

# Desarrollo de una colección de evaluación para personalización de periódicos digitales

## **Díaz Esteban, Alberto**

Centro de Estudios Superiores Felipe II  
Universidad Complutense de Madrid  
C/ Capitán 39, Aranjuez, Madrid 28300  
[adiaz@cesfelipesecondo.com](mailto:adiaz@cesfelipesecondo.com)

## **Gervás Gómez Navarro, Pablo**

Departamento de Sistemas Informáticos y Programación  
Facultad de Informática - Universidad Complutense de Madrid  
c/ Juan del Rosal, 8, Madrid 28040  
[pgevas@sip.ucm.es](mailto:pgevas@sip.ucm.es)

## **García Jiménez, Antonio**

Departamento de Comunicación y Ciencias Sociales  
Facultad de Ciencias de la Comunicación y del Turismo - Universidad Rey Juan Carlos  
Camino del Molino, s/n Fuenlabrada, Madrid 28943  
[an.garcia@cccom.urjc.es](mailto:an.garcia@cccom.urjc.es)

Los sistemas de personalización de contenidos web seleccionan, adaptan y generan contenidos de acuerdo a modelos de usuario que definen necesidades de información. En nuestro caso, el modelo de usuario utilizado está basado en intereses sobre secciones y términos que puedan aparecer en noticias de un periódico electrónico. Estos modelos incluyen también información sobre la realimentación del usuario. En este trabajo se presenta el proceso de construcción y las características de una colección de evaluación para este tipo de sistemas de personalización. Finalmente, el uso de medidas adecuadas que permitan establecer comparaciones entre distintos sistemas es discutido desde el punto de vista de la efectividad del sistema y desde el punto de vista del usuario.

**Palabras clave:** evaluación, colección, personalización, periódicos digitales, medidas

## **1. INTRODUCCIÓN**

Un sistema de personalización está basado en 3 funcionalidades principales: selección de contenidos, adaptación del modelo de usuario y generación de contenidos. Para que estas funcionalidades se realicen de manera personalizada deben estar basadas en información relacionada con el usuario que debe estar reflejada en su perfil o modelo de usuario [10].

La selección de contenidos obtiene los documentos más relevantes asociados a cada modelo de usuario. La adaptación de contenidos obtiene la representación de la realimentación mediante un algoritmo que utiliza los documentos realimentados explícitamente por el usuario. La generación de contenidos consiste en, una vez seleccionados los documentos que le interesan a un usuario,

generar una página web que contenga la información asociada a esos documentos de manera personalizada.

El modelo de usuario propuesto está basado en una combinación de intereses reflejados como categorías y términos. Puesto que el dominio elegido es el de los periódicos digitales las categorías son representadas por las secciones del periódico. Esta elección de modelo de usuario está basada en un estudio realizado sobre una serie de periódicos electrónicos [4]. Otro aspecto fundamental que debe ser almacenado en el modelo de usuario es una representación de la realimentación que va realizando cada usuario sobre el sistema. Este aspecto es el que permite al sistema irse adaptando a los intereses del usuario.

Las colecciones de evaluación son fundamentales para poder probar la efectividad de cualquier sistema, además aportan un marco

comparativo para poder establecer comparaciones entre distintas propuestas que solucionen el mismo problema.

En este trabajo, presentamos la construcción de una colección de evaluación para la personalización de contenidos web, así como las medidas que se pueden utilizar para establecer la efectividad del sistema y la satisfacción de los usuarios cuando se utilizan distintas propuestas de personalización.

## **2. COLECCIONES DE EVALUACIÓN**

En general, una colección de evaluación está compuesta por un conjunto de documentos con una estructura más o menos similar, que suelen tratar de contenidos sobre un mismo dominio, (p.ej: económico, periodístico, etc.), de un conjunto de tareas a realizar, y de los resultados asociados determinados manualmente por expertos humanos. P.ej: en recuperación de información, las tareas serían las consultas y los resultados serían los juicios de relevancia asociados a cada consulta.

