

Recomendación de objetos de aprendizaje basado en el perfil del usuario y la información de atención contextualizada

Gladys Carrillo^a, Xavier Ochoa^b

^a Universidad Oberta de Catalunya
gcarrillob@uoc.edu

^b Facultad de Ingeniería en Electricidad y Computación, Escuela Superior Politécnica del Litoral, Km. 30.5 Vía Perimetral, Guayaqui, Ecuador
xavier@cti.espol.edu.ec

Resumen. La selección de los objetos de aprendizaje más relevantes para cada usuario es un tema de investigación y desarrollo activo en el ámbito de las herramientas de aprendizaje en línea. Si bien los algoritmos de recomendación no son nuevos, su adaptación y uso sobre objetos de aprendizaje es un campo aún abierto donde se ensayan diversos enfoques tanto de modelamiento de perfiles como de la recomendación como tal. En el presente trabajo se ha desarrollado un prototipo de sistema de recomendación de objetos de aprendizaje que ha sido utilizado para experimentar con 3 diferentes algoritmos de recomendación basados en el perfil del usuario con el fin de determinar y comparar la precisión de sus resultados.

Palabras Clave: sistemas de recomendación, objetos de aprendizaje, mahout

1 Introducción

La educación a distancia ha existido desde hace mucho tiempo buscando siempre eliminar la barrera de la distancia geográfica para acceder a algún tipo de formación. Los medios utilizados han variado considerablemente desde el uso de la radio y el correo tradicional, hasta entornos de aprendizaje en línea utilizando el Internet.

Es indudable que el uso de Internet para esta labor ha mejorado ostensiblemente la cantidad tanto de la oferta académica disponible como de los usuarios de la misma, así como la digitalización de los contenidos y soportes educativos ha ayudado a una reducción de los costos de distribución de los mismos. Sin embargo, un aspecto a mejorar es la satisfacción del usuario (estudiante) respecto a los materiales utilizados durante su proceso de aprendizaje.

Además de la variedad y profundidad de conocimientos previos, los estudiantes pueden tener diferentes estilos de aprendizaje, unos pueden preferir aprender mediante la lectura, otros mediante gráficos, audios o videos. También se debe tomar en cuenta aspectos como el idioma.

Con todos estos factores y dada la cantidad de materiales que podrían estar

disponibles para cada tema de estudio, no todos los materiales almacenados en una plataforma de aprendizaje podrían ajustarse a las características individuales de cada usuario, es necesario entonces evaluar y aplicar métodos que permitan que los usuarios tengan disponibles los materiales que mejor se adecuen a su perfil.

El presente trabajo busca evaluar cuáles de los algoritmos utilizados para recomendación producen la mejor precisión y exhaustividad al ser utilizado en ambientes reales. El artículo está estructurado de la siguiente manera. En la sección 2 se presenta la revisión de la literatura donde se describen los algoritmos de recomendación así como también avances sobre recomendación de objetos de aprendizaje. En la sección 3 se describe la metodología utilizada donde se presentan los componentes del sistema de recomendación y los experimentos realizados para evaluar la eficiencia de los métodos de recomendación propuestos. En la sección 4 se presenta el análisis de los resultados obtenidos y finalmente se presentan las conclusiones del trabajo realizado.

2 Estado del Arte

Los sistemas de recomendación, sin importar el tipo de datos o dominio de aplicación, de manera general se clasifican, dependiendo del método utilizado, en: basados en contenido (*content-based*), colaborativos e híbridos.

2.1 Sistemas de Recomendación Basados en Contenido (*content-based*)

Los sistemas basados en contenido buscan utilizar la información disponible tanto del ítem como del usuario para calcular la recomendación más adecuada.

Pazzani y Billsus, en su artículo “*Content-Based Recommendation Systems*” [10], tratan los sistemas que presentan la recomendación de un ítem basada en una descripción del mismo y un perfil de los intereses del usuario, los mismos que deben tratar y definir 3 aspectos importantes que son el modelamiento del ítem, el modelamiento de las preferencias del usuario y el método o algoritmo para realizar la recomendación.

El modelamiento, representación y almacenamiento de los ítems puede presentar complicaciones y diferentes mecanismos de implementación dependiendo de si se trata de datos estructurados, donde cada ítem es representado por un conjunto conocido y limitado de atributos, o si por el contrario se trata de información más libre como páginas web, artículos de noticias, críticas de expertos, etc, requiriendo un procedimiento adicional para normalizar dicha información en datos estructurados. En el dominio de los entornos educativos, los objetos de aprendizaje tienen la particularidad de que suelen estar definidos a través de metadatos estructurados, de manera que se simplifica o elimina este proceso de modelamiento.

El modelamiento de los usuarios por su parte presenta una ventana mayor de posibilidades puesto que tanto el método como el resultado dependerá en gran medida

del procedimiento que se utilizará para realizar la recomendación, variando desde perfiles simples definidos por el mismo usuario a través de formularios, pasando por la definición del perfil a partir de la realización de un test al momento de registro e incluso llegando a modelar las preferencias del usuario a partir de sus interacciones previas con el sistema.

El método o algoritmo que se use para contrastar los ítems con los usuarios es por lo general el tema central de los diferentes estudios que exploraremos a continuación.

