

## Sistema de recomendación de contenidos audiovisuales: Algoritmo de inferencia semántica

Ávila J.<sup>1</sup>, Riofrío X.<sup>1</sup>, Palacio-Baus K.<sup>2</sup>, Espinoza-Mejía M.<sup>3</sup>, Saquicela V.<sup>3</sup>

<sup>1</sup> Escuela de Informática, Facultad de Ingeniería, Universidad de Cuenca, Av. 12 de Abril, Ciudadela Universitaria, Cuenca, Ecuador, EC010201.

<sup>2</sup> Departamento de Ingeniería Eléctrica, Electrónica y Telecomunicaciones, Universidad de Cuenca, Av. 12 de Abril, Ciudadela Universitaria, Cuenca, Ecuador, EC010201.

<sup>3</sup> Departamento de Ciencias de la Computación, Universidad de Cuenca, Av. 12 de Abril, Ciudadela Universitaria, Cuenca, Ecuador, EC010201.

Autores para correspondencia: {johnny.avilam, xavier.riofriom}@ucuenca.edu.ec, {kenneth.palacio, mauricio.espinoza, victor.saquicela}@ucuenca.edu.ec

Fecha de recepción: 21 de septiembre de 2014 - Fecha de aceptación: 17 de octubre de 2014

### RESUMEN

Este artículo presenta el análisis de un Algoritmo de Inferencia Semántica utilizado en un Sistema de Recomendación de Contenidos Audiovisuales en el contexto de la Televisión Digital. Los resultados obtenidos muestran que la inclusión de diferentes propiedades semánticas y sus combinaciones, influye directamente en la reducción del error absoluto promedio obtenido en la predicción de la calificación otorgada por un usuario a un ítem determinado. Además se ha determinado que la propiedad Actor tiene un impacto mayor con respecto a otras propiedades analizadas.

Palabras clave: Televisión digital, sistemas de recomendación semánticos, Web semántica, propiedades semánticas, ontologías.

### ABSTRACT

This article presents the analysis of a Semantic Inference Algorithm used in an Audiovisual Content Recommender System in the domain of Digital Television. The obtained results show that the inclusion of distinct semantic properties and their combinations lead to a considerable reduction of the Mean Absolute Error (MAE) obtained for user rating prediction computation and that the semantic property *Actor* has the major influence over recommendation estimation.

Keywords: Digital television, semantic recommender systems, Web semantic, semantic properties, ontologies.

## 1. INTRODUCCIÓN

En el contexto de los cambios tecnológicos de los últimos años, el apagón analógico, propiciado por la llegada de la Televisión Digital (DTV), ha llamado la atención de la sociedad por los cambios y nuevas características que se incorporan con los nuevos estándares de transmisión que se han adoptado a nivel mundial y que revolucionarán la forma de entretenerse con la televisión. Para el caso de Ecuador, el estándar Japonés-Brasileño ISDB-Tb fue adoptado en Marzo de 2010 (Superintendencia de Telecomunicaciones, 2010), y se ha programado un proceso de transición paulatina que garantiza que para el inicio del año 2019, todo el país contará con señales de televisión digital abierta. Según Alulema (2012), entre las novedades asociadas a la DTV se encuentra la interactividad, que permitirá que el televidente descargue aplicaciones relacionadas a diferentes servicios, como educación, salud,

democracia, entretenimiento, etc. Otras ventajas del estándar ISDB-Tb son: el aspecto de movilidad, que facilitará la recepción de televisión abierta en dispositivos móviles; el mejoramiento notable de la calidad de la imagen y el sonido transmitidos; y además, el mejor aprovechamiento del espectro radioeléctrico, que permitirá disponer de un mayor número de estaciones de televisión disponibles en el mismo ancho de banda (Nakahara *et al.*, 1996) contemplado en el modelo de televisión analógica, y con ello, un incremento considerable de la oferta de contenido televisivo. Si bien un aumento de la oferta televisiva podría parecer ventajoso en primera instancia, estudios como el documentado en Martin (2006), demuestran que esto no es necesariamente cierto, y que la satisfacción de un individuo es mayor cuando tiene pocas opciones para escoger. Para el caso de la televisión, se espera que una sobreoferta de contenidos audiovisuales abruma incluso a los más asiduos televidentes, por lo que es necesario contar con mecanismos que ayuden a disminuir el impacto de esta sobreoferta y contribuyan a enriquecer la experiencia televisiva.

Los Sistemas de Recomendación (SR) nacen como respuesta a la necesidad de contar con una herramienta de personalización de contenido, como se describe en Kantor *et al.* (2011). En general, estos sistemas utilizan información histórica del usuario para recomendarle ítems de su agrado. Las recomendaciones se basan normalmente en sus preferencias y apuntan a mejorar la experiencia del usuario automatizando la búsqueda de contenido. Aunque los SR ya se han utilizado exitosamente en diferentes dominios, incluyendo el comercio electrónico, las compras en línea y la distribución de noticias, su aplicación en el contexto de la DTV es relativamente nuevo, por lo que este proyecto persigue utilizar los SR para solventar y disminuir la sobrecarga de información a la que se enfrenta un usuario de televisión digital. Así, se abre las puertas de la investigación a través de técnicas novedosas como la incorporación de la web semántica.

En este artículo se extiende el trabajo realizado por Scuoteguazza & Juayek, (2012), con el objetivo de incorporar mecanismos de mitigación de los problemas comunes encontrados en los sistemas de recomendación y particularmente, encajando la propuesta en el contexto de la televisión digital terrestre. Adicionalmente se propone un procedimiento para evaluar el impacto que tienen ciertas propiedades relacionadas al contenido televisivo como los actores de una serie determinada, sus directores, género, etc., sobre las preferencias de los usuarios hacia el contenido televisivo. Así, puede determinarse con mayor precisión aquellos elementos preponderantes a incluirse en la estimación de las calificaciones que un usuario otorgaría a un ítem y con ello, la generación de recomendaciones de contenido.

