

DE LA CONSULTA DE CATÁLOGOS A LA GESTIÓN DE INFORMACIÓN: TENSIONES HACIA EL CAMBIO EN EL DISEÑO DE OPACs

FRANCISCO JAVIER GARCÍA MARCO
Facultad de Filosofía y Letras, Zaragoza

1. INTRODUCCIÓN

El momento actual se caracteriza por una auténtica revolución en el campo de las comunicaciones y la gestión de la información, consistente en la automatización y superación de las técnicas tradicionales soportadas por el papel y derivados. La disparidad entre lo nuevo y lo dado nos sitúa en un momento todavía transicional en el que se lucha por sacar el mejor partido a las enormes posibilidades de las nuevas tecnologías en medio de fuertes tensiones de adaptación. Ello queda reflejado en la investigación sobre OPACs, que hasta el momento presente se ha dirigido fundamentalmente, en una primera fase, a la automatización de las operaciones y usos tradicionales de las bibliotecas; y, en una segunda fase, a la incorporación gradual de nuevas técnicas de descripción, control y recuperación lingüística. Esta investigación ha producido ya enormes beneficios; sin embargo, tanto los problemas percibidos en la utilización cotidiana de los OPAC por los usuarios (condicionamientos de usuario: necesidades) como los recientes avances en su diseño y tecnología (condicionamientos de sistema: posibilidades) sugieren su continua actualización.

2. NUEVOS REQUERIMIENTOS DEL USUARIO AL SISTEMA DE RECUPERACIÓN

1. *Atención a los distintos tipos y niveles de usuarios.* Los usuarios que acuden a una biblioteca necesitados de información no constituyen un grupo homogéneo, sino que están muy diversificados en necesidades, habilidades generales y capacidades especializadas para la recuperación de in-

formación, y ello —entre otros factores— en función de su edad, cultura y adaptación tecnológica, que determinan a su vez su capacidad de simbolización y comunicación ante el OPAC. El mejor acercamiento al problema es, por tanto, presentar a los diferentes usuarios la posibilidad de utilizar diferentes códigos y canales de comunicación; sin olvidar que es tan importante suplir las carencias de ciertos grupos de usuarios (niños, estudiantes noveles, minusválidos, etc.) como permitir a los usuarios avanzados un completo uso de sus habilidades documentales.

2. *Acceso remoto e Información global (IG)*. Entendemos por «IG» la posibilidad de realizar búsquedas telemáticas de referencias documentales y documentos primarios en distintas fuentes y soportes descritos, indizados, referidos o reproducidos —íntegra o parcialmente— por medios electrónicos, y de contrastar y organizar los resultados de la consulta de dichas fuentes. El enorme crecimiento en la oferta de servicios de información —con la presión de acceso que ello ha generado sobre la demanda— y el desarrollo de las telecomunicaciones ha favorecido un intenso proceso de integración de servicios de información hasta ahora relativamente separados a consecuencia de características diferenciales que, a su vez, planteaban exigencias de especialización en su mantenimiento (servicios bibliográficos, bibliotecas de distintos tipos, archivos, oficinas de información, etc.). Ciertamente, desde la perspectiva del usuario este fraccionamiento resulta muy incómodo, reduce su capacidad de acceso directo a la información, y multiplica la inversión en recursos y tiempo al exigir costosos desplazamientos. Por ello, los servicios de información y aun las redes de centros que atienden a los diferentes grupos de usuarios se han visto progresivamente obligados a abrir sus sistemas de información a dichas necesidades. Como consecuencia de ello, las fronteras entre servicios de orientación bibliográfica y los servicios de lectura en sala, reproducción y préstamo se están difuminando. En la medida en que la biblioteca quiera ser un servicio de información además de un simple depósito accesible al usuario, el OPAC tradicional habrá de dar paso al terminal inteligente de teledocumentación, que proporcione acceso selectivo e integrado a sistemas remotos; permitiendo al usuario manejarse con plena autonomía en el mundo de la información, ampliando su espectro de búsqueda, y concediéndole la posibilidad de contrastar diversas fuentes de referencia y diversas posibilidades de consulta con especificación de sus respectivos costes y calidades. De hecho, se han puesto recientemente en actividad en bibliotecas importantes productos de acceso remoto a bases de datos y catálogos informatizados, aprovechando la base de comunicaciones electrónicas que proporciona de por sí la infraestructura del OPAC local. Los objetivos más inmediatos han sido la integración en redes interbibliotecarias —movimiento muy desarrollado en Norte-

