existentes (Guevara, 2008). A pesar de que las herramientas clásicas de información como las guías, audio-guías o páginas web pueden cubrir las necesidades básicas, el turista del siglo XXI demanda nuevas herramientas tecnológicas para obtener información con las características citadas (Guevara, et al., 2009). Es aquí donde los sistemas de recomendaciones y técnicas de inteligencia artificial (Borras, Valls, 2014), pueden ayudar a solucionar el problema de sobrecarga de información, y proveer a los usuarios recomendaciones valiosas de atracciones turísticas orientadas a sus preferencias personales, requerimientos y restricciones.

También se ha visto el uso de Smartphone para el manejo de las recomendaciones, con el desarrollo de sistemas de recomendación que usan un sistema de valoraciones el cual permite registrar los ratings de cada usuario a través de una aplicación móvil (Kenteris, et al., 2010). En su mayoría, estos sistemas de recomendación permiten determinar que sitios son mejores para visitar, basándose en las preferencias del turista y en los ratings de otros turistas (Smirnov, et al., 2013).

A lo largo de este artículo, se describirá el desarrollo de un sistema de filtrado de información de atractivos turísticos del caribe colombiano, utilizando técnicas híbridas de inteligencia computacional para el filtrado de información.

Esto con el fin de personalizar y mejorar la experiencia de turista, al sugerir los lugares y atractivos turísticos más adecuados al perfil del usuario. El documento se centrará en el diseño y desarrollo de un sistema de recomendación, las técnicas de recomendación usadas y los métodos de hibridación de dichas técnicas.

También se detallará el diseño y modelado de perfiles ontológicos que representen el concepto de atractivos turísticos, los cuales serán usados por el sistema de recomendación para optimizar el acceso y manejo de la información.

Los Sistema eclécticos tienen la característica de reunir ideas para agrupar los valores y tendencias en sistemas diversos. Es ecléctico, cuando combina varios métodos, pensamientos y esquemas con el mismo fin, para buscar la mejor manera de resolver los problemas.

El sistema de filtrado de información a desarrollar es ecléctico debido a que usa los algoritmos de recomendación: filtrado colaborativo (Collaborative Filtering), filtrado basado en contenido (Content-Based Filtering) y filtrado basado en conocimiento (Knowledge-Based Filtering); cuyos procesos de filtrados difieren entre sí en cuanto al algoritmo que usan y los parámetros que reciben, las cuales serán acopladas para producir las recomendaciones que se le proporcionaron al usuario con base en sus gustos y preferencias.

FUNDAMENTO TEÓRICO

Ontologías

El término ontología proviene de la filosofía y puede definirse como la rama correspondiente a la metafísica que se encarga del estudio de la naturaleza de la existencia. Este término se ha adoptado en la computación y se define como una especificación explícita y formal de una conceptualización compartida (Gruber, 1993).

Para explicar esta definición, se describirá por un lado el sentido de la conceptualización, el cual se refiere al modelo abstracto del dominio

de aplicación específico que se representa en la ontología, mientras que el término explícita se refiere a la necesidad de especificar los diferentes conceptos. Por otro lado, el aspecto formal se refiere a que debe indicarse la especificación mediante un lenguaje formalizado y la característica compartida alude al hecho de que una ontología debe representar el conocimiento aceptado por los usuarios que van a utilizarlo.

Las ontologías también permiten definir reglas para modelar el conocimiento que puede no ser representado por las clases, propiedades e instancias. Estas reglas tienen que ser fácilmente interpretadas por las personas y deben ser utilizadas de forma eficiente en cálculos computacionales. Una ontología no es un compendio del conocimiento objetivo, sino que depende del punto de vista de un grupo.

Sistemas de filtrado de información

Filtrado de información es un nombre usado para describir una variedad de procesos que implican la entrega de información a las personas que lo necesitan. Generalmente se trata de grandes cantidades de datos (Belkin y Croft, 1992). Las aplicaciones típicas se ocuparían de procesar gigabytes de texto, o cantidades muchos mayores de otros medios de comunicación. El filtrado se basa en descripciones de preferencias de información individuales o grupales, a menudo llamados perfiles. Dichos perfiles generalmente representan intereses a largo plazo.

Un sistema de filtrado de información es un sistema de información diseñado para datos no estructurados o semiestructurados, razón por la cual es importante el uso de ontologías en el desarrollo de la presente investigación.

En la actualidad los sistemas de filtrado de información han evolucionado y es posible encontrarlos en diversos ámbitos de aplicación como en el comercio electrónico, donde se han convertido en una herramienta fundamental para los proveedores en línea (Schafer, Konstan y Riedl, 2001) o en los servicios de información científica. Una categoría de sistemas de filtrado de información son los sistemas de recomendación.

