

Sistemas De Recomendaciones: Herramientas para el Filtrado de Información en Internet

Enrique Herrera-Viedma, Carlos Porcel, Lorenzo Hidalgo

¹ Depto. de Ciencias de la Computación e Inteligencia Artificial

Universidad de Granada

e-mails: viedma@decsai.ugr.es, cporcel@invest.ugr.es, lhidalgo@ugr.es

Resumen

Los sistemas de recomendaciones son herramientas que generan recomendaciones sobre un determinado objeto de estudio, a partir de las preferencias y opiniones dadas por los usuarios. El uso de estos sistemas se está poniendo cada vez más de moda en Internet debido a que son muy útiles para evaluar y filtrar la gran cantidad de información disponible en la Web con objeto de asistir a los usuarios en sus procesos de búsqueda y recuperación de información. En este trabajo realizaremos una revisión de las características y aspectos fundamentales relacionados con el diseño, implementación y estructura de los sistemas de recomendaciones analizando distintas propuestas que han ido apareciendo en la literatura al respecto.

1. Introducción

A menudo es necesario seleccionar una entre varias alternativas sin tener un conocimiento exacto de cada una de ellas. En estas situaciones, la decisión final puede depender de las recomendaciones de otras personas [14]. En recuperación de información (RI) los sistemas de recomendaciones (SR) son herramientas cuyo objetivo es asistir a los usuarios en sus procesos de búsqueda de información, ayudando a filtrar los ítems de información recuperados, usando recomendaciones propuestas sobre esos ítems. Dichas recomendaciones se generan a partir de las opiniones proporcionadas por otros usuarios sobre esos ítems en búsquedas previas o bien a partir de las preferencias del usuario objeto de la recomendación (en adelante, *usuario activo*), dando lugar a los dos grandes grupos de SR [19], *los colaborativos* y *los no colaborativos o basados en contenidos*. El uso de estos sistemas se está poniendo cada vez más de moda debido a su utilidad para evaluar y filtrar la gran cantidad de información disponible en la Web para asistir a los usuarios en sus procesos de búsqueda y recuperación de información [9, 11].

Algunos ejemplos reales de SR existentes en Internet son PHOAKS, Referral-Web, Fab, SiteSeer, GroupLens, etc., [1, 14]. En todos ellos se manifiesta un claro problema para representar la subjetividad e imprecisión asociadas típicamente a las opiniones o recomendaciones de los usuarios. La *Teoría de Conjuntos Difusos* constituye un marco de trabajo idóneo para representar la subjetividad e imprecisión a través del modelado

lingüístico y los conjuntos difusos [20]. En [10, 19] podemos encontrar algunos SR basados en lógica difusa.

La idea principal de este trabajo es presentar un estudio sobre los SR para la RI en Internet describiendo los aspectos más significativos de su diseño y problemas fundamentales con los que nos encontramos a la hora de diseñar un sistema de este tipo. Analizaremos los elementos que intervienen en el esquema de funcionamiento de un SR y los usaremos como criterios de clasificación.

El trabajo se estructura en cinco secciones. En la Sección 2 presentamos un estudio de los SR en general, presentando distintas taxonomías de SR según los criterios que tengamos en cuenta. En la Sección 3, nos centramos en aspectos de la evaluación de los SR, analizando distintas métricas que se pueden usar. En la Sección 4 proponemos posibles vías de trabajo futuras, para finalizar con nuestras conclusiones en la última sección.

2. Sistemas de Recomendaciones

En esta sección analizaremos una serie de aspectos y problemas generales que se deben tener en cuenta a la hora de diseñar un SR y mostraremos algunos ejemplos de SR.

2.1 Definición

En los últimos años, la Web ha experimentado un crecimiento enorme tanto en Web sites como en documentos Web, y además, el uso de Internet está cada vez más extendido. Por ello, cada vez se hace más difícil gestionar la excesiva cantidad de información a la que diariamente nos enfrentamos. Este problema se agrava cuando nos disponemos a realizar una búsqueda de información a través de Internet. Los SR se han ido consolidando como potentes herramientas para ayudar a reducir la sobrecarga de información a la que nos enfrentamos en los procesos de búsqueda de información. Ayudan a filtrar los ítems de información recuperados, usando distintas técnicas para identificar aquellos ítems que mejor satisfacen las preferencias o necesidades de los usuarios. Las recomendaciones se generan a partir de las opiniones proporcionadas por otros usuarios sobre esos ítems en búsquedas previas, o bien a partir del perfil del usuario [11].

