

Las fronteras de la automatización de sistemas de información y documentación

Francisco Javier García Marco

Universidad de Zaragoza

0.1. Resumen

Se analizan algunas de las principales fronteras que existen actualmente en la automatización de sistemas de información a partir de la presentación y discusión de los estudios publicados en el segundo número del volumen noveno de la revista *Scire*. Se abordan las bibliotecas digitales, la arquitectura de los sistemas automatizados de bibliotecas y las características de los servicios de información digital en diversos campos, señaladamente los medios de comunicación, la educación y la formación y el comercio y difusión del libro antiguo. Los servicios principales de los portales analizados son los de publicación, información, documentación y recursos, comunicación interactiva y alerta, que en conjunto crean un amplio espacio de servicio para los profesionales de la información y la documentación. Las tres fronteras de carácter general cuyas implicaciones generales se discuten son la clarificación conceptual —tanto a nivel ontológico como de procesos—, la integración de funciones, la normalización y el desarrollo de automatismos de segundo grado sobre los estándares consolidados.

Palabras clave: Sistemas de información y documentación. Automatización. Bibliotecas digitales. Sistemas automatizados de bibliotecas. Servicios de información digital. Medios de comunicación. Educación y formación. Libro antiguo. Normalización.

0.2. Abstract

Some of the main frontiers in information and documentation automation are considered. First, the results published in the second number of the ninth volume of the academic journal *Scire* are analysed. Digital libraries, library systems architecture and specific digital services —these especially in the fields of mass media, education and training and old book commerce and access— are the main topics

discussed. The main categories of services detected in portals and digital services are publication, information, resources, interactive communication and alerts. On the other hand, the general frontiers mainly emphasised are the conceptual clarification of processes and ontologies, functional integration and standardisation.

Keywords: Information and documentation systems. Automation. Digital libraries. Library systems. Digital information systems. Mass media. Education and training. Old books. Standardization.

1. Introducción

El conjunto de estudios que presentamos en este número tratan de la utilización de tecnologías de la información en beneficio de los sistemas de información y documentación. Se dividen en dos grupos. El primero está formado por un conjunto de estudios de carácter general que abordan problemas amplios —las bibliotecas digitales, la automatización de los flujos de trabajo de la biblioteca, el papel de las aplicaciones de gestión de flujos de trabajo en el entorno documental o la automatización y gestión en red de tesauros—. El segundo grupo de estudios consiste en varias presentaciones de aplicaciones concretas diseñadas para entornos específicos —archivos audiovisuales, portales periodísticos, comercio y difusión del libro antiguo, portales educativo o la autoformación de usuarios—.

2. Bibliotecas digitales

Las bibliotecas digitales son quizá la iniciativa actual más visible y que más interés despierta en los profesionales de la información y la documentación. Más allá de la gran variedad de servicios que se esconden tras esa denominación, el acceso telemático a fuentes de información primarias de calidad bien organizadas y sustentadas por un servicio profesional —esto es técnico y comprometido con el usuario, no tanto “de pago”— es, sin duda, su nota más distintiva.

Aprovechando la capacidad de procesamiento y almacenamiento masivo y barato de los ordenadores actuales y el bajo coste e inmediatez de las comunicaciones telemáticas a través de Internet, la biblioteca digital supone la evolución natural de las bases de datos referenciales en general y de los catálogos en particular: el acceso directo al documento primario a partir de la referencia o representación documental. De esta forma, la impresionante potencia de los catálogos automatizados en lo que respecta a la recuperación de información referencial se traslada al acceso al propio documento. También supone una culminación del proceso de automatización de la biblioteca.

Una vez completado el ciclo de la automatización de la biblioteca —sobre el papel, claro está—, la evolución de la automatización de las bibliotecas dejará probablemente de ser extensiva para volverse intensiva y subir de nivel. Esto es,

consistirá en añadir “inteligencia” y previsión a los sistemas automatizados de bibliotecas, tanto en lo que se refiere a su cuadro de mando como a la asistencia informática a los procesos técnicos más “intelectuales”, como la indización.