Betancur et al. [1] presenta un modelo de recomendación en el que los Objetos de Aprendizaje son representados a través de *metadatos* definidos en el Esquema LOM (Learning Object Metadata) por el Grupo de Trabajo IEEE P1484.12, mientras que el modelamiento de la recomendación es diferenciado dependiendo del usuario, para el caso del profesor el modelo se basa en los temas dentro de los cuales desarrolla el curso, mientras que para el caso del estudiante el perfil está formado por su información básica, preferencias y estilos de aprendizaje (identificados con el Test de Felder). Se presentan además dos tipos de recomendación, el primer tipo se enfoca en aumentar el conocimiento dentro del proceso de aprendizaje del estudiante en un curso determinado, tomando en cuenta los objetivos del curso para obtener del repositorio los objetos de aprendizaje que podría utilizar como complemento. El segundo tipo de recomendación examina el perfil del usuario en busca de sus estilos de aprendizaje y áreas de interés para ofrecer materiales independientes a los cursos en los que se encuentra registrado y también para promover el ingreso a otros cursos que se encuentran disponibles en el sistema de tutoría.

Casali et al. [2] también emplea LOM para la representación de los ítems (Objetos de Aprendizaje), además considera la importancia relativa de las preferencias del usuario como parte de la fórmula para realizar la recomendación, por lo que utiliza el modelo g-BDI (BGI graduado) para aplicar reglas que tomen en cuenta valoraciones graduadas (no bivaluadas) de las preferencias del usuario.

2.2 Sistemas de Recomendación Basados en Filtrado Colaborativo (*collaborative-filtering*)

El filtrado colaborativo es una técnica empleada por los Sistemas de Recomendación que utiliza la información de preferencias y calificación de un grupo de usuarios respecto a los ítems de un repositorio con el fin de predecir o inferir la preferencia de un usuario en particular sobre un ítem y a partir de esto generar una recomendación acertada.

El filtrado colaborativo presenta características y problemas específicos [12] tales como la *escasez de datos* referentes a calificaciones y preferencias lo que afecta el rendimiento del sistema siendo el problema más común el conocido como *arranque en frío* que se presenta con el ingreso de un nuevo usuario o ítem para el cual no existe suficiente información previa. Para este problema se propone usar el método de filtrado colaborativo híbrido como el medio para producir las predicciones de los nuevos usuarios o ítems. Otros desafíos del filtrado colaborativo son la *escalabilidad* del sistema, la *sinonimia* o similitud semántica de ítems, la *protección de la*

privacidad, entre otros.

Las técnicas de filtrado colaborativas pueden ser agrupadas en tres categorías[12]: los algoritmos *basados en memoria*, que usan una base de datos usuario-item para generar una predicción y se basa en que cada usuario es parte de un grupo con intereses similares, al identificar los vecinos similares de un nuevo usuario se pueden producir predicciones de preferencias de nuevos items, la similaridad entre usuarios puede calcularse por ejemplo con los algoritmos de correlación de Pearson, Spearman rank, Kendall o vector coseno. La segunda categoría son los algoritmos de filtrado basados en modelo, con los que se intenta que el sistema aprenda a reconocer patrones basados en datos de entrenamiento, entre los algoritmos para esta categoría están los modelos bayesianos, de clustering y basados en regresiones. La tercera categoría es el filtrado híbrido, que combina el filtrado colaborativo con otras técnicas de recomendación.

En [12] también se exponen métricas de evaluación para las técnicas de filtrado colaborativo, entre las más usadas están: error medio absoluto (MAE), recuperación y precisión y sensibilidad ROC.

Cechinel et al. [3] emplea la librería Apache Mahout para evaluar recomendaciones de recursos de aprendizaje generados por diferentes algoritmos de filtrado colaborativo del tipo memory-based también conocidos como user-based, aplicados sobre conjuntos de datos obtenidos del repositorio MERLOT. Se evaluó un conjunto de datos con la información de calificaciones que han dado los usuarios a los recursos, aplicando tres métricas de similaridad: Pearson, Euclidean y LogLikelihood, se mostró que el algoritmo LogLikelihood fue capaz de generar aproximadamente cuatro veces más recomendaciones que Pearson y Euclidean. Este trabajo expresa además la falta de pruebas y evaluaciones de sistemas de recomendación con datos capturados en tiempo real, por lo que se lo presenta como un primer intento a la evaluación de técnicas de filtrado colaborativo que puede ser usado como base para otras evaluaciones de recomendación.

2.3 Sistemas de Recomendación Híbridos

Los sistemas híbridos combinan las técnicas de filtrado colaborativo y basado en contenido con el fin de mejorar las recomendaciones resultantes o superar problemas como el arranque en frío [12] que presentan algunos de los métodos anteriores al ser utilizados por separado.

Kim et al. [6] describe un método que combina el filtrado basado en contenido y el filtrado colaborativo. El método parte de la obtención de la información para la creación de los perfiles de usuario. El perfil del usuario indica la información necesaria de las preferencias relacionadas a los items que son de interés del usuario. La creación del perfil puede ser realizada por un método de ponderación manual en el que cada usuario da un peso a los atributos de los items de acuerdo a sus gustos o puede ser automática donde la ponderación se calcula usando características de los atributos e información previa del usuario. Luego de creados los perfiles, éstos son agrupados mediante algoritmos de clustering para proveer la información con

contenido semántico, los grupos se forman usando el algoritmo de K-means. Luego estos clusters son tratados como items para formar una nueva matriz de usuario-item, la similaridad de los usuarios de esta nueva matriz se obtiene con una combinación de los algoritmos correlación coseno y Pearson. Finalmente predicciones para los nuevos usuarios se realizan aplicando sobre la nueva matriz creada un algoritmo de filtrado colaborativo.

En *Learning materials recommendation using good learners' ratings and content-based filtering* [5] se propone un sistema de recomendación para e-learning generado de la combinación de una recomendación basada en contenido y luego una recomendación colaborativa basada en las calificaciones dadas a los materiales por los buenos alumnos, aquellos que obtuvieron las mejores notas. Por medio de un experimento se ha comparado los resultados de la recomendación basada en contenido versus la recomendación híbrida aplicando la métrica del error medio absoluto (MAE), donde se demostró que al usar las calificaciones dadas por los buenos alumnos como suministro se ha mejorado la selección de items de gran calidad y también se ha incrementado el rendimiento en el aprendizaje.

