

Marcos Mora, Mari Carmen. “Interacción entre los usuarios y los catálogos de las bibliotecas: problemas actuales y posibles soluciones”. En: Rovira, Cristòfol; Codina, Lluís; Marcos, Mari Carmen; Palma, María del Valle. Información y documentación digital. Barcelona: IULA; Documenta Universitaria, 2004. ISBN 84-96367-09-6.

Interacción entre los usuarios y los catálogos de las bibliotecas: problemas actuales y posibles soluciones
Mari Carmen Marcos

1. Introducción

La Documentación —y concretamente la RI— ha centrado su visión tradicionalmente en las capacidades de búsqueda de los sistemas y ha dejado algo más de lado el punto de vista del usuario, sus capacidades perceptuales y cognitivas. Trataremos de adoptar las recomendaciones de la disciplina Interacción Persona-Ordenador (IPO) al desarrollo de OPACs más intuitivos que se adapten mejor a las necesidades de los usuarios.

Los catálogos de acceso público en línea son productos dirigidos a usuarios finales. Cuando su acceso era sólo local, la tipología de usuario podía limitarse en función del tipo de centro en el que estuviera la biblioteca y sus objetivos. Con la aparición de internet el panorama ha cambiado y ahora los usuarios de un OPAC son mucho más heterogéneos, tanto en el conocimiento del sistema informático concreto del catálogo como en sus objetivos al usarlo, pues entre ellos hay usuarios habituales y usuarios casuales, usuarios que acuden a la biblioteca a buscar los documentos que encuentran en el catálogo y usuarios que nunca irán a las dependencias físicas de ésta.

La IPO pretende que la tecnología sea fácil de usar, independientemente de las capacidades del usuario, pero teniendo en cuenta esta diversidad de conocimientos, por lo que apuesta por el diseño de sistemas que se adecuen a usuarios noveles y a usuarios “expertos”, de forma que para los primeros resulten intuitivos y para los segundos ágiles. Intentaremos dar pautas que ayuden a este objetivo.

Existe una buena cantidad de sistemas —en su mayoría prototipos— que emplean la visualización para representar información y facilitar su recuperación. Unos emplean el browsing para evitar al usuario plantear la búsqueda y darle a conocer el contenido de la base de datos (overview), otros optan por la agrupación (clustering) para presentar los documentos en función de su similitud en el momento de ofrecer los resultados (preview), y algunos de ellos idean metáforas que hagan más sencilla la comprensión de la visualización presentada.

La mayor parte de esos sistemas se han probado en bases de datos con documentos a texto completo, o bien con un campo de resumen, lo que permite contar con más palabras susceptibles de ser tomadas como términos de indización y por tanto tener más puntos de referencia para describir el contenido del documento que en una base de datos referencial, donde esa información sólo puede tomarse del título (con los problemas de ambigüedad que conlleva el lenguaje natural) o de campos creados por el documentalista para descripción de contenido, con descriptores o códigos de clasificación.

Vamos a reflexionar sobre posibles formas de aplicación de la IPO al caso concreto de los catálogos en línea y específicamente al acceso más complejo para el usuario: el acceso por materias.

2. El OPAC accesible en la web

Las bibliotecas han sido uno de los servicios que antes han comenzado a aprovechar las ventajas de disponer su información al alcance de personas conectadas a internet. En un primer momento, la forma de ofrecer ese acceso (que aún hoy se mantiene en muchos casos) era mediante una conexión remota Telnet, con el inconveniente de que es preciso conocer el lenguaje de interrogación y la interfaz de cada sistema en particular.

Con la aparición de la web a principio de los años 90, el panorama cambia y las bibliotecas migran sus catálogos a la nueva plataforma, estableciendo pasarelas que sirven de interfaz entre la página web de consulta y la base de datos del catálogo. Gracias a estas pasarelas (por ejemplo el CGI, Common Gateway Interface) el usuario consulta la base de datos desde su navegador web por medio de un formulario y recibe los resultados en esa misma pantalla, sin tener que utilizar el lenguaje de interrogación propio del sistema de recuperación de información que usa el OPAC. El funcionamiento sigue los siguientes pasos:

- Desde el navegador web, el usuario se conecta a un servidor HTTP que dispone de un formulario para la consulta de una base de datos.
- Los datos introducidos en el formulario son interpretados por la pasarela —por ejemplo un CGI— que lleva el servidor HTTP y confrontados con los datos de la base de datos. Esto es, la consulta hecha por el usuario es traducida al lenguaje de interrogación propio de la base de datos.
- Una vez que ha sido consultada la base de datos, los resultados de la búsqueda son presentados al usuario en la pantalla de su navegador.

Moscoso, Ortiz-Repiso y Ríos (1998) comparan estos dos tipos de entornos (Telnet y web) desde el punto de vista de la interfaz de usuario final y de las prestaciones que ofrecen. Hoy en día, el acceso mediante conexión Telnet prácticamente ha desaparecido, al menos para el usuario final, y ha dado paso a los OPACs accesibles desde web —ya sea con un protocolo Z39.50 o no, como ocurre aún en la mayoría de los casos— y en su mayoría integrados en el portal de la biblioteca.

El estudio de Moscoso (1998) analiza diecisiete OPACs accesibles en la web y apunta las nuevas necesidades que se imponen en este tipo de catálogos, debidas fundamentalmente a que aumenta el número de usuarios potenciales y por tanto resulta difícil establecer un perfil de usuario: cualquier persona conectada a la red es un posible usuario del OPAC. La autora indica que los nuevos desarrollos precisan que se elaboren herramientas más sencillas de usar, donde no haga falta memorizar comandos y aprender una sintaxis de búsqueda, que tengan una interfaz intuitiva y amigable. De estos puntos, se ha superado la memorización de comandos y la formulación de ecuaciones de búsqueda, si bien hay todavía mejoras por introducir.

