

HACIA UNA OPTIMIZACION DE LOS RECURSOS INTERNET EN LA EMPRESA

José Antonio Sánchez Montero*

Resumen: A través de la experiencia práctica de implantación de un sistema integrado de recursos de información Internet en un entorno empresarial, se describen los principales problemas para la recuperación de información de calidad en Internet, el diseño de esquemas de navegación en entornos hipermediales, y una metodología para la selección y gestión de recursos de información Internet.

Palabras Clave: Internet, Intranet, Hipermedia, Recuperación de información, Gestión de recursos de información Internet, World Wide Web.

Abstract: Based on a real experience of an Internet integrated information system within a business environment, this article analyzes the main problems for an Internet quality information retrieval, describes a hypermedia navigation framework and, finally, establishes a methodology for the selection and management of Internet information resources.

Keywords: Internet, Intranet, Hypermedia, Information retrieval, Internet information resources management, World Wide Web.

Dos artículos recientemente publicados en la revista *Information World Review* son un claro ejemplo del revolucionario impacto de Internet, y más concretamente, del World Wide Web en nuestra profesión. En el primero (1), Paul Blake, Vicepresidente ejecutivo de *Knight Rider* (Dialog y Datastar) señala el esfuerzo que están realizando para adaptar las bases de datos que comercializan a una estructura hipertextual compatible con plataformas Web, y en la que existirían «links» tanto internos, en cada base de datos, como externos entre las mismas, relacionándolas temáticamente. En el segundo (2), se describe como la SEC (*Securities Exchange Commission*), el organismo regulador del mercado de valores estadounidense, utiliza la red como una plataforma universal de difusión instantánea de la información de las compañías, con objeto de incrementar la eficiencia y transparencia del mercado de valores. Explica cómo tan sólo veinticuatro horas después de finalizar el plazo para que las compañías envíen la información requerida por la ley a la SEC, la información ya se encontraba disponible en el Web, tanto vía FTP como a través del servidor WWW.

Sin duda, estos dos ejemplos ilustran dos de las grandes potencialidades del World Wide Web: la navegación hipertextual, o hipermedial, de carácter asociativo, no lineal, por un lado, y su utilización a modo de gigantesca infraestructura digital, a través de la que podemos acceder a ingentes cantidades de información, por otro. La biblioteca o el centro de documentación de cualquier organización puede aprovechar, también, las ventajas de este entorno y comenzar a ofrecer nuevos servicios a nuestros usuarios basados en esta filosofía. El uso de Internet y del World Wide Web en nuestro ámbito personal de trabajo (3), supone un gran número de cambios e innovaciones que se traducen en una mejora de la eficacia en nuestras tareas: la introducción del correo electrónico, news y listas de discusión; la consulta online a nuestros proveedores con presencia en Inter-

* Comisión del Sistema Eléctrico Nacional (Madrid). Correo electrónico: csen1111@sarenet.es.
Recibido: 5-11-96.

net, o, incluso, a los propios autores y editores; la sensible mejora de la interfaz que supone la conexión a OPAC's de bibliotecas a través de pasarelas WEB que superan la ruda conexión Telnet; la extraordinaria ampliación del abanico de fuentes de información disponibles para responder a demandas concretas de nuestros usuarios, etc. En rasgos generales, para lograr una mayor optimización de nuestros recursos Internet, debemos tener en cuenta las siguientes cuestiones:

1. Aquéllas relacionadas con la recuperación puntual de información en el World Wide Web (satisfacción de necesidades de información concretas de nuestros usuarios).
2. La gestión de recursos de información interna en un entorno hipermedial.
3. La selección y gestión de recursos de información en Internet, así como la posibilidad de integrarlos con los recursos anteriores.

1. La recuperación de información en el WWW

Aunque el sistema de «*browsing*» es importante para descubrir nuevos recursos, la definición de mecanismos precisos de búsqueda y recuperación es un aspecto esencial para una satisfactoria evolución de la propia red. En la actualidad, los métodos empleados, directamente relacionados con los sistemas de indización usados por cada base de datos o buscador, sin ningún control del vocabulario, y en un entorno donde la información no está mínimamente estructurada, no están ofreciendo un resultado adecuado (4). Debemos, pues, saber combinar el uso de los distintos buscadores según se adecúen, en mayor o menor medida, a las características concretas de cada demanda de información.

