

TRAYECTORIA DE LA ORGANIZACIÓN DEL CONOCIMIENTO Y LA TECNOLOGÍA DE LA INFORMACIÓN EN RELACIÓN CON LAS ETAPAS DE LA COMUNICACIÓN CIENTÍFICA

*José Antonio Moreiro-González
Universidad Carlos III de Madrid*

1. INTRODUCCIÓN

La línea argumental viene marcada por la respuesta adaptativa de los vocabularios documentales a las necesidades que atravesó cada periodo histórico respecto al acceso a la información científica. En cada momento de cambio se advierte un paralelismo entre la generación y difusión de objetos de información por medio de tecnologías recién llegadas y la respuesta adecuada a partir del empleo de las mismas tecnologías que ocasionan los cambios. En cualquier tiempo, la tecnología sobre la que se apoya la creación de los nuevos tipos de soportes documentales y que impulsa su difusión, y cuyo acceso y control se complican inicialmente, apoya los métodos y procesos de respuesta a su tratamiento y recuperación.

Si nos centramos en el momento presente, muchos de los sistemas de representación son anteriores a la aparición de la comunicación cibernética, por lo que han estado obligados a pasar por una fase de adaptación a la nueva situación. Esta circunstancia ha llevado a organizar y representar el conocimiento con nuevos sistemas simbólicos, así como a reformular los precedentes. Y esto ha sucedido cada vez que se abrió un estado de cosas que variaba bastante lo establecido.

El primer objetivo de estas páginas consiste en considerar los momentos de inicio o cambio de los sucesivos movimientos en que la comunicación de la ciencia y el progreso de la tecnología llevaron a planteamientos innovadores y rupturistas en relación con el acceso a los objetos de información desde los vocabularios de representación y la organización del conocimiento.

Plantear las aportaciones hechas en los momentos cruciales desde lo novedoso y diferenciador y las consecuencias que tuvieron en la organización del conocimiento.

Asimismo, en inmediación temporal, atender dentro de un entorno determinado por la Web y por la interoperabilidad, a los cambios que han sufrido los conceptos y aplicaciones de los SOC preexistentes, para adaptarse a las nuevas formas de creación, difusión y recuperación de la información. Al tiempo que analizar las ideas de reciente aparición.

2. ORÍGENES DE LA ORGANIZACIÓN DEL CONOCIMIENTO

La primera guía para relacionar las búsquedas con el contenido de los documentos puede asignarse a Calímaco cuando organizó el Museo de Alejandría en ciento veintisiete apartados disciplinares para guardar u ofrecer la información por especialidades. Los *pinakes* surgieron al aplicar las categorías clasificatorias o modos de predicación de todos los seres (Aristóteles 2001) a los diferentes índices que encaminaban a las materias de las que trataban las colecciones de rollos (Casson 2001). Aunque también los distribuían por materias los *syllabus*, registros de identificación que, a la vez, permitían su búsqueda. Los antecedentes de los índices se relacionaron en tiempo y función con los primeros repertorios, pues Galeno en Pérgamo, otra ciudad helenística, reflejó la organización y las materias de su propia colección en *De libris propriis liber* y en *De ordine librorum suorum liber* (Terradas 1983). Por consiguiente, se establecían ya indizaciones genéricas en los primeros intentos de referenciar obras por sus contenidos. En estos casos mediante unas listas alfabéticas de autores o de materias que señalaban el lugar donde aparecían para facilitar el acceso. Cumplían, pues, los principios fundamentales de la indización documental de acuerdo con UNISIST (UNESCO 1983), al describir y caracterizar un documento mediante la representación de sus conceptos y transcribirlos en algún sistema de representación que, andando el tiempo, acabaríamos llamando lenguajes documentales y, después, vocabularios semánticos.