La solución se ha buscado en la propia forma de presentación de la información en la web: si el usuario está familiarizado con el entorno web, puede pensarse que lo mejor será diseñar OPACs que se asemejen a esta forma de presentar la información; esta

postura adoptada hasta el momento mejora la accesibilidad y el manejo del OPAC por parte de los usuarios en comparación con el acceso que hasta hace unos años se tenía mediante Telnet, pero va en detrimento de las capacidades de búsqueda ofrecidas por los catálogos en línea tradicionales (Moscoso, 1998).

Uno de los motivos por los que los OPACs deben estar accesibles a través de la web es que una gran cantidad de usuarios están familiarizados con este entorno, y por lo tanto, pueden aplicar sus conocimientos de navegación web al utilizar el OPAC.

La gran ventaja de este tipo de acceso a los catálogos es que éstos se presentan al usuario como una herramienta más en su navegador. Por ejemplo, si interesa buscar información acerca de algún aspecto de la biología molecular, podría resultar interesante abrir una ventana de navegación y consultar una base de datos especializada en la materia; abrir otra ventana para realizar una consulta en un buscador especializado; y una tercera para utilizar el catálogo de una biblioteca de una facultad de Biología. Y todo con la misma herramienta informática.

No debemos olvidar dos características de la web que pueden resultar de gran utilidad a los OPACs, si bien por el momento no están apenas explotadas: la hipertextualidad y la capacidad de incluir documentos multimedia. Por ahora, la hipertextualidad de los OPACs no responde a la hipertextualidad que estamos acostumbrados a tener en la web, donde un concepto lleva a otro relacionado de forma que se amplía la información sobre determinado objeto, sino que en el OPAC, pulsar en el nombre de un autor, supone ejecutar una nueva consulta donde el término buscado es ese autor que se ha pulsado. Tampoco se ha explotado todavía la capacidad multimedia de la web, pues los OPACs son claramente textuales y no incorporan ficheros de audio ni imágenes.

Como indican Ortiz-Repiso y Moscoso (1999), la tendencia que se observa es similar a lo ya ocurrido con los OPACs de primera generación con respecto a los catálogos en fichas manuales: cambia el medio, pero no la forma de organizar la información, puesto que se mantiene el modelo de catalogación (por ejemplo el formato MARC en lugar de otras opciones más flexibles como el lenguaje de etiquetado SGML o su derivado, el XML) y no se aprovechan las posibilidades de un medio hipertextual y multimedia. Esto ha supuesto mantener los mismos problemas de los catálogos de segunda generación en el nuevo medio.

Pero no todo son aspectos negativos. En los últimos años se está trabajando en arquitecturas de sistemas abiertos, se investiga la forma de mejorar el diseño de las interfaces de usuario, se ha incorporado el protocolo Z39.50, se comienzan a emplear las especificaciones del Dublin Core en la asignación de etiquetas meta en la catalogación, se está haciendo uso del hipertexto y se va incorporando la programación en Java en las interfaces de OPACs en web.

2.1. Surgen nuevos servicios

El catálogo ha sido tradicionalmente una herramienta de acceso a los documentos de una colección concreta. El nuevo concepto de catálogo incluye documentos accesibles en modo digital, como es el caso de las revistas electrónicas. Es evidente que la aparición de los documentos electrónicos accesibles en línea está haciendo cambiar el concepto de “fondo de biblioteca”, ya que la biblioteca puede dar acceso a más

documentos que los que se encuentran físicamente almacenados en sus dependencias, y llega a usuarios que de otra forma no llegaría.

De forma paralela a la incorporación de estos nuevos documentos, la biblioteca está asumiendo su papel de organizadora de información en el nuevo entorno, y ofrece servicios complementarios en su sitio web que pueden integrarse en el OPAC, como son los directorios temáticos con recursos web de interés. Por ejemplo, una biblioteca especializada en una materia podrá —y deberá— crear y mantener un directorio de recursos relacionados con su especialidad temática. Por lo tanto, al diseñar el sistema habrá que considerar esta característica, que si bien quizá no se cumple en el momento del diseño, podría incorporarse más adelante, por lo que debe ser prevista desde el principio.

Otro aspecto que debe contemplarse es el tipo de usuario que va a acudir al sistema y con qué objetivo. En función del tipo de biblioteca cambiará el modelo de usuario: universitario, investigador, ciudadano en general... y sus necesidades de información: académicas, investigadoras, de ocio, etc. La introducción de documentos a texto completo en el OPAC, así como otros servicios como directorios temáticos, amplía el abanico de usuarios propios de la biblioteca, de manera que muchos de estos usuarios no van a ser presenciales nunca o muy pocas veces. Por lo tanto, el diseño del sistema deberá contemplar un usuario muy variado y con diferentes necesidades informativas.

Con la aparición de la web surge una fuente de información muy accesible que brinda a las bibliotecas un rápido acceso a una cantidad enorme de documentos disponibles en páginas externas a ellas. El sitio web de la biblioteca comienza a entenderse como una puerta a información que, si bien no toda pertenece al propio fondo, sí es posible proporcionar acceso a ella, por lo que se plantea el concepto de acceso versus propiedad (Játiva, 2002; Frías, 1998; Peis y Fernández-Molina, 1996). Quiere decirse que desde el sitio web de la biblioteca se puede ofrecer toda la información propia y ajena: consulta en bases de datos externas, consulta del OPAC, acceso a artículos publicados en revistas suscritas que tengan versión electrónica a texto completo, enlaces a recursos web externos, etc. Se pretende lograr la integración de todos estos recursos en una sola interfaz de manera que el usuario obtenga la información sin necesidad de consultar cada fuente. El OPAC además comienza a brindar servicios que van más allá de la consulta del propio catálogo, puesto que da acceso a información externa (otras bases de datos, revistas-e, etc.) y a información relativa a la circulación de la colección (consulta del estado de préstamo de un documento, posibilidad de reservas, etc.).

Teniendo en cuenta todos estos factores (tipos de usuarios, forma de acceso, tipo de información) y el objetivo del catálogo, se pueden determinar algunas pautas de diseño que mejoren el uso del sistema por parte del usuario final.