Este artículo está organizado de la siguiente manera: La Sección 2 resume brevemente los Sistemas de Recomendación Semánticos como base de la presente propuesta, así mismo, recoge el estado del arte de varios enfoques que hacen uso de la semántica en el campo de la generación de recomendaciones. A continuación, la Sección 3 presenta la arquitectura del sistema implementado, describiendo sus componentes modulares y su interacción. La Sección 4 muestra los procedimientos utilizados para la evaluación del sistema, así como los descubrimientos y resultados obtenidos. Finalmente, la Sección 5 presenta conclusiones preliminares obtenidas a partir de la implementación del sistema.

## 2. SISTEMAS DE RECOMENDACIÓN SEMÁNTICOS - SRS

Los SRS se engloban dentro los *Sistemas de Recomendación Basados en Contenido*, que conjuntamente con los *Sistemas de Recomendación de Filtrado Colaborativo*, forman los dos tipos de SR definidos en la literatura (Kantor *et al.*, 2011). Se fundamentan en una base de conocimiento, en la que mediante herramientas de la web semántica se enriquece los perfiles de usuario haciendo uso de ontologías, como se sugiere en Espinoza-Mejía *et al.* (2014). Una ontología es un esquema conceptual con un orden determinado basado en etiquetas predefinidas, que abarca uno o varios dominios, con el objetivo de favorecer la comunicación de información entre diferentes sistemas y entidades. Lenguajes como RDF y OWL son ampliamente utilizados para describirlas. Además, se han propuesto métodos para enriquecer la información de las guías de programación electrónica (EPG) para mejorar las recomendaciones (Saquicela *et al.*, 2014). Los SRS explotan las características de las ontologías, en la que se pretende que la información tenga un significado bien definido y una estructura clara que

permita unir conceptos entre sí. De esta forma, los SRS permiten obtener mejores resultados que los sistemas de filtrado colaborativo, como se reporta en Sotelo *et al.* (2013), y por ello el interés de explotarlos.

Entre las propuestas más relevantes de SRS se encuentra el sistema AVATAR, ampliamente documentado en Fernández *et al.* (2006) y Lopez-Nores *et al.* (2010). Este se basa en una arquitectura híbrida, que emplea la estrategia colaborativa y la basada en contenido con el fin de inferir conocimiento combinando las semánticas del contenido televisivo y los perfiles del usuario. En AVATAR, se utiliza una ontología que describe el contenido utilizando OWL para almacenar clases, instancias y propiedades organizadas jerárquicamente e identificar a los recursos y relaciones más comunes en el dominio de la programación televisiva.

Cantador *et al.* (2008) presentan un modelo híbrido multi-capa basado en ontologías, donde el contenido a recomendar y las preferencias del usuario se representan con ontologías. En esta propuesta, se compara los intereses de los usuarios en función de un concepto semántico para encontrar similitudes entre dichos usuarios. Este modelo particiona los perfiles de usuario en grupos de interés relacionados para establecer varias capas de *Grado de Interés (DOI)* que proporcionan un modelo enriquecido de vínculos interpersonales.

Codina y otros, presentan en Codina & Ceccaroni (2010) un *SR* adaptativo enfocado al turismo, aunque puede extenderse a otros dominios. El sistema incluye mecanismos que permiten realizar inferencias en información incompleta, lo que permite reducir los efectos del *arranque en frío* como se documenta en Yu *et al.* (2004) y Adomavicius & Tuzhilin (2005). Su representación del conocimiento se basa en jerarquías y criterios de similitud semántica para encontrar los mejores ítems a ser recomendados para un usuario al igual que en AVATAR. Su arquitectura utiliza el paradigma Service Oriented Application (SOA) o Aplicación Orientada a los Servicios, lo que lo convierte en un sistema flexible en cuanto al dominio. Para hacer uso del módulo recomendador, solo es necesario consumir su servicio desde una interfaz pública que recibe una ontología de dominio como parámetro compatible predefinida en formato OWL o RDF.

### 3. SISTEMA PROPUESTO

#### 3.1. Algoritmo de inferencia semántica

El algoritmo analizado en este artículo encuentra conocimiento implícito buscando relaciones entre los recursos semánticos que comparten propiedades en común, según se documenta en Scuoteguazza & Juayek (2012). Así, se puede formar secuencias de recursos que permiten obtener una calificación para el contenido televisivo analizando cada una de las secuencias encontradas. Para ello se ejecutan los siguientes procedimientos:

##### Creación de las cadenas de secuencias

Consiste en encontrar relaciones entre los recursos conectando las propiedades semánticas que comparten entre sí. La Fig. 1 muestra el modelamiento para la película *“Piratas del caribe, la maldición de la perla negra”*.

A partir de este modelo, se busca contenidos televisivos que entre sus propiedades semánticas incluyan el recurso “Johnny Deep”, para crear una cadena con la primera película. En la propia descripción del recurso “Johnny Deep” se expresan explícitamente los roles que cumple este recurso en otros contenidos televisivos. El recurso “Johnny Deep”, se describe para obtener las conexiones mostradas en la Fig. 2. Así por ejemplo, se relaciona la película “Los piratas del caribe, la maldición de la perla negra” con las películas “Rango” y “Alicia en el país de las maravillas” puesto que en todas aparece el recurso “Johnny Deep” como actor. No solo se pueden conectar recursos de tipo película o del mismo tipo entre sí, sino que por ejemplo, se puede conectar la película “Rocky” con el reality show “El Retador” puesto que en ambos recursos aparece “Silvester Stallone” entre sus propiedades.