Los *directorios* tipo Yahoo o Magellan, basados en la *clasificación humana* de los recursos informativos de la red, no son capaces de seguir el vertiginoso ritmo de crecimiento de la red, afectándoles en gran medida problemas propios derivados de la subjetividad humana y del empleo de interminables categorías y subcategorías (5). Su mayor virtud es su precisión, aunque la posibilidad de «silencio» documental es alta.

Otra alternativa es la indización automática «*full-text*» realizada por robots (6), en la que los resultados de la búsqueda suelen ser ordenados en «rankings» de relevancia, basados en algoritmos matemáticos (normalmente la repetición del término buscado en el documento o en determinadas zonas del mismo (7)). El problema es el inverso: gran exhaustividad (resultados de miles de registros), pero el «ruido» producido es excesivo (8). La única posibilidad para limitar este problema es recurrir a buscadores que permitan acotar los resultados de nuestras estrategias de búsqueda mediante operadores booleanos y de proximidad. Altavista, Excite, Open-Text o Lycos pertenecen a este tipo de motores de búsqueda. Aunque todos utilizan sistemas de indización a texto completo, parecidos pero no idénticos, ofrecen diferentes posibilidades de recuperación. Altavista es la más valorada (9). Incluye la mayor parte de los recursos empleados tradicionalmente en la recuperación de información: operadores booleanos, de proximidad o adyacencia, truncamiento a la derecha y por caracteres, posibilidad de limitar por fechas, palabras contenidas en el título, etc. En el otro extremo, encontramos Lycos, que ni siquiera permite el uso de la lógica booleana. Open Text, por su parte, intenta solucionar el problema de la desestructuración de la información en su opción «Power Search», al permitir una interrogación por campos, muy «*sui generis*», de relativa utilidad, que recupera por título

(«Title» en html), primera cabecera («H1» en html), sumario de la página (creado automáticamente por el robot indizador a partir de las primeras 100 palabras), URL o en cualquier parte («anywhere»).

La última opción son los *meta-índices o motores de búsqueda múltiple*, que interrogan varios buscadores a la vez, eliminando, posteriormente, los duplicados. Son cada vez más y entre ellos figuran SavvySearch, All-in-One Search Page o Metacrawler (10). Aunque proporcionan aún mayor exhaustividad, no responden al problema de fondo: la precisión en los resultados.

La solución al problema es difícil, aún más cuando el crecimiento de la información es de tipo exponencial y, en gran parte, netamente comercial. Una opción interesante es la de los *sistemas distribuidos de indización*. La primera iniciativa, Aliweb (11), corrió a cargo del omnipresente Martin Koster. Se propuso una indización del contenido de las páginas por parte de los administradores de cada web. La indización se realizaría en un fichero normalizado, que sería recuperado por un robot que alimenta una base de datos global. Este fichero contendría una serie de campos al modo de una ficha catalográfica. El problema de la desestructuración de la información en la red sería mitigado al quedar normalizados los servidores, siguiendo el esquema de las bases de datos online tradicionales.

Una nueva aportación, en esta línea, fue la de OCLC (Online Computer Library Center, Inc.) con la organización de las «*Metadata Conferences*», celebradas en Dublín en marzo de 1995 y en la Universidad de Warwick (UK) en abril de 1996 (12). Su objetivo es conseguir definir un formato «amigable» para la catalogación distribuida de los recursos de la red, más rico que un índice creado automáticamente y menos complejo que el tradicional registro bibliográfico, así como la cooperación para promover una mayor interoperatividad semántica a través de lenguajes y disciplinas, de cara a lograr un efectivo control del vocabulario de indización.

En todo caso, la solución debe avanzar en este sentido, promoviendo una fórmula de tipo cooperativo y normalizada a nivel internacional (13), basada en la utilización de técnicas documentales, que elimine, a su vez, una excesiva navegación automática de robots que puedan saturar la red. La pregunta sería si no es demasiado tarde para ello.