2.2. Hacia la usabilidad

Los OPACs de lo que podríamos llamar la “generación web” deberán atender más al diseño y centrarse en el usuario final para adaptarse tanto a su comportamiento y su forma de conocer, como para “aprender” lo que éste considera relevante (con una buena retroalimentación por relevancia) y poder así ofrecer mejores resultados. En este sentido, un gran reto de los futuros OPACs es conseguir mayor usabilidad para los usuarios finales.

La usabilidad puede definirse de forma simple como la facilidad para ser usado. La norma ISO 9241-11 “Ergonomic requirements for office work with visual display terminals (VDTs)” en su parte 11 titulada “Guidance on usability”, la define así: “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use”.

Los principios de diseño y usabilidad generales y aquellos más concretos de la web son perfectamente aplicables a las interfaces de OPACs accesibles en línea. La bibliografía sobre diseño publicada en los últimos años —especialmente desde el nacimiento de la web— es muy amplia, de manera que remitimos a ella para ahondar en el tema.

Es cierto que se ha publicado abundante bibliografía tanto sobre diseño de interfaces como sobre OPACs, pero en cambio no hay mucho sobre el diseño de interfaces de OPACs, y lo que hay es —en su mayoría— anterior a las interfaces web para catálogos.

2.3. A modo de ejemplo: Cheshire II, una iniciativa de OPAC avanzado

Para finalizar este apartado, y como iniciativa interesante, destacamos Cheshire II, un SRI aplicado a una parte del catálogo de la biblioteca de la universidad californiana de Berkeley. El sistema se basa en una arquitectura cliente-servidor con implementaciones de los estándares actuales para RI, como Z39.50 y SGML/XML. En la búsqueda por materias, el servidor que recibe la petición de información lanza esta consulta a varios sitios para buscar documentos relacionados con la búsqueda planteada, computa el grado de similitud y devuelve una lista ordenada por relevancia.

Si no encuentra resultados que ofrecer, amplía los términos de consulta acudiendo a los encabezamientos de materia de la Library of Congress (LC). Su motor de búsqueda y su interfaz soportan la navegación entre enlaces hipertextuales generados automáticamente, las búsquedas “nearest neighbor” (por similitud) y la retroalimentación de relevancia. Se diseñó con la intención de ser un catálogo cercano o incluso ya de tercera generación, usando técnicas avanzadas de búsqueda para mejorar el acceso por materias, combatir el desbordamiento en los resultados y soportar un acceso en red a fuentes de información diversas usando protocolos normalizados.

3. Problemas no resueltos en el acceso a la información

A pesar de los avances logrados en los catálogos en línea en cuanto a su accesibilidad al estar integrados en la web, los usuarios se siguen encontrando todavía con las mismas dificultades que hace años: cómo localizar información de su interés cuando no conocen los documentos concretos que satisfacen esa necesidad, y cómo distinguir de entre los resultados obtenidos aquellos que son más relevantes para ese fin. No hay que olvidar que los OPACs son un servicio dirigido a muchos tipos de usuarios, y que se debe pensar en todos ellos al diseñarlos. En este apartado abordamos estos problemas.

3.1. Dificultad en el planteamiento de la búsqueda por materias

Se ha comprobado, en numerosos estudios empíricos, que una gran parte de las consultas que se realizan en los OPACs son por materias, y también que son las que más fracaso conllevan, tanto por dar cero resultados como por producir desbordamiento. En

un experimento llevado a cabo en el catálogo Melvyl (de la Universidad de California) se comprobó que el 82% de los usuarios obtenían cero resultados una vez o más en las búsquedas. A pesar de ello, más del 25% de ellos continuaban intentándolo diez o más veces hasta conseguirlo, y un 25% de éstos no lograba finalmente obtener ninguno. Borgman (1996) argumenta para este problema que en los OPACs falta incorporar conocimiento sobre el comportamiento de los usuarios al utilizar el catálogo y cree que se debería ofrecer más ayuda contextual.

Cuando un usuario final se enfrenta las primeras veces a la búsqueda temática en un sistema de información textual en base de datos —sea un catálogo automatizado de una biblioteca, un repertorio bibliográfico de una editorial, un motor de búsqueda o cualquier otro—, a menudo encuentra dificultades en todas las etapas del proceso. Destacamos tres:

1. Antes de plantear la consulta al sistema: ¿qué tengo que buscar, qué información necesito?
2. Durante la consulta al sistema: ¿qué términos utilizo para indicar al sistema lo que necesito? ¿cómo los combino para expresar la consulta con la mayor precisión posible?
3. Una vez obtenidos los resultados: ¿cómo distingo cuáles son los documentos más relevantes?

La primera cuestión es difícil que encuentre respuesta en el propio sistema. Este problema del usuario ha sido abordado en abundante bibliografía. Fernández y Moya (1998) recogen las reflexiones de distintos autores respecto al origen de la necesidad de información en una persona: se dice que es una situación problemática (Wersig, 1979), una necesidad visceral (Taylor, 1962), un defecto en un modelo mental (Marchionini, 1989) o una anomalía en el estado de conocimiento (Belkin, Oddy y Brooks, 1982a, 1982b). Éstos últimos explican que no se puede pretender que el usuario especifique su necesidad de información, no puede hacerlo puesto que se encuentra en un “estado anómalo de conocimiento” (ASK) y por lo tanto el acercamiento a su necesidad debe hacerlo el sistema. Este desconocimiento previo del usuario va a dificultar a su vez que determine la relevancia de los resultados obtenidos, ya que no tiene los conocimientos adecuados para valorar en un primer momento la utilidad de esos documentos.

Desde que la persona siente esa necesidad hasta que la presenta a un SRI pasa por un estado en el que se concienta sobre su falta de conocimiento. Primero la describe mentalmente en su cerebro y después la traslada a una expresión formal y delimitada que será la que refleje en el sistema. En la siguiente fase debe indicar al sistema esa necesidad de información. Si se da el caso de que el usuario conoce el autor o el título del documento, expresarlo resulta bastante sencillo.