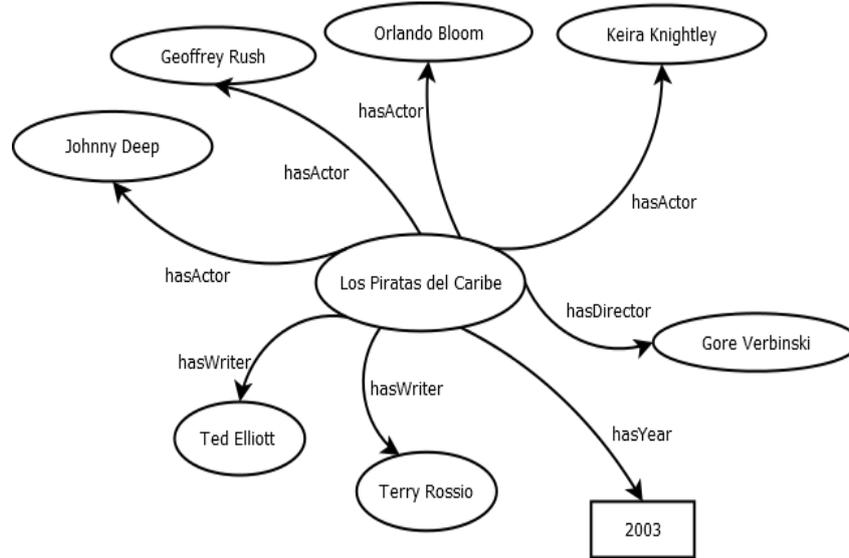


Figura 1. Ontología para la película pirata del caribe.

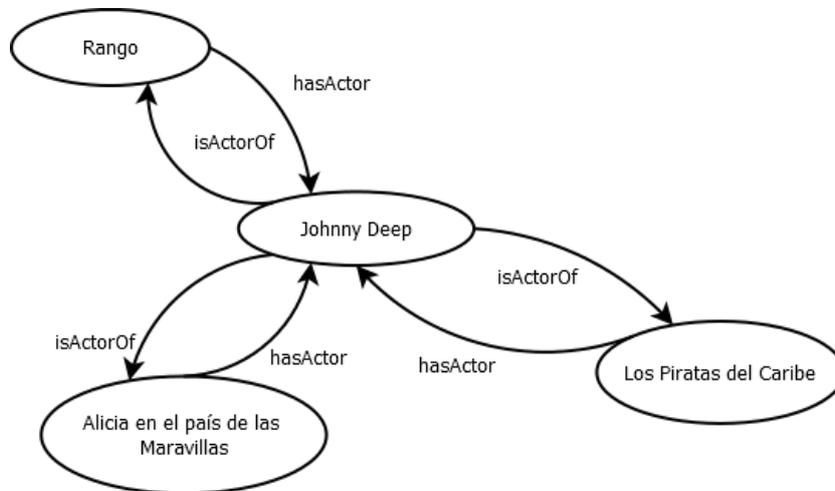


Figura 2. Descripción del recurso “Johnny Deep”.

La representación de la Fig. 2 tiene fines didácticos. En realidad, todos los recursos deben estar identificados con una *URI*. La Fig. 3 muestra la representación formal del recurso “Johnny Deep” en formato *RDF*.

En cada iteración, el algoritmo crea cadenas de conexiones o secuencias para relacionar contenidos entre sí, considerando aquellos conceptos de interés para el usuario, como se documenta en Ávila *et al.* (2014b). La primera conexión considera únicamente conceptos con un DOI superior a 3 (en la escala de calificaciones 1-5); a partir de la primera conexión se incluye sólo a los recursos con un DOI superior al recurso anterior, tomando en cuenta todos los tipos de propiedades semánticas: Escritor, Director, Etc. En una segunda etapa, el algoritmo busca recursivamente conexiones con más recursos con propiedades comunes. Así, suponiendo que la película “Alicia en el país de las maravillas” tiene un DOI de 5 para el usuario, mientras que la película “Rango” un DOI de 4, lo que resulta es que en la secuencia se agregará únicamente la primera película. Después de la segunda iteración y analizando todos los conceptos asociados, se crearán tantas conexiones como sea posible siempre y cuando el DOI del recurso sea mayor que el anterior y que un recurso no puede aparecer más de una vez. Una secuencia resultante se muestra en la Fig. 4.

```

<rdf:Description rdf:about="file:///home/marcos/workspace/Ontology/ONTOLOGIES/OntologyIMDB.owl#JohnnyDepp"
<isActorOf rdf:resource="file:///home/marcos/workspace/Ontology/ONTOLOGIES/OntologyIMDB.owl#tt0109707"/
<isActorOf rdf:resource="file:///home/marcos/workspace/Ontology/ONTOLOGIES/OntologyIMDB.owl#tt0113972"/
<rdf:type rdf:resource="file:///home/marcos/workspace/Ontology/ONTOLOGIES/OntologyIMDB.owl#Actor"/>
<isActorOf rdf:resource="file:///home/marcos/workspace/Ontology/ONTOLOGIES/OntologyIMDB.owl#tt0106387"/
<hasName rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Johnny Depp</hasName>
<isActorOf rdf:resource="file:///home/marcos/workspace/Ontology/ONTOLOGIES/OntologyIMDB.owl#tt0181833"/
<isActorOf rdf:resource="file:///home/marcos/workspace/Ontology/ONTOLOGIES/OntologyIMDB.owl#tt0138304"/
<isActorOf rdf:resource="file:///home/marcos/workspace/Ontology/ONTOLOGIES/OntologyIMDB.owl#tt0120669"/
<isActorOf rdf:resource="file:///home/marcos/workspace/Ontology/ONTOLOGIES/OntologyIMDB.owl#tt0162661"/
<isActorOf rdf:resource="file:///home/marcos/workspace/Ontology/ONTOLOGIES/OntologyIMDB.owl#tt0119008"/
<isActorOf rdf:resource="file:///home/marcos/workspace/Ontology/ONTOLOGIES/OntologyIMDB.owl#tt0099487"/
<isActorOf rdf:resource="file:///home/marcos/workspace/Ontology/ONTOLOGIES/OntologyIMDB.owl#tt0112883"/
<isActorOf rdf:resource="file:///home/marcos/workspace/Ontology/ONTOLOGIES/OntologyIMDB.owl#tt0108550"/
</rdf:Description>
    
```

Figura 3. Representación RDF del recurso “Johnny Deep”.