2 El desarrollo de una intranet: la gestión de recursos de información interna en un entorno hipermedial

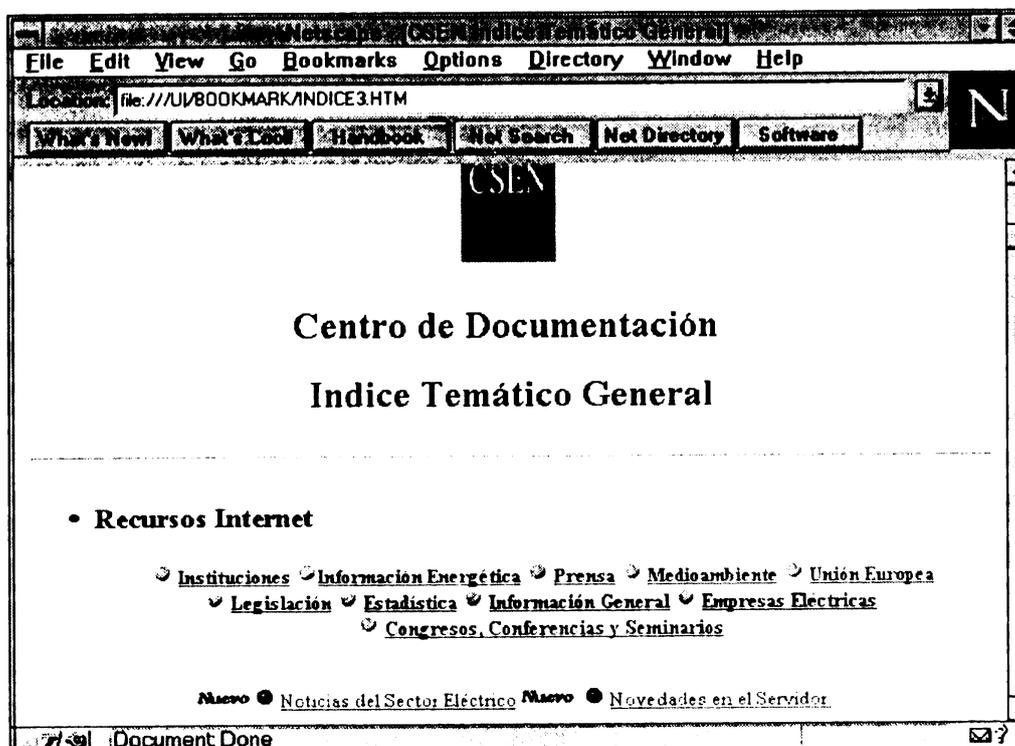
El concepto de *intranet* (14) hace referencia a la aplicación de las herramientas utilizadas en el World Wide Web a la configuración de la red local que permite la gestión de la información interna. La gestión de la intranet de nuestra empresa nos ofrece la oportunidad de convertir el centro de documentación en el articulador de la información de los distintos departamentos de nuestra organización, y, a la vez, en nexo con la exterior. Los únicos requisitos son el uso de una plataforma servidor (sobre Windows NT o UNIX), un programa cliente navegador o «web browser» (Explorer, Netscape, etc) instalado en los puestos de trabajo, y una organización de la información basada en documentos creados en HTML, que enlazarán con todos aquellos documentos, ficheros de texto, bases de datos, etc., que se estimen convenientes. Las ventajas propias de Internet se trasladarán, pues, al entorno inmediato de nuestra organización: navegación asociativa e intuitiva, amigabilidad de la interfaz, multimedia, integración con los recursos externos de información presentes en Internet, etc....

A pesar de no ser una cuestión en absoluto trivial, a veces se descuida la propia organización de la información. La navegación WWW a través de los diferentes nodos o bloques informativos se realiza de acuerdo con las preferencias o necesidades que se tengan en cada momento, característica que propicia la desorientación del usuario y la no asimilación del contenido cognitivo del documento. Es un problema vital y largamente tratado por los teóricos del hipertexto (15). La estructura hipermedial de un servidor Web ha de ser muy clara. Sin embargo, no es difícil encontrar ejemplos de servidores en Internet faltos de una mínima organización y en los que resulta fácil «perdersse». Para evitarlo, en el diseño de nuestro servidor debemos tener en cuenta una serie de principios:

- La existencia de *un índice general* que actúe como eje central articulador es fundamental (figura 1). La creación de índices en los sucesivos niveles informativos siempre es útil para el usuario.
- Los diferentes núcleos informativos han de conectarse al índice general, de tal forma que no exista ninguno aislado. Para ello, es conveniente, al menos, la presencia en cada página de *mecanismos de vuelta atrás* que refuercen el eje de navegación principal, conectando tanto con el índice general (normalmente la Home Page) como con el nivel informativo inmediatamente superior. Por supuesto, la implementación de otras rutas de navegación alternativas a la principal resultará enriquecedora para el usuario.