El problema surge cuando busca información sobre un tema, es decir, en la búsqueda por materias, y se agravará si se trata de un tema que desconoce, ya que tendrá dificultades tanto para plantear la consulta como para evaluar la relevancia de los resultados obtenidos. A esto se añade que a menudo el usuario plantea la consulta de acuerdo con lo que ya sabe sobre esa materia, y precisamente su necesidad de información está en lo que no sabe. Entonces, ¿cómo puede ayudar el sistema a que el usuario defina su necesidad de búsqueda y la exprese de la forma más adecuada? Un

buen diseño del acceso por materias permitirá al usuario encontrar su necesidad real de información y formular con precisión su consulta.

Para realizar la consulta correctamente y de esta manera obtener unos resultados satisfactorios en un sistema tradicional, deben cumplirse varios requisitos: por un lado, el usuario tiene que saber qué términos de indización se han usado en la base de datos, por otro, necesitaría saber cuáles de éstos se han elegido para indizar cada documento. Ambos requisitos son imposibles de conocer de antemano, por lo que es probable que se dé alguna de estas circunstancias:

- No se recupera ningún documento, lo que lleva a pensar que no hay ningún registro en esa base de datos sobre el tema que se busca.
- Se recuperan documentos que no responden a la necesidad de información.
- Se recibe como respuesta una cantidad excesiva de registros. Se ha comprobado que las personas tendemos a solicitar una información más genérica de la que realmente necesitamos; Ingwersen (1982) llama a este fenómeno “efecto etiqueta” (label effect). El resultado es el desbordamiento cognoscitivo con el consiguiente aumento de la dificultad para reconocer los documentos que resultan más relevantes.

3.2. Dificultad en la evaluación de los resultados obtenidos

Suponiendo que el usuario llega al tercer punto, está en la fase de evaluar los resultados obtenidos y reconocer los documentos que son relevantes a su consulta. La relevancia puede medirse en términos de precisión y exhaustividad, según se tome como referencia el conjunto total de documentos de interés para el usuario que hay en la base de datos, o el conjunto total de documentos recuperados por el usuario, respectivamente (Harter y Hert, 1997).

A pesar de existir fórmulas matemáticas que miden la relevancia, la subjetividad que entraña esta valoración hace que sea muy difícil determinar cuándo un resultado de la búsqueda es relevante y, en caso de que lo sea, en qué grado lo es. Es decir, que un documento puede ser más o menos relevante a una misma consulta dependiendo de quién sea el usuario y de sus propias circunstancias en un momento determinado. Y tampoco el usuario es siempre capaz de hacer esta valoración: si no conoce lo que quiere ¿cómo podrá evaluar su relevancia?

A la dificultad del propio usuario se suman las pocas facilidades que a su vez dan los SRI, que habitualmente no muestran suficiente información para que el usuario haga valoraciones y juzgue su utilidad. En el caso de los OPACs, se proporcionan los datos bibliográficos básicos y los encabezamientos de materias asignados al documento, pero no se relacionan los documentos con otros de temática similar ni se provee de ninguna información relativa a su contenido que pueda ayudar a dilucidar si resulta de interés a su necesidad de información.

Siguiendo el ejemplo de otros sistemas de información como las librerías en línea, SIRSI ha desarrollado iBistro, un OPAC vía web que está preparado para mostrar junto a cada título recuperado la portada del documento y da acceso al índice, al resumen y a una biografía del autor. En España se ha implantado recientemente (en el 2003) en las bibliotecas de las universidades Carlos III y Rey Juan Carlos, ambas en Madrid, si bien todavía no han aprovechado al máximo las capacidades del sistema. La figura 1 muestra

el sistema iBistro en una de las bases de datos de prueba, accesible desde <http://www.sirsi.com/Sirsiproducts/elibdemos.html>.

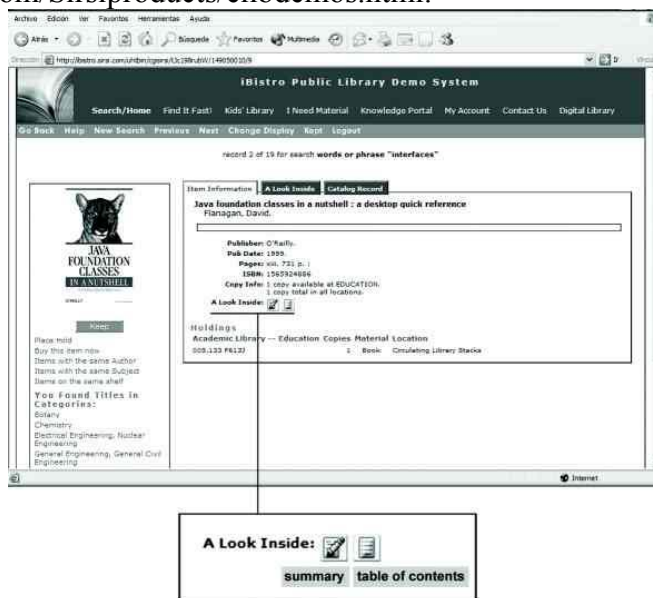


Figura 1. iBistro está preparado para proporcionar la imagen de la cubierta del libro, el resumen y el índice de los documentos

Otra dificultad más para que el usuario valore los resultados reside en que los OPACs habitualmente no ordenan los registros recuperados por orden de supuesta relevancia (coincidencia de términos de la consulta con términos de indización de los documentos), como ocurre en el caso del OPAC de la biblioteca de la Universidad de Zaragoza, por continuar con el mismo ejemplo del punto anterior. Esta carencia podría solucionarse empleando técnicas de la RI como han hecho los motores de búsqueda de la web, que incorporan, entre otros criterios de ranking, algoritmos de ordenación basados en la frecuencia de aparición de las palabras: a mayor número de palabras coincidentes entre la consulta y el documento, se presupone mayor relevancia.

