

Los múltiples afloramientos de la organización del conocimiento en el ciclo de la información documental

Francisco Javier García Marco

Universidad de Zaragoza (España)

Resumen

Se reflexiona sobre el papel de la organización del conocimiento en diferentes fases del ciclo de la información documental al hilo de la presentación de los artículos que componen el número 1 del volumen 12 de la revista *Scire: Organización y Representación del Conocimiento* (en.-jun. 2006), dedicado monográficamente al tema. El conjunto de estudios publicados en este monográfico muestra cómo la organización del conocimiento, lejos de ser una actividad circunscrita a una determinada fase de la cadena documental, es uno de los referentes de la ciencia de la información, que aflora en todos los aspectos del ciclo de la información documental. Así, la organización del conocimiento es fundamental en el análisis de contenido, la indización, el control de vocabulario, la construcción de vocabularios estructurados, la recuperación de la información y la interoperabilidad en los espacios de recuperación complejos, como la Internet. Tiene también una implicación de carácter central en el núcleo mismo de la teoría informativo-documental —puesto que la organización conceptual de la información documental remite a la organización del saber y la cultura y a su evolución—, y está presente en los desarrollos automatizados y también en las alternativas de carácter automático que se han propuesto en los últimos años. Se analiza, por fin, cómo afecta a cuestiones menos conocidas como el diseño de interfaces y la comunicación con el usuario en entornos automatizados, a la elaboración de estadísticas y la investigación operativa, y a la selección documental y bibliográfica. Por fin, se apuntan implicaciones de la organización del conocimiento en otras fases del ciclo de la información: creación, comunicación, evaluación, difusión y descubrimiento del conocimiento. Se recomienda que la investigación aplicada en organización del conocimiento con el objetivo de desarrollar técnicas y esquemas no se desvincule de la necesaria investigación básica sobre los procesos que se pretenden optimizar y de aquellos que le son subyacentes.

Palabras clave: Ciencias de la documentación. Teoría. Organización del conocimiento. Ciclo de la información. Ciclo del conocimiento. Perspectivas.

Abstract

A reflection is done on the role of Knowledge Organization (KO) as an activity and a discipline in the different phases of the information cycle, while reviewing the articles published in the number 2 of the volume 11 of *Scire* (Jan.-June 2005), a monographic number on the subject. The set of studies show how KO, far from being an activity circumscribed to a specific phase of the information cycle, is a reference for Information Science, and permeates in all its stages: it is fundamental in content analysis, indexing, vocabulary control, structured vocabularies design and maintenance, information retrieval, and interoperability in complex retrieval spaces like the Internet. It also has a central implication in the Information Science theoretical core —because the conceptual organization of documentary information relates ultimately to the organization of culture and its evolution—; and it is present in the semi —and whole— automatized developments and alternatives that have been proposed in the last decades. Besides, it is analyzed how KO affects less known subjects as interface design and user communication in information systems environments; in operative research, quantitative-based management and the elaboration of statistics; and in the selection of documents. Finally, the implications of KO in other phases of the information cycle are sketched: knowledge creation, communication, evaluation, dissemination and discovery. It is recommended that applied research in KO —done with the aim of developing techniques and KO systems— is compensated with basic research on the processes that are to be optimized and their underlying ones.

Keywords: Information Science. Theory. Knowledge Organization. Information cycle. Knowledge cycle. Perspectives.

1. Introducción

Aunque es correcto identificar la organización del conocimiento con los aspectos relacionados con el análisis de contenido, el control de vocabulario y el diseño y mantenimiento de vocabularios estructurados para la recuperación de la información, es también frecuente olvidar hasta qué punto los sistemas de organización del conocimiento están en el núcleo del funcionamiento de los sistemas de información y documentación, por lo que su influencia aflora en todas las fases del ciclo de información documental —el modelo central de las ciencias de la documentación—, y, por consiguiente, ocupa un lugar central en cualquier definición teórica de este conjunto de disciplinas.

Su relevancia teórica de alto nivel proviene del hecho de que la organización del conocimiento proporciona una perspectiva diferente para la articulación de los conceptos de la ciencia de la información, complementario al del modelo de comunicación de Shannon y Weaver (1949), y al del ciclo de la información, que supone una especificación de este en el campo de las disciplinas informativo-docu-

mentales y una síntesis con el modelo típico de las ingenierías, basado en el ciclo *input-processing-output*.