américa y una de las fronteras inmediatas de la política europea de bibliotecas—, la integración de CD-ROMs en el OPAC local, y la apertura de comunicaciones con redes de bases de datos. Un ejemplo del último caso es el Bell Atlantic Gateway que proporciona acceso a más de 30 hosts, a la cual están conectados los OPAC de bibliotecas universitarias como las Lehigh University Libraries (ASA)¹. Por último, los OPAC de numerosas bibliotecas han integrado servicios de información no bibliográficos, tales como información local, correo electrónico, foros de discusión, etc. En conclusión, los catálogos informatizados de las bibliotecas deberán ser accesibles mediante terminales tanto internas como externas al centro, y permitir la obtención de referencias, adquisición o préstamo de los documentos pertinentes desde dichas estaciones de trabajo a partir de muy diversas fuentes, así como abrir la biblioteca a otras necesidades de información de sus usuarios.

3. *Asistencia en el acceso temático a la colección.* El usuario se enfrenta con una serie de problemas ante el OPAC que impiden el aprovechamiento de las posibilidades que le ofrece la tecnología actual: exigencia de un buen conocimiento previo de la colección para la búsqueda temática efectiva, restricciones de comunicación consecuencia de la disparidad entre probabilidad de ejecución y disponibilidad por código, disparidad humana entre la capacidad de reconocimiento y de descripción, peligro de que la búsqueda aborte inadvertidamente a causa de fallos ortográficos, etc. Para Hildreth la raíz del problema radica en un paradigma anticuado de OPAC basado en la sola respuesta a preguntas que se suponen que van a ser específicas («Known-subject need, best-match, end product-oriented» information retrieval paradigm), pero que no da cuenta de necesidades más ambiguas o difíciles de definir².

4. *Gestión compleja de la información: integración de datos de diversas fuentes y presentación en pantalla.* Ante la necesidad ya planteada de discriminar y contrastar registros de distintas fuentes, organizar la información obtenida, flexibilizar la presentación en pantalla, etc., caben dos posibilidades. La primera de ellas es adaptar el sistema integrado de gestión bibliotecaria incorporando rutinas que lo permitan, tal y como se ha intentado por ejemplo en la segunda versión del OPAC del sistema interbi-

¹ JONHSON, Jean M. ASA on the Bell Atlantic Gateway: the Lehigh experience. En Martha E. Williams (ed.). *National Online Meeting 1990. Proceedings of the 11th National Online Meeting, New York, 13 May 1990*, Medford, New Jersey: Learned Information, Inc., 1990, p. 187-192.

² HILDRETH, Charles R. End users and structured searching of online catalogues: Recent research findings. En Robert FUGMANN (ed.). *Tools for Knowledge Organization and the Human Interface*, 2 vols., Frankfurt/Main: Indeks Verlag, 1991, vol. 2, p. 11.

bliotecario de Birmingham³. La segunda es ofrecer software en la terminal inteligente que se encargue de convertir la información extraída de diversas fuentes en un fichero organizado manipulable por el usuario. Sin entrar en las complejidades del tratamiento de texto, lo cierto es que un terminal inteligente deberá proporcionar utilidades semejantes al menos a las de los gestores de bases de datos (ordenar, disponer la información en distintas presentaciones, añadir, eliminar y modificar registros, darles salida mediante periféricos, etc.). Además sería conveniente que portaran utilidades específicas que le ayudaran a organizar su estrategia de consulta y lectura de acuerdo a la tipología documental y sus particulares intereses, según han sido explicitados en la petición de búsqueda. En este campo existen interesantes posibilidades de aplicación de la IA.

5. *Bilingüismo y multilingüismo*. Sin duda, uno de los inconvenientes más graves que se plantean en la transmisión de la información es la barrera lingüística, que complica extraordinariamente el acceso temático. Las soluciones ensayadas están sin duda más avanzadas allí donde el problema es más radical, es decir, en la comunidades multilingües y organismos internacionales. Las clasificaciones codificadas y los tesauros multilingües han sido las respuestas más eficaces al problema, aunque existen también listas de encabezamientos plurilingües, por ejemplo la del sistema Hermes⁴.