Un método similar es presentado Ruiz et al. [11], donde la recomendación se realiza a partir de una selección de objetos filtrando los mismos usando la información de la metadata del objeto de acuerdo al contenido de la búsqueda del usuario para luego pasar este resultado por un filtro colaborativo que considere las evaluaciones realizadas por otros usuarios. Sin embargo, se pretende mejorar un poco más la calidad del proceso de recomendación al analizar el perfil del usuario y verificar su historial y sus objetivos a corto y largo plazo de modo que los objetos que son presentados sean de valor para el usuario ya que representarán su estado cognitivo, así se podría evitar que el estudiante reciba un material que está muy avanzado para su nivel. El artículo muestra la aplicación de este método de recomendación en un repositorio de recursos educativos de lenguajes de programación, pero concluye que puede ser usado en otro dominio educativo.

Zapata et al. [14] proponen combinar varios tipos de filtrado para obtener una recomendación de objetos más ajustada al perfil de un usuario, tratando de utilizar toda la información que se tenga disponible sobre objetos y usuarios. Ésta información pasará por diversos métodos de clasificación como técnicas de filtrado y calificaciones, técnicas basadas en metadata y contenido, uso de los objetos de aprendizaje y similaridad del perfil del usuario. El método primero preselecciona objetos de aprendizaje del repositorio usando una búsqueda en base a los metadatos, luego pasa por los otros procesos de filtros para obtener una lista final que será la que mejor se adapte al usuario. Pruebas de funcionamiento han sido realizadas sobre un sistema llamado DELPHOS y se presenta un ejemplo donde se puede observar cómo una lista de objetos recomendados va cambiando a medida que se van aplicando los diferentes filtros.

Otros trabajos experimentan enfoques diferentes para mejorar el algoritmo de selección, como el uso de reglas de inferencia [7] como el medio para ayudar a un usuario en su búsqueda de objetos. También hay trabajos donde se emplean diferentes maneras de modelar los objetos, como la utilización de ontologías de dominio [8],

representando los metadatos en términos de conceptos y seleccionando los objetos en el repositorio mediante el algoritmo Spreading Activation, que recorre los arcos del grafo de la ontología y asigna una puntuación a los conceptos que están relacionados semánticamente con los parámetros de búsqueda del usuario. De esta consulta se obtiene una lista de objetos de aprendizaje a la que se aplica un método de recomendación primero filtrando por las preferencias del usuario y luego por el contenido de los objetos buscando la similitud con los objetos que han sido accedidos anteriormente por el usuario.

2.4 Atención Contextualizada

La atención contextualizada es la información de las interacciones del usuario con el sistema dentro del contexto de la tarea del usuario. Esta información también está siendo utilizada en el proceso de recomendación de algunos sistemas.

Ochoa, X. y Duval, E. [9] proponen y detallan el uso de metadata de atención contextualizada (CAM) recolectada en todas las fases del ciclo de vida de un objeto para la recomendación y ranking de objetos de aprendizaje. Primero se listan las acciones que deben ser registradas con un breve detalle de la información a almacenar con CAM durante cada fase del ciclo de vida del objeto. A continuación se propone las métricas para recomendación y ranking, las cuales usarán la metadata del objeto de aprendizaje y las acciones registradas con CAM. Se proponen 4 métricas: *Análisis de enlaces basado en Ranking* donde se utiliza la información de CAM como un grafo del cual se pueden obtener varios rankings como de popularidad, de autor, de ponderación, de factor de reuso y rango. *Medidas de Similitud para Recomendación* que pueden ser Similaridad de Objetos basada en el número de descargas, Similaridad de Objetos basada en Reuso, Similaridad de Usuarios basado en descargas. Similaridad de autores basado en reuso de componentes. *Ranking Personalizado* donde se puede generar un perfil de usuario fuzzy a partir de la información de CAM y combinarlo con el objeto, *Recomendación Contextual* para generar recomendación basada en lo que lo que el usuario está atendiendo al momento.

En *Context-aware Recommender Systems for Learning: a Survey and Future Challenges* [13] se trata de evaluar en los sistemas de recomendación para Innovación en Educación (Technology Enhanced Learning - TEL) el grado en el cual se ha incorporado el uso de información contextualizada sobre el usuario en el proceso de recomendación con el objetivo de construir sistemas inteligentes que pueden predecir mejor y anticipar las necesidades de los usuarios. Se presenta un framework de contexto que identifica las dimensiones relevantes de contexto para aplicaciones TEL y a partir de estas dimensiones se presenta un análisis de los sistemas existentes. Los resultados del estudio indican que existe mucho avance en el desarrollo de recomendadores context-aware en los últimos años.

En *Object Recommendation based on Usage Context* [4], Friedrich et al. parten de la hipótesis de que la similitud en uso da lugar a similitud de contenido y por tanto puede ser usada para recomendaciones. Se define la noción de perfil de contexto de uso para los objetos el cual puede ser generado a partir de un historial de uso y

contiene los objetos que fueron accedidos antes y después del objeto. Con esto se plantea que la recomendación puede ser mejorada comparando el historial de uso de un usuario con los perfiles de contexto de uso de los objetos. Para las pruebas hace uso del proyecto MACE (Metadata for Architectural Contents in Europe), las interacciones del usuario con este sistema (búsqueda, accesos, interacciones con objetos, etc) son monitoreadas y almacenadas usando CAM. Esta información es utilizada para elaborar perfiles de contexto de uso de los objetos y combinarlos por medio de cálculos de similitud con la información del usuario para recomendar objetos. Para la recomendación utiliza filtrado colaborativo item-based y un cálculo de similitud que puede ser de tres maneras diferentes: primero de contexto que está basado en los contextos de uso de los objetos, segundo la similitud de metadata que está basada en diferentes atributos de la metadata de los objetos (representados con LOM) y tercera una comparación manual de una selección de objetos basado en contenido.