Figura 1

Indice principal («Home Page») del Centro de Documentación de la Comisión del Sistema Eléctrico Nacional (CSEN)



- Por otro lado, también hay que evitar una *excesiva generación de enlaces*, que provocarán la pérdida del usuario.
- Un elemento importante para la navegación es conseguir un alto nivel de *consistencia en el diseño de la interfaz*. Debemos conseguir un alto nivel de homogeneidad en la funcionalidad y situación de los botones e iconos, empleo de colores, etc.
- Cuando el tamaño de nuestro servidor es considerable, resulta de utilidad la introducción de *mecanismos internos de búsqueda* y consulta directa de la información (sistemas del tipo SWISH o FFW) (16).
- Por último, una cuestión técnica: *no crear páginas demasiado extensas o sobrecargadas de imágenes*. Impiden la rapidez en el acceso y pueden provocar el rechazo del usuario a la visualización de la página, y la consiguiente quiebra de la organización informativa de nuestro servidor.

3 Selección y gestión en red de recursos Internet

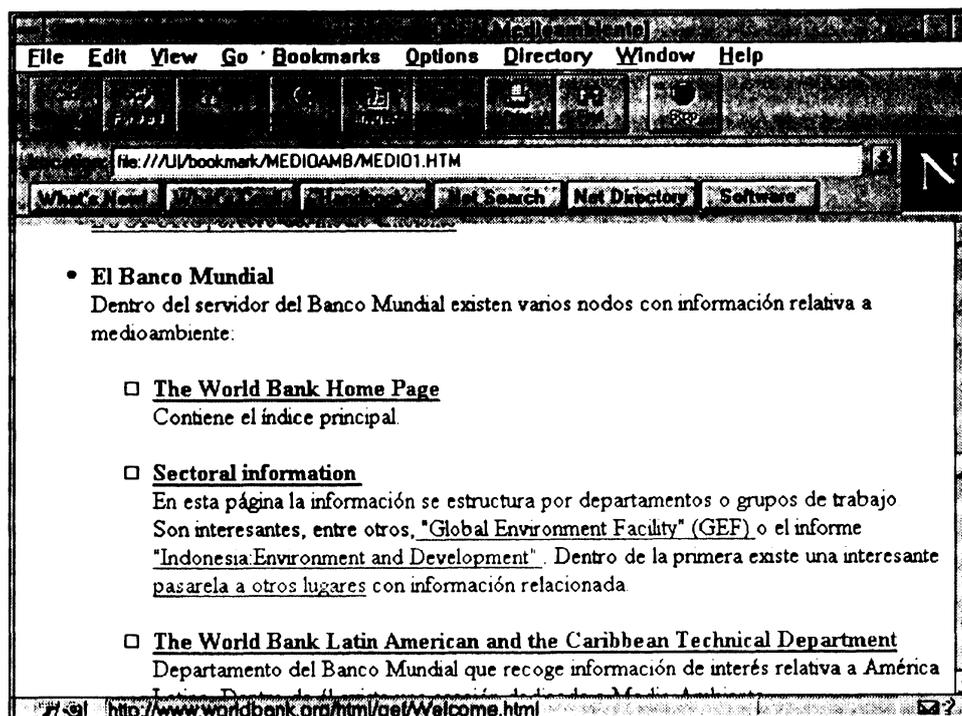
Como ya se ha mencionado, una de las ventajas de la gestión de información en una plataforma Web es la posibilidad de integrar los recursos de información internos con los externos, presentes en Internet, en un mismo entorno. La selección y gestión de recursos de información Internet, directamente relacionados con nuestro segmento de actividad, puede ser un buen servicio complementario, sobre todo en centros de documentación o bibliotecas especializadas, que actúan, así, como «filtro» necesario ante la acumulación de información no relevante para nuestros usuarios. Además, Internet enriquecerá nuestra colección mediante el uso de enlaces a artículos de revista, documentos, estadísticas, literatura gris, directorios e información de todo tipo, situados en cualquier servidor de la red. Las fases para la implementación de estas páginas serían:

