

Un panorama de la investigación en organización y representación del conocimiento

Francisco Javier García Marco
Universidad de Zaragoza

0.1. Resumen

Análisis de las tendencias de la investigación en representación y organización del conocimiento para la recuperación de información a partir de la contextualización y consideración crítica del conjunto de estudios publicados en el número monográfico de la revista *Scire* (vol. 8, nº. 1, enero-junio 2002) sobre *Perspectivas en representación y organización del conocimiento*.

Palabras clave: Tendencias de investigación. Producción científica. Representación del conocimiento. Organización del conocimiento. Recuperación de la información. Análisis crítico. Análisis cualitativo.

0.2. Abstract

Analysis of the research trends in the field of knowledge representation and organization for information retrieval taking as a base a critical appraisal and contextualization of the set of articles published in a monographic number of the journal *Scire* titled *Perspectives in representation and organisation of knowledge* (*Scire*, vol. 8, n. 1, January-June 2002).

Keywords: Research trends. Scientific production. Knowledge representation. Knowledge organisation. Information retrieval. Critical analysis. Qualitative analysis.

1. Introducción

El presente volumen monográfico ofrece un conjunto representativo de los esfuerzos de investigación en representación y organización del conocimiento que se están realizando en la actualidad. La representatividad de los artículos se

aprecia tanto por en temas tratados —de enorme actualidad— como en la procedencia disciplinar de sus autores, pues participan en este monográfico investigadores procedentes de la Filosofía, la Lingüística, la Informática, la Biblioteconomía, la Documentación y la Archivística.

2. Perspectivas interdisciplinares

Abren el libro tres capítulos que ofrecen sendas perspectivas netamente interdisciplinares de la organización del conocimiento; perspectivas de las que, en opinión de Emilio Delgado en una monografía recientemente publicada (2002, p. 229-231), están tan necesitadas las Ciencias de la Documentación.

El primero de ellos es un trabajo del filósofo y sociólogo Roberto Poli —uno de los especialistas destacados en el campo de las ontologías y director de la revista *Axiomata*—, que nos ofrece una completa panorámica de esta disciplina emergente. Es evidente que, a medida que el conocimiento crece —y en la actualidad lo está haciendo de manera exponencial en una sociedad específicamente organizada para su creación y explotación, la sociedad del conocimiento—, el problema de organizarlo se vuelve más importante y urgente. Roberto Poli muestra como una de las disciplinas seminales de la ciencia occidental —la ontología— vuelve a convertirse en foco de la atención de lingüistas, ingenieros, psicólogos, filósofos y otros científicos y tecnólogos por su importancia para iluminar y dilucidar numerosos problemas prácticos de los servicios de información digitales que se están convirtiendo en los pilares infraestructurales de la sociedad de la información.

En segundo lugar, el lingüista Carlos Inchaurrealde se ocupa de un tema de enorme interés para lingüistas y científicos de la información y la documentación en el marco políglota propiciado por la creciente internacionalización, de un lado, y por el renacer de las culturas regionales, del otro, que caracterizan el fenómeno de la globalización. Su trabajo aborda la representación léxica del conocimiento multilingüe. Su trabajo constituye una muestra de la relevancia de la investigación lingüística para la Ciencia de la Información. No en vano, la clasificación de las ciencias de la UNESCO —con mejor o peor fortuna, pero, en cualquier caso, con toda su autoridad— adscribe la Documentación al área de las ciencias del lenguaje.

Finalmente, cierra este primer conjunto de estudios, un extenso artículo en el que el bibliotecario y archivero Alejandro Delgado aborda críticamente las aportaciones realizadas desde la informática y la ingeniería en el campo de la Defensa a la representación ontológica en torno al desarrollo de los lenguajes de modelado IDEF5 —ICAM (Integrated Computer Aided Manufacturing) DEFINITION, y muestra sus enormes posibilidades en el nuevo contexto caracterizado por el renovado interés en la representación lógica y automatizable de la realidad.

Estos tres estudios muestran hasta qué punto es importante que los especialistas en información y documentación estén al tanto de los avances e investigaciones que se producen en otras disciplinas, superando la dinámica endógena a la que el sistema universitario y profesional conducen, tanto más cuanto el campo de las investigaciones sobre la información semeja un ecosistema inmenso y complejo por el número de “especies” que se mueven en él y por la maraña de relaciones que existen entre ellas. Por ello, planteamos para el debate público la necesidad de que una parte importante de la formación actitudinal de nuestros futuros investigadores se oriente a que sean capaces de buscar en su agenda, sus contactos y sus valores el equilibrio entre un firme compromiso disciplinar y una atenta, activa y modesta apertura a las aportaciones de otras disciplinas relacionadas. Hasta cierto punto, el crecimiento de la disciplina —favorecido por un sistema de educación superior que, con todos los defectos, le ha sido muy favorable— ha podido propiciar una cierta complacencia autorreferente e, incluso, un cierto discurso expansionista. Con todos sus defectos, esa nueva autoestima — que debería incluir también la percepción de los propios límites y el respeto por el trabajo de otros colegas— supone una excelente plataforma para profundizar en la colaboración interdisciplinar.