El artículo de Mercedes Martínez González (2003) plantea los retos principales a los que se enfrenta la investigación sobre bibliotecas digitales, que a su juicio son mejorar la recuperación — encontrar la información —, garantizar la calidad de la información y asegurar su perdurabilidad; y plantea el problema tanto en el nivel de las bibliotecas digitales específicas como en el del funcionamiento de la World Wide Web en su conjunto. Para Martínez el éxito de la biblioteca digital no es solo un problema técnico, sino que implica considerar su funcionamiento en el sistema social, lo cual se traduce inevitablemente también en un problema económico, de ahí que gran parte de su trabajo esté dedicado a los modelos económicos de mantenimiento de una biblioteca digital: el libre acceso, el préstamo interbibliotecario y el préstamo basado en un modelo de usuarios concurrentes limitados. En su análisis detecta el conflicto entre los editores comerciales y las asociaciones y grupos de investigadores y ciudadanos que promueven el acceso libre. Al final concluye (*ibidem*, p. 32):

La tecnología ofrece soluciones automáticas que pueden facilitar la implementación de ciertas tareas, inspiradas en los métodos de trabajo de los expertos. Pero también plantea nuevos interrogantes que necesitan respuestas. Es en dar respuesta a estas preguntas — allí donde los procesos automáticos no pueden tomar decisiones —, donde los profesionales del futuro deberán hacer mayores esfuerzos.

2. La automatización de los procesos técnicos en las unidades de información y documentación como gestión de flujos de trabajo

El trabajo de Tomás Saorín Pérez y Jesús González Lorca (2003) se complementa bien con el de Martínez en el sentido que frente a las bibliotecas puramente digitales — necesariamente limitadas en sus contenidos en el momento actual al centrarse en un solo formato — aborda las bibliotecas “físicas”, las que pretenden ofrecer un servicio integral al ciudadano, al universitario o a sus públicos especializados, independientemente del soporte. Estas bibliotecas, en el nuevo contexto digital, son — o serán — “bibliotecas híbridas”. La tabla II (*ibidem*, 47) sintetiza la visión de los autores sobre las principales transformaciones que están en marcha en la transición desde la biblioteca tradicional — de la década de los ochenta — a la biblioteca híbrida: personalización, gestión sistémica de los datos, acceso en texto completo, integración de los servicios y documentos de proveedores y distribuidores, carácter multimedia, colecciones digitales, punto de consulta en el escritorio, acceso a través de la WWW, orientación al descubrimiento de recursos (*knowledge discovering*), construcción de comunidades virtuales de usuarios, biblioteca en red y multiplicación de servicios. Pero independientemente

te de su contexto de aplicación, interesa la orientación del trabajo a la automatización de procesos. Los profesionales cuya tarea está relacionada con el análisis documental tienden a ver la automatización de bibliotecas como representación documental, como construcción del catálogo. Es decir, se centran en el producto y en el servicio, no tanto en los procesos. El artículo de Saorín y González tiene la virtud de enfocar la automatización de bibliotecas desde el punto de vista interno: la asistencia automatizada de los procesos técnicos que realiza el personal de la biblioteca, para lo cual la herramienta fundamental es la gestión de flujos de trabajo. Desde el punto de vista de los agentes humanos, el interfaz de todos los procesos automatizados es el portal, que ofrece distintas caras y servicios según una matriz que depende del tipo de persona que usa el sistema—bibliotecario o usuario— y del lugar de acceso —dentro o fuera de la biblioteca—.