Un SR está asociado con un conjunto de ítems $I = \{i_1, \dots, i_n\}$ y su objetivo es recomendar a los usuarios ítems de I que les puedan ser de interés [10, 19]. Por ejemplo, se podría diseñar un SR para la recomendación de películas; de hecho a lo largo de la literatura hemos encontrado numerosos ejemplos de SR de películas, tales como Film-Conseil, MovieFinder, Reel o MetaLens entre otros. En el caso de un SR para la recuperación de documentos los ítems serían los documentos almacenados.

La implementación de técnicas para el desarrollo de los SR está íntimamente relacionada con el tipo de información que se vaya a utilizar. Una primera fuente de información a tener en cuenta es el tipo de ítems con los que vamos a trabajar. Habrá situaciones en las que únicamente conozcamos un identificador de cada ítem. Por ejemplo en el caso de la recomendación de películas solemos conocer únicamente el título. En otras situaciones, dispondremos de más información sobre los ítems, a través de una serie de atributos. En el caso de la recomendación de películas, podrían ser el año en que se hizo la película, el género, el director, protagonistas, etc. En el caso de recuperación de documentos, la información con la que contamos serían los términos índice usados en su representación. En general, cuanto más sofisticada es la representación de los ítems mejor se puede desarrollar la actividad de los SR.

En cualquier caso, algunos aspectos que debemos considerar sobre las recomendaciones en el diseño de SR son [14]:

- Representación de las recomendaciones: Los contenidos de una evaluación pueden venir dados por un único bit (recomendado o no) o por comentarios de texto sin estructurar.
- Expresión de las recomendaciones: Las recomendaciones pueden ser introducidas de forma explícita o bien de forma implícita.
- Aspectos de identificación de la fuente: Las recomendaciones pueden realizarse de forma anónima, identificando la fuente, o bien usando un pseudónimo.
- Forma de agregar las evaluaciones: Se refiere a cómo vamos a ir agregando las evaluaciones disponibles sobre los ítems de cara a generar las recomendaciones. Más adelante entraremos en más detalle sobre este aspecto que es de suma importancia de cara a clasificar los SR.
- Uso de las recomendaciones: Las recomendaciones se pueden usar de distintas formas. Por ejemplo, se podrían mostrar los ítems en forma de lista ordenada según las recomendaciones de cada uno, o a la hora de visualizar los ítems que se muestre también su recomendación.

2.2 Estructura de los SR

A continuación presentamos los elementos fundamentales que intervienen en el esquema de funcionamiento de un SR. Dichos elementos los podemos usar como criterios de clasificación y son los siguientes:

- *las entradas / salidas del proceso de generación de la recomendación,*
- *el método usado para generar las recomendaciones, y*
- *el grado de personalización.*

2.2.1 Entradas / Salidas

Los SR para generar recomendaciones, usan las entradas del usuario activo, pero también información sobre los ítems o información del resto de usuarios del sistema, que actúan como colaboradores. En este sentido, la realimentación por parte de los usuarios es muy importante de cara a albergar una información más completa ante futuros procesos de generación de recomendaciones. La figura 2.1 refleja el proceso de generación de las recomendaciones

Para poder realizar una recomendación a un usuario, es necesario conocer algún tipo de información sobre sus preferencias. Además, dependiendo del tipo de sistema también necesitaremos información sobre los ítems a recomendar o información reunida sobre el resto de usuarios del sistema (comunidad de usuarios o colaboradores). Esta información que necesitamos para realizar las recomendaciones constituye la entrada o entradas del sistema.