¿De qué perspectiva se trata? Pues, bien, si el modelo de Shannon y Weaver organiza la actividad informativo-documental desde el concepto de comunicación, y el modelo del ciclo de información lo hace desde la noción de producción de un producto o servicio, la organización del conocimiento enfatiza el factor de memoria social, situando el problema informativo-documental en una perspectiva de larga duración que, por lo general, no es atendida por los estudios que —en el campo de la documentación— utilizan los otros dos modelos. Efectivamente, el objetivo de los esquemas más universales de organización del conocimiento es aplicar modelos epistemológicos de organización del saber al problema de organizar el mundo de los documentos de una manera coherente con las necesidades de información de los usuarios, unas necesidades que están gobernadas por procesos de desarrollo y reproducción cultural, y de especialización social, sin los cuales estas no se pueden explicar. Este objetivo exige una perspectiva culturoológica que, de manera natural, tiene que abordar el problema del conocimiento social en una perspectiva temporal amplia.

En los diferentes trabajos que conforman este monográfico se vislumbra ese carácter transversal de la organización del conocimiento dentro de la ciencia de la información, transversalidad que lo hace aflorar en diferentes aspectos de la estructura teórica de las ciencias de la documentación y en distintos momentos del ciclo de la información y del proceso de comunicación informativo-documental. Por otra parte, este conjunto de trabajos muestra la radical interdisciplinariedad de la organización del conocimiento, que, sin duda, puede producir desazón en los que buscan principios y técnicas que aplicar —y que, en gran medida, pueden prescindir de estos problemas, salvo que su centro de información tenga una larga historia—, pero que resulta ineludible para el científico y el académico.

2. Organización del conocimiento y teoría documental

Abre el monográfico un artículo de Alejandro Delgado Gómez (2006) que aborda un conjunto entreverado de problemas centrales en la teoría archivística: el conflicto entre macrodescripción y microdescripción; la oposición entre descripción archivística y bibliográfico-bibliotecaria; la disyuntiva entre la clasificación archivística basada en el principio de procedencia y orden original, y los metadatos descriptivos orientados a satisfacer los usos del documento; y, por debajo de todo, el conflicto de fondo que se produce entre el positivismo archivístico —heredado del positivismo historiográfico que subyace a la teoría archivística latina— y los paradigmas posmodernos de la descripción que sitúan la carga de la descripción del contenido en el autor y el lector, dentro de un esquema de comunicación que, en ocasiones, raya la deconstrucción. El autor aborda el tema desde la autoridad que le otorga su repetido contacto tanto con los problemas de la

automatización de la gestión de documentos como con los de preservación de registros electrónicos. Delgado encuentra en la propia tradición archivística claves para la síntesis, que es, en el fondo, la opción que propone, una opción que no renuncia a priori ni al clasicismo ni a la modernidad.

Desde nuestro punto de vista, interesa, en primer lugar, resaltar cómo la discusión sobre las herramientas de organización del conocimiento y de recuperación temática en los archivos conecta con profundos problemas de carácter epistemológico que pertenecen al espíritu de los tiempos, y que tienen un impacto decisivo en la totalidad del campo de las ciencias de la documentación. Efectivamente, el conflicto entre el clasicismo en la organización del conocimiento —representado actualmente por la confluencia del clasificacionismo y los tesauros bajo el paraguas conceptual del análisis facetado y las aportaciones de Ranganathan y sus seguidores— y el postmodernismo —personificado por la indización de autor y usuario y la asignación de *tags* en entornos colaborativos— es uno de los debates que caracterizan la disciplina en el cambio de siglo. Un nuevo clasicismo se vislumbra a la salida del túnel, y Alejandro Delgado marca la línea dentro del campo que le interesa, los archivos:

Es función del archivero sentar las bases —mediante aquella identificación y dilucidación del contexto de creación, gestión y conservación, y de las relaciones que se producen en estas fases— para que los documentos puedan utilizarse en diferentes contextos, establecer diferentes relaciones y adoptar diferentes significados. Todo ello se consigue mediante la asignación de metadatos que preceden a la existencia del documento, le acompañan durante toda su vida y perduran una vez que el documento se ha extinguido.

Este es el futuro del gestor de sistemas de organización del conocimiento —y de su responsabilidad, según va viendo—: diseñar y administrar sistemas y procesos con el fin de asegurar que las diferentes perspectivas —las de todas las personas implicadas en la creación, gestión y uso de la información— queden recogidas en las representaciones documentales. En esta gestión se ha de conseguir conectar a creador, gestor y lector-usuario real y potencial en un mismo ámbito de comunicación; y, en definitiva, al pasado —intención del autor y usos que fueron—, el presente —usos actuales contrastados— y el futuro —usos vislumbrados e imaginados—, colaborando en la tarea de trabar la cadena de la comunicación cultural, que es, sin remedio, una tarea de comunicación intergeneracional y de carácter histórico.