El conocimiento y los códices se conservaron durante la alta Edad Media en las bibliotecas monacales para lo que tenían que contar con sistemas de

ordenación y de acceso al contenido de los manuscritos. Se sirvieron del orden alfabético, como habían hecho antes los pitagóricos, se enumeraron los capítulos y las secciones y se incluyeron índices. Listas de palabras, por lo común en orden alfabético que aparecen al final de un ejemplar localizando en sus páginas un nombre o un asunto. Facilitaban encontrar la información deseada evitando consultar el libro entero. Aquellos índices han llegado en uso hasta nosotros, no sustituían a la información, se limitaban a señalarla. Incluso en aquellas bibliotecas monásticas se establecieron los índices de concordancias terminológicas bíblicas y de las obras en ellas custodiadas de mayor interés, sin duda antecedentes de los índices permutados y, por tanto, de la indización con palabras-clave libres.

Aún después, en las bibliotecas episcopales, de inserción urbana y más abiertas al intercambio y circulación de obras, con influjo inmediato sobre las universitarias que, para atender a las mayores necesidades de sus usuarios, tenían que reflejar las materias impartidas en las diferentes escuelas en la manera práctica de organizar sus colecciones. Precisamente en la biblioteca de un monasterio romano se empleó un precedente de los índices de materias aplicado a su colección de Derecho canónico (Beare 2007). Las tareas que en aquellas bibliotecas realizaban los bibliotecarios aparecen recogidas en las normas de San Isidoro (1951), mientras que Richard de Bury en 1473 (Bury 1972) atendió a las desarrolladas en las bibliotecas universitarias. Por más que, para nuestros propósitos, no pasasen de ser clasificaciones genéricas por materias. Fueron entonces muy utilizados los índices de nombres de los autores contribuidos en los repertorios biobibliográficos que recogían la biografía de los escritores con la lista de sus obras.

Con la aparición de la imprenta se facilitó la obtención de copias ilimitadas de cada libro, por lo que se desbordaron las anteriores relaciones a fondos exclusivos de una sola biblioteca. Cada libro, tras el sumario de sus contenidos, ofrecía índices de los nombres y de las materias tratadas. Pero, no bastaba, pues la multiplicación de originales creció tan rápido que se necesitaron unas listas que informasen de las existencias y que satisficiesen el afán cultural de los humanistas y al aumento de lectores. Desde el siglo XVI las bibliografías inventariaron los impresos existentes bajo denominaciones como *bibliotheca*, *index*, *inventarium* o *repertorium* ... (Millares, 1971). Esta variedad de sinónimos no oculta que se tratase, en definitiva, de recursos referenciales para conocer y acceder a la existencia y consulta de aquellos. En menos de un siglo hubo ya barridos totales de la producción gracias a los registros de las obras para ser sometidas a depósito

legal y a la censura, cuyas listas se publicaban como índices de obras prohibidas. Para favorecer su consulta, las bibliografías ofrecían índices de nombres y de materias que fuesen interesantes para formar una entrada o incluso para vender con acierto una obra. Así había índices de nombres propios, de nombres geográficos, tipográficos y cronológicos. Pero también índices de materias e índices analíticos que, para su elaboración, requerían de un dominio terminológico de la especialidad analizada.