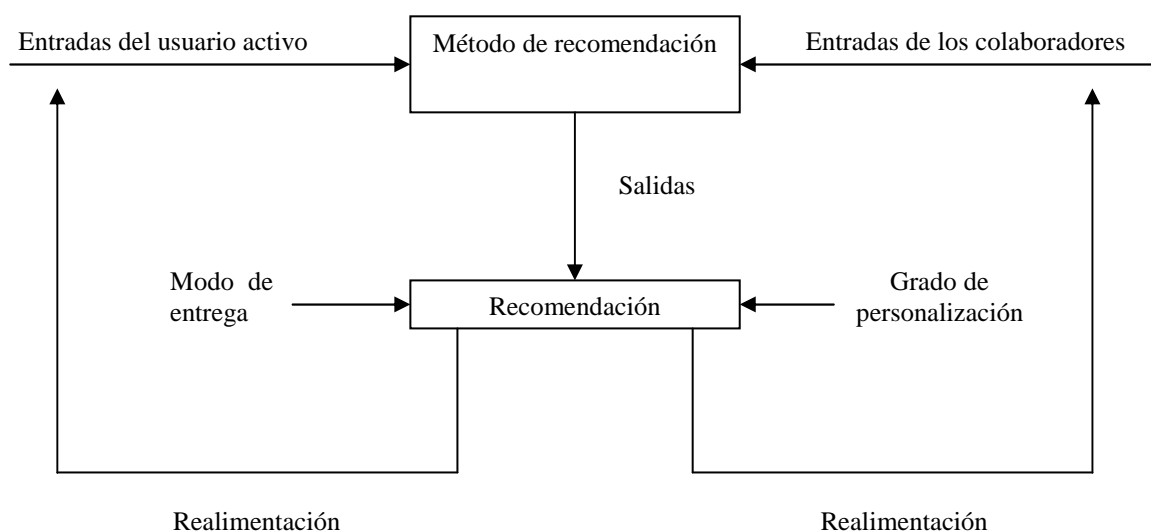


Figura 2.1. Esquema del proceso de generación de una recomendación

La información sobre los usuarios puede venir dada de dos formas [19] que no tienen por qué ser mutuamente exclusivas: **por extensión o intencionalmente**. Por **extensión** se refiere a información que se tenga sobre las experiencias pasadas del usuario con respecto a los ítems encontrados. Es lo que también conocemos como **navegación implícita** pues el usuario no es consciente de estos seguimientos. Por información expresada **intencionalmente** se entiende alguna especificación de los ítems deseados por los usuarios. También se le llama **navegación explícita** y consiste en que el usuario expresa intencionalmente (de forma explícita) al SR información sobre sus preferencias.

La **salida del sistema** está constituida por las recomendaciones generadas por el sistema, que variarán dependiendo del tipo, cantidad y formato de la información proporcionada al usuario. Algunas de las formas más comunes de representar la salida son las siguientes:

- **Sugerencia o lista de sugerencias** al usuario de una serie de ítems.
- Presentar al usuario **predicciones** del grado de satisfacción que se asignará al ítem concreto. Estas estimaciones pueden ser presentadas como personalizadas al usuario o como estimaciones generales del conjunto de colaboradores.
- Cuando la comunidad de usuarios es pequeña o se conocen bien los miembros de dicha comunidad, podría ser útil **visualizar las valoraciones individuales** de los miembros que permitiría al usuario activo obtener sus propias conclusiones sobre la efectividad de una recomendación.

Independientemente de estos formatos de salida, puede resultar muy interesante incluir una breve **descripción o explicación** sobre el ítem recomendado a modo de justificación del porqué de dicha recomendación.

2.2.2 Método de generación de recomendaciones

En esta sección describimos una serie de métodos de recomendación que se usan habitualmente en los SR, pero debemos tener en cuenta que no son mutuamente exclusivos entre sí, sino complementarios, es decir, que en un mismo SR podríamos usar uno o varios de estos métodos. Enunciamos en principio los tres métodos más simples:

- i. **Recuperación pura o recomendación nula**, en la que el sistema ofrece a los usuarios una interfaz de búsqueda a través de la cual pueden realizar consultas a una base de datos de ítems. Se trata, pues, de un sistema de búsqueda por lo que técnicamente no es un método de recomendación, aunque ante los usuarios aparece como tal.
- ii. Otros sistemas usan **recomendaciones seleccionadas manualmente** por expertos, como por ejemplo editores, artistas o críticos en el caso de recomendaciones de películas o cd's de música. Los expertos identifican ítems basándose en sus propias preferencias, intereses u objetivos, y crean una lista de ítems que esté disponible para todos los usuarios del sistema. A menudo acompañan estas recomendaciones de comentarios de texto que puedan ayudar a los usuarios a evaluar y entender la recomendación.
- iii. En otros casos, los sistemas ofrecen **resúmenes estadísticos** calculados en función de las opiniones del conjunto de usuarios, por lo que tampoco son personalizados. Por ejemplo, se podrían tener en cuenta el porcentaje de usuarios

a los que ha satisfecho o han comprado un artículo, número de usuarios que recomiendan un ítem, o una evaluación media de todos los usuarios con respecto al ítem.