3 Metodología

Las herramientas de aprendizaje en línea poseen o acceden a repositorios de objetos de aprendizaje que deben poner a disposición de los estudiantes de acuerdo a la asignatura o tema de estudio. Es lógico pensar que mientras más grandes sean dichos repositorios, mayores posibilidades tienen los estudiantes de obtener mejores objetos para su aprendizaje, sin embargo, se presenta el inconveniente de tener que realizar la búsqueda de los mismos, así como de diferenciar cuales de todos ellos podrían ser más adecuados para cada usuario.

El método utilizado para afrontar dicho problema es el desarrollo de sistemas de recomendación de objetos de aprendizaje que, empleando uno o varios algoritmos determinados, realice una comparación entre el perfil del estudiante y las características e historial de uso de los objetos con el objetivo de proveer el subconjunto ordenado de los mismos que aporte un mayor apoyo al proceso de aprendizaje.

Para el presente trabajo de investigación se utiliza el repositorio de objetos de aprendizaje del proyecto IGUAL (Innovation for Equality in Latin American Universities), el mismo que tiene por objetivo proveer una herramienta de aprendizaje en línea donde los estudiantes puedan acceder a los materiales que le ayuden a complementar su proceso de aprendizaje. Este repositorio ha sido implementado empleando las siguientes herramientas de software libre:

- Hadoop¹ para el almacenamiento distribuido de los archivos.
- 4Store para el almacenamiento de la metadata con RDF

A partir del mencionado repositorio se ha desarrollado un prototipo de sistema de recomendación de objetos de aprendizaje con el que se realizó una evaluación experimental con el objetivo de comparar los resultados obtenidos mediante la

1 <http://hadoop.apache.org/>

aplicación de diferentes algoritmos de recomendación.

3.1 Arquitectura del Sistema de Recomendación

El componente principal del prototipo implementado consiste en una Interfaz de Programación de Aplicaciones, API por sus siglas en inglés, desarrollada en Java, que se encarga de la comunicación con el repositorio del Proyecto IGUAL. Esta API implementa, entre otras, las funciones que ejecutan los algoritmos de recomendación seleccionados.

Otro componente del prototipo es una interfaz web desarrollada con PHP que provee dos funcionalidades, la primera es la creación del perfil de usuario mediante el formulario de registro y la aplicación del Test de Felder, la segunda funcionalidad es la búsqueda de objetos de aprendizaje, para lo cual utiliza una función de la API que se encarga del filtrado de los objetos por el patrón de búsqueda y luego de la selección secuencial y posterior ejecución del método de recomendación.

Perfil del Usuario

El perfil del usuario estará formado por los datos básicos que incluyen el nombre, apellido, usuario, clave, dirección de correo electrónico e idioma y por los estilos de aprendizaje recolectados a través del Test de Felder. Este test consiste en un cuestionario de 44 preguntas cuyo análisis clasifica los estilos de aprendizaje de una persona en cuatro dimensiones:

- Tipo de información: Sensitivo - Intuitivo
- Tipo de estímulos preferenciales: Visual - Verbal
- Forma de procesar y comprender la información: Secuencial – Global
- Forma de trabajar con la información: Activo – Reflexivo

Cabe destacar que por razones de escalabilidad futura, ante la posibilidad de utilizar diferentes métodos o pruebas para complementar el perfil del usuario, se ha decidido almacenar la información obtenida del Test de Felder en una estructura de datos separada.