Es importante la existencia de una colección de evaluación para poder evaluar un sistema de manera sistemática, más allá de mostrar el funcionamiento para unos determinados casos concretos que no muestran la efectividad real. Además, aporta un marco sobre el cual poder efectuar comparaciones entre distintas propuestas.

Existen colecciones de evaluación para distintas tareas relacionadas con la recuperación de información, algunas de ellas han surgido en el contexto de congresos tipo competición en los cuales se les suministra a los distintos participantes una misma colección y cada uno de ellos tratan de acercarse lo más posible a los resultados correctos. Quizá el contexto más conocido son las conferencias TREC [13] sobre recuperación de información. Otro tipo de colecciones han surgido en otros contextos diferentes, normalmente como necesidad para la evaluación de algún nuevo tipo de sistema.

Sin embargo, hay que hacer un esfuerzo a la hora de realizar estas colecciones para que sean utilizables, en el sentido de que la información que contengan sea suficientemente representativa para el caso de estudio, y reutilizables, es decir, que puedan ser utilizadas por personas diferentes a sus desarrolladores sin necesitar un gran esfuerzo de adaptación.

La utilización de lenguajes de marcado (SGML, HTML, XML, ...) para la definición y almacenamiento de estas colecciones es quizás

la mejor solución para estandarizar estas colecciones [8].

Una de las fuentes más interesantes y de mayor accesibilidad es la Web [6]. En realidad es una fuente inagotable de documentos que podemos utilizar como colección de evaluación para nuestro sistema, aunque para tener la colección completa deberíamos obtener los juicios de relevancia asociados lo cual no es una tarea nada trivial.

En todo caso, la idea de utilizar documentos de la Web como fuentes para una colección choca con la heterogeneidad extrema de los documentos HTML [8]: no comparten el mismo formato, ni aún procediendo de la misma fuente de información. Esto requiere un esfuerzo extra para preprocesar estos documentos y obtener la colección inicial "limpia", es decir, sin etiquetas y sólo con la información que nos interese.

## **3. COLECCIÓN DE EVALUACIÓN PARA PERSONALIZACIÓN**

Las colecciones de evaluación para personalización tienen un inconveniente principal respecto a otras colecciones de evaluación: se necesitan juicios de relevancia distintos para cada usuario. Además, como se pretende que esta personalización sea adaptativa con el paso del tiempo, es necesario recoger juicios de realimentación del usuario (positivo/negativo/nulo) sobre los documentos que recibe.

En este caso, la colección de evaluación está enfocada a sistemas de personalización de periódicos digitales, contiene información sobre modelos de usuario, sobre la relevancia de las noticias para los distintos usuarios y sobre la realimentación explícita producida por cada uno de ellos. Esta colección se puede aplicar a cualquier otro sistema que utilice técnicas similares de personalización o filtrado de información.

A continuación se describen brevemente los distintos pasos seguidos para la construcción de la colección:

En primer lugar, puesto que el dominio seleccionado es el de los periódicos electrónicos es necesario construir un programa que se conecte a la web del periódico y se baje los archivos correspondientes. En nuestro caso, el periódico utilizado fue ABC, durante los días del 6 al 10 de mayo de 2002. El número de noticias bajadas cada día fue 128, 104, 87, 98 y 102.

Se puede acceder a las noticias a partir de la página de cada sección, la cual almacena enlaces a todas las noticias contenidas en dicha sección [8]. En nuestro caso, se utilizan las siguientes 8 secciones: nacional, internacional, deportes, economía, sociedad, cultura, gente y opinión.