Estos métodos de generación de recomendaciones, por su simplicidad no son considerados propiamente métodos de generar recomendaciones en la literatura. Para generar las recomendaciones hay dos posibilidades comúnmente aceptadas que dan lugar a dos grandes grupos de sistemas de SRs [10, 19]: los SR colaborativos y los no colaborativos o basados en contenidos (también conocidos como reclusivos).

- Los SR *no colaborativos* realizan las recomendaciones usando únicamente las preferencias del usuario activo y los atributos de los ítems a recomendar. Estos sistemas usan correlaciones entre ítems para identificar ítems asociados frecuentemente a un ítem por el que el usuario ha mostrado interés y por tanto recomendarle dichos ítems al usuario. Como ejemplo de esta idea, supongamos que tenemos un SR de libros y tenemos dos libros “La clave está en Rebeca” y “La isla de las tormentas”, ambos del mismo autor Ken Follet, pero que además ambos son de intriga y están ambientados en la 2ª Guerra Mundial; por tanto, se podrían considerar en cierto sentido similares, es decir, existe una alta correlación entre ambos. Por esa razón, si un usuario de nuestro sistema está interesado en “La clave está en Rebeca”, podemos recomendarle también la lectura de “La isla de las tormentas”.
- Por otro lado, la mayoría de SR existentes son *colaborativos* [1, 14]. Un SR se dice colaborativo si usa la información conocida sobre las preferencias de otros usuarios para realizar la recomendación al usuario que la precise. Los SR colaborativos identifican usuarios cuyas preferencias sean similares a las de otros usuarios dados y recomiendan a los primeros los elementos que hayan satisfecho a los otros; de esta forma, si dos usuarios U_1 y U_2 , comparten el mismo sistema de valores (tienen las mismas preferencias) y al usuario U_1 le ha satisfecho un ítem i , probablemente este ítem también satisfaga al usuario U_2 por lo que deberíamos recomendárselo. Por ello, en estos SR la definición de medidas de similitud entre preferencias es un punto crítico. La situación puede ser representada como una matriz de usuarios e ítems, donde cada celda representa la valoración de un usuario con respecto a un ítem concreto. Así visto, el problema consiste en predecir valores para las celdas que estén vacías. En la tabla 2.1 podemos ver un ejemplo de matriz que representa valoraciones de usuarios con respecto a una serie de películas.

	Gladiator	Leyendas de pasión	Spiderman	Braveheart
Daniel	5	2	4	5
María	2	5		3
Luis	3	2	4	2
Susana	?	1	4	5

Tabla 2.1. Ejemplo de matriz de valoraciones

Los algoritmos que se suelen usar para implementar las técnicas de filtrado colaborativo se llaman **métodos basados en vecindad**. Funcionan seleccionando un conjunto apropiado de usuarios, según la similitud de los mismos con respecto al usuario activo, y usan las valoraciones de dichos usuarios para generar la valoración del usuario activo. Concretamente, los tres pasos a seguir para realizar esto son los siguientes:

- i. Medir la similitud de todos los usuarios con respecto al usuario activo.
- ii. Seleccionar un subconjunto de usuarios cuyas valoraciones se van a usar y por tanto, tendrán influencia en la generación de la predicción para el usuario activo.
- iii. Normalizar las puntuaciones de los distintos usuarios y calcular una predicción a partir de algún tipo de combinación pesada de las puntuaciones asignadas al ítem por los usuarios seleccionados en el paso anterior.