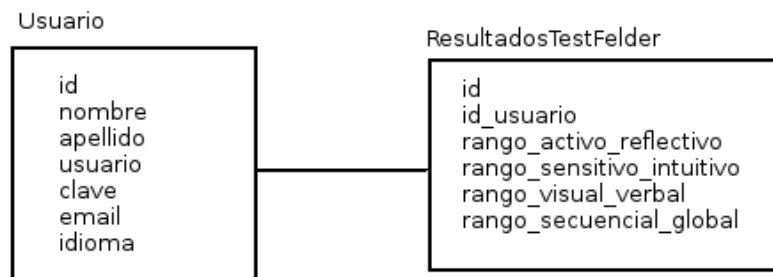


Fig. 1. Estructura de Datos para el Perfil de Usuario

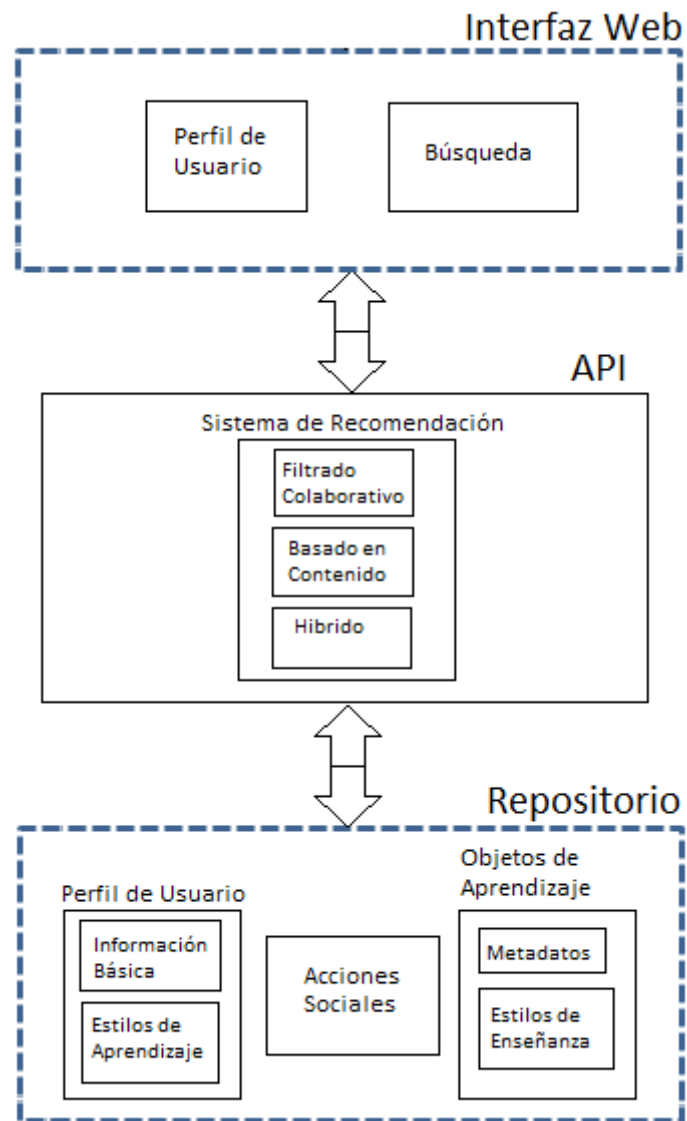


Fig. 2. Arquitectura del Sistema de Recomendación

Objetos de Aprendizaje

Los objetos de aprendizaje están localizados en el repositorio del proyecto IGUAL, el mismo que ha sido implementado en una arquitectura distribuida utilizando Hadoop y son descritos con metadatos en formato LOM, además cada objeto de aprendizaje posee información de los estilos de enseñanza de acuerdo al modelo de Felder. Las dimensiones de los estilos de enseñanza son:

- Contenido: Concreto-Abstracto
- Presentación: Visual-Verbal
- Organización: Inductivo-Deductivo
- Participación del estudiante: Activo-Pasivo
- Perspectiva: Secuencial-Global

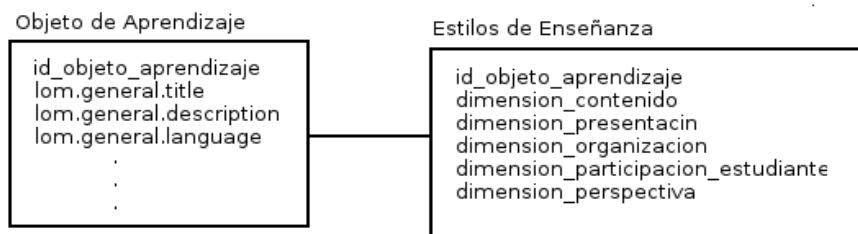


Fig. 3. Estructura de Datos para Objeto de Aprendizaje y Estilos de Enseñanza

Acciones Sociales

Las acciones sociales están representadas por las interacciones que realiza un usuario con los objetos de aprendizaje como recomendar, calificar o definir gusto por los mismos. Estas interacciones representan el grado de satisfacción o preferencia que un usuario tiene sobre cierto objeto y están almacenadas en el repositorio de la siguiente manera:



Fig. 4. Estructura de Datos para Acciones Sociales

Estas preferencias tienen dos funciones principales. En primer lugar sirven como fuente de datos en el proceso de recomendación de objetos para otros usuarios por medio del método de filtrado colaborativo. La segunda función es servir como fuente de información para comprobar cómo valora el usuario el objeto y de esta manera analizar si el método de recomendación por el cual este objeto fue seleccionado

satisface las necesidades del usuario.

Métodos de Recomendación

Recomendación basada en contenido

La técnica empleada para la recomendación basada en contenido consiste en un método simple de asignación de puntos de acuerdo a la similitud de características compartidas entre el objeto de aprendizaje y el perfil del estudiante.

En primer lugar se compara el idioma del objeto de aprendizaje contra el idioma del usuario, si estos son iguales se suman dos puntos al peso del objeto. A continuación se analizan los estilos de enseñanza que se han definido para los objetos y los estilos de aprendizaje del estudiante. En la siguiente tabla se muestra las características examinadas y la puntuación correspondiente.

Tabla 1: Valoración para la Recomendación Basada en Contenido

Objeto de Aprendizaje Estilos de Enseñanza	Alumno Estilos de Aprendizaje	Puntos sumados
Visual	Valor del rango Visual/Verbal ≤ -5	1
Verbal	Valor del rango Visual/Verbal ≥ 5	1
Secuencial	Valor del rango Secuencial/Global ≤ -5	1
Global	Valor del rango Secuencial/Global ≥ 5	1
Activo	Valor del rango Activo/Reflexivo ≤ -5	1
Pasivo	Valor del rango Activo/ Reflexivo ≥ 5	1
Concreto	Valor del rango Sensorial/Intuitivo ≤ -5	1
Abstracto	Valor del rango Sensorial/Intuitivo ≥ 5	1

Al final los objetos a recomendar se obtienen ordenándolos de mayor a menor de acuerdo al puntaje obtenido, de esta manera los objetos que mejor se ajustan al perfil del usuarios serán los primeros en mostrarse.

Recomendación basada en filtrado colaborativo

Para esta recomendación se usarán las acciones sociales realizadas por los usuarios sobre los objetos de aprendizaje, con el fin de recomendar al estudiante los materiales que han resultado de mayor interés para otros usuarios y que él aún no ha revisado. Este proceso de recomendación usará Mahout², un proyecto de la Fundación Apache compuesto por un conjunto de librerías de código abierto para aprendizaje automatizado.