6. *Obtención de los documentos primarios*, organizando y gestionando la estrategia de localización física (préstamo de biblioteca local, servicios de préstamo interbibliotecarios, adquisición telemática de documentos), bajo el control de la entidad global (biblioteca, servicio de documentación de la institución) de acuerdo a parámetros tales como el costo económico y el nivel del usuario. Aquí son también posibles rutinas automáticas de evaluación.

Concluyendo, se echa de menos un grupo de productos que permitan el acceso a múltiples bases de datos con múltiples lenguajes de interrogación y de indización, que asesoren eficazmente al usuario en el proceso de recuperación, y que permitan una gestión avanzada de la información obtenida. Precisamente los nuevos avances tecnológicos han venido a proporcionar respuestas a esta situación por medio de productos como los gateway o los terminales inteligentes con asistencia automática en la búsqueda.

³ WELLINGS, Richard. Providing online access to different views of the library catalogue: further development of BLCMP's OPAC. *Program*, oct. 1990, vol. 24, no. 4, p. 343-356.

⁴ HOLMES, David; BULGER, Derrick. A day in the life of a public terminal—a transaction analysis of an online catalogue terminal in a bilingual analysis. *Canadian-Journal-of-Information-Science*; Dec. 1988, vol. 13, no. 3/4, p. 21-33. Sistema Hermes de la Ottawa University.

3. VARIABLES DEL SISTEMA GLOBAL QUE INCIDEN EN LA RECUPERACIÓN

Seis aspectos fundamentales del sistema de información condicionan decisivamente el proceso de recuperación de la información: el entorno de trabajo; las posibilidades técnicas del hardware que determinan a su vez los límites del sistema; la estructura lógica que organiza los datos; los sistemas de indización utilizados para la descripción lingüística de los documentos; y las posibilidades operativas de interrogación:

1. *El entorno de trabajo.* No se deben descuidar los aspectos materiales —de menor importancia formal pero de gran importancia práctica— del acceso a la información tales como el entorno físico de trabajo, la disposición de los periféricos, etc. Además, como se ha repetido insistentemente, el acceso automático no agota ni mucho menos los problemas de accesibilidad en la biblioteca.

2. *Soluciones tecnológicas adoptadas (Hardware y sistemas operativos).* Quizá el avance más importante en este campo ha sido la tendencia a adoptar sistemas informáticos abiertos, para liberar a la biblioteca o al servicio de información de la onerosa dependencia de un fabricante y/o de un diseñador de software. Dentro de la tendencia creciente a optar por sistemas estándares se inscribe la creciente adopción por un importante número de centros del sistema operativo UNIX, cuyas ventajas son evidentes para la biblioteca al tratarse de un entorno multiusuario y multiárea. Otro factor importante en este sentido han sido el giro que han tomado algunas bibliotecas desde los grandes ordenadores (*mainframes*) y terminales colgados a la opción por el uso de terminales inteligentes, llegando, en algunos casos extremos, al abandono de las grandes redes en beneficio del ordenador personal y el CD-ROM combinando con discos flexibles para reactualizar los discos duros. De hecho hay un importante debate abierto sobre la compatibilización o competencia entre OPAC y CD-ROM.

3. *La estructura lógica de organización de los datos:* Las ventajas que ofrece el principio de separación entre la base de datos y el gestor para asegurar la portabilidad de la información —asegurando de paso la posibilidad de cambiar de gestor sin grandes traumas— ha asegurado la rápida adopción de los formatos compatibles MARC y abierto el camino hacia la adopción de un UNIMARC. Como no podía ser de otra manera, ha habido importantes problemas: a) la enorme extensión de los formatos de descripción provocan el abandono local de campos y la utilización de subconjuntos, que no son siempre predecibles debido a la ausencia de normalización en la determinación de niveles de descripción; b) además, los

OPAC en servicio desaprovechan información importante que portan muchos estándares de almacenamiento, como es manifiesto en los distintos MARC; c) por fin, la complejidad de la categorización en formatos MARC ha llevado a cierta insatisfacción en los investigadores que trabajan en optimizar la accesibilidad al registro, lo que ha conducido a la formulación de propuestas de reforma de las reglas de descripción⁵.