Sistemas de recomendación

Los sistemas de recomendación surgen como respuesta a la sobrecarga de información presente en numerosos dominios, que dificulta a los usuarios identificar los productos (artículos comerciales, contenidos de televisión, cursos educativos, etc.) que son relevantes para ellos. Como su nombre indica, este tipo de sistemas se encargan de recomendar o sugerir a los usuarios que lo utilizan, productos relacionados a sus preferencias, de entre la gran cantidad de opciones disponibles (Burke, 1999).

Las diferentes estrategias de personalización que emplean estos sistemas se basan en la información que recopilan en perfiles personales (Vicente, 2012). Por lo tanto, es necesario recopilar información personal del usuario y crear un perfil de cada uno de ellos. La obtención de ésta información y la creación de ese perfil, es una parte muy importante de los sistemas de recomendación, ya que es la que determina la calidad de la recomendación.

Aunque es posible encontrar diferentes clasificaciones de sistemas de recomendación, se pueden distinguir en la mayoría de la literatura los siguientes tipos, en función de las técnicas que utilizan:

Basado en contenido: Genera las recomendaciones basándose en las preferencias que indica el usuario y analizando los ítems detalladamente.

Colaborativo: Localiza similitudes en los gustos con otros usuarios, mediante las valoraciones similares, y recomienda ítems que hayan gustado a éstos usuarios afines.

Demográfico: Emite las recomendaciones clasificando al usuario dentro de su grupo demográfico (misma edad, sexo, salario, etc.). Realiza recomendaciones acordes al perfil del usuario, relacionándolo con algunos de los estereotipos existentes. La organización de los estereotipos suele ser una estructura jerárquica, de manera que se permitan realizar razonamientos de inferencia (Pazzani, 1999), su principal

inconveniente es la dificultad de adquirir estos datos demográficos por parte de los usuarios y el recelo de estos al sentir que se invade su intimidad (Pérez, 2008).

Basado en conocimiento: El sistema localizará un problema anterior similar al dado por el usuario en este momento y lo resolverá de la misma forma.

Híbridos: Son aquellos que combinan dos o más de los sistemas anteriores. El uso de técnicas de hibridación permite mejorar los resultados independientemente del método utilizado para realizar la hibridación (Burke, 2002). Este es el tipo de sistema en el que se basa nuestra propuesta. Profundizando un poco más, se pueden encontrar diferentes métodos de hibridación (Álvarez, 2011): De peso, conmutación, mezclas, combinación de características, cascada, aumento de características y meta nivel.

Grupales: Genera las recomendaciones teniendo en cuenta las preferencias de todo el grupo que viaja junto, proponen adaptar los resultados del sistema a los gustos e intereses de un grupo entero (Leiva, Guevara y Rossi, 2012).

METODOLOGÍA

El proyecto está enmarcado en un tipo de investigación aplicada, pues consiste en la implementación de métodos de minería de datos e inteligencia computacional, con el objetivo de desarrollar un sistema de recomendación de atractivos turísticos colombianos. Posee una parte descriptiva ya que se registró detalladamente todo el sistema de recomendación, que usa técnicas eclécticas de inteligencia computacional para el filtrado de información, y ontologías para representar los atractivos turísticos utilizados.

Recolección de datos

Se realizaron búsquedas en bases de datos y servicios web de turismo en Colombia, que proporcionan información acerca de atractivos turísticos en la región caribe, una de las principales ventajas de estos servicios, es que todas las entradas tienen un valor específico con respecto al

turismo. En total se registraron 400 sitios conocidos por el ministerio de industria y turismo así como por las páginas web dedicadas a promocionar este sector de la economía colombiana.

También se recolectaron datos adicionales los cuales se usaron para inicializar la matriz de recomendaciones y poder entrenar el sistema de recomendación. Se necesitó encuestar usuarios reales que valoraran y asignaran ratings a diferentes sitios turísticos. Las encuestas realizadas estaban conformadas por el listado de tipos de atractivos turísticos a considerar. Se le pidió a cada persona encuestada que le asignara un puntaje de valoración entre 1 y 5 a cada tipo de sitio turístico. Para abarcar una muestra demográfica más diversa, se viajó a las ciudades de Cartagena, Barranquilla, Santa Marta y Sincelejo, y se encuestaron a personas en la calle de forma aleatoria. En total se encuestaron a 150 personas entre todas las ciudades.

Procedimiento

Teniendo en cuenta que el sistema a implementar conlleva un proceso de minería de datos, se utilizó la metodología CRISP-DM Cross Industry Standard Process for Data Mining (Shearer, 2000), la cual trata de un modelo de proceso de minería de datos que describe los enfoques comunes que utilizan los expertos en minería de datos. CRISP-DM divide el proceso de minería de datos en seis fases principales a saber:

1. **Comprensión del negocio:** En esta fase se hizo un análisis de las características que debería tener el sistema de filtrado de información cuando se aplica al contexto de recomendar atractivos turísticos, con el objetivo de identificar que atractivos turísticos considerar en el sistema. Una de las intenciones del sistema es fomentar la pasión del usuario por viajar, conocer lugares nuevos y fomentar el turismo, por ésta razón, se tuvieron en cuenta tipos de atractivos turísticos que exigieran de una actividad a realizar o tuvieran un valor específico que los hiciera únicos, lo cual fomentará en el usuario el deseo de hacer un viaje para conocerlos. No se tuvieron en cuenta discotecas, tampoco restaurantes o bares, a menos que estos tuvieran algún valor cultural adicional.