3. REVOLUCIÓN CIENTÍFICA, TERMINOLOGÍA ESPECIALIZADA Y LENGUAJES DOCUMENTALES

Una de las vías sobre las que se estableció la Ciencia Moderna fue el modo de comunicar y circular el conocimiento. En el siglo XVII se superó el libro como soporte documental habitual cuando aparecieron las publicaciones periódicas para comunicar los avances y descubrimientos científicos, al tiempo que para atender a las necesidades informativas de los investigadores. La revolución científica necesitaba un medio de expresión más rápido en su redacción y confección que las monografías y más preciso y regular que las conferencias y debates en las Sociedades Científicas. Por más que se debiera a estas el impulso de su edición como fórmula para difundir los nuevos conocimientos. Aunque la primera de ellas aparecida en 1665, el *Journal des Sçavants* se debió al empeño de Denis de Sallo, la segunda y de carácter pleno científico, *Philosophical Transactions* se editó por la Royal Society de Londres tan sólo desde unos meses después, en 1666. Las publicaciones periódicas se convirtieron desde el primer momento en el mejor medio para difundir los conocimientos científicos. Fue causado por una doble razón. Desde su aparición las revistas contenían información original formada por trabajos científicos breves. Y, además, información referencial que cumplía la misión de permitir el acceso a documentos primarios. Función que en la actualidad cumplen las bases de datos o las revistas de resúmenes. Estas dos características, vehículo y depósito de información, son exigencias de una comunidad científica que amplió mucho su número y sus investigaciones que formalizó en nuevos canales de transmisión del conocimiento en el siglo XVII.

Las revistas contenían información sobre los descubrimientos e investigaciones, junto a información que daba noticia de la existencia de aquellos, exigencias de una comunidad científica que se incrementó mucho en ese siglo y que necesitaba nuevos canales de transmisión de conocimientos. Las publicaciones

periódicas iniciaron la transferencia dinámica de la información científica, al situarse como el núcleo básico del análisis y la difusión bibliográficos. Cada número de las revistas contaba con un índice y, pronto, cada título también. Además, esta vía de ampliación de los impresos trajo una modificación radical en la forma en que las bibliografías se confeccionaban, pues se obligaron a aparecer, a su vez, de forma periódica, regular y normalizada, con barridos exhaustivos y progresivo seguimiento de la producción por especialidades (Millares 1973). Con estas modificaciones cobraba carta de naturaleza la información científica sobre la que se desarrollaba la ciencia experimental y sobre la que se fijó el eje de lo que sería la Documentación a finales del siglo XIX (Mueller 1995).

Entre los siglos XVI y XIX las ciencias naturales alcanzaron un sistema de unidades de medida y de referencias universales. Entre ellas, la de un lenguaje al servicio preferente de la comunicación entre expertos, lógico y muy formalizado, donde los términos empleados era referentes ideales y homogéneos, del que se había eliminado cualquier posibilidad de sinonimia o de polisemia (Krieger; Bocorny 2004). Si la ciencia se comunicaba dentro de estos márgenes, los sistemas de información especializados tenían también que llegar a establecer los términos como unidades de conocimiento para ser operativos. Para contextualizar el sentido en que cada término debía ser comprendido Roget, en su *Thesaurus of English Words and Phrases* (1883), lo situó mediante las relaciones que mantiene con otros términos. Mostrándose así que la categorización de un dominio se hace de acuerdo con las normas terminológicas, mientras que luego, en los lenguajes controlados, la selección de descriptores resultará de las normas documentales. Avanzado en siglo XIX, se abrió la aspiración de los lenguajes controlados a representar de forma normalizada los conceptos de los dominios del conocimiento mediante términos unívocos que guardasen los principios lógico-semánticos. Inicialmente esos lenguajes favorecieron el acceso a la información por materias, precediendo a los lenguajes combinatorios que aparecerían mediado el siglo XX.

La comunicación del conocimiento ha exigido formalizaciones de la información de las que se ha beneficiado su adquisición, desarrollo y transferencia. Desde una perspectiva histórica, en Estados Unidos, el acceso a la información por materias abrió la trayectoria espectacular que tendrían los lenguajes documentales en el siglo XX (Izquierdo; Moreno 1994). Esta afirmación se entiende al considerar que, en la misma fecha de 1876, se publicaron dos obras trascendentes, la Clasificación decimal de Dewey y las *Rules for a dictionary catalog* de Cutter. La primera fijó la metodología y