El siguiente artículo, también de Jesús González Lorca (2003), incide en el problema más general de la teoría de los flujos de trabajo y de las implicaciones para el mundo de la información y documentación. La gestión automatizada de flujos de trabajo surgió precisamente en el ámbito de la gestión de procesos administrativos —la Ofimática—, que, como bien sabemos, están normalmente apoyados en documentos. Tras una cuidada presentación de carácter general, González incide en dos ideas importantes para nuestro campo: en primer lugar, que muchos procesos de gestión de documentos —tanto en bibliotecas como archivos o centros de documentación— se benefician o se podrían beneficiar de la automatización de flujos de trabajo, aunque no siempre se abordan estos procesos de forma sistemática; en segundo lugar, que la gestión de flujos de trabajo es una parte importante de la gestión de documentos de la cual no dan cuenta las aplicaciones tradicionales de gestión de bases de datos documentales, que a veces se utilizan erróneamente con este fin.

Juan Antonio Pastor Sánchez y Francisco Javier Méndez (2003) presentan el análisis completo de un sistema de gestión de tesauros diseñado para funcionar en la web y para servir a comunidades virtuales en la indización cooperativa de recursos de Internet: iSGAT (Sistema de Gestión Automatizada de Tesauros para Internet). El Grupo de Investigación en Tecnologías de la Información de la Universidad de Murcia viene realizando desde hace más de una década una importante labor en la investigación sobre la gestión automatizada de tesauros, parte de la cual hemos revisado en un trabajo reciente (García, 2002). Este proyecto construye sobre las aportaciones anteriores aborda la frontera de Internet, tanto en lo que se refiere a la necesidad de hacer disponibles estas herramientas en red para facilitar un trabajo que es eminentemente colaborativo —aunque no por ello menos necesitado de control responsable—, como por su utilidad para indizar recursos de Internet.

3. Ámbitos de aplicación en los servicios de información WWW

El resto de los trabajos del número 9:2 de *Scire* presentan aplicaciones de las nuevas tecnologías de la información a diferentes campos aplicativos: la documentación de medios, el mundo del libro antiguo y la documentación educativa.

La documentación de medios de comunicación de masas es uno de los campos más dinámicos de aplicación, especialmente en el caso de la prensa escrita que ha reaccionado con presteza y eficacia al reto de Internet. El artículo de Àngels Jiménez, María Eulàlia Fuentes y Alfons González (2003) se dedica a mapear el desarrollo y características de los sitios de los medios televisivos, y especialmente de un servicio eminentemente documental, los archivos de televisión en línea. Plantean, en primer lugar, las características que definen el perfil de los nuevos medios: interactividad; información multimedia; inmediatez y, por tanto, superación del carácter periódico de la noticia; y personalización de los contenidos. Seguidamente, muestran el retraso general de las televisiones españolas respecto a los líderes internacionales y sus excepciones. Sin duda, el desarrollo de estos servicios en España está lastrado por el alto coste que supone almacenar y servir vídeo digital, muy superior al del texto. En tercer lugar, dibujan el panorama de los archivos audiovisuales de las principales televisiones extranjeras, horizonte de los desarrollos que sin duda han de producirse en nuestro país. Por fin, muestran el surgimiento y desarrollo de archivos audiovisuales para el apoyo a la producción de programas retrospectivos, normalmente proporcionados como servicios de pago; lo cual indica también el amplio lugar que existe para el desarrollo de empresas de mediación documental entorno a los nuevos medios.

Rosana López Carreño y José Vicente Rodríguez Muñoz (2003) dedican su trabajo al estudio de los servicios de valor añadido (SVA) de los portales periodísticos españoles. Los servicios de valor añadido son servicios nuevos que ha parecido en los medios digitales y que suponen un avance respecto a los tradicionales. A partir del análisis de un total de diecisiete portales —los de los rotativos nacionales y una representación de la prensa regional— identifican ocho categorías de SVA: acceso a la información, interactivos, participación, multimedia, personalizados, entretenimiento, comerciales y complementarios. Los SVA más consolidados resultaron ser los pertenecientes a las categorías de acceso a la información e interactivos, seguidos muy de cerca por los de participación. Estos servicios muestran la madurez de los portales de los medios de comunicación de la prensa escrita y el valor de los servicios que pueden ofrecer los documentalistas de prensa.