2. Comprensión de datos: Se realizó la búsqueda de los atractivos turísticos en la región caribe utilizando las técnicas de recolección de información descritas anteriormente. Estos se registraron en un archivo en el cual se les ordeno por nombre, departamento, municipio y tipo de atractivo turístico. Luego se estudiaron diferentes técnicas de recomendación, aplicándolas al contexto de atractivos turísticos, teniendo en cuenta la cantidad de elementos que se utilizaron en el sistema de filtrado de información, así como la información recolectada para cada atractivo. Primero se realizó una matriz de recomendaciones, en la que cada celda corresponde al rating que un usuario le dio a un atractivo turístico. Esta matriz se usó como base para implementar el algoritmo de Filtrado Colaborativo (Collaborative Filtering). También se usó Filtrado Basado en Contenido (Content Based Filtering), donde el perfil de cada atractivo turístico está conformado por los tipos de atractivo turístico con los que este asociado. Se usó Filtrado Basado en Conocimiento (Knowledge Based Filtering) para tomar ventaja del dominio en el que se intenta recomendar. Para representar este conocimiento, se usó un árbol de taxonomía en el cual se clasificaron todos los tipos de atractivos turísticos, sus relaciones y sus dependencias. Por último, se usó un algoritmo de hibridación para combinar los puntajes de los tres recomendadores anteriores, usando una Suma Ponderada (Weighted Sum).

3. Preparación de datos: A fin de construir el conjunto de datos en el formato que serán usados en el sistema, se hizo el diseño del perfil ontológico que representa el concepto de 'atractivo turístico', y luego se modelaron todos los atractivos turísticos recolectados usando el perfil ontológico definido. Este perfil ontológico está conformado por nombre, departamento, municipio y tipo de atractivo turístico. Los tipos de atractivos turísticos se modelaron usando un árbol de taxonomía en el cual también se describen las relaciones y dependencias entre todos los tipos de actividades turísticas. También se hizo el diseño de la arquitectura del sistema de recomendación. Específicamente, se diseñaron las bases de datos y los formatos en que se guarda en memoria toda la información usada, así como la interacción del aplicativo móvil con el sistema y como el usuario

accederá a sus recomendaciones.

4. Modelado: Durante esta etapa se desarrolló el método de búsquedas de ontologías que usa el sistema de recomendación, luego se desarrollaron las técnicas de recomendación elegidas anteriormente y, por último, se realizó el proceso de acoplamiento de estas técnicas teniendo en cuenta el método de ecléctico propuesto. Se desarrollaron las diferentes técnicas de recomendación descritas anteriormente, usando el lenguaje Java. Cada algoritmo de recomendación se desarrolló de forma que funcionaran independientemente, donde cada uno genera una lista de recomendaciones y estima ratings. Finalmente, se realizó el proceso de acoplamiento de estas técnicas, usando una suma ponderada como lo hace el método de hibridación para combinar los ratings estimados de los algoritmos anteriores. De esta forma se finalizó el desarrollo del motor de recomendaciones del sistema.

5. Evaluación: En esta fase, se realizaron pruebas para evaluar la eficacia de las recomendaciones y calibración de las técnicas de recomendación. Las pruebas se hicieron en conjunto con el desarrollo de cada técnica para así asegurar su correcto funcionamiento. Después de acoplar todas las técnicas, se hizo una calibración del recomendador ecléctico para verificar la eficacia de las recomendaciones. Se usó Normalized Discounted Cumulative Gain (NDCG) como medida de evaluación (Wang et al., 2013), la cual es muy popular a la hora de evaluar recomendadores.

6. Despliegue: Se desarrolló el aplicativo móvil y se hizo la integración de este con el sistema de recomendación. Por último, se realizaron pruebas funcionales de integración y de aceptación para asegurar la calidad del aplicativo. Las pruebas de integración están asociadas al acoplamiento del aplicativo móvil con el sistema de recomendación. Las pruebas de aceptación se hicieron al aplicativo una vez terminado, estando este en su fase final de desarrollo. Para ello, se abordaron varias personas a las cuales se les solicitó que usaran la aplicación para luego valorar y dar sus opiniones acerca de las recomendaciones recibidas y de la aplicación

en general. Por medio del aplicativo, los usuarios acceden a sus recomendaciones al tiempo que el sistema se alimentará de la información provista por los usuarios para así mejorar, con el pasar del tiempo, la eficacia de sus recomendaciones.