² <http://mahout.apache.org/>

El uso de Mahout para la recomendación con filtrado colaborativo incluye los siguientes pasos:

1. Proporcionar el modelo de datos que consiste en un archivo de texto que representa las preferencias de los usuarios a los objetos de aprendizaje, cada línea está expresada de la siguiente manera:
id_usuario, id_objeto, calificación
2. Seleccionar el algoritmo de similaridad, Mahout implementa varios algoritmos como Correlación de Pearson, Medida Coseno, Coeficiente de Tanimoto, Log Likelihood , entre otros.
3. Establecer el tamaño del vecindario del usuario U_1 para determinar los usuarios con gustos similares a U_1 y de esta manera obtener mejores recomendaciones.
4. Inicializar el recomendador en el que se especifica el modelo de datos, el algoritmo de similaridad y el vecindario.
5. Finalmente los objetos recomendados se obtienen al indicar sobre cuál usuario se desea realizar la recomendación.

```
//1. Modelo de datos
FileDataModel model = new FileDataModel(new File(sourcefile));
//2. Algoritmo de similaridad
UserSimilarity userSimilarity = new PearsonCorrelationSimilarity(model);
//3. Vecindario del usuario
UserNeighborhood neighborhood = new NearestUserNeighborhood(3,userSimilarity, model);
//4. Inicializar el recomendador
Recommender recommender = new GenericUserBasedRecommender(model, neighborhood, userSimilarity);
//5. Generar las recomendaciones de objetos
List<RecommendedItem> recommendations =recommender.recommend(userId, numItems);
```

Fig. 5. Uso de la librería Mahout en la Recomendación Colaborativa

Recomendación Híbrida

Esta recomendación es una combinación de los dos métodos anteriores. En primer lugar se aplica el método de recomendación basado en contenido para obtener una lista de objetos de aprendizaje de los cuales se buscarán todas las acciones sociales que los usuarios han realizado sobre ellos. A continuación el método de recomendación de filtrado colaborativo es aplicado utilizando esta entrada de datos.

3.2 Evaluación Experimental

Esta fase consiste en la aplicación de los métodos de recomendación implementados y la recolección de los datos que sirven para la comparación y evaluación de los mismos. Consta de 2 tipos de experimentos:

Experimento Sintético

El objetivo de este experimento es comparar la lista de objetos de aprendizaje recomendados por cada uno de los métodos implementados en el prototipo, con aquellos sugeridos por un usuario experto en la materia.

A un profesor de la materia de Fundamentos de Programación se le envió una lista de 10 temas de estudio de la materia y se le solicitó que seleccionara del repositorio aquellos objetos de aprendizaje que estén relacionados con cada tópico, ordenándolos de acuerdo a si el estudiante prefiere materiales Visuales o si tiene preferencia por materiales de tipo Verbal.

Adicionalmente, se utiliza el prototipo para ejecutar cada método de recomendación bajo el perfil de cada tipo de estudiante (preferencias verbales o visuales), esperando que los resultados obtenidos sean similares a los propuestos por el usuario experto.

Experimento Natural

Este experimento tiene como finalidad medir los niveles de aceptación de los resultados de cada método de recomendación, obtenidos a partir de la interacción de usuarios reales (alumnos) con el sistema. Para esto se hace uso de las acciones sociales como like, recomendar y calificar, disponibles en la misma interfaz de búsqueda.

La función de la API se encarga de alternar los métodos de recomendación en orden secuencial, de manera transparente para el usuario, quien no tiene conocimiento sobre el método utilizado en cada momento, lo que brinda neutralidad al momento de analizar cuál de los métodos recibe mejor aceptación.

Para la ejecución de este experimento se solicitó la colaboración de estudiantes de la materia Fundamentos de Programación, quienes tuvieron libertad respecto a los tópicos sobre los cuales realizaron las búsquedas de objetos de aprendizaje.

4 Análisis de Resultados

4.1 Resultados Experimento Sintético

Para evaluar los objetos recomendados por el tutor experto y aquellos recomendados por el sistema de recomendación se realiza el cálculo de la variable *Precisión*, mientras que para medir la similaridad del orden de los objetos recomendados se ha utilizado la *Distancia de Kendall Tau*.

$$\text{Precisión} = \frac{\text{Número Documentos Relevantes Recuperados}}{\text{Número Total de Documentos Recuperados}}$$

Para este análisis el número de documentos relevantes corresponde al número de objetos que están en el grupo de los objetos recomendados por el tutor experto.

R1: Recomendación Filtrado Colaborativo

R2: Recomendación Basada en Contenido

R3: Recomendación Híbrida

Tabla 2. Resultados y Análisis de la búsqueda de materiales sobre *manejo de memoria*

Tutor Experto		Sistema de Recomendación					
Preferencia Objetos Verbales	Preferencia Objetos Visuales	Preferencia Objetos Verbales			Preferencia Objetos Visuales		
		R1	R2	R3	R1	R2	R3
126	114	114	126	126	114	114	114
127	126	126	127	127	126	126	126
114	127	127	114	114	127	127	127
Precisión		100%	100%	100%	100%	100%	100%
Distancia Kendall Tau		0.67	0	0	0	0	0

Tabla 3. Resultados y Análisis de la búsqueda de materiales sobre *arreglos*

Tutor Experto		Sistema de Recomendación					
Preferencia Objetos Verbales	Preferencia Objetos Visuales	Preferencia Objetos Verbales			Preferencia Objetos Visuales		
		R1	R2	R3	R1	R2	R3
1	19	19	1	1	19	19	19
19	35	35	19	19	35	35	35
35	1	1	35	35	1	1	1
Precisión		100%	100%	100%	100%	100%	100%
Distancia Kendall Tau		0.67	0	0	0	0	0