4. *Lenguajes documentales.* Los OPAC han aportado un ambiente único para favorecer la mutación de los sistemas lingüísticos de descripción y recuperación de la información. La variedad de sistemas de representación semántica añade ciertamente problemas logísticos a la recuperación. Sin embargo, la diversidad de información que portan los formatos normalizados —asociada a los nuevos sistemas de catalogación basados en la importación de registros y adecuación al entorno local— ofrece grandes posibilidades en espera del desarrollo de lenguajes de descripción internacionales, ya que los distintos lenguajes documentales son susceptibles de funcionar como bases terminológicas sobre las que expandir los puntos de acceso de forma automática. De hecho, la gran ventaja ofrecida por los OPAC en el acceso a la información ha sido la posibilidad de utilizar simultáneamente distintas vías sin multiplicar el trabajo de mantenimiento y consulta de ficheros. La diversidad de puntos de acceso del registro MARC ha favorecido una fuerte tendencia actual a utilizar todas las vías posibles de acceso temático —por ejemplo, encabezamientos controlados en unión de búsqueda en texto libre en el campo del título—, en consonancia con un principio de mínimos bien establecido en la investigación sobre acceso temático: no se puede dejar al usuario sin una probabilidad alta de encontrar al menos un registro relevante a su petición. Paralelamente, la informatización de los catálogos ha potenciado las posibilidades de control del vocabulario de recuperación e indización y el mantenimiento de ficheros de autoridades, acercando cada vez más —por vía de desplazamiento, evolución o de compatibilización— los sistemas de encabezamiento de materias a los lenguajes controlados más evolucionados. Al contrario de lo que se preveía, la adopción de OPACs ha reforzado y simplificado la utilización de clasificaciones y códigos de clasificación, puesto que el ordenador posibilita un fácil acceso alfabético, secuencial y jerárquico a la propia clasificación, solventando los problemas que el fichero sistemático impone al usuario no especializado. En Europa el sistema más usado es la CDU. Sus ventajas para la recuperación automatizada son bien conocidas: es un sistema transnacional, su código

⁵ OLIVING Berg. Current Problems with the MARC/ISBD Formats in relation to Online Public Access of Bibliographic Information, *International Cataloguing & Bibliographic Control*, January-March 1991, p. 12-14.

alfanumérico lo hace independiente de las lenguas concretas y, además, porta en escasos caracteres una gran cantidad de información. Todo ello ha provocado su adopción como sistema de recuperación temática en diversos sistemas⁶. Otra opción interesante ha sido ampliar los términos de búsqueda multipalabra introduciendo en el registro MARC descriptores del índice o la tabla de contenidos de la publicación. Además, el formato MARC porta importantes informaciones bibliográficas que no se suelen utilizar en la recuperación pero pueden ser infinitamente útiles a los usuarios (tipo de documento, etc.). Por último, no debemos olvidar que el desarrollo de la «navegación» en bases de datos gracias al hipertexto facilita enormemente el acceso preferente, asociativo y jerárquico, frente a la tradicional disposición en lista que forzaban los soportes en papel. Igualmente el hipertexto rompe la tradicional disyuntiva entre búsqueda de cadenas y términos controlados y búsqueda de palabras clave discretas, al poder referir ambos vocabularios uno al otro gracias a los diccionarios invertidos. No obstante, se ha demostrado que puede ser útil permitir al usuario acceder al examen secuencial del fichero topográfico, que como es sabido, obedece en la bibliotecas bien organizadas a una clasificación temática (HILDRETH, 19).