3. Organización del conocimiento, lectura, análisis de contenido e indización

Dos artículos del fértil entramado de grupos de investigación de Marília (UNESP, Brasil) presentan otro de los frentes en los que la organización del conocimiento realiza una aportación decisiva: el análisis de contenido.

Una reflexión previa —que es, a mi entender, clave en la dilucidación del papel de la organización del conocimiento en el análisis de contenido— tiene que ver con la relación entre análisis de contenido y análisis temático. Sabido es que el análisis de contenido incluye en nuestro ámbito latino tanto la indización como el resumen. Si bien el papel de la organización del conocimiento es claro en la indización, resulta oscuro para muchas personas en el caso del resumen. La conexión más evidente es, sin duda, el análisis temático, que es fundamental tanto para una operación documental como para la otra. Y es que el resumen documental se construye sobre la dinámica entre tema —lo previo— y rhema —lo nuevo—. Así, su primera parte —la indicativa— es, fundamentalmente, un análisis temático, y la segunda —la identificación de lo nuevo, de los resultados, conclusiones y recomendaciones— es imposible de forma cabal sin la primera. Por consiguiente, el análisis temático no es solo el resultado del análisis de contenido previo, sino que es también su condición previa. La imbricación entre análisis temático y de contenido resulta ser uno de los universales documentales en los que aflora de nuevo la organización del conocimiento, pues el objetivo de esta disciplina es intentar descubrir las estructuras en las que se organiza el conocimiento social e inventar otras que —derivadas de las primeras— permitan organizar el mundo de los documentos desde el punto de vista temático.

Precisamente, los trabajos de Fujita y Rubi (2006) y Moraes y Guimarães (2006) inciden en la interconexión entre análisis temático y de contenido en los textos científicos y narrativos. En “Modelo de lectura profesional para la indización de textos científicos”, Mariângela Spotti Lopes Fujita y Milena Polsinelli Rubi proponen un modelo de lectura profesional orientado a la indización de documentos científicos que sintetiza los modelos más utilizados en indización de contenido: el basado en preguntas a partir de los pronombres interrogativos —heredado de la tradición periodística, en una interpretación un tanto libre del aforismo de Lasswell (1948)—y, por otra parte, el lingüístico-textual utilizado en la elaboración de resúmenes, que consiste en la identificación de las partes de la estructura textual de los documentos que contienen las respuestas para un tipo documental dado. El modelo fue estudiado empíricamente mediante la metodología de protocolos verbales en dos centros de referencia: la Subred Nacional de Información en Ciencias de la Salud (BIREME) y del Centro de Informaciones Nucleares. En el artículo, las autoras publican también el manual para la formación de indizadores.

El artículo de João Batista Ernesto de Moraes y José Augusto Chaves Guimarães sobre el “Análisis documental de contenido de textos literarios narrativos: en busca del diálogo entre las concepciones de *aboutness/meaning* y del recorrido temático/recorrido figurativo” aborda la indización de este tipo de documentos a partir de una noción fundamental en la lingüística textual: la tematicidad intrínseca del texto. La investigación de los autores busca acercar los hallazgos realizados

por los especialistas en organización del conocimiento a los de la lingüística textual, y para ello realizan un análisis comparativo entre las dimensiones de *aboutness* y *meaning* previstas por Begthol (1986) y los recorridos temático y figurativo estudiados por Greimas (1971) y sus seguidores. La relación entre estos dos tipos de recorrido resulta fundamental también en el análisis de la obra gráfica (García y Agustín, 1999; Pinto, García y Agustín, 2002), en los que, recurriendo esta vez a Panofsky (1962, 1970) se detectan tres planos: descripción, identificación e interpretación. Un estudio comparativo entre ambas perspectivas del análisis de contenido de la narrativa y de la obra gráfica bajo el objetivo pragmático de la recuperación aportará sin duda conclusiones interesantes para la teoría de la organización del conocimiento.

Desde una perspectiva diferente, el interesante trabajo de Luis Miguel Moreno Fernández, Mónica Izquierdo Alonso y Antonio Jesús Vázquez Montávez (2006) aborda el siguiente paso, esto es, el momento en que la indización deja de ser análisis de contenido para concretarse en el producto: los términos de indización que servirán para la recuperación. El artículo parte del análisis de la consistencia para ir abordando los diferentes aspectos que determinan la calidad de la indización. El planteamiento teórico se complementa, paso a paso, con un estudio empírico sobre las bases de datos LISA, PASCAL y VOCED. En él, aplicaron en primer lugar las fórmulas de Hooper y Rolling, y luego relacionaron los resultados con otros factores clave del proceso de indizar —corrección, exhaustividad y especificidad—, y calcularon el índice de Kappa. Se trata de una metodología muy valiosa que sitúa la consistencia como una condición *sine qua non* de la calidad, pero que la aborda en conjunto con los demás factores que condicionan la indización.