Posteriormente, hay que extraer del formato HTML de los archivos bajados los elementos que sean interesantes para nuestra colección. En este caso, se extraen el título, el autor, la sección, el texto del artículo y el enlace al texto completo. Hay que resaltar que el tratamiento de esta información no es trivial, puesto que aunque las páginas web se generan automáticamente en el periódico mediante motores ASP, el aspecto final del código HTML generado es bastante caótico. Afortunadamente, la búsqueda de las etiquetas título, autor y cuerpo permite la localización de la información relevante sin muchas dificultades. La etiqueta *entradilla* no ha sido utilizada porque no aparecía en todas las noticias. Sin embargo, este esquema no es válido para otros periódicos electrónicos ya que la forma de generar la información es diferente.

El formato final en el que se almacenan las noticias es como ficheros de texto: una línea para el título, otra para los autores, y el resto para el cuerpo de la noticia.

En segundo lugar, es necesario obtener la definición de los intereses iniciales de varios modelos de usuario para poder empezar a ejecutar el sistema y obtener la primera personalización de contenidos. En nuestro caso, se ha dispuesto de 11 modelos de usuario correspondientes a 11 personas diferentes. En nuestro caso, los intereses iniciales se almacenan sobre secciones y términos. Estos intereses son introducidos por el usuario manualmente utilizando una escala de 0 a 3 para asignar un peso a cada uno de ellos.

La Tabla I muestra el número de secciones y términos elegidos por los usuarios en sus perfiles iniciales. En general, los usuarios eligieron varias secciones y varios términos. Estos modelos contenían, en media, 5 secciones y 4 términos. Es interesante resaltar que un usuario (número 2) no seleccionó ningún término, mientras que el resto introdujeron 3 términos como mínimo y el máximo número de términos fue 7. También es un dato significativo que 2 usuarios (números 5 y 6) eligieran todas las secciones como relevantes,

mientras que el mínimo de secciones elegidas fue 2.

Usuario	NumSecciones	NumTérminos
0	2	3
1	4	3
2	4	0
3	7	5
4	4	4
5	8	3
6	8	6
7	7	7
8	2	4
9	5	5
10	6	3
media	4.75	3.58

Tabla I. Número de secciones y términos elegidos por los usuarios.

La tabla II muestra los pesos asignados a cada una de las secciones. Se puede observar que las secciones más importantes para los usuarios son internacional y cultura, un 82% de los usuarios (9 de los 11) las eligieron, aunque esta última tiene un mayor promedio (0.69) en los valores de los pesos asignados por los usuarios. Las menos elegidas son opinión y gente, 45% de los usuarios (5 de los 11), aunque la de menor peso promedio es opinión (0.27). También se puede resaltar que los usuarios variaron sus pesos entre las distintas secciones, es decir, no utilizaron sólo el criterio me interesa / no me interesa, lo cual demuestra que el sistema de asignación de pesos es atractivo para los usuarios.

En cuanto a los pesos de los términos cabe destacar que todos los usuarios asignaron peso 1 a sus términos excepto 3 usuarios que escogieron para su último término peso 0.66.

En tercer lugar, es necesario obtener los juicios de relevancia de cada uno de los usuarios respecto a todas las noticias, durante los 5 días. Para ello, se le envía cada día a cada usuario un mensaje de correo electrónico con la siguiente información asociada a cada noticia: título, autor, sección, resumen del artículo y enlace a la noticia completa. En este caso, el resumen es el que ofrecen los periódicos electrónicos, esto es, las primeras frases del artículo. Cada usuario debe juzgar como relevante o no relevante cada una de esas noticias. Para establecer ese juicio el usuario puede visualizar el texto completo si la información presentada inicialmente no es suficiente.