Como ejemplo, consideremos de nuevo la tabla 2.1; podemos predecir que a Susana le gustará la película “Gladiator”. Observamos que Daniel es el que tiene preferencias más parecidas a Susana, puesto que ambos tienen unas valoraciones muy similares de las películas que ya han visto. Por tanto la valoración de Daniel sobre la película “Gladiator” tendrá gran influencia en la predicción que hagamos a Susana sobre dicha película. Por el contrario, María y Luis tienen opiniones más dispares con respecto a Susana, por lo que tendrán una influencia mucho menor en las recomendaciones que se hagan a dicho usuario.

Los SR colaborativos tienen importantes ventajas con respecto a los no colaborativos:

- i. Dan un mayor soporte para el filtrado de ítems cuyo contenido no es fácil de analizar por procesos automatizados,
- ii. la posibilidad de filtrar ítems basándose en su calidad o preferencias, y
- iii. la posibilidad de realizar recomendaciones válidas, pero que no esperábamos, lo cual puede resultar de gran utilidad.

Sin embargo, también presentan inconvenientes, como por ejemplo que no trabajan bien a la hora de filtrar información para necesidades de contenido específicas, o cuando el número de usuarios es bajo. Por ello, en muchas ocasiones la mejor opción es adoptar un **enfoque híbrido** entre colaborativo y no colaborativo y de esta forma disfrutar de las ventajas de ambos.

2.2.3 Grado de personalización

Según su grado de personalización, también podremos clasificar los SR:

1. Cuando los SR proporcionan las mismas recomendaciones a todos los usuarios, son clasificados en este ámbito como **no personalizados**. Dichas recomendaciones estarán basadas en selecciones manuales, resúmenes estadísticos u otras técnicas similares.
2. Los SR que tienen en cuenta la información actual del usuario objeto de las recomendaciones, proporcionan **personalización efímera**, puesto que las recomendaciones son repuesta al comportamiento y acciones del usuario en su sesión actual de navegación.
3. Los SR que ofrecen el mayor grado de personalización son los que usan **personalización persistente** ofreciendo recomendaciones distintas para distintos usuarios, incluso cuando estén buscando el mismo ítem. Estos sistemas están basados en el perfil de los usuarios, por lo que hacen uso de métodos de filtrado colaborativo, filtrado basado en contenidos o correlaciones entre ítems.

2.3 Problemas asociados a los SR

En los SR aparecen una serie de problemas que habrá que considerar en su diseño [1, 14] y que describimos a continuación:

- Primero, una vez que se ha establecido un perfil de intereses, es fácil considerar las evaluaciones suministradas por otros. Sin embargo, en algunos casos se hace necesario recurrir a incentivos para la provisión de recomendaciones, puesto que los usuarios no suelen estar muy dispuestos a colaborar proporcionando información personal sobre sus preferencias. Estos incentivos podrían consistir en forzar al usuario a introducir sus preferencias a cambio de recibir recomendaciones, o bien asignarle otro tipo de compensaciones.
- Un segundo problema a solucionar es que si cualquiera puede recomendar, los propietarios de determinados productos podrían generar recomendaciones positivas de los mismos y negativas de otros. En los SR también habrá que tener en cuenta aspectos de privacidad y debido a que algunas personas no quieren que se conozcan sus hábitos o preferencias, se permite la participación anónima o bajo un pseudónimo.
- Otro problema es que el mantenimiento de un SR es costoso, por lo que se hace necesario pensar en modelos de negocio que se podrían usar para generar ingresos suficientes para cubrir dichos costes. Un modelo posible es que los receptores de las recomendaciones paguen una suscripción, es decir, pagar por usar. Un segundo modelo podría ser la inclusión de publicidad. Un tercer

modelo es cobrar una cuota a los propietarios de los elementos que deseen sean evaluados. Sin embargo, los dos últimos modelos pueden presentar problemas de corrupción, en el caso de importantes empresas con un gran peso publicitario.

- Un último problema es dotar a los SR de mejores técnicas de representación de las preferencias o recomendaciones de los usuarios que nos permitan captar verdaderamente su concepto del objeto recomendado y mejorar la interacción SR-usuario.