RESULTADOS

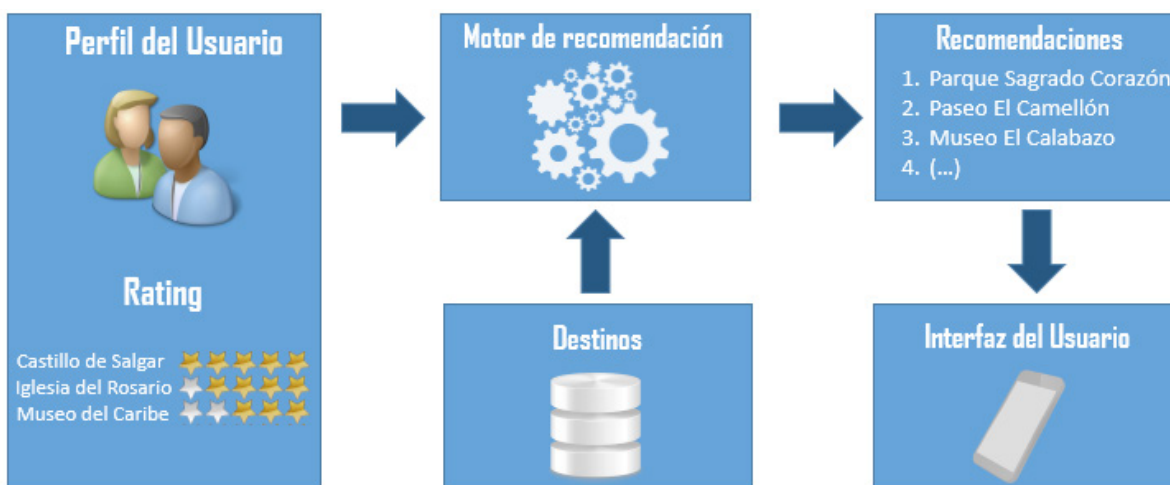
De conformidad con el cumplimiento del objetivo el resultado principal es el desarrollo del sistema ecléctico de recomendación de atractivos turísticos cuya estructura se representa en la figura 1.

Para presentar las recomendaciones a un

usuario, la aplicación primero necesita aprender más acerca del usuario al recolectar información y resumiéndola en un perfil de usuario. Cada usuario puede especificar sus previas experiencias de viaje al darle ratings a las atracciones turísticas que haya visitado en el pasado. Esta información le permite al sistema de filtrado de información construir el perfil del usuario.

El sistema recibe ratings y atracciones turísticas de la base de datos. Los usuarios interactúan con el sistema a través de la aplicación móvil. Las recomendaciones se entregan al usuario, quien puede dar feedback adicional en estas recomendaciones.

Figura 1. Estructura básica del sistema de recomendación



Fuente . Autores (2017)

Los siguientes pasos pueden ser identificados en el flujo de la información:

1. Se crea la consulta del usuario: Se carga la lista de atractivos turísticos junto con el perfil del usuario.
2. Predicción de ratings: Se usan diferentes algoritmos de recomendación que calculan las predicciones de los ratings para la lista de atractivos turísticos.
3. Se combinan los ratings: Las predicciones de los ratings, calculadas por lo diferentes algoritmos de recomendación, se combinan en un predictor de ratings híbrido.

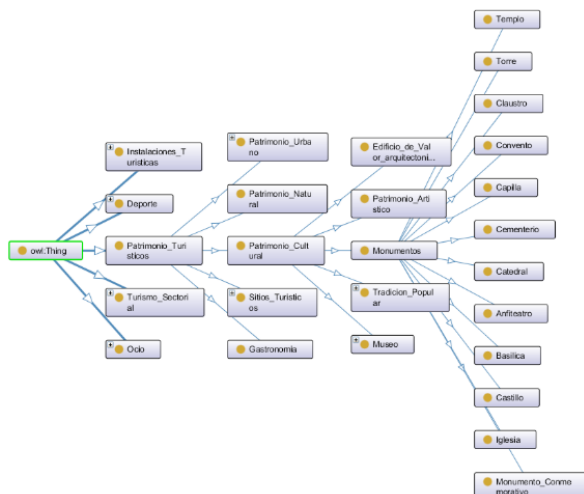
4. Entrega de recomendaciones: Los atractivos turísticos con el mayor puntaje son presentados al usuario como su lista final de recomendaciones.

Sin embargo, no hubiese sido posible el desarrollo del sistema ecléctico sin los siguientes productos: Perfil ontológico para atractivos turísticos, el motor de recomendación y el aplicativo móvil. Por consiguiente cada uno de estos componentes se describe a continuación.

Perfil ontológico para atractivos turísticos

Para crear el perfil ontológico de los atractivos turísticos, se tuvo en cuenta el nombre, departamento, municipio y tipo de actividad turística.

Figura 2. Porción del árbol de taxonomía



Fuente . Autores (2017)

Las entidades en una taxonomía se organizan en una estructura jerárquica conectadas a través de una relación "es-un", donde los niveles de clasificación más específicos corresponden a niveles inferiores. En el árbol de taxonomías, las entidades representan conceptos comunes de un dominio de turismo, como son instalaciones turísticas o patrimonios culturales. Los nodos hojas de la taxonomía son actividades recomendadas a los usuarios, es decir, el concepto más específico.