4. Organización del conocimiento y lenguajes documentales

Sin duda, el aspecto que se identifica normalmente con la organización del conocimiento es el de la construcción, mantenimiento y conexión de lenguajes documentales. Pues bien, en su *Revision of British Standards BS5723 and BS6723 for brief progress report*, Alan Gilchrist (2006) da cuenta de uno de los avances más importantes que están en curso en este campo: la reforma de las normas nacionales e internacionales sobre tesauros.

Como señala Gilchrist, hace tiempo que las normas de tesauros monolingües y multilingües vienen requiriendo una actualización. La primera norma en ser renovada ha sido la estadounidense, bajo el título de *Guidelines for the construction, format, and management of monolingual controlled vocabularies* (ANSI, 2006). En este documento de 172 páginas se sintetiza la doctrina sobre la necesidad del control de vocabulario; se presentan los objetivos, conceptos, principios y estructura de los tesauros; se definen las normas de tratamiento terminológico —selección, alcance y forma de los términos simples y complejos—; las relaciones; las

técnicas de presentación; los aspectos relacionados con la interoperabilidad; y los aspectos relacionados con su construcción, prueba, mantenimiento y gestión. Como se puede observar, la norma americana se limita a los tesauros multilingües, aunque aborda también las listas de descriptores y las taxonomías.

Más ambicioso es el proyecto británico. La norma británica supone varios avances notables sobre la norteamericana, en especial la consideración de la interoperabilidad lingüística, el análisis de facetas —central en la moderna teoría de la clasificación y los tesauros—, la inclusión de los sistemas de clasificación como herramientas complementarias de los tesauros y puntos clave de una estrategia de interoperabilidad entre los lenguajes documentales en conjunto, y la consideración de otros lenguajes como las listas de encabezamientos de materia y las ontologías, además de las listas de descriptores y las taxonomías también tratadas en la última edición de la ANSI Z39.19. Actualmente está sobre la mesa una propuesta británica para acometer la reforma de las normas ISO 2788 e ISO 5964, que se votará en el verano de 2007 y que es apoyada por el grupo español de Vocabularios Estructurados para la Recuperación de Información del comité 50 de la Asociación Española de Normalización.

La perspectiva de una norma que aborde la totalidad de los lenguajes documentales constituye una perspectiva interesantísima para el teórico de la organización del conocimiento, en la medida en que contribuye a una teoría “unificada” de los lenguajes documentales que permitirá, en definitiva, restañar las actuales brechas existentes entre los defensores de los lenguajes precoordinados y los de los postcoordinados, respectivamente, y, en menor medida, de los sistemáticos y alfabéticos. El nuevo entorno de la Internet, en el que todos están presentes y que, por otra parte, exige la interoperabilidad entre ellos, es la evolución que se encuentra detrás de estos movimientos.

Sin embargo, es necesario mantener la modestia ante estas grandes perspectivas. Al fin y al cabo, estos documentos, por más que sean editados por agencias de normalización, tienen el estatus de recomendaciones, y solo pueden pretenden recoger lo mejor de lo llevado a cabo hasta ahora, sin presuponer ni coartar los avances que se producirán en el futuro. En palabras de Gilchrist (2006): “It should be stressed here that it is only by historical accident that these publications appear as standards; and, as such, they differ markedly from normal standards. And while they are guidelines rather than standards, nor are they textbooks. In essence they are attempts to discover best practice through a process of consensus building, made far harder by the existence of different communities of information professionals working in this area”.

En cualquier caso, como señala también Gilchrist, la publicación de normas sólidas y comprensivas en el campo de los sistemas de organización del conocimiento constituye una gran oportunidad para hacer visible la gran experiencia acu-

mulada por los profesionales de la información y la documentación en la resolución de los problemas relacionados con la representación y organización del conocimiento, y, quizá, una de las últimas antes de que este campo de investigación sea absorbido por otros, excluyendo —en el peor de los casos— a una comunidad científica que lleva más de un siglo trabajando en el tema con excelentes resultados.

5. Organización del conocimiento y recuperación de la información

Precisamente, tres de los trabajos presentados abordan el problema de la interoperabilidad que, como se ha visto, es central para explicar los desarrollos recientes que se están produciendo en el campo de la organización del conocimiento. Dos de ellos se ocupan de la cuestión en el entorno abierto de la Internet y el otro lo hace en un entorno corporativo, si se quiere, de intranet.