En cuanto a la presentación de los resultados, advertimos que en la mayoría de los OPACs se hace en forma de listado, de manera que se obliga al usuario a revisar todos los registros sin poder obtener información sobre la relación que esos documentos guardan entre sí. Este inconveniente ha llevado a la investigación en RI a plantear sistemas basados en la agrupación (clustering) y en la visualización de la información.

3.3. Dificultad en la reformulación de la consulta

Desde los primeros años de la Documentación el fenómeno de la relevancia ha suscitado gran interés debido a la problemática que conlleva el propio concepto —¿qué es relevante y qué no lo es?—. Entre los primeros trabajos importantes se encuentran los de Saracevic (1975), que revisa el concepto en la bibliografía publicada hasta el año 1975 y propone una clasificación, o los posteriores de Froehlich (1994) en JASIS o de Buckley, Salton y Allen (1994) en el congreso de la ACM SIGIR, u otros más recientes como los de Mizzaro (1998) y Harter y Hert (1997). A pesar de que la bibliografía sobre la relevancia es muy abundante, la práctica resulta harto compleja y, de hecho, los sistemas no logran fácilmente este objetivo.

Mizzaro (1998) propone clasificar lo que se entiende por relevancia en función de cuatro factores que influyen en ella:

1. Cuál es la fuente de información
 - a. La referencia bibliográfica de un documento (documento secundario).
 - b. El documento en sí mismo (documento primario).
 - c. La información obtenida tras la asimilación de la información del documento (el estado de conocimiento del usuario).
2. Cuál es el problema que tiene el usuario que le hace recurrir a un sistema de RI para subsanarlo (idea expresada anteriormente por otros autores como Taylor, 1962)
 - a. Su necesidad real.
 - b. Su percepción de la necesidad.
 - c. Su forma de expresar en un lenguaje natural la necesidad percibida.
 - d. La forma en que se indica al sistema la necesidad expresada por el usuario en un lenguaje de búsqueda (la consulta o query).

En esta secuencia desde a hasta d se producen algunos problemas:

- En cada paso desde la necesidad real (a) hasta la forma en que el usuario lo expresa al sistema (d) puede producirse un error, de manera que la consulta efectuada al sistema no coincida con exactitud con la necesidad real del usuario.
- El usuario tiene dificultades para percibir la necesidad tal y como es realmente, puesto que se trata de buscar una información que desconoce, por lo que no siempre sabe qué busca. Es el llamado ASK (Anomalous State of Knowledge) que explican Belkin, Oddy y Brooks (1982a, 1982b).
- Para pasar del lenguaje natural con que el usuario expresa la necesidad percibida (c) a un lenguaje que el sistema sea capaz de reconocer (d), a menudo es necesaria la ayuda de un intermediario porque el usuario desconoce ese lenguaje de búsqueda que debe emplear.

3. En qué momento se produce la necesidad de información.
En función del momento, los resultados obtenidos no serán igual de relevantes, ni probablemente la necesidad real tampoco, ni su manera de percibirla ni de expresarla. Quiere decirse que un mismo usuario puede sentir una necesidad de información que expresa de la misma manera en distintas ocasiones y, en cambio, un mismo documento obtendrá una vez un alto ranking de relevancia para él y otra vez quizá no.

4. Los componentes
 - a. Tema sobre el que se busca información (qué).
 - b. Tarea que se va a realizar con esa información recibida (para qué).
 - c. Contexto. Por ejemplo, los documentos recuperados que ya se conocían, otros que no se comprenden por su dificultad o por su idioma, los que no da tiempo a leer, no se dispone de dinero para pagar por ellos, etc.

Añadimos a estos factores uno fundamental: quién es el usuario. Aquí se englobarían diversos aspectos como su nivel de conocimientos (su especialización en el tema, los idiomas que lee...), sus expectativas (qué espera encontrar), el uso que va a hacer de la información (para qué la quiere), sus gustos y preferencias (en cuanto a los autores, a los idiomas...), su forma de entender la información que se le proporciona, etc.

El SRI ideal debería abandonar el modelo clásico de relevancia, entendido como la obtención de referencias bibliográficas que respondan a la consulta efectuada por el usuario al sistema (query) en un momento determinado en cuanto al tema buscado, y optar por un modelo en el que se trate de resolver la necesidad real del usuario a lo largo del tiempo ofreciéndole información que tenga en cuenta tanto el tema sobre el que se busca como las tareas para las que se va a emplear la información y otros datos de contexto. Estas son las dos formas de entender la relevancia, dependiendo del punto de vista que se adopte: la relevancia para el sistema y la relevancia para el usuario. Tradicionalmente se ha tomado el primer punto de vista, ahora se tiende a centrarse en el usuario, que es al fin y al cabo quien precisa esa relevancia (Mizzaro, 1998).

En su artículo, Mizzaro (1998) propone que se diseñen interfaces que presenten la información clasificada en función de su relevancia de manera que el usuario pueda arrastrar los documentos (representados por ejemplo mediante puntos en el espacio) en virtud de los factores de relevancia que más le interesen en cada momento. La propuesta se basa en que un determinado documento puede ser muy relevante en cuanto al tema, pero no en cuanto a la tarea para la que se va a aplicar. Esto pone de nuevo de manifiesto la diferencia entre “relevancia para el sistema” y “relevancia para el usuario”, y la necesidad que hay de idear sistemas que presenten los documentos en función de ésta segunda.

Es fundamental que el usuario pueda reformular la consulta una vez obtenidos los resultados y así afinar el planteamiento para obtener documentos que se ajusten mejor a sus necesidades. Estas técnicas son las de retroalimentación por relevancia (relevance feedback): el ciclo de interacción en el que el usuario selecciona los documentos que considera relevantes, el sistema utiliza las características de los documentos seleccionados para revisar la consulta y vuelve a ejecutarla. Esta retroalimentación introduce posibilidades en el diseño, entre ellas las que permiten al usuario iniciar y controlar el proceso.