Tabla 1. Perfil ontológico de un atractivo turístico

NOMBRE	BOCAS DE CENIZAS
Departamento	Atlántico
Municipio	Barranquilla
Tipos	Playa, Patrimonio Natural

Fuente . Autores (2017)

Del árbol resultante, solo se consideraron los nodos hoja para clasificar los atractivos turísticos. Luego para catalogar cada atractivo turístico, se hizo un análisis de cada uno de ellos teniendo en cuenta la descripción del atractivo turístico y las

actividades que se pueden realizar en ese lugar o cerca de este. La clasificación de los atractivos turísticos y la creación del árbol de taxonomías constituyen componentes importantes en la implementación de los algoritmos de filtrado basado en contenido y basado en conocimiento.

Motor de recomendación

El recomendador ecléctico es la parte más crucial del sistema, a fin de encontrar las mejores recomendaciones, combina tres algoritmos de filtrado de información como lo son: filtrado colaborativo, filtrado basado en conocimiento y filtrado basado en contenido. La relación de cada uno de ellos se describe a continuación.

El algoritmo de filtrado colaborativo hace uso del conjunto de ratings que los usuarios les han dado a los atractivos turísticos. Los usuarios pueden asignarles ratings a los atractivos turísticos usando una puntuación entre 1 y 5, donde 5 corresponde al mayor grado de satisfacción.

El algoritmo de filtrado colaborativo realiza dos pasos para predecir los ratings de un atractivo turístico para un usuario. Primero selecciona el conjunto de atractivos turísticos más similares a este ítem. Estos vecinos son seleccionados usando una función de similitud con base en los ratings que definen el perfil de los atractivos turísticos. Luego se calcula la predicción del rating para cada ítem calculado con base en los ítems más similares.

Este enfoque es muy conocido en "Machine Learning" y "Minería de datos" con la abreviación de kNN (k-Nearest-Neighbor) k Vecinos más Cercanos, el cual describe un algoritmo que selecciona los k ítems mas similares.

Como fue mencionado, para hacer esto se necesitó una función de similitud basada en los ratings de los atractivos turísticos. Debido a que el vector de ratings de cada ítem está conformado por valores dentro de un rango (1 a 5), se optó por usar la función de similitud de Pearson. Sea r_x el vector de ratings del ítem X y S_{xy} usuarios que calificaron ambos ítems X y Y.

(1)

$$sim(x, y) = \frac{\sum_{s \in S_{xy}} (r_{xs} - \bar{r}_x)(r_{ys} - \bar{r}_y)}{\sqrt{\sum_{s \in S_{xy}} (r_{xs} - \bar{r}_x)^2} \sqrt{\sum_{s \in S_{xy}} (r_{ys} - \bar{r}_y)^2}}$$

Habiendo calculado la medida de similitud, los k vecinos más cercanos pueden ser seleccionados. Para un k específico se denota el vecindario del atractivo turístico i para el usuario x como $N(i, x)$. Note que este vecindario consiste de los ítems más similares j que el usuario x ha valorado. Por lo tanto, este vecindario es diferente para cada usuario. Para la función de puntuación (scoring function) se usa una suma ponderada (weighted sum). Sea r_{xi} la similitud de los ítems i y j , el rating del usuario x al ítem j y r_{xj} ítems similares a i que también fueron valorados por el usuario x .

(2)

$$r_{xi} = \frac{\sum_{j \in N(i, x)} s_{ij} r_{xj}}{\sum_{j \in N(i, x)} s_{ij}}$$

Finalmente, si i tiene menos de 5 ítems, se considera que el algoritmo de filtrado colaborativo no es suficientemente preciso. En estos casos, se recurre a otro algoritmo que lo remplace, como el de filtrado basado en contenido discutido a continuación.

Los algoritmos basados en contenido miran las características que describen un ítem particular y tratan de encontrar similitudes entre estas características y el perfil del usuario. Para el algoritmo aquí propuesto, se usaron las etiquetas definidas en el perfil ontológico las cuales especifican el tipo de atractivo turístico. El perfil ontológico, definido para cada atractivo turístico, se utilizó como el perfil de cada ítem en el algoritmo. Finalmente, el perfil de cada uno de estos ítems es comparado con el perfil del usuario usando la función de similitud de coseno. Este valor es transformado al rango [1-5] y usado como la estimación del rating de ese ítem. Sea r_x el vector de ratings del ítem X .

(3)

$$sim(x, y) = \frac{\sum_i r_{xi} r_{yi}}{\sqrt{\sum_i r_{xi}^2} \sqrt{\sum_i r_{yi}^2}}$$

Por otro lado, el filtrado basado en conocimiento difiere del algoritmo basado en contenido en la forma que procesa información que es única al dominio de atractivos turísticos, mientras que los algoritmos anteriores podrían ser fácilmente usados en otros dominios para generar recomendaciones.