Tabla 4. Resultados y Análisis de la búsqueda de materiales sobre *operadores*

Tutor Experto		Sistema de Recomendación					
Preferencia Objetos Verbales	Preferencia Objetos Visuales	Preferencia Objetos Verbales			Preferencia Objetos Visuales		
		R1	R2	R3	R1	R2	R3
8	20	10	10	10	10	44	44
9	21	100	9	9	100	41	41
10	41	11	8	8	11	21	21
11	44	136	11	11	136	20	20
20	8	155	20	20	155	11	11
21	9	156	21	21	156	10	10
41	10	20	44	44	20	8	8
44	11	21	41	41	21	9	9
Precisión		50%	100%	100%	50%	100%	100%
Distancia Kendall Tau		-	0.18	0.18	-	0.39	0.39

Tabla 5. Resultados y Análisis de la búsqueda de materiales sobre *switch*

Tutor Experto		Sistema de Recomendación					
Preferencia Objetos Verbales	Preferencia Objetos Visuales	Preferencia Objetos Verbales			Preferencia Objetos Visuales		
		R1	R2	R3	R1	R2	R3
15	28	15	15	15	15	28	28
28	43	163	28	28	163	43	43
43	15	28	43	43	28	15	15
Precisión		67%	100%	100%	66%	100%	100%
Distancia Kendall Tau		-	0	0	-	0	0

Tabla 6. Resultados y Análisis de la búsqueda de materiales sobre *condicionales*

Tutor Experto		Sistema de Recomendación					
Preferencia Objetos Verbales	Preferencia Objetos Visuales	Preferencia Objetos Verbales			Preferencia Objetos Visuales		
		R1	R2	R3	R1	R2	R3
11	27	11	11	11	11	27	27
27	28	27	27	27	27	28	28
28	43	28	28	28	28	43	43
43	11	43	43	43	43	11	11
Precisión		100%	100%	100%	100%	100%	100%
Distancia Kendall Tau		0	0	0	0.5	0	0

Tabla 7. Resultados y Análisis de la búsqueda de materiales sobre *variables*

Tutor Experto		Sistema de Recomendación					
Preferencia Objetos Verbales	Preferencia Objetos Visuales	Preferencia Objetos Verbales			Preferencia Objetos Visuales		
		R1	R2	R3	R1	R2	R3
16	107	107	16	16	107	107	107
33	19	16	33	33	16	19	19
19	16	19	107	107	19	16	16
107	33	33	19	19	33	33	33
Precisión		100%	100%	100%	100%	100%	100%
Distancia Kendall Tau		0.67	0.17	0.17	0	0	0

Tabla 8. Resultados y Análisis de la búsqueda de materiales sobre *punteros*

Tutor Experto		Sistema de Recomendación					
Preferencia Objetos Verbales	Preferencia Objetos Visuales	Preferencia Objetos Verbales			Preferencia Objetos Visuales		
		R1	R2	R3	R1	R2	R3
13	114	128	13	13	128	128	128

14	128	114	14	14	114	114	114
114	37	13	128	128	13	37	37
128	13	14	114	114	14	13	13
37	14	37	37	37	37	14	14
Precisión		100%	100%	100%	100%	100%	100%
Distancia Kendall Tau		0.5	0.1	0.1	0.3	0.1	0.1

Tabla 9. Resultados y Análisis de la búsqueda de materiales sobre *punteros*

Tutor Experto		Sistema de Recomendación					
Preferencia Objetos Verbales	Preferencia Objetos Visuales	Preferencia Objetos Verbales			Preferencia Objetos Visuales		
		R1	R2	R3	R1	R2	R3
119	120	119	119	119	119	120	120
120	119	120	120	120	120	119	119
Precisión		100%	100%	100%	100%	100%	100%
Distancia Kendall Tau		0	0	0	0	0	0

Tabla 10. Resultados y Análisis de la búsqueda de materiales sobre *sentencias de iteración*

Tutor Experto		Sistema de Recomendación					
Preferencia Objetos Verbales	Preferencia Objetos Visuales	Preferencia Objetos Verbales			Preferencia Objetos Visuales		
		R1	R2	R3	R1	R2	R3
7	45	45	7	7	45	45	45
15	7	7	15	15	7	7	7
45	15	15	45	45	15	15	15
Precisión		100%	100%	100%	100%	100%	100%
Distancia Kendall Tau		0.67	0	0	0	0	0

Tabla 11. Resultados y Análisis de la búsqueda de materiales sobre *formato de cadenas*

Tutor Experto		Sistema de Recomendación					
Preferencia Objetos Verbales	Preferencia Objetos Visuales	Preferencia Objetos Verbales			Preferencia Objetos Visuales		
		R1	R2	R3	R1	R2	R3
2	46	117	117	117	117	46	46
117	2	46	2	2	46	117	117
46	117	2	46	46	2	2	2
Precisión		100%	100%	100%	100%	100%	100%
Distancia Kendall Tau		0.66	0.34	0.34	0.66	0.34	0.34

De los resultados de las pruebas se puede notar que en todos los casos la recomendación basada en contenido y la recomendación híbrida tiene una precisión del 100%. Para el caso de la recomendación basada en filtrado colaborativo, al ser la muestra pequeña no se genera ninguna recomendación y el resultado es el generado por el patrón de búsqueda. De acuerdo a los resultados de la recomendación el método híbrido no pudo generar la recomendación usando información colaborativa lo que se ve reflejado en la igualdad de las recomendaciones R2 y R3

En cuanto a la similitud del orden de los objetos, medida con la Distancia de Kendall Tau, se puede ver que en la mayoría de los casos coinciden, siendo las más parecidas las recomendaciones por el método basado en contenido e híbrido

4.1 Resultados Experimento Natural

En este experimento participaron 44 estudiantes, 65% de ellos tiene preferencia por objetos visuales, 4.5% preferencia por objetos verbales y 29.5% se encuentra en un rango de equilibrio entre estas dos dimensiones. Los estudiantes realizaron pruebas con el repositorio que contiene 168 objetos de aprendizaje.