Usuario	Opinión	Nacional	Internacional	Economía	Sociedad	Cultura	Deportes	Gente
0	0	1	0	0	0	0	1	0
1	0	0.33	0.66	1	0	0	0	0.33
2	0	0	1	0	0	1	0	1
3	0.33	0.66	0.66	1	0.33	1	1	0
4	1	0	0	0	1	1	0.66	0
5	0.66	0.66	0.66	1	0.66	0.33	0.33	0.33
6	0.66	0.66	0.66	0.66	0.33	1	0.33	0.66
7	0.33	0.66	1	0.66	0.66	1	0.33	0
8	0	0	0.66	0	0	0.66	0	0
9	0	0.66	0.66	0	1	0.66	0	1
10	0	0.33	0.66	1	1	0.66	0.33	0
Media (valores)	0.27	0.45	0.60	0.48	0.45	0.69	0.36	0.30
Media (veces)	0.45	0.73	0.82	0.55	0.64	0.82	0.64	0.45

Tabla II. Pesos asignados a las distintas secciones por los usuarios

Asociado a cada una de las noticias existen dos iconos que, en este caso, tienen el significado de relevante/no relevante (1/0). Cada usuario debe pulsar en estos iconos para emitir su juicio, esta pulsación invoca a un programa JSP que actualiza la información correspondiente a cada usuario. Los juicios de relevancia de los usuarios no se deben basar sólo en su perfil inicial sino que deben de ser independientes, precisamente los intereses que no sean acordes al modelo inicial deberían ser capturados a través de la realimentación.

La tabla III muestra el número de noticias indicadas como relevantes por cada uno de los usuarios durante cada uno de los 5 días que duró el experimento. El número medio de noticias relevantes por día varía entre 25 y 34, con 28 como valor medio. Los valores sufren una variación considerable, entre 5 noticias relevantes (10-Mayo, usuario 1) y 80 noticias relevantes (7-Mayo, usuario 8).

Hay una diferencia significativa entre un conjunto de usuarios más exigentes, que juzgan pocas noticias como relevantes, y usuarios juzgan muchas noticias como relevantes. El primer grupo lo constituyen los usuarios 0, 1, 2, 3, 4 y 10, que juzgan, en media, 17, 11, 11, 15, 15 y 18 noticias como relevantes. El segundo grupo es el formado por los usuarios 5, 6, 8 y 9, que juzgan, en media, 32, 56, 68 y 43 noticias como relevantes.

Día	6-5	7-5	8-5	9-5	10-5	Media
Noticias	128	104	87	98	102	103.8
Usuario						
0	18	16	11	17	21	16.6
1	23	8	10	9	5	11.0
2	16	8	10	8	14	11.2
3	20	15	9	10	22	15.2
4	30	12	16	6	9	14.6
5	30	35	33	27	37	32.4
6	72	72	43	53	41	56.2
7	39	21	20	24	10	22.8
8	77	80	72	60	51	68.0
9	23	47	41	42	61	42.8
10	22	16	17	17	16	17.6
Media	33.6	30.0	25.6	24.8	26.1	28.0

Tabla III. Número de noticias relevantes por usuario y por día

En cuarto lugar, es necesario obtener los juicios sobre la realimentación del usuario sobre cada una de las noticias, durante los 5 días. Para ello, se le envía cada día un segundo mensaje con el mismo formato que el anterior excepto que los iconos representan realimentación positiva o negativa. Cada usuario puede pulsar en uno de estos iconos o no pulsar en ninguno (1/-1/0). En este caso el usuario puede no querer realimentar ni positiva ni negativamente sobre alguna noticia. La pulsación sobre los iconos actualiza, a través del programa JSP correspondiente la información del usuario. En realidad se consideran como realimentación positiva aquellas noticias consideradas como

relevantes en el juicio anterior y se pide al usuario que realmente negativamente sobre aquellas noticias sobre las que no quiere saber nada de entre las 20 primeras de las que recibe. Esta simplificación es debida a que el comportamiento real del usuario no va a ser realimentar sobre todas las noticias que reciba.

La tabla IV muestra el número de noticias realimentadas, tanto positiva como negativamente (R+ / R-), por cada uno de los usuarios durante cada uno de los 5 días que duró el experimento. El usuario 2 es el más exigente. El 8 el menos exigente, no realimenta negativamente ninguna noticia.