Las ontologías pueden ser útiles para representar el conocimiento en cualquier campo (Grossman, Domingos, 2004). En un sistema de recomendación, el dominio ontológico permite la clasificación de ítems a ser recomendados. En el sistema aquí propuesto se consideró que cada atractivo turístico es una instancia de uno (o varios) de las clases de nivel más bajo del árbol de taxonomías, en las cuales el usuario puede estar interesado.

La estructura de la ontología puede ser usada para transferir información a través de los nodos usando una propagación de los valores de los nodos hoja hacia arriba en el árbol, y luego distribuyéndolos a todos los nodos del árbol (ver figura 2). Para inicializar los nodos hoja se usa el historial de ratings del usuario y se crea un perfil que contenga el valor de preferencia estimado para cada tipo de actividad turística. Estos valores se usan para inicializar los nodos hoja, y aquellos tipos a los que el usuario no haya proporcionado ningún tipo de información, se ignoran y sus valores serán calculados luego.

Para distribuir los puntajes a través de todo el árbol, primero se estiman las preferencias del usuario para los conceptos más específicos. Por cada conjunto de nodos "hermanos" que tengan un puntaje asociado, se calcula el promedio de sus ratings. Debido a que realmente no se conoce la preferencia del usuario para los conceptos más generales o hermanos, se desea conservar esta incertidumbre al propagar los valores. Para esto, se usa un factor de propagación el cual determina

que tanta información se pierde al propagar una preferencia desde los nodos más bajos. Los valores propagados se calculan de la siguiente forma. Sea el conjunto de nodos hijo del nodo p , el puntaje del nodo x y el factor de propagación, donde $W_p < 1$.

(4)

$$r_{padre} = W_p * \frac{\sum N_p r_{hijo}}{\sum N_p}$$

Para redistribuir los puntajes en todo el árbol se deben propagar desde los conceptos más generales y así llegar a todos los nodos que en el paso anterior no se les asignó ningún valor. Para definir la función de propagación hacia abajo, se consideraron dos factores. En primera instancia se desea que, al propagar los valores, aquellos que ya tenían un valor asignado, conserven este valor sin alterar. Para ello se usa el mismo factor de propagación que en el paso anterior para recuperar esa información. Pero también se tiene en cuenta la incertidumbre de conocer estos valores de forma explícita y darles menor preferencia al considerarlos para producir recomendaciones. Por

esta razón se mantiene un factor de incertidumbre que permite reflejar este comportamiento. Sea el puntaje del nodo x , el factor de propagación, donde y el factor de incertidumbre, donde $w_i < 1$.

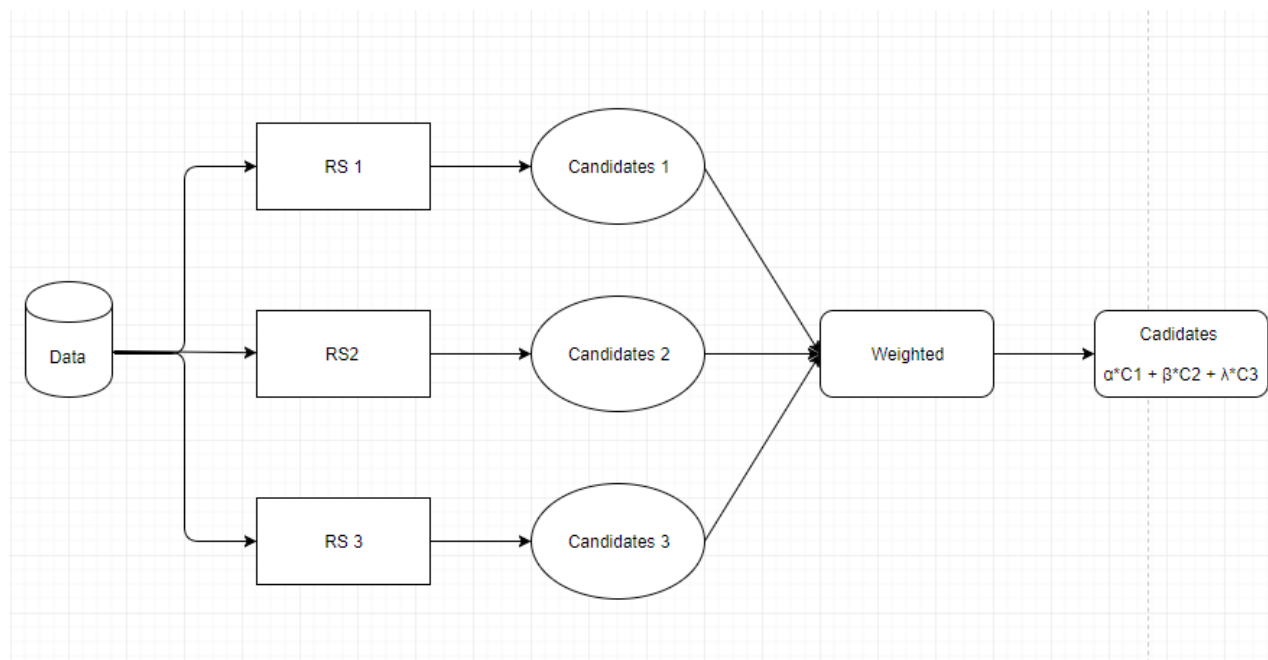
(5)