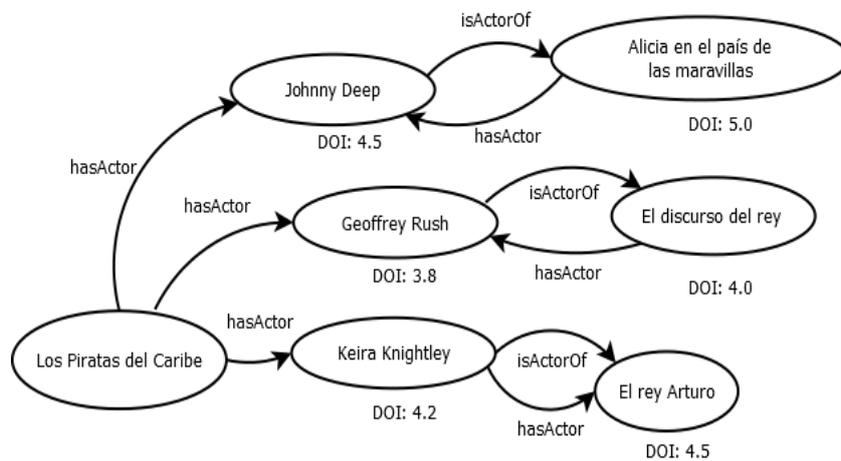


Figura 4. Creación de la secuencia en la segunda iteración.

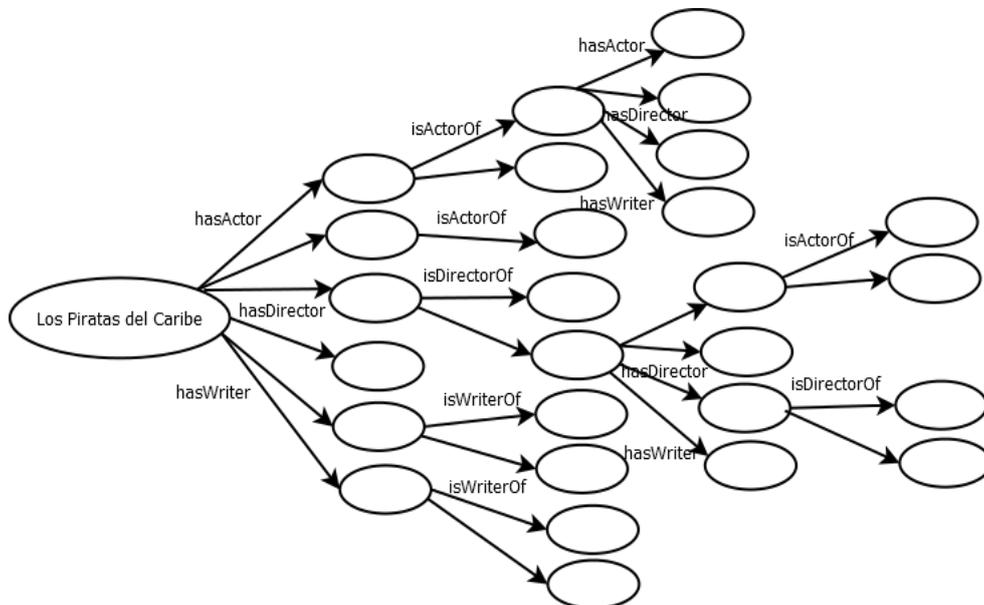
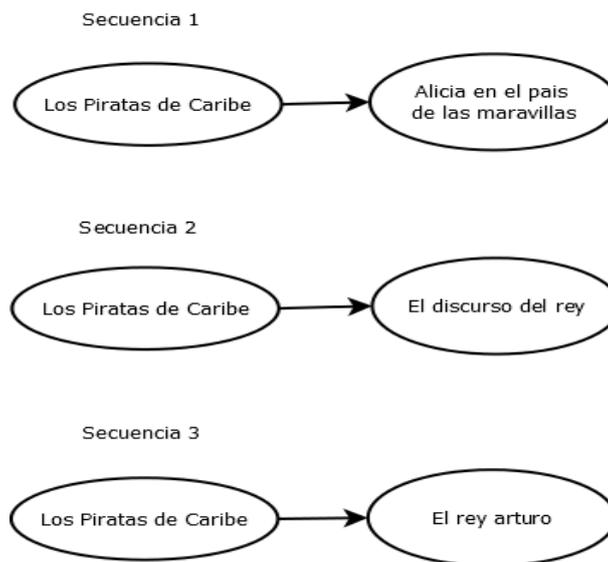


Figura 5. Ejemplo de secuencias de un contenido televisivo.

La adición de recursos a la secuencia terminará cuando ya no haya nuevos conceptos o se llegue a un límite de profundidad pre-establecido, conocido como *path*. Al finalizar el proceso recursivo, se tendrá una secuencia de conexiones similar al ilustrado en la Fig. 5 donde las posiciones impares (1, 3, 5,...,2n+1) siempre se encontrarán recursos de tipo contenido televisivo (películas, novelas, series), mientras que las posiciones pares (2,4,6,...,2n) tendrán recursos de tipo propiedad (Actores, Directores, Escritores).

### Búsqueda de relaciones rho-path

A partir de una cadena de secuencias completa, se busca todas las relaciones *rho-path* existentes; dos recursos semánticos tienen una relación *rho-path* cuando es posible establecer por lo menos un camino que los conecte (Scuoteguazza & Juayek, 2012). Por ejemplo, la película “Los piratas del caribe, la maldición de la perla negra” tiene una relación *rho-path* con la película “Alicia en el país de las maravillas”, mediante el actor “Johnny Deep”. Para encontrar todas las relaciones *rho-path*, se descompone la cadena de secuencias en secuencias individuales. Las nuevas secuencias incluyen únicamente a los recursos de tipo programación televisiva. En la Fig. 6 se observa las secuencias creadas, (relación *rho-path*); en el ejemplo, la longitud de la relación es de dos, pero la longitud y el número de relaciones varían dependiendo de la longitud de las secuencias originales.



**Figura 6.** Relaciones *Rho-Path*.

### Búsqueda de relaciones rho-join

Existe una relación *rho-join* entre dos recursos siempre y cuando los dos sean instancias de la misma clase de unión, por ejemplo, existe una relación *rho-join* de longitud 1 entre todos los recursos que son de tipo película de acción, y una relación *rho-join* de longitud 2 entre todos los contenidos que son de tipo series sin tomar en cuenta su género. Según el modelo de AVATAR, si dos recursos tienen tanto una relación *rho-join* como una *rho-path* se infiere la relación *rho-path* entre los contenidos en cuestión.