Día	6-5		7-5		8-5		9-5		10-5		Media	
Usuario	R+	R-	R+	R-	R+	R-	R+	R-	R+	R-	R+	R-
0	18	3	16	2	11	5	17	3	21	3	16.6	3.2
1	23	5	8	6	10	3	9	2	5	9	11.0	5.0
2	16	15	8	19	10	19	8	18	14	16	11.2	17.4
3	20	2	15	3	9	2	10	1	22	2	15.2	2.0
4	30	5	12	1	16	2	6	2	9	0	14.6	2.0
5	30	5	35	1	33	1	27	0	37	1	32.4	1.6
6	72	4	72	5	43	3	53	4	41	2	56.2	3.6
7	39	6	21	4	20	3	24	2	10	1	22.8	3.2
8	77	0	80	0	72	0	60	0	51	0	68.0	0.0
9	23	0	47	2	41	7	42	5	61	1	42.8	3.0
10	22	7	16	5	17	4	17	4	16	6	17.6	5.2
Media	33.6	4.7	30.0	4.4	25.6	4.5	24.8	3.7	26.1	3.7	28.0	4.2

Tabla IV. Número de noticias realimentadas cada día por cada usuario, de manera positiva o negativa (R+ / R-).

#### 4. MÉTRICAS DE EVALUACIÓN

Una vez construida la colección se puede someter al sistema a distintas combinaciones de los parámetros de los distintos procesos y obtener los resultados asociados, estos resultados serán ranking de noticias que deberán ser comparados con los juicios de relevancia asociados aplicando la métrica adecuada.

Las métricas más conocidas son recall y precisión, donde el recall indica la proporción entre documentos recuperados relevantes y documentos relevantes, y la precisión indica la proporción entre los documentos recuperados relevantes y los documentos recuperados. Estas métricas se suelen utilizar cuando existe distinción entre documentos relevantes y no relevantes, ya sea por un valor de relevancia o por una posición en un ranking, superiores a un determinado umbral. Otras medidas conocidas son: fallout, E y  $F_1$  [2; 11].

Cuando lo que se compara es un ranking con juicios de relevancias binarios, las métricas más adecuada son recall y precisión normalizados, y longitud de búsqueda esperada [9; 11].

Además, hay que comprobar si los resultados son estadísticamente significativos utilizando algún test de significancia estadística.

En nuestro caso hemos considerado sign-test con muestras apareadas, a un nivel de significación del 5% ( $p \leq 0.05$ ). Esta decisión está basada en el hecho de que no se hace ninguna suposición sobre la distribución subyacente, y que, dados los diferentes procesos de normalización que se aplican, a distintos niveles, se deben tener en cuenta los valores relativos en lugar de las magnitudes absolutas para establecer la significancia estadística [11].

#### 5. OTRAS MEDIDAS DE EVALUACIÓN

Durante los últimos años, en la investigación alrededor de la recuperación de información, está creciendo el cuerpo de trabajos centrados en el análisis de la apreciación de los usuarios sobre los sistemas de recuperación de información. Y son diversos los modelos y los autores, como Ingwersen [7], Belkin [1] y Spink [12] que entroncan con esta línea de trabajo que se suele denominarse corriente cognitiva [14].

Esta perspectiva parte de la consideración de la recuperación de información como un proceso interactivo en el que una parte esencial es el usuario. En este sentido, su principal pretensión es la de incluir en el análisis todos

los factores que influyen en el uso de los sistemas de recuperación de información. Factores que en un estudio tradicional se dejarían al margen y que están relacionados con el interés que tiene el usuario por el tema preguntado, el nivel de conocimientos que posee o con su motivación. La interacción entre estos factores, cómo expresa el usuario sus necesidades o problemas de información, las técnicas y procedimientos que se usan y los cambios en los estados de conocimiento son cuestiones que se convierten en nuevos objetos de estudio.