- *Búsqueda y selección de los servidores Web de interés*. Un buen punto de partida, para un primer contacto con los recursos informativos de nuestro ámbito, sería la visita a la *World-Wide Web Virtual Library* (17) pertinente, o a cualquiera de los directorios o clasificaciones existentes. La propia naturaleza asociativa de la red propicia el salto a otros servidores relacionados, simplemente siguiendo los «links». Además, la existencia de páginas de recopilaciones de enlaces relacionados temáticamente con el nuestro (tipo «Related Homepages», «Links to Other Sites», etc.) facilitan nuestra labor investigadora. Frente al «browsing» de la red, siempre podemos recurrir a los buscadores para lograr una mayor exhaustividad.
- *Organización y clasificación temática* de los recursos seleccionados con arreglo a los intereses de nuestros usuarios. El instrumento idóneo para facilitar esta tarea es la *gestión de bookmarks* en nuestro navegador. Dependiendo de la versión empleada, el programa resulta más o menos operativo para este trabajo. Una vez organizados, podemos exportar directamente a nuestra futura página HTML mediante la opción «Export bookmarks».
- Proceso de *conversión a formato HTML*, elaboración de las páginas y organización hipertextual de las mismas (ver apartado 2).
- Una vez definida la estructura hipertextual se pueden utilizar dos *formas de conexión o carga en la red de área local o intranet*: establecer un puntero desde la

página principal de nuestra intranet al índice principal de recursos Internet en aquellos PC's con acceso a la red, o bien, configurar directamente el archivo que contiene el índice principal de nuestras páginas como "Home Page" para el inicio de cada sesión de conexión a Internet (punto de partida, con direcciones de interés, para la navegación de nuestros usuarios).

- Intentar dotar de *valor añadido* a nuestro producto, por ejemplo mediante la aplicación de técnicas documentales: utilización de un pequeño resumen de tipo informativo del contenido del servidor Web referenciado, que evite una vacía y simple lista enumerativa y, además, oriente a nuestros usuarios; en el caso de que recojamos grandes servidores, podemos «desglosar» su contenido, estableciendo, además de la Homepage, enlaces directos con sus distintas secciones, información útil a nuestros usuarios, etc. (figura 2).

Figura 2
Ejemplo de servicios de valor añadido en una página Web
(clasificación, resumen, desglose, etc)



Por último, es necesario un trabajo regular para el *mantenimiento y la actualización de los links*. Esta tarea se puede automatizar con la utilización de algunos programas (18) como por ejemplo «NetCarta WebMap», que, entre otras funcionalidades, verifica, mediante mapas, cada uno de los links de nuestro servidor, identificando los desactivados, y posibilita el cambio de propiedades de cualquiera de nuestras páginas (modificación de etiquetas, fechas, URL's, headings, etc.) (19).

4 Ejemplo práctico: la gestión de recursos Internet en la Comisión del Sistema Eléctrico Nacional (CSEN)

Resulta sorprendente la gran cantidad de recursos informativos existentes en Internet que giran alrededor del concepto de «*public utilities*» o empresas de servicio público (agua, gas, electricidad, telecomunicaciones, etc.) y para el sector eléctrico, en particular, ámbito temático sobre el que se desarrolla la actividad reguladora de la CSEN. Ejemplos de instituciones reguladoras, como el servidor Web de la California Public Utilities Commission (CPUC), actualizado a diario, y que recoge todos los informes, estadísticas y legislación del proceso de reestructuración del sector eléctrico californiano; de organismos gubernamentales, como la Energy Information Administration, dependiente del Departamento de Energía de Estados Unidos (DOE), o de grupos de investigación como el Harvard Electricity Policy Group o el North American Electric Reliability Council, que incluyen extensos «FTP sites», con bibliografía y documentos de interés, constituyen fuentes de información de primera magnitud para la CSEN.