- Uniform Resource Identifier Resource Description Framework

### Cálculo del DOI para un recurso de programación televisiva

Una vez obtenidas todas las relaciones *rho-path* y *rho-join*, se valora el contenido televisivo mediante la fórmula 1, que realiza un promedio ponderado del DOI de los recursos, donde los recursos más alejados o

las relaciones más largas son semánticamente menos influyentes que los recursos más próximos al recurso objetivo (Scuoteguazza & Juayek, 2012):

$$DOI_i = \frac{\sum_{i=1}^n DOI(R_i) / length(R_i)}{\sum_{i=1}^n 1 / length(R_i)} \tag{1}$$

donde:

- $n$ : es el número total de relaciones rho-path y rho-join encontradas.
- $DOI(R_i)$ : representa el DOI del usuario sobre el último recurso de una relación.
- $length(R_i)$ : representa la longitud de la relación rho-path o rho-join.

### 3.2. Características generales del sistema

#### Sistema base

Este SR toma como base el trabajo realizado por Juayek & Scuoteguazza (2012), en el cual se evalúan los sistemas de recomendación audiovisuales basados en técnicas inteligentes a través de un SR basado en AVATAR (Yu *et al.*, 2004; Adomavicius & Tuzhilin, 2005). El presente trabajo extiende el mencionado SR modificando su algoritmo núcleo e incorporando una serie de módulos enfocados a contrarrestar los problemas comunes de los SR y a establecer un marco adecuado para su evaluación, con características similares al documentado en Ávila *et al.* (2014a).

#### Modificaciones realizadas

Este sistema mejora el esquema de la base de datos utilizada, particularmente para el manejo de grandes volúmenes de datos. Además, se ha controlado el problema de la disminución del rendimiento asociado al exceso de información presente en ciertos ítems, a través de limitar el número máximo de instancias de una propiedad determinada. Esta limitación se implementa cuando se utiliza la API-OMDB que se invoca para el enriquecimiento de los perfiles. Finalmente, se optimizó el código, mediante funciones genéricas que pueden ser reutilizadas en diferentes contextos.

#### Modelo de programación

El SR ha sido programado modularmente de tal forma que resulte sencillo organizar el código agrupando las funciones y las clases de acuerdo al papel que desempeñan. El código entero está dividido en seis paquetes principales. En la Fig. 7 se muestra la arquitectura del sistema desde la perspectiva de la programación.

Los módulos de *Perfiles de Usuario*, *Perfiles de Estereotipo* y *Ontología de Programación* son los encargados del manejo de los datos semánticos, contienen funciones para la creación, enriquecimiento, lectura y escritura de los perfiles en la ontología.

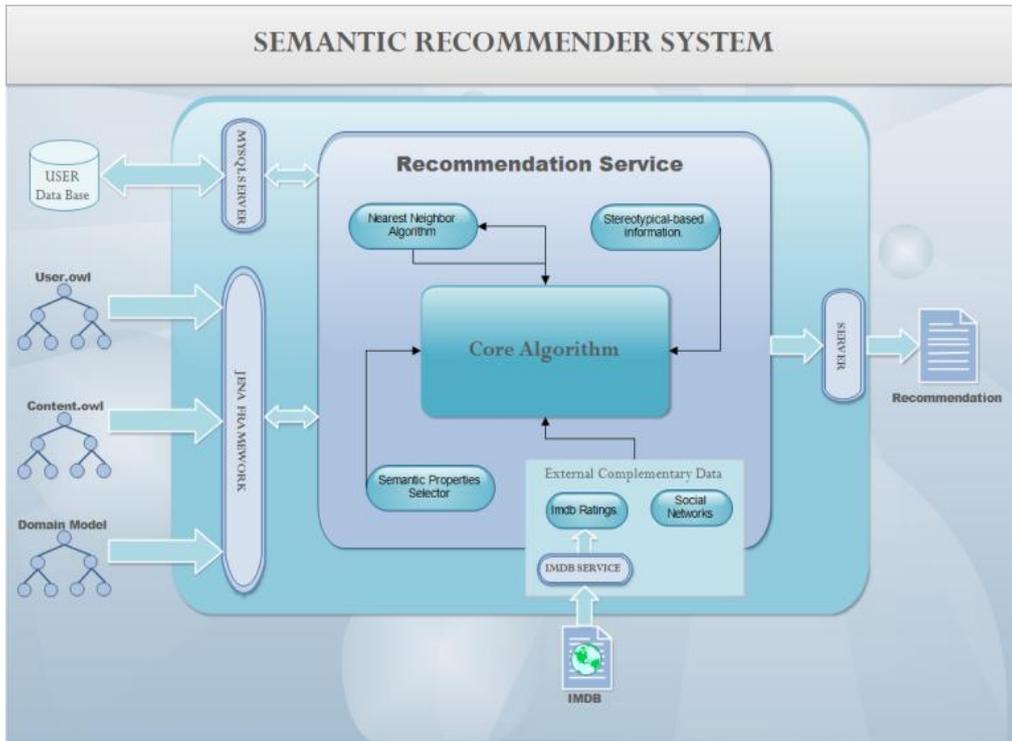


Figura 7. Arquitectura del Sistema desde la perspectiva de la programación.

*Recomendador* contiene los algoritmos necesarios para realizar la recomendación. Específicamente, realiza los procedimientos para evitar la sobrecarga informática y poder así emitir una recomendación destinada a un usuario determinado a través del *Servicio de Recomendación*.

### 3.3. Arquitectura del sistema

El sistema de recomendación semántico propuesto en el proyecto está estructurado según se muestra en la Fig. 8. Este diseño posibilita la generación de una recomendación para un usuario partiendo de un conjunto de ítems dado, y además se incluyen varios módulos adicionales que permiten realizar evaluaciones paramétricas, según se puede profundizar en Ávila *et al.* (2014b).