2.4 Ejemplos de SR

En esta sección vamos a analizar algunos ejemplos de SR existentes en Internet [14] que han servido de base para numerosos estudios:

- *PHOAKS*: Se trata de un sistema experimental para solucionar el problema de encontrar información relevante y de alta calidad en la Web, usando el enfoque colaborativo en el que los usuarios recomiendan determinados ítems a otros usuarios. PHOAKS trabaja reconociendo, concordando y redistribuyendo automáticamente recomendaciones de recursos Web extraídos de mensajes de noticias.
- *Referral Web*: Numerosos estudios muestran que una de las formas más efectivas de divulgar información y conocimientos dentro de una determinada organización es a través de una red informal de colaboradores o amigos. Referral Web se basa en la idea de combinar “redes sociales” con el filtrado colaborativo, entendiendo por “redes sociales” grupos de personas vinculadas por determinadas actividades profesionales.
- *FAB*: Sistema orientado a la recomendación de URL que combina el uso de información por extensión con el enfoque colaborativo.
- *Siteseer*: Recomienda páginas Web relevantes y usa las listas de favoritos y la organización de registros, como una declaración implícita de intereses respecto al contenido subyacente y se va midiendo el grado de solapamiento con las de otros usuarios.
- *GroupLens*: El proyecto GroupLens diseña, implementa y evalúa un sistema de filtrado colaborativo para Usenet, un servicio de listas de discusión con un alto volumen de negocio en Internet.

Anteriormente, en el apartado 2.1 definíamos cinco aspectos o dimensiones a considerar a la hora de diseñar un SR. Pues bien, en la tabla 2.2 mostramos cómo se clasifican con respecto a dichos aspectos los sistemas que hemos comentado como ejemplo.

	Contenidos de la recomendación	Tipo de entrada	Identificación de la fuente	Modo de agregación	Uso de las recomendaciones
GroupLens	Numérico: 1-5	Explícita	Seudónima	Suma pesada basada en acuerdos anteriores de los recomendadores	Visualización junto a los artículos en las vistas de resúmenes
Fab	Numérico: 1-7	Explícita	Seudónima	Suma pesada junto con análisis de contenidos	Selección / Filtrado
ReferralWeb	Recomendación de una persona o documento	Extraída de datos públicos	Atribuida	Reunir cadenas referidas a la persona deseada	Visualización
PHOAKS	Recomendación de una URL	Extraída de envíos Usenet	Atribuida	Un voto por persona (por URL)	Visualización ordenada
SiteSeer	Recomendación de una URL	Extraída de listas de favoritos	Anónima	Frecuencia de mención de la URL	Visualización

Tabla 2.2. Caracterización de algunos SR

3 Evaluación de los SR: Métricas

A pesar de que este tipo de sistemas están tan extendidos, aún no hay estándares definidos para su evaluación, sino que cada grupo de investigación ha ido aplicando diferentes técnicas y métricas de evaluación, sin que en muchos casos existan relaciones entre ellas. Esta diversidad de métricas conlleva tres problemas fundamentales [5]:

- i. Si dos investigadores distintos, evalúan sus sistemas con distintas métricas, los resultados no son comparables.
- ii. Si la métrica usada no está estandarizada, podemos pensar que los investigadores han elegido la métrica más adecuada de cara a obtener los resultados deseados.
- iii. Sin una métrica estandarizada, cada investigador debe realizar un esfuerzo extra en identificar o desarrollar una métrica apropiada.

Para medir el rendimiento de un SR colaborativo, hay tres pasos fundamentales que son identificar a alto nivel los objetivos del sistema, identificar las tareas específicas que permiten alcanzar esos objetivos e identificar las métricas a nivel de sistema. La evaluación a nivel del sistema se realizará en casos en los que los investigadores identifiquen indicadores que se puedan medir y que evidencien correlación con la efectividad del sistema independientemente de la interacción de los usuarios. Este tipo

de evaluación es la más usada en Filtrado de Información, porque ofrece un análisis económico y fácilmente repetible, donde los datos de los usuarios son recogidos una sola vez.

Cleverdon [3, 5] identifica cinco **métricas** a tener en cuenta:

- **Retardo.** Intervalo de tiempo transcurrido desde que se hace la demanda hasta que se da la respuesta.
- **Presentación.** El formato físico de la salida del sistema.
- **Esfuerzo del usuario.** El esfuerzo, intelectual o físico que se demanda del usuario.
- **Exhaustividad.** Capacidad del sistema de presentar todos los ítems relevantes.
- **Precisión.** Capacidad del sistema de ocultar ítems que no sean relevantes.