5. *Los lenguajes de recuperación.* Los lenguajes de interrogación se originaron como un subconjunto de los lenguajes de programación de los gestores de bases de datos. Actualmente, la integración de sistemas y la creación de una red mundial de la información han puesto en candelero el problema de la multiplicidad de lenguajes y dialectos de recuperación, dado el gran número de productores y distribuidores. Ello ha originado un esfuerzo consistente de proponer estándares independientes del lenguaje de programación que faciliten a los usuarios la búsqueda en línea, fruto del cual han surgido iniciativas de integración de la CEE, la ANSI, la ISO, etc. En el momento actual —también en esto de carácter transicional— se camina también hacia la adopción de lenguajes de recuperación normalizados al menos a nivel macroregional: SQL, CCL, LINCE, etc. Por ello, y durante algunos años, los desarrolladores de sistemas centrarán una buena parte de sus esfuerzos en crear módulos de comunicación entre los distintos lenguajes. En Europa la CEE lleva una intensa política de promoción del estándar CCL (Common Command Language) de cara a la construcción de un espacio europeo de la información. Existe un movimiento actual hacia la construcción de traductores que, independiente-

⁶ HUG, Hannes; META, Walser. Retrieval in the ETH Database Using the UDC. En Robert FUGMANN (ed.). *Tools for Knowledge Organization and the Human Interface*, 2 vols., Frankfurt/Main: Indeks Verlag, 1991, vol. 2, p. 202-209.

mente del lenguaje original del productor, favorezcan la adopción del CCL como lenguaje de búsqueda intracomunitario.

5. FACILITACIÓN DE LA INTERACCIÓN HOMBRE/MÁQUINA A TRAVÉS DEL INTERFACE DE USUARIO

La complejidad de los lenguajes de las bases de datos relacionales o documentales que se han utilizado para gestionar su información ha provocado una fuerte tendencia al desarrollo de interfaces que faciliten al usuario su diálogo con el sistema, cuestión absolutamente vital en el caso del acceso público. A veces, este esfuerzo se ha realizado incluso a espaldas de los fabricantes de los gestores.

Uno de los grandes avances en accesibilidad ha sido la adopción de interfaces gráficos de usuario, orientados a objetos e inteligibles intuitivamente sin necesidad de estudio de las complejas formalizaciones lógicas del lenguaje de recuperación. El interface gráfico de usuario (IGU) se debe concebir como algo separado de la estructura lógica y del lenguaje de recuperación, así cabrá la posibilidad de adaptar el sistema a los distintos tipos de usuarios con la máxima exigencia en la recuperación. Se han hecho algunas propuestas teóricas en cuanto al modo de interacción entre el usuario y el sistema. Los dos niveles mínimos son un tránsito por pantallas a través del menú para los usuarios inexpertos y acceso al lenguaje de recuperación para los usuarios más expertos, que sin ninguna duda ofrece mayores posibilidades de pertinencia y exhaustividad en la recuperación. La amigabilidad del sistema se ha solido centrar en factores como el acceso metafórico o simbólico, y, con menos ambiciones, en la claridad de la presentación, pero incluso detalles pequeños como el intervalo de desuso deben ser cuidadosamente establecidos.

Un aspecto fundamental del interface de usuario es la documentación aneja sea on-line o impresa. Lo ideal sería un sistema de ayuda basado en el principio de las zonas de aprendizaje posible de WIGOTSKI o similar, capaz de tutorizar el tránsito de niveles en usuarios deseosos de conseguirlo y que, en todo caso, parta realmente de los conocimientos mínimos de los presuntos usuarios. No se les debe presuponer ninguna capacidad de manejo del ordenador ni de búsqueda documental. La ayuda debería ser provista tanto para cada una de los objetos u opciones como en la forma de un manual en línea, y mejor, integrando ambos aspectos. En la realización del manual de ayuda no se debería olvidar que la búsqueda en línea no es sino un aspecto más de la búsqueda bibliográfica y documental en cuyos principios se basa. Es decir, el manual de ayuda debería proporcionar un marco general de la búsqueda documental al usua-

rio, e incluso distintos niveles de ayuda (novato, experimentado, avanzado). A pesar de todo, el bibliotecario referencialista será siempre el auxiliar indispensable en la búsqueda.