Por su parte, Manuel José Pedraza (2003) y José Luis Herrera Morillas (2003) realizan sendos trabajos que muestran la ubicuidad del impacto de Internet a través del aparentemente recóndito mundo del libro antiguo. Aparentemente recóndito, porque el libro antiguo es un elemento importante en las sociedades, sobre

todo, como símbolo de afiliación social y cultural que se expone y se enseña — y del que se presume — en las colecciones particulares y públicas; símbolo valioso que como tal transmite su prestigio a su poseedor, igual que ocurre con los objetos de arte y arqueológicos. En este sentido, los trabajos de Pedraza y Herrera son temáticamente complementarios. Pedraza nos muestra el mundo del comercio electrónico del libro antiguo en un completo y cuidado análisis, en el que estudia la oferta, la demanda, el medio —el catálogo—, el objeto —el libro— y el precio. Como documentalistas nos interesa resaltar el valor del catálogo como instrumento pivotal de la venta, más valioso cuanto más amplio, y, por tanto, abierto a iniciativas colectivas semejantes, salvando las distancias, a las que se dan en el campo de las bibliotecas. Por lo demás, el catálogo en Internet resulta más barato de producir y difundir, aunque supone un canal de venta limitado, porque no todos los compradores lo utilizan, aunque su número crece constantemente. Por su parte, Herrera estudia el libro antiguo en las instituciones públicas del Estado —concretamente en el Catálogo Colectivo del Patrimonio Bibliográfico Español y los de las principales bibliotecas universitarias españolas— a través del ejemplo de los impresos antiguos zaragozanos. Además de procurar una amplia perspectiva de la historia de la impresión de libros en Zaragoza, Herrera muestra como el catálogo bibliográfico es también —además de un instrumento de control patrimonial— un instrumento de difusión —sin perjuicio y más bien en favor de su conservación—, especialmente si se añaden los facsímiles digitales, como ocurre, entre otros casos, de las universidades de Granada y Barcelona.

Por fin, sendos trabajos de Felipe Zapico Alonso y José Antonio León Moreno (2003), por un lado, y de Patricia Hernández Salazar (2003), por el otro, abordan webs destinados a la educación desde distintas perspectivas. Zapico y León presentan el excelente Portal de Educación de Extremadura (ERT-Extremadura.org), cuyo objetivo es servir a la comunidad educativa extremeña proporcionando una visión global de las iniciativas existentes, configurando un punto de encuentro, así como una biblioteca digital de recursos didácticos, apoyando especialmente las actividades de formación de los agentes de la red y promover el uso de las tecnologías de la información y la comunicación en la región. Se trata de un excelente ejemplo de la estrategia de la Junta de Extremadura de proporcionar servicios informáticos y telemáticos avanzados para toda la población utilizando tecnologías abiertas y de libre uso, y de la colaboración de profesionales de la información y la documentación en este tipo de proyectos.

Finalmente, el artículo de Hernández —cuyo tema de interés, aun dentro del ámbito de lo educativo, es más específico— presenta la experiencia de diseño, implementación y puesta en marcha de un tutorial en línea para la formación de humanistas en el uso de las tecnologías de la información en su investigación; es decir, se trata del desarrollo de una herramienta para la formación de usuarios

académicos. Interesan de este trabajo, entre otros aspectos, la sustentación pedagógica del recurso en la filosofía del aprendizaje significativo de Ausubel; el cuidadoso análisis de las necesidades, comportamientos de búsqueda y habilidades de los usuarios; y la utilización de las técnicas del guión narrativo y de los mapas de conceptos para diseñar sus contenidos. La detallada presentación del marco teórico y de la metodología del proyecto son un aporte importante al trabajo de formación de usuarios mediante publicaciones interactivas en Internet.