De estas métricas, podemos destacar las dos últimas. Se calculan a partir de una tabla de contingencia que clasifica los ítems con respecto a las necesidades de información distinguiendo dos grupos: *relevantes* o *no relevantes*. Además, también se clasifican los ítems según se hayan recomendado al usuario (*seleccionados*) o no (*no seleccionados*). Con estas cuatro categorías, construimos la tabla de contingencia 2.3.

	Seleccionados	No Seleccionados	Total
Relevantes	N_{rs}	N_{rn}	N_r
Irrelevantes	N_{is}	N_{in}	N_i
Total	N_s	N_n	N

Tabla 2.3. Tabla de contingencia

La *precisión* (P) se define como la proporción de ítems relevantes seleccionados (N_{rs}) con respecto al total de ítems seleccionados (N_s), es decir, mide la probabilidad de que un ítem seleccionado sea relevante:

$$P = \frac{N_{rs}}{N_s}$$

La *exhaustividad* (R) se define como la proporción de ítems relevantes seleccionados (N_{rs}) con respecto al total de ítems relevantes existentes (N_r), es decir, representa la probabilidad de que un documento relevante sea seleccionado.

$$R = \frac{N_{rs}}{N_r}$$

Otro aspecto importante a tener en cuenta en la evaluación de un sistema es su *cobertura*, que indica el porcentaje de ítems para los que el sistema podría generar una recomendación. Es una métrica importante puesto que en ocasiones los SR no son capaces de generar recomendaciones para determinados ítems debido a la ausencia de datos u otro tipo de restricciones.

4 Futuras líneas de trabajo

En esta sección destacaremos algunas de las líneas que consideramos más interesantes de cara a nuestros futuros trabajos en el desarrollo de SR para la recuperación de información:

1. Una posible línea de investigación estaría relacionada con estudiar aspectos relacionados con la credibilidad de las recomendaciones en los SR colaborativos, para considerar o no la validez de dichas recomendaciones.
2. Una segunda línea de investigación sería definir nuevos métodos de evaluación de la calidad informativa de los documentos y de los Web sites que los almacenan para conseguir generar recomendaciones más consistentes y completas.
3. El estudio del modelado lingüístico difuso de los SR tanto para representar los ítems, como la información sobre las preferencias de los usuarios, así como las recomendaciones finales que se obtengan de modo que la información se represente de una forma más natural y así mejorar la interacción SR-usuario.
4. Por último, una línea especialmente interesante, es el diseño de SR para aplicaciones concretas en la Web. En particular nos interesa el diseño de SR para recomendar los documentos y los Web sites en función de su calidad informativa.

5 Conclusiones

En este trabajo hemos realizado un análisis del paradigma de los SR, como herramientas de gran utilidad de cara a ayudar a los usuarios en sus procesos de búsqueda y recuperación de información. Según el proceso de filtrado en que se basa la generación de recomendaciones, hemos distinguido entre dos grandes grupos de SR, los colaborativos y no colaborativos, llegando a la conclusión de que en muchas ocasiones, la mejor opción es adoptar un enfoque híbrido y así aprovechar las ventajas de ambos. Hemos completado nuestro estudio presentando algunas de nuestras líneas futuras de trabajo para el desarrollo de SR en RI.

Queremos destacar que el desarrollo de SR lingüísticos y de SR para aplicaciones concretas en la Web son dos importante líneas de trabajo futuras, ya que como comprobamos diariamente la incorporación de nuevos usuarios a la Web, la creación de nuevos sitios Web, la creación de nuevos documentos Web, o el comercio electrónico,

son hechos o actividades en constante expansión, y por tanto, contar con sistemas que ayuden a los usuarios a buscar información y tomar decisiones en este entorno es un tema de crucial importancia.