Bates (1990) defiende una postura intermedia en la que el sistema ayude a la búsqueda automática en un nivel estratégico (como ocurre con las cámaras de fotos en las que el usuario decide si dejar que sea automática o usarla con todas las opciones manuales).

La forma de que el usuario indique en la interfaz de resultados su valoración de la relevancia con la intención de mejorar su consulta suele consistir en marcar en la lista de títulos los que considera relevantes—incluso a veces es posible indicar el grado en que los considera relevantes— y también puede tener la posibilidad de marcar como no relevantes aquellos que no se lo parezcan. Estos datos son muy útiles para los algoritmos que aprenden a obtener documentos relevantes. Una vez que el sistema recoge las marcas hechas por el usuario, reformula la consulta modificando los pesos de los términos utilizados y puede hacer dos cosas: formular directamente la consulta u ofrecer una lista de términos al usuario para que marque los que le interesa buscar. Por último, da los nuevos resultados.

Koeneman y Belkin (1996) hicieron un experimento con varios niveles de retroalimentación y llegaron a la conclusión de que es preferible que el usuario controle los términos que se usarán para obtener mayor relevancia dándoles una lista para elegir. En su artículo explican en qué consistió: 64 alumnos de lengua inglesa debían buscar en una base de datos de casi 75.000 artículos del Wall Street Journal dos temas distintos.

Se les pedía que formularan una consulta con la que se obtuvieran los 30 documentos más relevantes (es decir, que crearan el perfil de búsqueda para esos temas). Con el estudio se quiso comprobar:

- si la búsqueda por comparación, el resultado ordenado por relevancia, la RI a texto completo combinados con un lenguaje sin operadores podrían ser usados de forma efectiva por usuarios finales sin formación en RI,
- hasta qué punto es interesante permitir la retroalimentación basada en la relevancia de los resultados obtenidos para los usuarios inexpertos en sistemas interactivos de RI,
- si es útil que el usuario conozca e intervenga activamente en esa retroalimentación y
- cómo influyen los diferentes niveles de interacción en la búsqueda de información en el número de consultas que crean y en la manera de hacerlas.

Se les ofreció cuatro interfaces de retroalimentación por relevancia: en la más básica el usuario introduce cada término de búsqueda y una vez obtenidos los resultados puede volver a hacer la búsqueda cambiando términos “a mano”; en la segunda, una vez hecha la búsqueda, el usuario puede marcar los documentos relevantes para que el sistema busque los más similares, o bien indicar que busque los similares a los primeros de la lista; la tercera añade a la versión anterior la lista de términos que se van a añadir a la consulta tras seleccionar los documentos más relevantes; y la última permite manipular esa lista antes de enviar la nueva consulta.

El resultado general fue que los que usaron los sistemas de retroalimentación obtuvieron mejores resultados que los que no los usaron. Además, en la entrevista posterior lo manifestaron así: preferían manipular los términos. Es decir, que los sistemas de RI para usuarios finales deben ser diseñados para permitir la interacción en la formulación de la consulta y la colaboración entre los usuarios y los programas.

En los últimos años, la retroalimentación de relevancia se está usando en los buscadores de la web para predecir qué documentos va a querer probablemente el usuario en el futuro, lo que se llama asistentes semiautomáticos o agentes, y suelen usar técnicas de aprendizaje (Baeza-Yates y Ribeiro-Neto, 1999). Unos funcionan preguntando al usuario por su objetivo y haciéndole elegir los términos relevantes, otros van recogiendo las acciones del usuario y hacen inferencias basadas en ellas. Kozierok y Maes (1993) hicieron un sistema de este tipo para organizar automáticamente el correo electrónico de un usuario, estudiando sus acciones y preguntándole antes de efectuarlas.

El sistema Letizia controlaba el comportamiento del usuario mientras navegaba y leía las páginas web: si la guarda en marcadores son relevantes, si no se pincha ningún enlace se considera que no lo son. El tiempo durante el que usuario se detiene en la lectura de una página también determinará una mayor o menor relevancia. Con esos criterios el sistema puede aprender a recomendar páginas del interés del usuario. También parece interesante estudiar los criterios de relevancia de grupos de personas para idear sistemas dirigidos a colectivos con intereses similares.

Baeza-Yates y Ribeiro-Neto (1999) llaman “pseudo-retroalimentación por relevancia” a aquella aplicada por los sistemas que asumen que los documentos con mayor puntuación en el ranking son los más relevantes y los usan para aumentar los términos de la consulta y reformularla. Al parecer, esto resulta de utilidad en consultas con un

gran número de resultados. Otra posibilidad para la retroalimentación podría consistir en tomar los resultados obtenidos, hacer un clustering y utilizar el resultado como nueva consulta, pero esta acción no se ha evaluado suficientemente como para determinar sus ventajas.

Como ejemplo de implementación de la retroalimentación por relevancia en un catálogo de biblioteca queremos mencionar Okapi, desarrollado en City University (Londres). Se trata de un sistema de recuperación vectorial, es decir, que recurre a los pesos de frecuencia de los términos para recuperar los documentos y mostrarlos por orden de relevancia. El usuario tiene la posibilidad de marcar los documentos más relevantes que ha recuperado y reformular la búsqueda, que automáticamente se lanzará al sistema ampliando los términos de consulta a aquellos más frecuentes en los documentos marcados como relevantes —a lo que se denomina expansión de la consulta— (Beaulieu y Jones, 1998). Recientemente se ha puesto a prueba el sistema en la base de datos para benchmarks TREC (Text REtrieval Conference) (Robertson, Walker y Beaulieu, 2000).