Lamentablemente por la cantidad limitada de objetos de aprendizaje y de usuarios, el método de recomendación basado en filtrado colaborativo no generó resultados en la mayoría de su ejecución. También hay que anotar que el método de recomendación de filtrado usando Mahout requiere que el archivo con la información colaborativa contenga registros del usuario al que se va a realizar la recomendación. Esto también originó que muchas de las recomendaciones híbridas tengan como resultado el equivalente a la recomendación basada en contenido.

A pesar de estos inconvenientes se pudo observar lo siguiente:

- De los patrones de búsqueda que devolvían entre 2 y 4 materiales, el 97% de los materiales fueron seleccionados y el 58% recibió algún tipo de actividad

social.

- Gran parte de los usuarios descargó los primeros objetos de la lista recomendada.
- Los resultados de la recomendación basada en contenido fueron las más aceptadas.

5 Conclusiones

Durante la realización del presente proyecto se utilizó de manera exclusiva herramientas de software libre, lo que reafirma la viabilidad de éstas en tareas de soporte a investigaciones. Además, la existencia y uso de librerías especializadas, como el caso de Mahout para la evaluación de algoritmos de recomendación, permite ahorrar tiempo y esfuerzo en el desarrollo de los prototipos y permiten enfocarse en otros aspectos de la investigación.

Aunque la aplicación de los métodos de recomendación colaborativa no arrojó los resultados esperados, se puede atribuir esta situación a la poca cantidad de materiales disponibles, así como a la escasa actividad previa de parte de los usuarios, que imposibilitó la creación del perfil basado en su comportamiento con el sistema.

Los buenos resultados obtenidos por el método de recomendación basado en contenido durante la etapa de experimentación avalan no solo el algoritmo como tal, sino también la utilización del Test de Felder para agregar al perfil del usuario información referente al estilo de aprendizaje.

Se confirma la ventaja de utilizar el método de recomendación híbrido sobre la recomendación colaborativa en los casos de arranque en frío, puesto que mientras el método colaborativo simplemente muestra los resultados en el orden obtenido por el filtro, los items recomendados por el algoritmo híbrido llevan al menos cierta información relacionada al usuario provista por la recomendación basado en contenido.

Referencias

1. Betancur, D., Moreno, J., Ovalle, D., Modelo para la recomendación y recuperación de objetos de aprendizaje en entornos virtuales de enseñanza/aprendizaje. Revista Avances en Sistemas e Informática Vol 6 N° 1. (2009)
2. Casali, A., Gerling, V., Deco, C., Bender, C, Sistema inteligente para la recomendación de objetos de aprendizaje. Revista Generación Digital Vol. 9 No. 1. Edición 16. (2011)
3. Cechinel, C., Sicilia, M., Sánchez S., García, E., Evaluating collaborative filtering recommendations inside large learning object repositories. Information Pocesing & Management. (2012)
4. Friedrich, M., Niemann, K., Scheffel, M., Schmitz, H., Wolpers, M., Object Recommendation based on Usage Context. Workshop about Context-aware Recommendation for Learning at the STELLAR Alpine Rendez-Vous (2009)

5. Ghauth, K. I., Abdullah, N. A., Learning materials recommendation using good learners' ratings and content-based filtering. *Educational Technology Research and Development* 58, no. 6. (2010)
6. Kim, B. , Li, Q. , Park, Ch., Kim, S. , Kim, J., A new approach for combining content- based and collaborative filters. *Journal of Intelligent Information Systems*. (2006).
7. Lemire, D., Boley, H., McGrath, S., Ball, M., Collaborative Filtering and Inference Rules for Context-Aware Learning Object Recommendation. *Interactive Technology and Smart Education* , Vol. 2 Iss: 3, pp.179 – 188. (2005)
8. Machado , N., Montoyo , A., Recomendación de objetos de aprendizaje almacenados en repositorios lor@server según las preferencias del usuario. *Proceedings of the 5th WSEAS International Conference on E- ACTIVITIES* . (2006)
9. Ochoa, X., Duval, E., Use of contextualized attention metadata for ranking and recommending learning objects. *Proceedings of the 1st international workshop on Contextualized attention metadata: collecting, managing and exploiting of rich usage information CIKM 2006*, pp. 9-16. (2006)
10. Pazzani, M., Billsus, D., Content-Based Recommendation Systems. *The Adaptive Web: Methods and Strategies of Web Personalization, Lecture Notes in Computer Science*, Vol. 4321. (2007)
11. Ruiz A., Jiménez, G. Gómez, M., Personalización en Recomendadores Basados en Contenido y su Aplicación a Repositorios de Objetos de Aprendizaje. *IEEE-RITA Vol. 5, Núm. 1*. (2010)
12. Su, X., Khoshgoftaar, T., A Survey of Collaborative Filtering Techniques. *Advances in Artificial Intelligence Vol 2009*. (2012)
13. Verbert, K., Manouselis, N., Ochoa, X., Wolpers, M., Drachsler, H., Bosnic, I., Duval, E., Context-aware Recommender Systems for Learning: a Survey and Future Challenges. *IEEE Transactions on Learning Technologies Vol. 99*. (2012)
14. Zapata, A., Menendez, V., Prieto, M., Romero, C., A Hybrid Recommender Method for Learning Objects. *Design and Evaluation of Digital Content for Education Proceedings published by International Journal of Computer Applications*. (2011)