Desde la posición habitual de los estudios sobre recuperación se han subrayado diferentes aspectos negativos, tales como la ambigüedad en la definición de los conceptos empleados, la falta de aproximaciones prácticas, la dependencia del modelo algorítmico y la ausencia de fórmulas para extrapolar los datos, más allá del individuo, a comunidades o a grupos. No obstante, desde nuestro punto de vista, es una línea de investigación que aporta nuevas perspectivas enriquecedoras al estudio de la recuperación y filtrado de información.

En este trabajo, estamos más interesados, aunque en parte bajo el paraguas teórico cognitivo, en los métodos de evaluación de los diferentes procesos de interacción que se producen en el manejo de sistemas de información.

Una muestra de esto la tenemos en el artículo elaborado por Spink [12], en el que se presenta un estudio centrado en el usuario que busca analizar la efectividad de un motor de recuperación de información, basándose en el impacto de las interacciones de las que es protagonista cada usuario sobre sus necesidades de información y en las fases de búsqueda de información. En segundo lugar, también pretende precisar la usabilidad del sistema.

Destaca en este estudio el modelo de evaluación, que se fundamenta en tres elementos: uno primero que está vinculado al tiempo, es decir, a los movimientos y cambios que tienen lugar durante los episodios de búsqueda de información, en términos de tácticas y términos empleados, objetivos propuestos, incertidumbre generada, o necesidad de información; uno segundo que está relacionado con los episodios de búsqueda interactiva y que recurre, con afán integrador, a diferentes estudios cognitivos; y en tercer lugar, el conjunto de acciones que engloba, por ejemplo, las decisiones y juicios que se realizan

durante la búsqueda de información, especialmente en torno a la consideración de la relevancia.

Como instrumentos de trabajo destacan dos cuestionarios, uno que se aplica antes del momento de la búsqueda y otro que debe ser cumplimentado con posterioridad a ésta, con los que se puede medir los cambios que se producen durante la interacción con el sistema de recuperación de información, en este caso un motor de búsqueda en Internet. También se recurre a la grabación de la interacción para recoger los comentarios que van surgiendo.

El objetivo es estudiar el comportamiento de los usuarios a partir de diferentes variables. Con respecto a la necesidad de información las variables son las características del usuario, su estilo cognitivo, cómo exponen su problema y necesidad de información, su nivel de conocimiento, la fase en la que se encuentra de búsqueda de información y el nivel de incertidumbre. También hay variables referidas a los comportamientos de búsqueda de información, a la exposición y análisis de las preguntas, a las características de los intermediarios, a la estrategia, características y procesos de búsqueda, a los ítems recuperados y formas de entrega de la información, a la relevancia y utilidad de la información recibida, a las características del usuario después de la búsqueda, a la exposición de la necesidad de información, al nivel de conocimiento, a la fase de búsqueda de la información, y nivel de incertidumbre. Con respecto a la usabilidad, se mide fundamentalmente la satisfacción general, el tiempo de aprendizaje, la cantidad de información proporcionada y el diseño y la adaptación de la pantalla.

Con los datos extraídos, se realiza tanto un análisis cuantitativo, preocupado por el análisis estadístico de los datos del cuestionario, como un análisis cualitativo, que es el que predomina, especialmente cuando los datos que se recogen sobre el usuario son textos de cierta dimensión. Sin entrar en las conclusiones concretas, puesto que nuestra intención directa es más bien metodológica, cabe destacar que, por un lado, se comprueba que este método de evaluación funciona adecuadamente si lo que se quiere es mejorar el sistema de recuperación. De igual forma, se observa que es útil para el objetivo planteado en un primer momento que no es otro sino advertir los cambios que se producen en el usuario mientras trabaja con el sistema.