Como el concepto de retroalimentación de relevancia es desconocido para los usuarios y además difícil de implementar, los motores de búsqueda de la web disponen en ocasiones de una opción que, aunque no llega a ser una retroalimentación tal como la hemos explicado —quizá se podría entender más bien como una nueva consulta—, sí tiene como objetivo obtener resultados relevantes por comparación con alguno que, a criterio del usuario, lo es. Es la opción conocida como “more like this”: una vez obtenido un conjunto de documentos, el sistema ofrece la posibilidad de crear de forma automática una nueva consulta a partir de las características de uno de los recuperados que el usuario considere relevante. No suele obtener una buena relevancia, puesto que se necesitarían más criterios y basarse en más de un resultado.

Si concretamos la relevancia al ámbito de los OPACs, tenemos que tener en cuenta sus peculiaridades en relación con las bases de datos a texto completo, por ejemplo con los buscadores para la web, si bien en ambos casos encontramos semejanzas. Los buscadores utilizan como uno de los varios criterios de el número de apariciones de los términos de búsqueda en el texto completo y las partes de éste donde están, mientras que en el caso de los registros de OPACs cuentan con menos texto pero en cambio tienen la información muy estructurada en campos, y cuentan con algunos específicos para la descripción de contenido: materias y clasificación, lo que supone una ventaja para obtener resultados más relevantes.

Por este motivo vamos a proponer complementar la ordenación de los resultados siguiendo el criterio de la relevancia tradicional (el que se basa en la frecuencia de aparición) que establece el sistema y proponer como alternativa la presentación basada en el clustering. De esta forma, además de seguir un orden secuencial (de más a menos relevante) aparecerían los resultados en un mapa de asociaciones. Para esta propuesta nos basamos en la idea de que los resultados no son relevantes por sí mismos, sino que eso es algo que decide cada usuario en cada momento en función de sus necesidades de información puntuales. Como ya hemos señalado, ante una misma búsqueda de un mismo usuario en dos momentos diferentes, el criterio de relevancia de los documentos puede variar.

4. Diversidad de opciones para la diversidad de usuarios

Podemos dividir los tipos de usuarios en función de dos variables relacionadas con su conocimiento previo:

- la concreción de su necesidad de información
- su experiencia en el manejo del sistema de RI

Así, podemos establecer cuatro tipos generales de usuarios (véase la tabla 2):

Los usuarios de tipo A parten de una situación privilegiada, pues conocen con claridad su necesidad de información, están familiarizados con la colección y tienen un buen manejo de las opciones de búsqueda que permite el sistema. Para estos usuarios (entre los que se pueden encontrar bibliotecarios del propio centro o de otros centros) deben proveerse métodos rápidos de búsqueda que le permitan ir directamente a los documentos objeto de búsqueda, si bien no será el usuario mayoritario del sistema.

	Con necesidad de información bien definida	Con necesidad de información poco definida
Experto en el manejo del SRI	A	B
Novel en el manejo del SRI	C	D

Tabla 2. Tipología de usuarios

Los usuarios de tipo B también conocen bien la colección y la forma de consulta del OPAC, pero no tanto los objetivos de su búsqueda. Serían usuarios habituales que buscan información sobre determinado tema pero no están seguros de qué documentos podrían interesarles, quizá porque tienen poco conocimiento de la materia. Para ellos hay que habilitar una función de búsqueda similar al “paseo” por las estanterías, es decir, una opción de browsing temática que les evite plantear una consulta y en cambio les dé como resultado documentos cercanos a su necesidad de información. En estos casos en los que el usuario no conoce con exactitud qué necesita encontrar, irá identificando de entre los documentos que obtenga aquellos que considere de interés.

Los usuarios de tipo C y D no están muy familiarizados con el OPAC que van a utilizar, lo que supone que no conocen bien la colección de la biblioteca y menos aún el lenguaje de consulta que usa el sistema. Los de tipo C al menos tienen una idea clara de su necesidad de información, pero cuentan con el problema de no saber formular la consulta al sistema. Para ellos deben habilitarse métodos que les permitan llegar a los documentos de una forma directa pero sin pasar por una consulta que implique el manejo de operadores. Una interrogación sencilla con uno o dos términos de búsqueda podría darles un primer resultado sobre el que luego afinar.

Los usuarios de tipo D probablemente han llegado a él de una forma casual y van a querer ojear qué contiene, por si es de su interés. No hay que dejar de lado este usuario potencial, sino cubrir su necesidad de información de la manera más efectiva posible. La forma idónea de darle a conocer la colección es ofrecerle una visión general de ésta y permitirle navegar por los distintos temas para que llegue a documentos de su interés. Al no conocer el sistema de búsqueda, el browsing será la mejor opción para este usuario casual, puesto que no se le puede exigir que plantee los términos de búsqueda en su consulta.

5. Un prototipo que intenta dar soluciones

El prototipo CHILE (Ardila, Marcos, Baeza, 2004), cuyo nombre responde a Computer-Human Interaction Librarian Experience, pretende mejorar la interacción que se da entre las personas y los sistemas de consulta de las bibliotecas. Para ello, se da la posibilidad de búsqueda mediante consulta y ojeo, y la presentación de la información en un mapa visual que ofrece tanto una visión general de la colección como del conjunto de documentos recuperados.

Su interfaz se presenta de forma visual, tanto para mostrar toda la colección como la parte correspondiente al perfil de un usuario registrado, así como los resultados obtenidos tras la consulta. La colocación de la información en la pantalla es indicativa de la cercanía temática entre los documentos. El sistema de visualización elegido es una variante de los mapas auto-organizativos (SOM), donde la ubicación de los objetos o grupos de objetos depende de cálculos vectoriales sobre las posiciones de los objetos ya existentes en el mapa.