Cabe además extraer algunas lecciones sobre los servicios de información web a partir de experiencias tan distintas como los portales periodísticos, los dedicados al libro antiguo o los educativos. Todos estos portales se apoyan como instrumentos fundamentales en cuatro pilares: en primer lugar, los servicios de publicación —noticias y novedades, clasificadas o no por tipos de usuarios—; en segundo lugar, servicios de información “enciclopédica” —esto es, que reúnen los datos más utilizados por los usuarios del sistema convenientemente organizados para su fácil consulta—; en tercer lugar, los servicios de documentación y recursos —que giran entorno a los catálogos y, en su caso, el acceso a los recursos digitales o digitalizados—; en cuarto lugar, los de comunicación interactiva —correo electrónico, foros, etc.—; y, en quinto lugar, los servicios de alerta que afectan a la información de novedades generada en todas las categorías anteriores. Publicación, información, documentación y recursos, comunicación interactiva y alerta parecen ser, pues, los cinco pilares de la Internet como medio de comunicación, lo cual crea un amplio espacio de servicio para los profesionales de la información y la documentación.

4. Conclusiones

De los trabajos presentados por los autores del número podemos extraer los principales retos de la automatización de sistemas de información y documentación. Lo más notable es que los retos clásicos de la automatización de sistemas de información y documentación mantienen todo su interés y validez, más allá de los procesos actualmente más en boga, señaladamente las bibliotecas y los sistemas de información digitales, y el movimiento para asegurar la recuperación consistente mediante metadatos.

El primero de ellos sigue siendo la clarificación conceptual de procesos y ontologías, esto es, de los modelos de datos y de procesos que permiten el desarrollo de programas automatizados de propósito general o específico. Hemos visto que es importante atender tanto a la formalización de procesos como a la de modelos de datos, y que debe considerarse al mismo nivel. Más aún, cabría decir que son los procesos los que soportan las ontologías, en el sentido en que los modelos de la realidad se diseñan siempre con un objetivo, en una suerte de arco sensorio de gran alcance que explica la particularización conceptual concreta

en la que se parcela un determinado dominio de la realidad. El segundo reto está muy relacionado con el anterior y es la integración de funciones, su especialización, modularización e interconexión.

El tercer reto sigue siendo la estandarización, esto es, la necesidad de procurar la sostenibilidad en el tiempo y en el espacio de las soluciones haciéndolas compatibles e integrándolas. La globalización imparable hace impensable otra estrategia, en la medida en que todas las soluciones quedan expuestas en la aldea global con sus ventajas e inconvenientes a una evolución acelerada. Quedarán nichos, pero incluso dichos nichos serán de un tamaño considerable y solo resistirán en la medida en que se apoyen en una diferencia real.

Aunque todavía queda un largo proceso por delante, la generalización del código libre —no por ello “gratis”— en casi todos los tipos de actividades va a cambiar radicalmente la filosofía y los costes de la automatización de sistemas de información y documentación. Aunque las soluciones más punteras seguirán siendo normalmente privadas, los mecanismos de indización, recuperación, protección de derechos, representación de documentos, control de procesos administrativos, descripción de “objetos” —personas, registros de conocimiento, entidades, procesos...—, etc., van a ser públicos en un plazo más corto que largo. El movimiento de promulgación de estándares y modelos de aceptación general que ha promovido la Internet es un ejemplo clarísimo de esa tendencia. En el proceso de nivelación generalizado de los códigos informáticos dentro del sistema social y económico que está en marcha, quedará también un gran margen al desarrollo comercial, especialmente en lo que se refiere a soluciones que ahorren tiempo, esfuerzo y recursos.