**Figura 8.** Arquitectura del sistema.

#### Entradas y salida del sistema

El sistema tiene como parámetros de entrada tres ontologías, las cuales son: *User ontology* la cual consta del perfil ontológico de cada usuario, *Content ontology* que corresponde a la programación o contenido y consta de un conjunto de ítems expresados de manera semántica, y además, *Domain ontology* que define la estructura en la que se basan las dos ontologías mencionadas anteriormente. Se cuenta también con una entrada de naturaleza no-semántica: la base de datos de los usuarios en la cual se almacena datos básicos tales como edad, género, etc.

La recomendación final (salida del sistema) consta de una lista ordenada de sugerencias de alternativas de *entretenimiento* para cada usuario. Esta lista presenta desde los ítems con mayor compatibilidad, hasta aquellos menos relacionados a los intereses del usuario.

#### Núcleo del sistema

Este es el bloque central del sistema, aquí se agrupan las clases y funciones principales para vincular los perfiles de usuario, las programaciones televisivas y generar una recomendación final. En el núcleo se dispone de dos algoritmos: *Algoritmo de recomendación semántico por dispersión* y *Algoritmo de Recomendación con Inferencia Semántica* explicados detalladamente en Ávila *et al.* (2014b). En este trabajo se analiza el algoritmo de Inferencia Semántica.

### Módulos complementarios

El sistema incluye cuatro módulos que incorporan funciones e información adicionales al algoritmo núcleo.

- *Stereotypical-based Information*: Módulo para resolver los problemas de *arranque en frío*, descrito en la Sección 2, para lo cual, se agrupa a los usuarios nuevos o con un número reducido de calificaciones de acuerdo a ciertos factores demográficos similares, a manera de “Estereotipos” para los que se genera recomendaciones dirigidas a todos los miembros del grupo (*filtrado pasivo*).
- *Semantic Properties Selector*: Este módulo permita realizar una variación de los algoritmos núcleo para que cualquier algoritmo sea capaz de recibir como parámetros de entrada un conjunto de propiedades semánticas determinado, considerando únicamente las propiedades ingresadas. Este módulo permite analizar el impacto que causa cada propiedad en la estimación de una recomendación.
- *Nearest Neighbor Algorithm*: Este módulo posibilita la generación de una predicción diferente a la del sistema principal y que permita obtener resultados que puedan ser comparados. Su funcionamiento se basa en la utilización de una técnica llamada “K-Vecinos Cercanos” o KNN (Rodríguez *et al.*, 2013), que permite encontrar “usuarios semejantes” y en consecuencia, perfiles de usuarios con gustos similares. El algoritmo se detalla en Ávila *et al.* (2014b).
- *External Complementary: Data* Para la incorporación de este módulo, se parte de la premisa de que una predicción obtenida mediante nuestro SRS se puede enriquecer mediante información externa, no necesariamente de naturaleza semántica. Para ello, se plantea mejorar la exactitud de las predicciones generadas al combinar sus resultados con información externa, la cual está disponible en la web, tal como aquella extraída de IMDB o de redes sociales siguiendo, procedimientos como el propuesto en Espinoza-Mejía *et al.* (2014).

## 4. PROCEDIMIENTO DE EVALUACIÓN DEL SISTEMA Y RESULTADOS

A través de la experimentación se analiza el comportamiento del algoritmo de inferencia semántica según se varían los datos de entrada y ciertos parámetros de ajuste. Este procedimiento se basa en cuatro etapas diferentes:

- Análisis y especificación de resultados a obtener.
- Preparación del algoritmo para cada experimento.
- Elección y construcción del conjunto de datos para la creación de las ontologías y las pruebas.
- Análisis de resultados y conclusiones de cada prueba.

La precisión del algoritmo de recomendación se mide promediando el error absoluto entre la calificación contenida en el conjunto de datos conocidos y la calificación obtenida por el recomendador. Dicho promedio es conocido como Mean Absolute Error MAE y se expresa en la ecuación 2:

$$MAE = \frac{1}{N} \sum_{u,i}^{U,I} |P_{u,i} - R_{u,i}| \quad (2)$$

donde:

- *U*: conjunto de los usuarios examinados.
- *I*: conjunto de ítems evaluados.
- *N*: corresponde al total de *ratings* comparados.
- *P*: es la predicción obtenida por el sistema.
- *R*: es la calificación otorgada por el usuario.

#### 4.1. Conjunto de datos utilizado para las pruebas

Para la evaluación del algoritmo se ha utilizado la base de datos libre desarrollada para el sistema MovieLens (Herlocker, 2014), que se orienta principalmente a proveer información de libre disponibilidad con fines académicos e investigativos. En ella se dispone un conjunto de usuarios y sus calificaciones para diferentes películas. Sus características específicas son:

- 6040 usuarios, con un identificador numérico (id) individual en el rango de 1 al 6040, cada usuario tiene especificado su género, edad y ocupación.
- 3952 películas, con un identificador numérico (id) individual enumerado entre 1 y 3952, de las cuales se además especifica el título y el género.
- La base de datos posee más de un millón de ratings, ya que cada usuario tiene al menos 20 calificaciones realizadas y además, existen usuarios con más de ochocientas calificaciones. Cada usuario califica una película en un escala de entre 1 a 5 estrellas.

#### 4.2. Resultados obtenidos

En Ávila *et al.* (2014a) se hace un análisis de la influencia de las propiedades semánticas en los algoritmos de recomendación utilizando un algoritmo de dispersión, en dicho trabajo se concluye que al incrementar el número de propiedades semánticas el error en la recomendación tiende a reducir en una determinada proporción; a continuación se realiza un análisis similar utilizando el algoritmo de inferencia semántica.

##### Análisis de la influencia de las propiedades semánticas en la estimación de las predicciones utilizando el algoritmo de inferencia semántica

En esta prueba se analiza el comportamiento del error en el algoritmo de inferencia semántica según se incluyan o excluyan propiedades semánticas de un ítem, siguiendo los lineamientos establecidos en Ávila *et al.* (2014a). La hipótesis planteada sugiere que al incrementar el número de propiedades semánticas con las que el algoritmo realiza los cálculos, el error se reducirá en una determinada proporción.