En el primero, *Interoperable Subject Retrieval in a Distributed Multi-Scheme Environment: New Developments in the HILT Project*, Dennis Nicholson y Emma McCulloch (2006) presentan la evolución y perspectivas del proyecto High-Level Thesaurus (HILT) sobre interoperabilidad en la recuperación temática (<http://hilt.cdlr.strath.ac.uk/>). Se trata de un interesantísimo proyecto que, en relación con el concepto de lenguajes de interconexión, ha promovido un *gateway* terminológico estructurado utilizando la clasificación decimal de Dewey a modo de espina. La metodología de la investigación es modélica. El diseño del prototipo se acometió después de un estudio de campo que estableció la existencia en el Reino Unido de un consenso a favor del mapeo de esquemas como metodología para conseguir la interoperabilidad. El prototipo se ha realizado sobre estándares abiertos, y actualmente está en desarrollo una versión para la comunicación de máquina a máquina sin intervención humana, basada en SRW y SKOS-Core extendido, con el horizonte de futuro de una versión para la implementación distribuida del sistema.

El segundo trabajo, titulado *Information Architecture: creating the foundations for interoperability at Scottish Natural Heritage* y escrito por Jennifer Boyle (2006), presenta un estudio de caso de gran interés sobre la importancia de la interoperabilidad conceptual y terminológica, no solo tecnológica. La autora relata cómo la necesidad de trabajar en la dirección de la Agenda de Modernización Gubernamental condujo al Scottish Natural Heritage (SNH) a la compra e implementación de un Sistema de Gestión de Documentos y Registros Electrónicos corporativo, y detalla todos los pasos dados para que el sistema fuera operable entre los diferentes departamentos de la agencia. Una de las tareas más importantes fue el desarrollo de un Master Authority File (fichero máster de autoridades), capaz de integrar los vocabularios de las diferentes aplicaciones existentes en el SNH.

Por fin, Carlos García Figuerola Paniagua, José Luis Alonso Berrocal, Ángel Francisco Zazo Rodríguez y el Emilio Rodríguez Vázquez de Aldana (2006) pre-

sentan un cuidadoso estudio sobre “El uso de metadatos en las webs universitarias españolas”. En él se confirman cuestiones ya sabidas sobre la utilización de esquemas de metadatos en la Internet: fundamentalmente, que se usan muy poco. Aunque trabajos anteriores de los autores (Figuerola *et al.*, 2006) confirman el efecto positivo de contemplar el uso de metadatos en los modelos vectoriales, la escasa implantación de la especificación Dublin Core u otras semejantes implica una enorme dificultad para aplicar técnicas de recuperación de información basadas solamente en metadatos.

Es interesante, en este sentido, resaltar que los vocabularios estructurados se utilizan fundamentalmente en entornos estructurados en sí mismos que valoran el control de la información mediante metadatos y control de vocabulario y que son conscientes de los logros en eficacia y eficiencia que esta estrategia aporta a medio y largo plazo —organizaciones empresariales y de las administraciones públicas como el SNH y, por supuesto, bibliotecas, archivos, centros de documentación y museos—. Suponen la existencia de una fuerte organización humana detrás, que se refleja en la organización de la información. Sin embargo, gran parte de la información proporcionada en la Internet escapa a este rigor —la Internet en su conjunto es un medio inherentemente no estructurado, aunque puedan darse en ella estructuras informacionales altamente organizadas—, por lo que ambos modelos —el basado en el control y estructuración del vocabulario de búsqueda, y el modelo de síntesis entre el booleano y vectorial que se suele utilizar para buscar en texto libre— son, a efectos prácticos y teóricos, complementarios y complementables.

6. Aplicaciones específicas de la organización del conocimiento

El siguiente bloque de estudios lo conforman dos trabajos que se ocupan de la organización del conocimiento en entornos específicos, pero que proporcionan también conclusiones de carácter general.

En el primero de ellos, Yolanda Martín González y Ana Belén Ríos Hilario (2006) se ocupan en un trabajo titulado “Buscando a Nemo...” de la organización y recuperación de la información en los catálogos en línea destinados al público infantil. Los aspectos relacionados con la organización del conocimiento afloran en la interfaz de usuario a través de la terminología del interfaz, del lenguaje gráfico —presentación de la información e iconos— y de los lenguajes de términos utilizados en la recuperación, que pueden alejarse muchísimo del utilizado por los niños, especialmente si se considera el vector edad. De alguna manera, el estudio del lenguaje de recuperación —incluyendo sus aspectos procesales e icónicos— debería ser incorporado a una teoría de la organización del conocimiento, junto con el metalenguaje propiamente dicho de esta disciplina.