3. Recuperación de la información y construcción de tesauros

El siguiente bloque de trabajos aborda en diferente medida los temas clásicos de la organización y representación del conocimiento en los sistemas de información y documentación: el análisis de contenido, la representación del usuario y el desarrollo de lenguajes documentales y de autoridades.

En el primer capítulo de este bloque, Carlos García Figuerola, Ángel Francisco Zazo y José Luis Alonso abordan un capítulo ya clásico de la representación del usuario y del documentalista en los sistemas de recuperación de la información: la realimentación por relevancia. A través de la retroalimentación por relevancia resulta posible implementar en los sistemas de recuperación de información parte de las habilidades tradicionalmente aportadas por los mediadores humanos y, en definitiva, representar conocimiento experto. Los investigadores salmantinos explican los fundamentos de las técnicas subyacentes y ofrecen un estudio realizado con metodología experimental, rara en el panorama bibliotecológico español.

Seguidamente, se presentan dos trabajos que abordan la construcción de tesauros desde perspectivas nuevas. Las limitaciones de la estructura de los tesauros tradicionales frente a las posibilidades ofrecidas por las nuevas tecnologías y los retos planteados por los enormes espacios de recuperación de información que ha propiciado la Internet viene siendo señalada en la investigación sobre el tema desde hace décadas (Dahlberg, ; García Marco, 2002). Sin embargo, la

construcción de tesauros en el ámbito hispano raramente ha traspasado las relaciones clásicas, sin arpar los límites de las normas UNE 50-106-90 y 50-125.

La situación está cambiando rápidamente, y los dos trabajos que analizamos a continuación son un buen ejemplo de ello. El primero, a cargo de un equipo interdisciplinar de documentalistas e informáticos de la Universidad Carlos III de Madrid —compuesto por Miguel Ángel Marzal García-Quismondo, Pilar Beltrán Orenes, Jorge Luis Morato Lara, Juan Llorens Morillo, José Antonio Moreira González y Sonia Sánchez Cuadrado— aborda la ampliación de los conjuntos de términos tradicionalmente usados en los tesauros —normalmente sintagmas nominales—, investigando las posibilidades de los verbos como descriptores. El trabajo del equipo de la Universidad Carlos III afronta también una de las líneas de investigación más importantes en la construcción de tesauros —su automatización— y muestra en este sentido también como la construcción de tesauros de verbos puede facilitar la construcción automática de tesauros documentales clásicos.

El segundo capítulo presenta la experiencia de otro grupo de investigación madrileño interuniversitario —compuesto por Félix del Valle Gastaminza, Antonio García Jiménez, Inmaculada Chacón, José Ramón Laiseca, Cristina de la Vega, María Eugenia Martín Giraldo, Blanca Gil Urdiciain, Antonio Hernández y David Rodríguez— responsable de la construcción de un tesoro para la cadena de televisión Telecinco. El proyecto se caracteriza por el cuidadoso análisis de las relaciones semánticas dentro del tesoro, en línea con el estuendo trabajo realizado por Antonio García Gutiérrez (1998) y su equipo en la concepción y desarrollo del *Tesoro de Patrimonio Histórico Andaluz*. Así pues, los dos artículos sobre tesauros abordan la renovación de la construcción de tesauros en el ámbito español tanto en lo que se refiere a su vocabulario como a las relaciones representadas en ellos.

4. Análisis de contenido

El artículo de Madalena Martins Lopes Naves está dedicado al análisis de los factores informativos y cognitivos que afectan al proceso de análisis de contenido durante la indización de los documentos. Partiendo de las ideas de Clare Beghtol (1986) y María Pinto (1994) así como de un detallado estudio del estado de la cuestión dibuja un modelo del proceso del análisis de contenido orientado a la indización y propone una metodología para su estudio basada en la realización de entrevistas abierta y el análisis de protocolos verbales, que son muy prometedores para el análisis científico de los muchos fenómenos de interés dentro de las Ciencias de la Documentación que no permiten, en el estado actual del conocimiento, un abordaje más formalizado y, mucho menos, experimental.