$$r_{hijo} = W_i * \frac{r_{padre}}{w_p}$$

Finalmente, para generar la lista de recomendaciones, se construye el perfil del usuario conformado por todos los nodos hoja del árbol, y el perfil de cada uno de los atractivos turísticos es comparado con el perfil del usuario usando la función de similaridad de coseno.

Mientras los tres algoritmos de filtrado de información anteriores pueden predecir los ratings de atracciones turísticas y hacer recomendaciones por su propia cuenta, combinar estos resultados permite utilizar todas las fuentes de información diferentes a fin de obtener un mejor recomendador, eso se logra mediante una suma ponderada de los tres ratings estimados.

Figura 3. Esquema general del recomendador ecléctico.



Fuente . Autores (2017)

Los diferentes índices son cf para filtrado colaborativo, cbf para filtrado basado en contenido y kbf para filtrado basado en conocimiento.

(6)

$$\hat{r}_{hibrido} = (w_{cf} * \hat{r}_{cf}) + (w_{cbf} * \hat{r}_{cbf}) + (w_{kbf} * \hat{r}_{kbf})$$

(7)

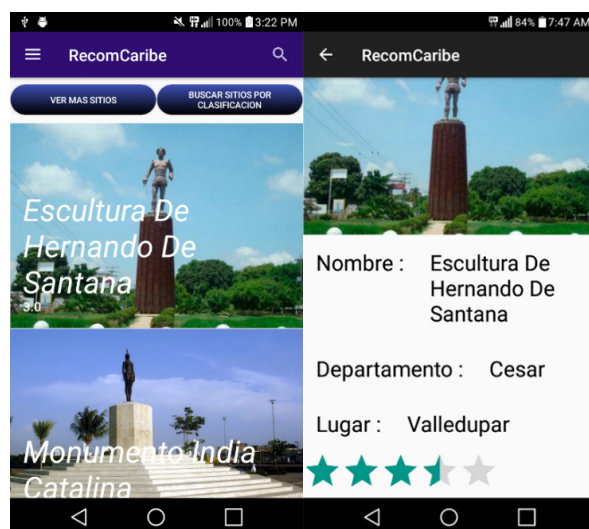
$$w_{cf} + w_{cbf} + w_{kbf} = 1$$

Estos pesos (w) no son estáticos. Cuando no se logra encontrar suficientes vecinos para filtrado colaborativo, se coloca como cero. En este caso, y se distribuyen el peso para satisfacer la segunda ecuación.

Aplicativo móvil

El sistema de recomendación está disponible para los usuarios a través la aplicación móvil RecomCaribe. La aplicación cuenta con un sistema de registro, permite la búsqueda de sitios, calificación de los mismos y consulta de atractivos turísticos, entre otras funcionalidades.

Figura 4. Captura de pantalla de la interfaz principal



Fuente . Autores (2017)

Esta aplicación se desarrolló con Android Studio, que es un entorno de desarrollo integrado para el sistema operativo Android lanzado por Google. Se utilizó el patrón de arquitectura modelo vista controlador (MVC) en tanto las herramientas ofrecidas son las que mejor se adaptan para trabajar en conjunto con el sistema operativo de Google.

Para la calificación de los sitios, la aplicación cuenta con múltiples interfaces que le permiten al usuario puntuar fácilmente. La interfaz (figura 4) cuenta con una barra de texto donde el usuario puede buscar sitios turísticos por su nombre.

Dos botones que permiten buscar por categoría y acceder a sitios que el programa recomienda según los intereses de cada usuario.

CONCLUSIONES

Este trabajo se presenta como una alternativa de solución al problema de recomendación en el dominio del turismo, con el desarrollo de un sistema ecléctico de filtrado de información para recomendar atractivos turísticos del caribe colombiano, que se ajusten a los intereses del turista al mostrarle los diferentes lugares turísticos

y atracciones del sitio, y al mismo tiempo sugerir los más adecuados para el gusto del usuario, satisfaciendo así sus preferencias.

Además, se pudo comprobar que el sistema

desarrollado resultó más eficaz que usar los algoritmos de filtrado de forma independiente, lo corroboran los resultados de la métrica NDCG (Normalized Discounted Cumulative Gain).

Tabla 2. Evaluación de la eficacia de filtrado.