Para este experimento se utilizó la información de usuarios disponible en MovieLens. Específicamente, se ha elegido aleatoriamente un conjunto de 100 usuarios que tengan un rango de entre 200 y 220 ítems calificados, de los cuales, se dividió en un 80 % para la creación de los perfiles de usuario y un 20 % para la realización de las pruebas, garantizando así la independencia de los conjuntos de entrenamiento y prueba y al mismo tiempo la validez del experimento. Además, se utiliza el módulo selector de propiedades semánticas, descrito en Ávila *et al.* (2014b), y se introducen ocho combinaciones diferentes de propiedades semánticas en igual número de pruebas. Se incluyó la propiedad *Género* (Genre) en todas las pruebas, con el objetivo comparar posteriormente los resultados obtenidos con los resultados obtenidos haciendo uso del algoritmo de dispersión documentado en Ávila *et al.* (2014a). El sistema implementado calcula la calificación que un usuario otorgaría a un ítem determinado, a manera de una predicción y con el objetivo de generar las recomendaciones en base a aquellos ítems cuya calificación estimada sea la más alta. Con fines de evaluación, esta predicción es comparada con la calificación real otorgada por el usuario, ya ésta que consta en la información disponible del conjunto de pruebas. Las diferentes combinaciones de propiedades semánticas utilizadas en el algoritmo se muestran en la Fig. 9, que también exhibe los resultados del MAE obtenido para cada una mediante la ecuación 2.

Como se puede observar, la ejecución del algoritmo incluyendo únicamente la propiedad *Género* denota un error promedio de aproximadamente 16.6 % en la predicción. A medida que se incrementa el número de propiedades semánticas el error tiende a reducirse, en una proporción que depende de la propiedad. Se puede observar por ejemplo, que al agregar la propiedad *Actor* a la propiedad *Género*, el error se reduce al 13.86 % lo cual representa una reducción mayor que cuando se evalúa el algoritmo con las propiedades *Género*, *Director* y *Escritor* lo cual produce un error de aproximadamente el 15 %. Se observa también que la propiedad *Director* en todos los casos ofrece una mayor reducción del error que la

propiedad *Escritor*. Al introducir todas las propiedades semánticas disponibles se obtiene el menor promedio de error.

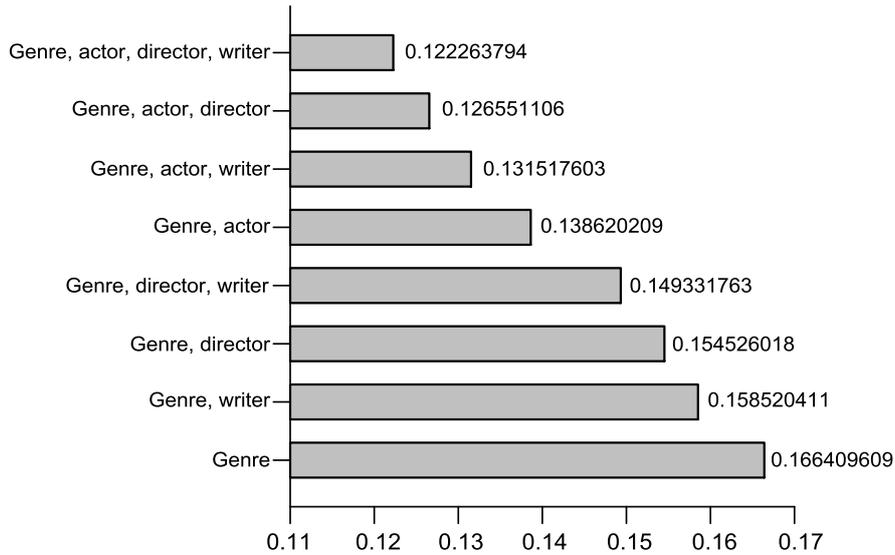


Figura 9. MAE obtenido para cada combinación.

La Fig. 10 muestra el promedio de error de cada usuario involucrado en la prueba. En ella puede notarse el comportamiento del MAE para cada una de las combinaciones de propiedades semánticas utilizadas, las mismas que se distinguen por colores y que se han denotado por sus siglas.

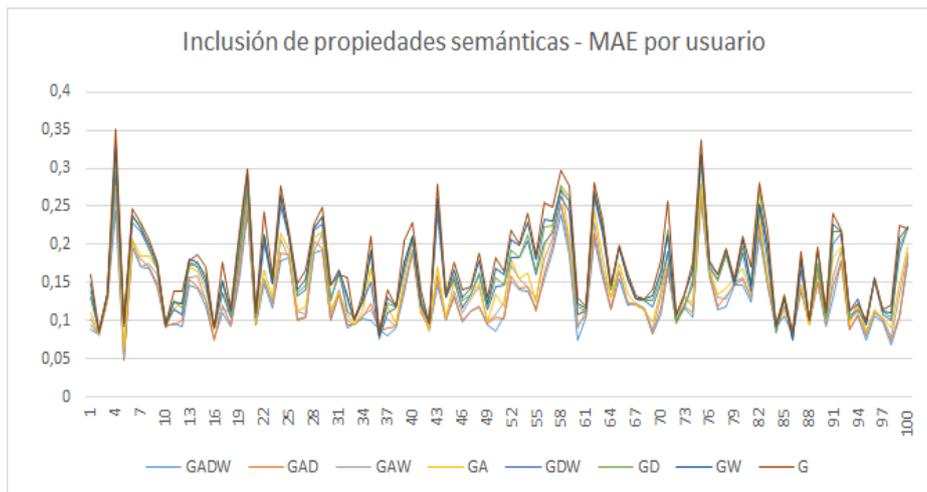


Figura 10. Promedio de error por usuario en las pruebas efectuadas.

## 5. CONCLUSIONES Y TRABAJO A FUTURO

Con los resultados obtenidos de la experimentación se puede confirmar que la inclusión de diferentes propiedades semánticas en el algoritmo de inferencia semántica contribuye en la reducción del error obtenido en la predicción de una calificación, sin embargo, existen ciertas propiedades que tienen una mayor influencia. Por ejemplo, la propiedad *Actor* denota una mayor influencia y se destaca como la propiedad que redujo el error en mayor proporción, en comparación con las propiedades *Director* y

*Escritor* combinados. Esto sugiere que la mayor reducción del error se logra con la adecuada combinación de propiedades más no necesariamente con un mayor número de ellas.