El Centro de Documentación de la CSEN se planteó la posibilidad de hacerlo extensivo a sus usuarios aprovechando la implantación de nuevas cuentas de acceso a Internet en la institución. En la actualidad, existen 19 puestos de consulta, distribuidos por todos los departamentos, para una plantilla aproximada de un centenar de personas. La red de área local corre sobre Microsoft Windows NT (versión 3.51) y el navegador utilizado es Netscape (versión 2.0). Para ello, y como parte de un proyecto más extenso de intranet (comunicación interdepartamental a través de formularios, acceso al catálogo de la biblioteca, documentos internos, etc.), se diseñó un acceso corporativo a Internet a través del Centro de Documentación en el que se siguieron, básicamente, los pasos explicados en el apartado 3:

- Se contaba con un gran número de servidores de interés, fruto de meses de navegación, recopilados y clasificados como «bookmarks».
- Una primera *clasificación* provisional establecida durante el proceso de selección de bookmarks fue sustituida por una organización de los recursos atendiendo a las necesidades informativas de nuestros usuarios. Se optó en todo momento por una clasificación simple y «amigable», con el menor número posible de subcategorías. En un primer nivel se dividieron en: Instituciones, Estadística, Medio Ambiente, Información Energética, Empresas Eléctricas, Unión Europea, Prensa, Legislación, Congresos, Conferencias y Seminarios e Información General. Las secciones de mayor interés se fueron subdividiendo con objeto de ofrecer una mayor especificidad, pero siempre teniendo en cuenta las ideas de simplicidad y afinidad con las necesidades de nuestros usuarios. Por ejemplo: Instituciones se subdividió en: Reguladoras, Investigación y Defensa de la Competencia; Empresas Eléctricas lo hizo en Directorios, Grandes Empresas Internacionales y Empresas Españolas, y así sucesivamente.
- Para la conversión a lenguaje *HTML* se utilizó, en un primer momento, la opción «Save as» del menú «File» para el archivo «bookmark.htm» que Netscape va creando automáticamente a partir de los bookmarks personales. Esta opción ahorra esfuerzo al crear una página en formato HTML con el contenido de este archivo que respeta su disposición original: la clasificación, establece «headings», incluye la dirección IP, etc. En actualizaciones posteriores, se utilizó la opción

- «Export bookmarks» para ir incluyendo nuevos enlaces.
- Hasta el desarrollo definitivo de la intranet, se optó por configurar el índice principal de las páginas como Home Page en los puestos de acceso directo a Internet (para Netscape la opción «Home Page Location» de «Appearance» en el menú «Preferences”).
- Se procuró incorporar al diseño de las páginas *servicios de valor añadido*, del tipo de los comentados anteriormente: generación de índices, resúmenes informativos, desglose de servidores para enlaces específicos a información de utilidad, etc.
- Como complemento, también se abordó el tema de la *formación de usuarios*, con el objetivo de convertirlos en usuarios finales de la red. Se realizaron varias sesiones de introducción a Internet, en las que se explicaron temas como el uso del correo electrónico, navegación WWW, búsquedas de información en la red, como realizar un FTP o utilización de Netscape.
- Por último, se ha intentado diseñar un sistema de «feedback» centro de documentación-usuario, gracias al correo electrónico interno, que consiste en el diseño de un formulario por el que los usuarios proponen la inclusión de nuevos Webs de interés y un servicio de novedades mensual en el que el centro informa de los nuevos enlaces añadidos.

De esta forma, el Centro de Documentación se ha convertido en el intermediario para la navegación Internet del personal de la CSEN. En la actualidad, son más de 200 los enlaces, en áreas de interés para la CSEN, recopilados en nuestras páginas, a través de los cuales, se ofrece al usuario una amplia guía para el acceso a la información relevante presente en Internet, independientemente de la posibilidad individual de acceso mediante el uso de bookmarks personales.

5 Conclusiones

El uso de una herramienta como Internet abre una gran cantidad de nuevas posibilidades de acceso e integración de la información que no podemos ignorar. Por un lado, es posible acceder a una gran cantidad de recursos informativos con un bajo coste económico, y por otro, el desarrollo de una intranet nos permite integrarlos dentro de nuestra colección bajo una misma interfaz, fácil e intuitiva, al mismo tiempo que organizamos y estructuramos los recursos informativos de nuestra organización. Además, podemos reforzar nuestro papel de intermediarios en el acceso a la información, realizando esa tan necesaria criba que significa seleccionar aquellos recursos de utilidad para nuestros usuarios, que faciliten la navegación por el maremagnum de información no relevante existente en la red. Sin embargo, el camino por recorrer es aún muy largo, no sólo en lo que respecta al arduo problema de la recuperación de información, sino en la cuestión de dotar a la red de contenidos de calidad. Debemos apostar por Internet y colaborar activamente en la resolución de sus problemas, aportando años de experiencia de nuestra profesión en gestión y recuperación de información.