5. Estudios especializados

Cierran el número tres artículos que destacan por ofrecer interesantes perspectivas en ámbitos muy especializados. El trabajo de Rosa Santos se inserta dentro de la tradición empírica en la elaboración de tesauros, pero aborda un ámbito de extraordinario interés por la importancia que adquiere en él el vector temporal: el desarrollo de sistemas de descriptores para la recuperación de información de carácter histórico. Además, el artículo de Rosa Santos muestra una aplicación menos conocida de los tesauros, pero que es parte de una tradición muy antigua dentro de la investigación documental en nuestro país (1), a saber, su empleo como base para organizar conceptualmente un tema de investigación, para fundamentarlo firmemente sobre un corpus documental y para la eventual explotación sistemática e informétrica de su apoyatura crítica. En el trabajo de Rosa Santos se muestra como el tesoro puede ser un instrumento al servicio del papel central que la clasificación ocupa en el desarrollo del pensamiento científico.

Del artículo del equipo de la Universidad de Extremadura compuesto por sobre la *Determinación de campos onomásticos en las bases de datos descriptivas de imagen en movimiento* me interesa destacar sobre todo dos aspectos. En primer lugar, aborda un tema injustamente considerado menor dentro de la disciplina del control del vocabulario: el control de autoridades de nombres personales y entidades. A nadie se le escapa que personas y entidades son temas fundamentales y recurrentes en los procesos de recuperación de la información. Un examen más detallado del problema muestra que las relaciones entre las entidades personales, grupales y jurídicas son tan complejas como las que ocurren en el resto de los temas “comunes”. En segundo lugar, el trabajo de nuestros compañeros de Extremadura enfoca el problema dentro del contexto del análisis multimedia, donde se presenta de forma especialmente aguda el problema de la relación entre los signos y el contenido comunicativo, y, en torno a él, los fenómenos de la referencia, la denotación, la connotación y la designación (García Marco, 1996). Se trata de un enfoque de extraordinario interés en la línea de otros recientes, como el que aparece en este mismo número monográfico a cargo de Félix de Valle y sus colaboradores, o también el reciente estudio de María Pinto, Javier García Marco y Carmen Agustín (1992).

Finalmente, el artículo de José Luis Herrera Morillas sobre *El patrimonio bibliográfico a través de los museos y exposiciones virtuales: análisis y descripción del contenido*, además de su interés inmediato para el estudio del patrimonio bibliográfico y para su promoción, ofrece sobre todo, dentro del marco de este volumen, una metodología para el análisis de contenido, descripción y resumen de recursos web. El modelo utilizado por José Luis Herrera coincide notablemente con el del resumen indicativo: al comienzo, una primera parte anotativa precisa el tema del recurso y el alcance con el que ha sido tratado —circunscrip-

ción temporal, geográfica o temática, amplitud del recurso, etc.—; seguidamente, se indican con detalle y de forma sistemática los temas tratados; finalmente, se resume la metodología con la que se han presentado. Precisamente por la especificidad del material tratado —recursos web, temática bibliográfica—, el artículo de José Luis Herrera muestra la validez que el modelo clásico de resumen indicativo sigue teniendo para el abordaje de los nuevos materiales digitales y materiales. Este hecho nos confirma en nuestra creencia de que existen aspectos lógicos y epistemológicos en los fenómenos que son objeto del estudio de la Documentación cuyo alcance sobrepasa a los meramente tecnológicos —sin poner en cuestión, claro está, la retroalimentación que existe entre los diferentes niveles de estratificación de lo real—, y que en definitiva garantiza la continuidad científica en el área a pesar del revolucionario impacto que han tenido las tecnologías de la información.

6. Conclusiones

En definitiva, el conjunto de estudios que se presentan en este número monográfico sobre organización y representación del conocimiento constituyen una instantánea del momento actual de la investigación sobre el tema, pues:

- a) abordan toda la jerarquía conceptual de la disciplina,
 - desde los aspectos filosóficos y lingüísticos, cuya naturaleza es eminentemente teórica,
 - pasando por los modelos empíricos y lógicos utilizados en el desarrollo de las herramientas informáticas que están en la raíz de los esfuerzos de representación del conocimiento experto, y ...
 - descendiendo hasta los proyectos de desarrollo de tesauros, sistemas de autoridades y resúmenes, que constituyen la vertiente aplicada — y probablemente más genuina— de nuestra disciplina;
- b) lo hacen tanto en los aspectos generales como de forma aplicada en los más variados campos del conocimiento y del desempeño profesional: medios de comunicación, estudios bibliográficos; y, finalmente, ...
- c) insertan el estudio de las últimas novedades en nuestra disciplina —fundamentalmente relacionadas con la explosión de la información digital en Internet— en su mejor tradición filosófica, científica e historiográfica.