Aparte de observar diferentes niveles de cambio en cada usuario, se apreció que este factor no es correlativo a la precisión de la búsqueda. En este sentido, entendemos que estos estudios pueden tener implicaciones para las medidas tradicionales sobre la evaluación de los sistemas de recuperación y filtrado de información. Hasta tal punto que Spink propone una nueva medida derivada de los cambios producidos respecto a la necesidad o problema de información del usuario. Con el *cambio en el problema de información* se pretende advertir cómo un usuario progresa hasta conseguir resolver su necesidad de información, lo que se deduce restando la fase de necesidad de información antes de su interacción de recuperación de información a la fase de necesidad de información después de interacción.

Esta aproximación ya ha sido empleada con éxito en otros trabajos relacionados con los sistemas de información personalizada vinculados a periódicos digitales [13]. En estos trabajos se realizó, en primer lugar, un análisis cualitativo, por un lado vinculado a entrevistas con los usuarios-evaluadores, que aportaban una visión interesante sobre los elementos de mejora del sistema, tanto en el apartado de recuperación como en el de usabilidad.

En segundo lugar, se llevó a cabo un estudio cuantitativo, con tres variantes: una evaluación orientada al usuario, una evaluación objetiva del sistema a partir de los perfiles de los usuarios y una evaluación del sistema propiamente dicho.

Con la evaluación orientada al usuario, se medía cuestiones tales como el tiempo de respuesta, el grado de dificultad que plantea el trabajo con el sistema, la eficacia, la forma de presentación conceptual y física, etc. El instrumento de trabajo fue un cuestionario cerrado que se respondía exige en cuatros niveles de respuesta (25%, 50%, 75%, 100%) o mediante la dualidad Sí/No. Las preguntas giraban alrededor de cuatro facetas básicas: una primera referida a las prestaciones del sistema en temas como el grado de satisfacción con los distintos elementos de la interfaz, la capacidad que de personalización del sistema o la capacidad para acceder a la información y las posibilidades de búsqueda de información. En segundo lugar, se examinaba las categorías y las secciones, como formas adecuadas de modelado del usuario, en torno a aspectos como si el sistema proporciona información clara sobre el

sistema de categorías/secciones o si el modelo de categorización es el más adecuado o si es incompleto. En tercer lugar, se valoraban los resúmenes que se aportaban, por ejemplo en lo que se refería al grado de adecuación con respecto al contenido o al perfil del usuario. Finalmente, se intentaba superar la medida sobre la relevancia de la información, tomando en consideración estudios anteriores por ejemplo con la calidad del contenido, el grado de cercanía, o el conocimiento que le aporta al usuario.

En resumen, se puede hablar de tres niveles que hay que tener en cuenta a la hora de evaluar los sistemas objeto de nuestro estudio y para lo cual, metodológicamente hablando, habría que emplear diferentes métodos de investigación. En primer lugar, hay precisar las características y naturaleza del sistema de información de que se trate y la perspectiva que se quiera que predomine. En segundo lugar, una aproximación a los estudios centrados tanto en las medidas tradicionales de evaluación como en la aplicación de nuevas medidas, también con afán de comprobar su validez. Todo esto sin olvidar, en tercer lugar, la aplicación de una investigación centrada en el usuario.

En último lugar, cabe destacar que este tipo de estudios requiere una nueva visión sobre las condiciones que debe seguir las colecciones de estudio. En primer lugar, hay que comprobar que la colección sea representativa de todas las formas discursivas con las que se trabaje. En segundo lugar, hay que determinar si es capaz de permitir el estudio del usuario y su interacción, resolviendo dudas por ejemplo relacionadas con el número de días que se debe tomar en consideración, si se ha de utilizar más de una colección, etc. Y, finalmente, si la colección es la más adecuada en función de las formas de medida tanto de la relevancia de la información recibida como de la efectividad del sistema de información.

## 5. CONCLUSIONES

Aunque el número de días y de usuarios para el cual se ha realizado la colección es pequeño, permite una comparación inicial entre distintas propuestas de personalización. De hecho, los sistemas más parecidos utilizan colecciones de documentos durante únicamente 3 días [2; 5].