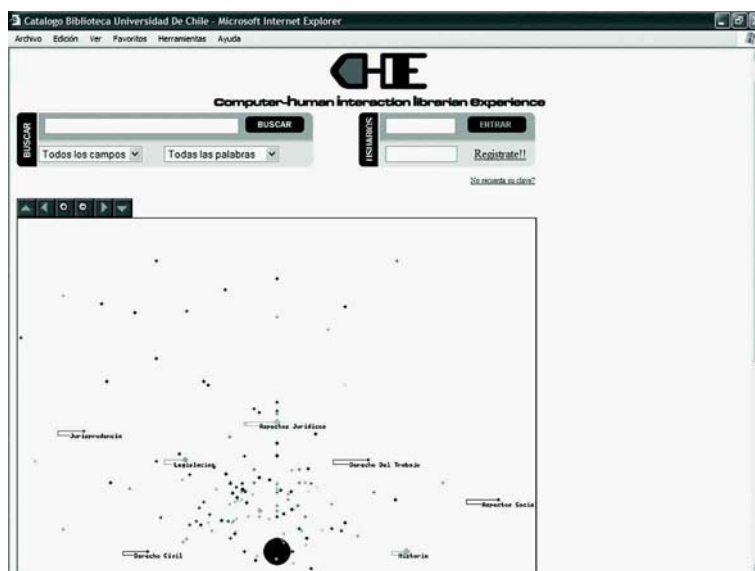


Figura 2. Prototipo CHILE: visión general de la colección en forma de mapa

Tal y como se muestra en la figura 2, el prototipo CHILE permite visualizar libros clasificados por materias. La dificultad radica en los pocos puntos de referencia que hay en los registros bibliográficos de los OPACs para detectar semejanzas temáticas entre documentos, a diferencia de otras bases de datos que contienen el texto completo de los documentos, o al menos un campo de resumen. A pesar de ello, el resultado de agrupar por materias y clasificación resulta positivo.

6. Bibliografía

- Ardila, A.; Marcos, M. C.; Baeza, R. (2004). "CHILE: un prototipo visual de catálogo de biblioteca". *Actas de Interacción 2004* (Lérida, mayo de 2004), en prensa.
- Baeza-Yates, R.; Ribeiro-Neto, B. (1999). *Modern Information Retrieval*. Reading (Ma): Addison Wesley.

- Beaulieu, M.; Jones, S. (1998). Interactive searching and interface issues in the Okapi best match probabilistic retrieval system. *Interacting with Computers*, 10, 237-248.
- Belkin, N.; Oddy, R.; Brooks, H. (1982a). ASK for information retrieval. Part I: Background and theory. *Journal of Documentation*, 38:2, 61-71.
- Belkin, N.; Oddy, R.; Brooks, H. (1982b). ASK for information retrieval. Part II: Results of a design study. *Journal of Documentation*, 38:3, 145-164.
- Borgman, C. (1996). Why are online catalogs still hard to use? *Journal of the American Society for Information Science*, 47:7, 493-503.
- Buckley, C.; Salton, G.; Allen, J. (1994). The Effect of Adding Relevance Information in a Relevance Feedback Environment. En: Croft, W.; van Rijsbergen, C. (eds.). *Proceedings of the Seventeenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. Springer-Verlag, London, 292-301.
- Fernández Molina, J. C.; Moya Anegón, F. de. (1998). *Los catálogos de acceso público en línea: el futuro de la recuperación de información bibliográfica*. Málaga: Asociación Andaluza de Bibliotecarios.
- Froelich, T.J. (1994). Relevance reconsidered: towards an Agenda for the 21st Century: Introduction to the Special Topic Issue on Relevance Research. *Journal of the American Society for Information Science*, 45, 124-134.
- Harter, S.; Hert, C. (1997). Evaluation of information retrieval systems: approaches, issues, and methods. *Annual Review of Information Science and Technology*, 32, 3-94.
- Ingwersen, P. (1982). Search procedures in the library analysed from the cognitive point of view. *Journal of Documentation*, 38:3, 165-191.
- ISO 9241 (1992-2000). Ergonomic requirements for office work with visual display terminals (VDTs), <http://www.iso.ch/>
- Játiva Miralles, M.V. (2002). Opac-portal: una nueva forma de ofrecer los recursos y servicios de la biblioteca. *El Profesional de la Información*, 11:6, 442-453.
- Koenemann, J.; Belkin, N. (1996). A Case for Interaction A Study of Interactive Information Retrieval Behavior and Effectiveness. *CHI '96*.
- Mizzaro, S. (1998). How many relevances in information retrieval? *Interacting with Computers*, 10, 303-320.
- Marcos, M. C. (2004). Interacción en interfaces de recuperación de información: concepto, metáfora y visualización. Gijón: Trea.
- Moscoso, P. (1998). Análisis y evaluación de catálogos automatizados de acceso público en entorno web. *Revista Española de Documentación Científica*, 21:1-2, 57-75.
- Moscoso, P.; Ortiz-Repiso, V.; Ríos, Y. (1998). Telnet versus web en el acceso a la información catalográfica: estudio comparativo del caso español. *VI Jornadas Españolas de Documentación : Los Sistemas de Información al Servicio de la Sociedad (Valencia, 29, 30 y 31 octubre 1998)*, http://fesabid98.florida-uni.es/Comunicaciones/p_moscoso.htm
- Ortiz-Repiso Jiménez, V.; Moscoso, P. (1999). Web-based OPACs: between tradition and innovation. *Information Technology & Libraries*, 18:2, 68-77.
- Peis, E.; Fernández Molina, J.C. (1996). Los OPACs supermercados de información bibliográfica. *Jornadas Bibliotecarias de Andalucía*, 468-475.
- Robertson, S.; Walker, S.; Beaulieu, M. (2000). Experimentation as a way of life: Okapi at TREC. *Information Processing & Management*, 36:1, 95-108.
- Saracevic, T. (1975). Relevance: a review of and a framework for the thinking on the notion in information science. *Journal of American Society of Information Science*, 26:6, 321-343.

Taylor, R. (1962). The process of asking questions. *American Documentation*, 13:4, 391-397.

Wersig, G. (1979). The Problematic Situation as a Basic Concept of Information Science in the Framework of the Social Sciences: A Reply to Belkin. En: International Federation for Documentation. *Theoretical Problems of Informatics: New Trends in Informatics and its Terminology*. FID 568. Moscow: VINITI, 48-57.