En su “Clasificación y análisis de páginas web comerciales extremeñas”, María del Rosario Fernández Falero, María Antonia Hurtado Guapo y Diego Peral Pa-

checo (2006) nos muestran, además de un interesante panorama de la evolución de la sociedad de la información en Extremadura en el sector de la empresa, uno de los usos frecuentemente descuidados de la clasificación que, sin embargo, resulta ser uno de los más importantes; a saber: el hacer posible el análisis cuantitativo y estadístico. Efectivamente, las clasificaciones permiten la categorización disjunta que facilita la realización de los estudios estadísticos descriptivos. Fernández, Hurtado y Peral dan un paso más y, frente a otras opciones en el campo de la economía, utilizan la CDU, una clasificación universal de carácter bibliográfico, para dibujar con éxito la evolución del sector del comercio extremeño en la Internet. Esta aproximación permite traer de una manera especial a la luz la función estadística de la clasificación, fundamental para resumir la realidad y permitir inferencias. Esta función suele ser olvidada por expertos en recuperación que no son conscientes de su valor operativo; en el campo de las bibliotecas permite, por ejemplo, analizar los préstamos, programar las adquisiciones o dimensionar estructuras y equipos humanos.

7. Evaluación de sistemas de información del conocimiento: la perspectiva del usuario

El artículo de Vera Regina Casari Boccato y Mariângela Spotti Lopes Fujita (2006) sobre la “Evaluación del vocabulario controlado DeCS en logopedia desde la perspectiva del usuario: estudio de observación de la recuperación de información con protocolo verbal” aborda un tema fundamental en la organización del conocimiento: la evaluación de los sistemas. No se trata solo de comparar parámetros bien establecidos y de indudable utilidad como la tasa de reenvío o el tamaño del vocabulario, sino también de abordar la satisfacción del usuario en su utilización. Y es que la experiencia del usuario —experiencia positiva, o satisfacción— es el patrón por excelencia respecto al cual se debe evaluar la calidad de cualquier producto o servicio. El artículo de Boccato y Fujita aborda el análisis cualitativo de esa experiencia por medio de una metodología que las autoras han utilizado en varias ocasiones con éxito en el estudio de la indización: los protocolos verbales. La metodología permite hacer aflorar y justificar adecuadamente la necesidad de abordar el tratamiento de la logopedia en los Descriptores en Ciencias de la Salud (DeCS).

8. Automatización de los procesos de organización del conocimiento

Por fin, el último conjunto de estudios que forman la monografía tiene que ver con la construcción de modelos automáticos de organización del conocimiento, fundamental para dar cuenta de grandes conjuntos de documentos que no permiten el tratamiento manual —y, de manera muy especial, de la Internet—, para desarrollar herramientas de asistencia al indizador y también para facilitar la representa-

ción formal y la prueba empírica de los modelos teóricos propuestos por los especialistas.

Los dos primeros son fruto del trabajo del Grupo de Investigación en Recuperación de Información Avanzada (REINA) de la Universidad de Salamanca, formado por José Luis Alonso Berrocal, Carlos García Figuerola Paniagua, Ángel Francisco Zazo Rodríguez y Emilio Rodríguez Vázquez de Aldana. Este grupo se ha convertido en uno de los principales grupos de investigación españoles que trabajan en las líneas de las Text Retrieval Conferences (TREC), y recientemente han consolidado su trabajo en el estudio de la recuperación de la información en la Internet. En este monográfico, Alonso, García y Zazo (2006b) presentan y describen detalladamente, por un lado, el robot de código abierto que han desarrollado para desarrollar muchos de sus experimentos, SACARINO (Sonda Automática para la Recuperación de Información en la Web). El otro trabajo (Alonso, García y Zazo, 2006a) muestra los resultados de una implementación local de PageRank en SACARINO y de su comparación con el PageRank de Google.

El tercer artículo relata una investigación sobre el “Reconocimiento de títulos de artículos no concordantes con el contenido a través de la utilización de palabras clave” realizada por Manoel Palhares Moreira y Sergio Murilo Stempliuć (2006). Como bien saben los clasificadores e indizadores, un problema fundamental de la clasificación asistida de documentos mediante taxonomías o clasificaciones documentales es la detección de los títulos que no representan correctamente el contenido del documento. Moreira y Stempliuć han realizado una aproximación a partir del cálculo de la coocurrencia de las palabras clave en el título, en la indización, en los resúmenes y, subsidiariamente, en el texto completo. Los resultados preliminares sitúan el problema en un contexto abordable y abren nuevas perspectivas a la investigación.