En este sentido, el cuarto reto en marcha es el proceso de añadir inteligencia y automatismos de orden mayor a las tecnologías que actualmente se están popularizando —bases de datos relacionales, indización de texto e imagen, análisis estadístico—. A pesar de que los “sistemas expertos” han perdido la visibilidad y el aura de hace unos años, herramientas de este tipo se están generalizando en muchos ámbitos, ahora bien en ámbitos más reducidos de los inicialmente pensados y como partes modestas de sistemas informáticos más amplios, especialmente aprovechando la potencia de las bases de datos relacionales multimedia como instrumentos de almacenamiento del conocimiento “específico”. Por otra parte, es necesario prevenir contra la “pseudo-inteligencia” que suplanta al agente humano sin auténticas garantías de calidad en los procesos en aras de una mal entendida “usabilidad”, sin negar tampoco el mérito de este tipo de herramienta, a las que hay que prestar la debida atención tanto en sus luces como en sus sombras.

En fin, clarificación conceptual —tanto en lo que se refiere a las entidades de análisis como a los procesos—, integración de funciones, normalización y

profundización en la representación de procesos de razonamiento complejos son tres de los procesos generales más importantes que orientan la investigación y el desarrollo en la automatización de sistemas de información hacia mediados de la primera década del siglo XXI.

Sin embargo, todos ellos superan el ámbito de lo técnico. Su solución no puede dejar al margen el análisis de las condiciones en las que estos procesos se insertan dentro del sistema social al que sirven, e, igualmente, la consideración de los factores humanos implicados, como veíamos en el primer número del volumen noveno (García, 2003). Una de las conclusiones más importantes del volumen noveno de Scire en su conjunto es que el impacto social de las nuevas tecnologías de la información va mucho más allá de la automatización de los procesos y de las técnicas y herramientas utilizadas, y es precisamente en ese más allá donde, como concluye Martínez González (2003, p. 32) “los profesionales del futuro deberán hacer mayores esfuerzos”. Ciertamente estos esfuerzos están en marcha ya, y los trabajos que se presentan en este volumen nos animan en la dirección correcta.

5. Referencias

- García Marco, Francisco Javier (2002). La literatura científica sobre lenguajes poscoordinados en España: de la divulgación del concepto a la Internet. // Documentación de las Ciencias de la Información. 25 (2002) 291-319.
- González Lorca, Jesús(2003). Automatización dinámica de entornos documentales mediante tecnología workflow. // Scire. 9:2 (jul.-dic. 2003) 61-83.
- Hernández Salazar, Patricia (2003). La formación de usuarios de tecnologías de información: un prototipo para humanistas. // Scire. 9:2 (jul.-dic. 2003) 169-186.
- Herrera Morillas, José Luis (2003). Acceso y difusión a los libros antiguos a través de Internet: el caso de los impresos en Zaragoza. // Scire. 9:2 (jul.-dic. 2003) 145-160.
- Jiménez, Àngels; Fuentes, M^a Eulalia; González, Alfons (2003). Los archivos audiovisuales en los medios de comunicación digital. // Scire. 9:2 (jul.-dic. 2003) 99-110.
- López Carreño, Rosana; Rodríguez Muñoz, José Vicente (2003). Los servicios de valor añadido en los portales periodísticos: una propuesta de clasificación. // Scire. 9:2 (jul.-dic. 2003) 111-122.
- Martínez González, María Mercedes (2003). Bibliotecas digitales: retos y libre acceso. // Scire. 9:2 (jul.-dic. 2003) 19-35.
- Pastor Sánchez, Juan Antonio; Martínez Méndez, Francisco Javier (2003). iSGAT: Gestión colaborativa de tesauros en Internet. // Scire. 9:2 (jul.-dic. 2003) 85-98.
- Pedraza Gracia, Manuel José (2003). El libro antiguo: demanda, oferta y comercio en Internet. // Scire. 9:2 (jul.-dic. 2003) 123-144.
- Saorín Pérez, Tomás; González Lorca, Jesús (2003). Dentro de los portales bibliotecarios: flujo de trabajo en la automatización de bibliotecas. // Scire. 9:2 (jul.-dic. 2003) 37-59.

Zapico Alonso, Felipe; León Moreno, José Antonio (2003). RTE-Extremadura.org: un portal de acceso al Sistema Educativo de Extremadura. // *Scire*. 9:2 (jul.-dic. 2003) 161-168.