6. LAS POSIBILIDADES DEL PROCESAMIENTO LINGÜÍSTICO AUTOMÁTICO

El procesamiento del lenguaje natural ha dejado de ser ciencia ficción para convertirse en uno de los campos más prometedores de la ciencia de la Información. Actualmente está siendo aplicado también a los OPAC. La idea consiste en utilizar las regularidades lingüísticas bien establecidas como algoritmos deterministas válidos para el procesamiento simbólico que realiza el ordenador: a) tratamiento de errores, variantes y signos ortográficos; b) eliminación de palabras no pertinentes (Stop-list); c) análisis morfológico que permita la extracción de raíces, método seguro para facilitar la expansión de la búsqueda tanto por equivalencia de la cadena de caracteres como para posibilitar secundariamente la búsqueda de nuevos términos asociados, equivalentes o relacionados jerárquicamente; d) análisis semántico con el apoyo de ficheros de sinónimos, sinónimos documentales, antónimos, etc. y otras bases terminológicas (diccionarios y léxicos generales y especializados); e) análisis sintáctico, traduciendo la sintaxis natural a expresiones booleanas; f) análisis sociolingüístico, ampliando las relaciones y unidades en los vocabularios de búsqueda tomando como base la coocurrencia de asociaciones empíricas entre términos en una población de usuarios. Es importante volver a señalar que la utilización de varios léxicos documentales favorece automáticamente la expansión del vocabulario de búsqueda y de las propias relaciones léxicas susceptibles de ampliar el ámbito de la búsqueda.

Un acercamiento importante ha sido la adopción de métodos semejantes —aunque inversamente aplicados— a los de la indización automática para favorecer la expansión del vocabulario de búsqueda. Así por ejemplo, el sistema OKAPI se basa en la ponderación de los términos de búsqueda (siempre que sean más de dos) de forma inversamente proporcional a su frecuencia en la base de datos y en la suma de los pesos de los términos empleados en la pregunta de tal forma que se recupera el registro si este supera un límite presupuesto y presenta los registros recuperados en orden directamente proporcional al peso de cada uno. Una vez encontrado al menos un ítem es posible expandir la búsqueda a través de su código de clasificación, autor, serie, encabezamiento, extrayendo a su vez nuevos términos del registro recuperado para expandir la expresión de búsqueda original. Okapi utiliza además expansión automática de la búsqueda en caso de búsqueda nula sustituyendo la intersección implícita por un operador de suma lógica.

7. CONCLUSIONES

Visto lo anterior, podemos resumir las tendencias en el desarrollo de OPAC en los siguientes puntos:

1. Los sistemas están cada vez más dirigidos a la atención de las necesidades del usuario, no sólo a la gestión bibliotecaria y al control bibliográfico, y ello mediante la provisión de acceso a distintos niveles según las necesidades y capacidades de los usuarios, y tutorizando en la medida de lo posible el tránsito entre niveles de uso. Dentro de esta tendencia, ha cobrado especial incidencia en el interface de usuario como medio de facilitar el diálogo hombre máquina, mediante el uso de los lenguajes orientados a objetos con relevancia simbólica para el usuario.

2. Los OPAC —sin perder importancia su funcionalidad de descripción y gestión de la colección local— se están convirtiendo progresivamente en accesos para tareas de información global, en parte como justificación de las altas inversiones informáticas en bibliotecas en términos de servicio público y en parte como medio de una información bibliográfica más eficaz. De forma inversa, habrá que preparar los OPAC locales para ser accedidos remotamente. Por todo ello, es necesario dotarlos de posibilidades de gestionar la información proveniente de diversas fuentes.

3. La utilización de estándares a todos los niveles se ha demostrado la base a partir de la cual intentar las compatibilizaciones de sistemas, aspecto fundamental en la racionalización de un mercado en vías de macro-regionalización y mundialización: a) en la estructura lógica de los registros mediante la adopción de formatos MARC o del normalizado UNIMARC, b) en la adopción de lenguajes de recuperación normalizados (CCL en Europa), c) incluso en los lenguajes de programación y sistemas operativos.

4. Adopción de sistemas informáticos abiertos como soporte del sistema de gestión de información.

5. Existe un fuerte movimiento hacia la compatibilización e interconexión de los diferentes sistemas de descripción y acceso temático como medio de aprovechar distintas fuentes de información; así como utilización de la simulación automática de comportamientos inteligentes (IA) en la asistencia a la recuperación temática.

6. Se están dando los primeros pasos en la automatización —aun limitada en su eficacia— de algoritmos de búsqueda.