Además, esta colección se puede utilizar para evaluar tanto la selección, como la adaptación como la generación de contenidos.

Para evaluar esta última se puede comparar los resultados obtenidos con los contenidos completos frente a los que se obtienen con distintas maneras de generar los contenidos finales de los documentos.

Adicionalmente hay que tener en consideración factores como la validez de los métodos de análisis y de recogida de la información, la identificación de los problemas y necesidades de información de los usuarios, la escala de medida utilizada, los estados de búsqueda y recepción de información, los niveles y formas de interacción de los usuarios, su satisfacción propiamente dicha, etc. [12; 14].

## BIBLIOGRAFÍA

- [1] BELKIN, N. J. (2001): "Iterative exploration, design and evaluation of support for query reformulation interactive information retrieval", *Information Processing & Management*, vol. 37, p. 403-434.
- [2] BILLSUS, D. y PAZZANI. M.J. (2000): "User Modeling for Adaptive News Access", *User Modeling and User-Adapted Interaction Journal* 10(2-3), pp. 147-180.
- [3] DÍAZ, A., GERVÁS, P. y GARCÍA, A. (2000): "Evaluating a User-Model Based Personalisation Architecture for Digital News Services", *Research and Advanced Technology for Digital Libraries* (Proceedings 4th European Conference on Digital Libraries, ECDL 2000), Lectures notes in computer science, vol. 1923, Springer, pp. 259-268.
- [4] GARCÍA, A., CHACÓN, I., DÍAZ, A. y GERVÁS, P. (2000): "Nuevos sistemas de información: tendencias y evaluación", *Cuadernos de Documentación Multimedia*, nº9, <http://www.ucm.es/info/multidoc/multidoc/revista/num9/prensa/jime-chacon.htm>.
- [5] GATES K.F., LAWHEAD, P.B. y WILKINS, D.E. (1998): "Towards an adaptive WWW: a case study in customized hypermedia". *The New Review of Hypermedia and Multimedia*, 4.
- [6] GREFENSTETTE, G. (1999): "The WWW as a Resource for Example-Based MT Tasks". *ASLIB'99, Translating and the Computer*, 21, London, UK, Nov 1999.
- [7] INGWERSEN, P. (1996): "Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory", *Journal of Documentation*, v. 52, nº 1., pp. 3-50.
- [8] MARTÍNEZ, F., UREÑA, A. y GARCÍA, M. (2001): "WWW como Fuente de Recursos Lingüísticos para su Uso en PLN". *Procesamiento del Lenguaje Natural*, 27. Septiembre 2001.
- [9] MIZZARO, S. (2001): "A New Measure Of Retrieval Effectiveness (or: What's Wrong With Precision And Recall)". In: T. Ojala (ed.): *International Workshop on Information Retrieval (IR'2001)*, Infotech Oulu, pp. 43-52.
- [10] MIZARRO, S. y TASSO, C. (2002): "Ephemeral and Persistent Personalization in Adaptive Information Access to Scholarly Publications on the Web". *Proceedings of the 2nd International Conference on Adaptive Hypermedia and Adaptive Web Based Systems*, Málaga, España, Mayo 2002.
- [11] SALTON, G. (1989): *Automatic Text Processing: The Transformation, Analysis and Retrieval of Information by Computer*, Addison-Wesley Publishing, Reading, Massachusetts, US, 1989.
- [12] SPINK, A. (2002): "A user-centered approach to evaluating human interaction with Web search engines: an exploratory study", *Information Processing & Management*, vol. 38., nº 3, pp. 401-426.
- [13] Text REtrieval Conference (TREC) Home Page: <http://trec.nist.gov/>
- [14] VARGAS-QUESADA, B, MOYA, F. DE y OLVERA, M. D. (2002): "Enfoques en torno al modelo cognitivo para la recuperación de información: análisis crítico", *Ciencia da Informaçao*, v. 31, nº 2, pp. 107-119.