PARTICIÓN	MÉTRICA NDCG			
	KB	CBF	CF	ECLÉCTICO
0	0,799860277	0,797931694	0,81125408	0,828348284
1	0,798120459	0,793074252	0,810034675	0,811086631
2	0,800124278	0,79861241	0,808325665	0,81805667
3	0,791583194	0,792187162	0,79923241	0,820050831
4	0,79179714	0,790618217	0,80990731	0,81018134

Fuente . Autores (2017)

Para realizar las pruebas, se dividió la matriz de recomendaciones en 5 particiones de igual tamaño. Luego se escogieron 4 de estas particiones para ser el conjunto de entrenamiento (Training Set) y la partición restante para ser el conjunto de pruebas (Test Set). Este proceso se repitió para cada partición y para cada algoritmo de recomendación. Los resultados se pueden observar en la tabla 2, estos puntajes de NDCG son bastante positivos teniendo en cuenta que el máximo puntaje posible es 1, por lo que se observa que el enfoque ecléctico obtiene mejores puntajes que los otros recomendadores.

Uno de los propósitos al desarrollar este proyecto es que sea utilizado por todas aquellas personas que deseen visitar la región caribe colombiana, siendo que quieran conocerla en profundidad o simplemente quieran realizar actividades turísticas. Así, el principal beneficio que aporta este proyecto es la posibilidad de generar visitas turísticas adaptadas al perfil del usuario, simplificando la tarea de planificar un viaje o un día de turismo.

REFERENCIAS BIBLIOGRÁFICAS

Álvarez, C. A. G. (2011). Sistema De Recomendación Y Planificación Turística De La Ciudad De Valencia Vía Web.

Belkin, N. J., & Croft, W. B. (1992). Information filtering and information retrieval: Two sides of the same coin?. Communications of the ACM, 35(12), 29-38.

Borras, J., Moreno, A., & Valls, A. (2014). Intelligent tourism recommender systems: A survey. Expert Systems with Applications, 41(16), 7370-7389.

Burke, R. (1999, July). Integrating knowledge-based and collaborative-filtering recommender systems. In Proceedings of the Workshop on AI and Electronic Commerce (pp. 69-72).

Guevara, A. (2008). Sistemas de información de destinos turísticos integrados (SIDTI). I Jornada de investigación en turismo. Universidad de Sevilla.

Guevara, A., Aguayo, A., Gómez, L., Caro, J. L., Leiva, J. L., Hornos, M. J., & Fernández, M. (2009). Sistemas informáticos aplicados al turismo. Editorial Pirámide.

Grossman, D., & Domingos, P. (2004, July). Learning Bayesian network classifiers by maximizing conditional likelihood. In Proceedings of the twenty-first international conference on Machine learning (p. 46). ACM.

- Gruber, T. R. (1993). A translation approach to portable ontologies [Электронный ресурс] (No. 5, p. 2). TR Gruber//Knowledge Acquisition.
- Kenteris, M., Gavalas, D., & Mpitziopoulos, A. (2010, June). A mobile tourism recommender system. In Computers and Communications (ISCC), 2010 IEEE Symposium on (pp. 840-845). IEEE.
- Leiva Olivencia, J. L., Guevara Plaza, A., & Rossi Jiménez, C. (2012). Sistemas de recomendación para realidad aumentada en un sistema integral de gestión de destinos. Revista de análisis turístico, (14), 69-81.
- Pazzani, M. J. (1999). A framework for collaborative, content-based and demographic filtering. Artificial Intelligence Review, 13(5-6), 393-408.
- Pérez Cordon, L. G. P. (2008). Modelos de recomendación con falta de información. Aplicaciones al sector turístico (Doctoral dissertation, Universidad de Jaén).
- Schafer, J. B., Konstan, J. A., & Riedl, J. (2001). E-commerce recommendation applications. In Applications of Data Mining to Electronic Commerce (pp. 115-153). Springer US.
- Shearer, C. (2000). The CRISP-DM model: the new blueprint for data mining. Journal of data warehousing, 5(4), 13-22.
- Smirnov, A., Kashevnik, A., Ponomarev, A., Shilov, N., Schekotov, M., & Teslya, N. (2013, November). Recommendation system for tourist attraction information service. In Open Innovations Association (FRUCT), 2013 14th Conference of (pp. 148-155). IEEE.
- Vicente, M. (2012). Estrategias para el incremento de la fiabilidad Y confianza en sistemas de recomendación Colaborativa. Un enfoque semántico para la personalización De la publicidad en comercio electrónico.
- Wang, D., Park, S., & Fesenmaier, D. R. (2012). The role of smartphones in mediating the touristic experience. Journal of Travel Research, 51(4), 371-387.
- Wang, D., Xiang, Z., & Fesenmaier, D. R. (2016). Smartphone use in everyday life and travel. Journal of Travel Research, 55(1), 52-63.
- Wang, Y., Wang, L., Li, Y., He, D., Chen, W., & Liu, T. Y. (2013, May). A theoretical analysis of NDCG ranking measures. In Proceedings of the 26th Annual Conference on Learning Theory (COLT 2013).