9. Conclusiones

El conjunto de estudios publicados en este monográfico muestra cómo la organización del conocimiento, lejos de ser una actividad circunscrita a una determinada fase de la cadena documental, es uno de los referentes de la ciencia de la información, que aflora en todos los aspectos del ciclo de la información documental. La organización del conocimiento es fundamental en el análisis de contenido, la indización, el control de vocabulario, la construcción de vocabularios estructurados, la recuperación de la información y la interoperabilidad en los espacios de recuperación complejos como la Internet. Tiene también una implicación de carácter central en el núcleo mismo de la teoría informativo-documental —puesto que la organización conceptual de la información documental remite a la organización del saber y la cultura y a su evolución—, y está presente en los desarrollos automatizados y también en las alternativas de carácter automático que se han propuesto

en los últimos años. Se ha visto, por fin, cómo afecta a cuestiones menos conocidas como el diseño de interfaces y la comunicación con el usuario en entornos automatizados, a la elaboración de estadísticas y la investigación operativa, y a la selección documental y bibliográfica.

Pero es que su importancia tampoco acaba ahí. Además de los momentos del ciclo estudiados en este monográfico, existen otros en los que los sistemas de organización del conocimiento juegan un papel fundamental. Nos limitaremos a dar algunos ejemplos más: así, en la creación de nuevo conocimiento, los mapas conceptuales guían la recopilación de información, la realización del estado de la cuestión y la identificación de las lagunas y conexiones interdisciplinares; en la expresión y comunicación del conocimiento, resultan claves para expresar los conceptos en los términos utilizados por las diferentes audiencias objetivo; en la edición, la asignación temática —interdisciplinar— es fundamental para asegurar el adecuado funcionamiento de los procesos de revisión por pares; en la difusión editorial, el marketing exige comenzar con la identificación y categorización de los clientes potenciales, actividad en la que la determinación de los tópicos de interés resulta ser un aspecto clave; en el descubrimiento de recursos, los observatorios y responsables de adquisición trabajan también con mapas de conocimiento...

En fin, la estructura temática del conocimiento —la realidad que subyace a la organización del conocimiento como disciplina— es un referente de todos los procesos relacionados con el mismo, que se puede manifestar de muchas maneras —intuitiva o razonada, sistematizada o emergente, directa o automatizada, etcétera—, pero que siempre está presente. No es el único aspecto —ni mucho menos— que afecta a la información documental, pero sí se puede afirmar sin lugar a dudas que sitúa a la organización del conocimiento en el seno del núcleo de disciplinas ineludibles para describir, explicar y mejorar el mundo de la transferencia del conocimiento.

En este sentido, conviene insistir en huir de perspectivas técnicas de carácter ingenuo: la organización del conocimiento no puede ser solo un esfuerzo voluntarista y técnico, sino que, sobre todo, debe partir de la investigación sobre cómo los conocimientos se organizan en la realidad. Así como la automatización es un invento relativamente reciente bajo el control de la razón humana —aunque sometido también a los procesos naturales y sociales en el seno de los cuales se desenvuelve—, la organización del conocimiento —la estructuración temático-conceptual— es un proceso que ha acompañado la evolución de los seres humanos y el lenguaje. Por lo tanto, es un proceso que debe ser comprendido, no solo manipulado. Del mismo modo que las tecnologías del mundo físico se basan en el conocimiento de las leyes de la física, de la química y de la biología, las tecnologías del conocimiento se deben apoyar en el descubrimiento y la comprensión de las leyes del conocimiento psicológico y social.