Una vez que se han registrado los resultados del algoritmo de inferencia semántica, éstos pueden compararse con los obtenidos con el algoritmo de dispersión, documentados en Ávila *et al.* (2014a), con el objetivo de determinar el algoritmo de mejor desempeño y que se plantea como trabajo futuro, incluyendo además otras propuestas presentadas en el mismo dominio. Se espera que la incorporación del módulo KNN permita reducir el MAE obtenido, por ejemplo mediante la creación de un sistema híbrido o conmutado en el que intervengan los algoritmos mencionados en este artículo y sobre los cuales se implemente un mecanismo de votación y selección de recomendaciones.

Adicionalmente, la incorporación de i) guías de programación televisiva enriquecidas con el procedimiento propuesto en Saquicela *et al.* (2014), y de ii) un perfil de usuario cuyas preferencias televisivas se obtienen mediante la extracción de información desde redes sociales según la propuesta de Espinoza-Mejía *et al.* (2014), permitirá obtener un sistema de recomendación más robusto y preciso, que posibilite consolidar esta propuesta en el dominio de la televisión digital terrestre y sobre el cual deberá realizarse evaluación adicional para la obtención de nuevos resultados.

## AGRADECIMIENTOS

Este trabajo forma parte del proyecto: “Aplicación de Tecnologías Semánticas para Disminuir la Sobrecarga de Información en Usuarios de TV digital”, auspiciado por la Dirección de Investigación de la Universidad de Cuenca - DIUC.

## REFERENCIAS

- Adomavicius G., A. Tuzhilin, 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6), 734-749.
- Alulema D., 2012. La televisión digital terrestre en el ecuador es interactiva. *EIDOS Universidad Tecnológica Equinoccial*, (5), 12-19.
- Ávila J., X. Riofrío, M. Espinoza-Mejía, V. Saquicela, K. Palacio-Baus, H. Albán, 2014a. Análisis de la influencia de las propiedades semánticas en los sistemas de recomendación. *Revista Politécnica (EPN)*, 34(2), 68-76.
- Ávila J., X. Riofrío, K. Palacio, 2014b. *Sistema de recomendación de contenido para tv digital basado en ontologías*. Tesis pregrado, Escuela de Ingeniería de Sistemas, Facultad de Ingeniería, Universidad de Cuenca, 136 pp.
- Cantador I., A. Bellogín, P. Castells, 2008. A multilayer ontology-based hybrid recommendation model. *AI Communications*, 21(2), 203-210.
- Codina V., L. Ceccaroni, 2010. Taking advantage of semantics in recommendation systems. In: Alquézar, R., A. Moreno, J. Aguilar-Martin (Eds.). IOS Press, *CCIA, Frontiers in Artificial Intelligence and Applications*, 210, 163-172.
- Espinoza-Mejía, M., V. Saquicela, K. Palacio-Baus, H. Albán, 2014. Extracción de preferencias televisivas desde los perfiles de redes sociales. *Revista Politécnica (EPN)*, 34(2), 66-74.
- Fernández, Y.B., J.J. Pazos Arias, M.L. Nores, A.G. Solla, M.R. Cabrer, 2006. Avatar: An improved solution for personalized tv based on semantic inference. *IEEE Transactions on Consumer Electronics*, 52(1), 223-231.
- Herlocker J., J. Konstan, 2014. Movielens dataset. Disponible en <http://files.grouplens.org/datasets/movielens/ml-100k-README.txt>.
- Kantor, P.B., L. Rokach, F. Ricci, B. Shapira, 2011. *Recommender systems handbook*. Springer, 842 pp.

- Lopez-Nores, M., Y. Blanco-Fernandez, J.J. Pazos-Arias, R.P. Díaz-Redondo, 2010. Property-based collaborative filtering: A new paradigm for semantics-based, health-aware recommender systems. *IEEE 5th International Workshop on Semantic Media Adaptation and Personalization (SMAP)*, 98-103 pp.
- Martin, A., 2006. *Older adulthood, education and social change (Australia, New Zealand)*. PhD thesis, ResearchSpace Auckland. Disponible en <https://researchspace.auckland.ac.nz/handle/2292/88>. 323 pp.
- Nakahara, S., S. Moriyama, T. Kuroda, M. Sasaki, S. Yamazaki, O. Yamada, 1996. Efficient use of frequencies in terrestrial isdb system. *IEEE Transactions on Broadcasting*, 42(3), 173-178.
- Rodríguez, J.E.R., E.A.R. Blanco, R.O.F. Camacho, 2013. Clasificación de datos usando el método k-nn. *Vínculos*, 4(1), 4-18.
- Saquicela, V., M. Espinoza-Mejía, K. Palacio-Baus, H. Albán, 2014. *Enriching electronic program guides using semantic technologies and external resources*. Computing Conference (CLEI), XL Latin American, 8 pp.
- Scuoteguazza, A., M. Juayek, 2012. *Evaluación de sistemas recomendadores de contenidos audiovisuales basados en técnicas inteligentes*. Universidad de Montevideo, Montevideo, Uruguay.
- Sotelo, R., M. Juayek, A. Scuoteguazza, 2013. A comparison of audiovisual content recommender systems performance: Collaborative vs. semantic approaches. *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 5 pp.
- Superintendencia de Telecomunicaciones, 2010. Informe para la definición e implementación de la televisión digital terrestre en el Ecuador. Disponible en [http://www.supertel.gob.ec/pdf/publicaciones/informe\\_idt\\_mar26\\_2010.pdf](http://www.supertel.gob.ec/pdf/publicaciones/informe_idt_mar26_2010.pdf), 108 pp.
- Yu, K., A. Schwaighofer, V. Tresp, X. Xu, H.-P. Kriegel, 2004. Probabilistic memory-based collaborative filtering. *IEEE Transactions on Knowledge and Data Engineering*, 16(1), 56-69.