Este conjunto de trabajos nos envían un mensaje inequívoco sobre los valores que deben presidir el desarrollo futuro de la investigación en representación y organización del conocimiento: la interdisciplinariedad, la acumulación crítica del saber, la integración y sistematización de principios y modelos —tanto a nivel interno de la disciplina, como externo con el resto de las ciencias y tecno-

logías—, la búsqueda de la consistencia lógica y empírica, y el equilibrio entre desarrollo teórico y aplicación práctica.

Esos son, precisamente, los valores que caracterizan a la actividad científica en general. Este hecho nos confirma en la idea de que, a pesar de las limitaciones de nuestra actividad científica —que a nadie se le escapan y que tienen que ver, entre otros factores con la juventud del área científica y su propia interdisciplinariedad (Delgado, 1992)—, la investigación científica en las Ciencias de la Información y la Documentación marcha por el buen camino.

7. Notas

- (1) Algunas de las primeras tesis de la Documentación española, dirigidas por José María López Piñero en el Instituto de Estudios Documentales e Históricos sobre la Ciencia incluyen la realización de tesauros no solo como herramienta de control documental, sino como instrumento para el análisis bibliométrico, esto es, para el análisis de contenido de la producción bibliográfica española en Medicina. Es el caso de las tesis de Sopena Monforte (1972) —concebida con un concepto muy moderno del tesoro, pues incluía el tratamiento terminológico mediante un glosario en la línea de lo que actualmente varios expertos españoles denominan *tesauros conceptuales*— y la de Torres Beneyto (1977).

8. Referencias

- Asociación Española de Normalización y Certificación (1990b). Directrices para el establecimiento y desarrollo de tesauros monolingües. Madrid: AENOR, D.L. 1990. 47 p. UNE 50-106-90. Equivalente a ISO 2788-1986.
- Asociación Española de Normalización y Certificación (1997b). Documentación. Directrices para la creación y desarrollo de tesauros multilingües = Documentation. Guidelines for the establishment and development of multilingual thesauri = Documentation. Principes directeurs pour l'établissement et le développement de thesaurus multilingues. Madrid: AENOR, 1997. 77 p. Normas UNE 50125. Norma equivalente a ISO 5964:1985.
- Beghtol, Clare (1986). Bibliographic classification theory and text linguistics: aboutness, intertextuality, and the cognitive act of classifying documents. // *Journal of Documentation*. 42 : 2 (June 1990) 84-113.
- Dahlberg, Ingetraut (1995). Current trends in knowledge organization. // García Marco, Francisco Javier. Organización del conocimiento en sistemas de información y documentación. 1 (1995) 7-26.
- Delgado López-Cózar, Emilio (2002). La investigación en biblioteconomía y documentación. Gijón: Ediciones Trea, 2002.
- García Marco, Francisco Javier (1996). Fundamentos de tratamiento y recuperación de la información. Zaragoza: Proyecto Scire, Universidad de Zaragoza, 1996.
- García Marco, Francisco Javier (2002). La literatura científica sobre lenguajes poscoordi-

- nados en España: de la divulgación del concepto a la Internet. // *Documentación de las Ciencias de la Información*. 25 (2002) 291-319.
- García Gutiérrez, Antonio Luis (1998). *Principios de lenguaje epistemográfico: la representación del conocimiento sobre Patrimonio Histórico Andaluz*. Sevilla : Instituto Andaluz del Patrimonio Histórico, 1998. 91 p. (Cuadernos Técnicos).
- Pinto Molina, María. (1994). Interdisciplinary approaches to the concept and practice of Written Documentary Content Analysis (WTDCA). // *Journal of Documentation*. 50 : 2 (June 1994) 111-133.
- Pinto Molina, María; García Marco, Francisco Javier; Agustín Lacruz, María del Carmen (2002). *Indización y resumen de documentos digitales y multimedia: técnicas y procedimientos*. Gijón : Trea, 2002
- Sopena Monforte, José Ramón (1972). *Thesaurus, codificación y glosario de diagnósticos endocrinológicos* [Tesis doctoral]. Valencia : Universidad de Valencia, Facultad de Medicina, 1972
- Tesoro de Patrimonio Histórico Andaluz: versión 0 / Antonio García Gutiérrez; coordinador, Antonio Martín Pradas; compiladores, Juan Agudo Torrico ,[y otros/as]. Sevilla: Junta de Andalucía, Consejería de Cultura, Instituto Andaluz del Patrimonio Histórico, Editorial Comares, 1998. 1035 p.
- Torres Beneyto, Jesús (1977), *Thesaurus de diagnósticos cardiológicos* [Tesis doctoral]. Valencia: Universidad de Valencia, Facultad de Medicina, 1977.
- Unesco. *Nomenclatura internacional de la Unesco para los campos de ciencia y tecnología: segunda versión en español* / Grupo de Trabajo de Información y Documentación de la Comisión Española de la UNESCO. Madrid: Unesco. Comisión Nacional de España, 1987.