Referencias

- Alonso Berrocal, José Luis; García Figuerola Paniagua, Carlos; Zazo Rodríguez, Ángel Francisco (2006). Mejoras en la recuperación de información en la Web mediante el tratamiento de la información de los enlaces. // *Scire: Representación y Organización del Conocimiento*. 12:1 (en.-jun. 2006) 195-207.
- Alonso Berrocal, José Luis; García Figuerola Paniagua, Carlos; Zazo Rodríguez, Ángel Francisco (2006). SACARINO (Sonda Automática para la Recuperación de Información en la Web), un robot para recorrer y procesar la Web. // *Scire: Representación y Organización del Conocimiento*. 12:1 (en.-jun. 2006) 209-222.
- Begthol, C. (1986). Bibliographic classification theory and text linguistics: aboutness analysis, intertextuality and the cognitive act of classifying documents. // *Journal of Documentation*. 42:2 (June 1986) 84-113.
- Boyle, Jennifer (2006). Information Architecture: creating the foundations for interoperability at Scottish Natural Heritage. // *Scire: Representación y Organización del Conocimiento*. 12:1 (en.-jun. 2006) 123-137.
- Casari Boccato, Vera Regina; Lopes Fujita, Mariángela Spotti (2006). Evaluación del vocabulario controlado DeCS en logopedia en la perspectiva del usuario: estudio de observación de la recuperación de información con protocolo verbal. // *Scire: Representación y Organización del Conocimiento*. 12:1 (en.-jun. 2006) 177-193.
- Delgado Gómez, Alejandro (2006). El principio de contextualidad y relación de los documentos: una aproximación tentativa. // *Scire: Representación y Organización del Conocimiento*. 12:1 (en.-jun. 2006) 23-46.
- Ernesto de Moraes, João Batista; Chaves Guimarães, José Augusto (2006). Análisis documental de contenido de textos literarios narrativos: en busca del diálogo entre las concepciones del *aboutness/meaning* y del recorrido temático/recorrido figurativo. // *Scire: Representación y Organización del Conocimiento*. 12:1 (en.-jun. 2006) 71-82.
- Fernández Falero, María del Rosario; Hurtado Guapo, María Antonia; Peral Pacheco, Diego (2006). Clasificación y análisis de páginas web comerciales extremeñas. // *Scire: Representación y Organización del Conocimiento*. 12:1 (en.-jun. 2006) 161-176.
- García Figuerola Paniagua, Carlos; Alonso Berrocal, José Luis; Zazo Rodríguez, Ángel Francisco; Rodríguez Vázquez de Aldana, Emilio (2006a). El uso de metadatos en las webs universitarias españolas. // *Scire: Representación y Organización del Conocimiento*. 12:1 (en.-jun. 2006) 139-149.
- García Figuerola Paniagua, Carlos; Alonso Berrocal, José Luis; Zazo Rodríguez, Ángel Francisco; Rodríguez Vázquez de Aldana, Emilio (2006b). Web page retrieval by combining evidence. // *Accessing Multilingual Information Repositories: 6th Workshop of the Cross-Language Evaluation Forum, CLEF 2005, Vienna, Austria, 21-23 September, 2005, Revised Selected Papers*. 880-887. Springer GMBH, 2006. *Lecture Notes in Computer Science*, 4022.
- García Marco, Francisco Javier; Agustín Lacruz, María del Carmen. El análisis de contenido de las reproducciones fotográficas de obras artísticas. // Valle Gastaminza, Félix del (ed.). *Manual de documentación fotográfica*. Madrid: Síntesis, 1999. 133-168.

- Gilchrist, Alan (2006). Revision of British Standards BS5723 and BS6723 for brief progress report. // *Scire: Representación y Organización del Conocimiento*. 12:1 (en.-jun. 2006) 97-105.
- Greimas, Algirdas Julián (1971). *Semántica estructural: investigación metodológica/versión española de Alfredo de la Fuente*. Madrid: Gredos, DL 1971.
- Lasswell, H. (1948). *The Structure and Function of Communication in Society*. // Bryson, L. (ed.). *The Communication of Ideas*. New York: Harper. 37-51.
- Lopes Fujita, Mariângela Spotti; Polsinelli Rubi, Milena (2006). Modelo de lectura profesional para la indización de textos científicos. // *Scire: Representación y Organización del Conocimiento*. 12:1 (en.-jun. 2006) 47-69.
- Martín González, Yolanda; Ríos Hilario, Ana Belén (2006). Buscando a Nemo...: la organización y recuperación de la información en los catálogos en línea destinados al público infantil. // *Scire: Representación y Organización del Conocimiento*. 12:1 (en.-jun. 2006) 151-160.
- Moreno Fernández, Luis Miguel; Izquierdo Alonso, Mónica; Vázquez Montávez, Antonio Jesús (2006). La consistencia de la indización en documentación: los casos de LISA y PASCAL. // *Scire: Representación y Organización del Conocimiento*. 12:1 (en.-jun. 2006) 83-96.
- Nicholson, Dennis; McCulloch, Emma (2006). Recuperación temática interoperable en un entorno distribuido multi-esquema: nuevos desarrollos en el proyecto HILT. // *Scire: Representación y Organización del Conocimiento*. 12:1 (en.-jun. 2006) 107-122.
- Palhares Moreira, Manoel; Murilo Stempliu, Sergio (2006). Reconocimiento de títulos de artículos no concordantes con el contenido a través de la utilización de palabras clave. // *Scire: Representación y Organización del Conocimiento*. 12:1 (en.-jun. 2006) 223-243.
- Panofsky, Erwin (1962). *Studies in the Visual Arts*. Nueva York: Harper & Row, 1962.
- Panofsky, Erwin (1970). *Meaning in the visual arts*. [Middlessex]: Penguin, [1970].
- Pinto Molina, María; García Marco, Francisco Javier; Agustín Lacruz, María del Carmen. Indización y resumen de documentos digitales y multimedia: técnicas y procedimientos. Gijón: Trea, 2002.
- Shannon, Claude; Weaver, Warren (1949). *The mathematical theory of communication*. Urbana: The University of Illinois, 1949.