Referencias

1. BLAKE, P. Knight-Rider offerings break with tradition. *Information World Review*, 1996, núm. 117, págs. 4-5.
2. Companies for the common people: an impossible dream? *Information World Review*, 1996, núm. 117, págs. 13-14.
3. CORNELLA, A. Las empresas en la era de Internet. *Revista Electrónica Net Conexión*. Se encuentra accesible en <http://www.site.es/NET/hemero/articulos/tema.htm>.
4. Para una información bastante exhaustiva de los buscadores de Internet se puede consultar el WEB de CODIBUCE (Conferencia de Directores de Bibliotecas Universitarias Españolas) en la siguiente dirección: <http://dalila.ugr.es/~felix/g9/g9.htm>.
5. TAYLOR, D. Yahoo!: a hotlist cornucopia. *Internet World*, marzo 1995, págs. 24-26. LESTER, D. Yahoo!: profile of a web database. *Database*, diciembre 1995, págs. 46-50.
6. CHEONG, F.-C. *Internet agents: spiders, wanderers, brokers and bots*, Indianapolis: New Riders, 1996. WILLIAM, J. (et al.). *Bots and other internet beasts*, Indianapolis: Sams. Net, 1996.
7. EVANS, R. Beyond boolean: relevance ranking, natural language and the new search paradigm. *Proceedings of the 15th National Online Meeting 1994*, New York: Learned Information, 1994, págs. 121-128.
8. RAEDER, A. Cataloguing the WEB. *Searcher*, 1995, vol. 3, núm. 7, págs. 12-19.
9. La revista *PC Magazine Online*, en su edición de julio de 1996, la erigió como la más completa, potente y precisa herramienta de recuperación en el Web. El informe está accesible en: <http://www.pcmag.com/issues/1513/pcmg0080.htm>.
10. LEONARD, A. J. Search engines: where to find anything on the net. Está accesible en: <http://www.cnet.com/Content/reviews/Compare/Search/>.
11. KOSTER, M. Aliweb-Archie-like indexing in the Web. Proceedings of the First World Wide Web Conference, Geneva, 1994. *Computer Networks and ISDN systems*, 1994, vol. 27, núm. 23, págs. 175-182.
12. Información exhaustiva sobre estas conferencias se encuentra en el servidor de OCLC: <http://www.oclc.org:5046/oclc/research/conferences/conferences.html>. Y, en general, sobre este tema en: <http://www.ukoln.ac.uk/metadata/interoperability/>.
13. SCHNEIDERMAN, R. A. Why librarians should rule the net. *E-Node*, 1996, volumen 1, número 4. Para suscribir esta revista electrónica, mandar un mensaje a: enodelist@garnet.berkeley.ed («subscribe e-node» en el cuerpo del mensaje).
14. TRAMULLAS, J. Internet e Intranet: la creación de servicios de información en las organizaciones. Este artículo presentado en las Jornadas Nacionales de Información y Documentación Empresarial INDOEM 96, celebradas en Murcia en mayo de 1996 está accesible en: <http://jabato.unizar.es/intranet/intranet.htm>.
15. CARIDAD, M.; MOSCOSO, P. *Los sistemas de hipertexto e hipermedios: una nueva aplicación en informática documental*. Madrid-Salamanca: Fundación Germán Sánchez Ruipérez. Madrid: Pirámide, 1991, págs. 123-136.
16. Encontramos información sobre Swhis (Simple Web Indexing System for Humans), así como la posibilidad de hacer ftp del programa en: <http://www.eit.com/software/swish/>. Sobre FFW (Freetext search For Web) existe información en: <http://www.nota.no/produktier/ffw/ffw.html>.
17. Auténtica encrucijada de caminos en Internet, referencia recursos informativos de cualquier área. La dirección de su índice de materias es: <http://www.w3.org/pub/DataSources/bySubject/Overview.html>.
18. EITO BRUN, R. Una nueva forma de recuperar información: los robots personales. *Information World en español*, 1996, núm. 46, julio-agosto.
19. El Web de NetCarta Corporation se encuentra accesible en la dirección: <http://www.netcarta.com>. Permite el uso gratuito de la aplicación por un período de treinta días.