Referencias

- [1] G. Bafoutsou, G. Mentzas. *Review and Functional Classification of Collaborative Systems*. International Journal of Information Management, 22 (2002), 281-305.
- [2] K.D. Bollacker, S. Lawrence, C.L. Giles. *Discovering Relevant Scientific Literature on the Web*. IEEE Intelligent Systems, volume 15, number 2 (2000), month March/April, 42-47.
- [3] C. Cleverdon, M. Keen. *Factors Determining the Performance of Indexing Systems*, Vol. 2--Test Results. ASLIB Cranfield Res. Proj., Cranfield, Bedford, England, 1966.
- [4] N. Good, J.B. Schafer, J.A. Konstan, A. Borchers, B. Sarwar, J. Herlocker, J. Riedl. *Combining Collaborative Filtering with Personal Agents for Better Recommendations*. Proceedings of the 1999 National Conference of the American Association of Artificial intelligence, 439-446.
- [5] J.L. Herlocker. *Understanding and Improving Automated Collaborative Filtering Systems*. A thesis submitted to the faculty of the graduate school of the University of Minnesota by Jonathan Lee Herlocker.
- [6] J.L. Herlocker, J.A. Konstan, A. Borchers, J. Riedl. *An Algorithmic Framework for Performing Collaborative Filtering*. In Proceedings of the 22nd annual international ACM SIGIR conference on Research and Development in Information Retrieval (1999), 230-237.
- [7] J.L. Herlocker, J.A. Konstan, J. Riedl. *An Empirical Analysis of Design Choices in Neighborhood-Based Collaborative Filtering Algorithms*. Information Retrieval, 5 (2002), 287-310.
- [8] J.L. Herlocker, J.A. Konstan, J. Riedl. *Explaining Collaborative Filtering Recommendations*. ACM 2000 Conference on Computer-Supported Collaborative Work.
- [9] E. Herrera-Viedma, E. Peis. *Evaluating the Informative Quality of Documents in SGML Format Using Fuzzy Linguistic Techniques Based on Computing with Words*. Information Processing & Management, 39(2) (2003), 233-249.

- [10] E. Herrera-Viedma, F. Herrera, L. Martínez, J.C. Herrera, A.G. López. *Incorporating Filtering Techniques in a Fuzzy Linguistic Multi-Agent Model for Information Gathering on the Web*. Fuzzy Sets and Systems. Por aparecer.
- [11] E. Herrera-Viedma, L. Olvera, E. Peis, C. Porcel. *Revisión de los sistemas de recomendaciones para la recuperación de información*. Tendencias de investigación en organización del conocimiento (Trends in knowledge organization research), José Antonio Frías, Ed. Crispulo Travieso, Universidad de Salamanca, (2003), 507-513.
- [12] P. Perny, J.D. Zucker. *Collaborative Filtering Methods based on Fuzzy Preference Relations*. Proceedings of the conference EUROFUSE-SIC'99 (1999), Budapest, 25-28.
- [13] P. Perny, J.D. Zucker. *Preference-based Search and Machine Learning for Collaborative Filtering: the Film Conseil Movie Recommender System*. Information – Interaction – Intelligence, 13 (2001), Vol. 1, N. 1.
- [14] P. Resnick, H.R. Varian, Guest Editors. *Recommender Systems*. Communications of the ACM, 40 (3) (1997), 56-89.
- [15] J.B. Schafer. *MetaLens: A Framework for Multi-source Recommendations*. A thesis submitted to the faculty of the graduate school of the University of Minnesota by John Benjamin Schafer.
- [16] J.B. Schafer, J.A. Konstan, J. Riedl. *Recommender Systems in E-Commerce*. ACM Conference on Electronic Commerce (EC-99), November 3-5 (1999), Denver, CO.
- [17] J.B. Schafer, J.A. Konstan, J. Riedl. *MetaRecommendation Systems: User-controlled Integration of Diverse Recommendations*. ACM Conference on Information and Knowledge Management (CIKM-02), November 5-7, 2002, McClean, VA.
- [18] K. Swearingen, R. Sinha. *Beyond Algorithms: An HCI Perspective on Recommender Systems*. Recommender Systems. Papers and Notes from the 2001 Workshop New Orleans, LA. In conjunction with ACM SIGIR Conference on Research and Development in Information Retrieval.
- [19] R.R. Yager. *Fuzzy Logic Methods in Recommender Systems*. Fuzzy Sets and Systems, Volume 136, Issue 2 (2003), 133-149.
- [20] L.A. Zadeh. *The concept of a linguistic variable and its applications to approximate reasoning*. Part I, Information Sciences, 8 (1975) 199-249. Part II, Information Sciences, 8 (1975) 301-357. Part III, Information Sciences, 9 (1975) 43-80.