estructura de los sistemas de clasificación en cuanto lenguajes de estructura jerárquica y precoordinada. Mientras que la de Cutter sigue vigente en los *Encabezamientos de materia*, de carácter precoordinado también, pero de estructura asociativa y con el control de vocabulario de aplicación específica a las materias a indizar que caracterizará a los lenguajes controlados. La Clasificación de Dewey buscó una expresión enciclopédica del conocimiento humano, que descompuso y encajó dentro de unas divisiones decimales. Graduaba así las disciplinas científicas a partir del ordenamiento propuesto por Bacon (1975) para los niveles más amplios, mientras que, en los niveles intermedios, siguió las propuestas científicas positivistas de dividir y clasificar hasta alcanzar los niveles más específicos. La clasificación de Dewey dotó de adaptabilidad a sus anotaciones, pero la rigidez estructural la hacía menos amigable para los usuarios que los encabezamientos que consiguieron proyectarse mejor hacia los lenguajes controlados.

Los avances tecnológicos de la revolución industrial permitieron fijar o emitir los mensajes a través de nuevos medios técnicos de información. El sonido se gravaba en discos, la imagen en fotos y en películas, la radio y luego la televisión comunicaban con usuarios distantes. La vía impresa quedaba superada como medio único para registrar y comunicar los mensajes y junto a ella el concepto original de la bibliografía en cuanto lista de impresos. Se buscó un término que abarcara toda la variedad de soportes y se consideró que documento reunía los nombres de los productos de la imprenta: libro, revista, actas, informes, ... a la vez que generalizaba la referencia a los nuevos objetos de conocimiento. La gestión y el procesamiento de estos documentos había que abordarlos desde la practicidad en su tratamiento, sin olvidar los requisitos de uso y aprovechamiento finales ni su gestión tecnológica. Por lo que, tanto la variedad de soportes atendidos como su procesamiento tecnológico para atender a usuarios científicos, técnicos, empresariales, institucionales, educativos, etc. quedaban recogidos bajo el nombre de Documentación.

La fundamentación disciplinar de la Documentación, cuando empezaba el siglo XX, afirmó los recursos semánticos como vía metodizada y concordada para organizar y controlar la complicada existencia de los documentos. La obra de los bibliógrafos La Fontaine y Otlet es un referente inevitable en la organización del conocimiento sobre la que volcaron sus convicciones positivistas e internacionalistas. Su principal realización fue el *Instituto Internacional de Bibliografía* para crear un *Repertorio Bibliográfico Universal* (RBU) índice del conocimiento universal donde se registraba cada documento existente y para

cuya organización había que contar con una Clasificación Decimal que facilitase la imprescindible cooperación internacional (Ayuso 1998). Desde la última década del siglo XIX adaptaron la Clasificación Decimal de Dewey.

De forma que la organización detallada del conocimiento se situaba en el inicio de su pensamiento. Sin poderse evitar que en su propuesta pesase mucho la búsqueda de un lenguaje de comprensión internacional, que evitase el imperialismo de las lenguas de mayor difusión y que se pudiese aplicar a todas las materias tratadas en los documentos. Pero este proyecto de enciclopedia global requería de unos lazos que asociasen todas las materias que allí se plasmaban. Era un momento en que la modernización de los estudios filológicos abordaba el estudio científico de la lengua, que repercutió en el planteamiento de vías de representación del contenido documental. Aunque fuese mediante el empleo de unos lenguajes codificados, cuya agrupación se conoce como Sistemas de Clasificación, que tenían como finalidad ordenar los conceptos de los numerosos objetos de información generados en el avance científico y técnico consecuente a la Revolución Industrial. Hijos de su tiempo, se diseñaron desde métodos racionales y positivistas que consiguieron representar el conocimiento a través de códigos pactados que mantenían un estrecho paralelismo con los lenguajes científicos. En las clasificaciones de uso documental actuaban principios lingüísticos pues, aunque las nomenclaturas se reemplazaban por códigos, había que efectuar análisis semántico y representar las materias mediante signos convenidos (Gomes 2009). De forma que las materias científico-técnicas y la equivalencia, hiponimia o diferencia entre ellas se expresaba en la CDU con cifras. Se consideraba así la capacidad del mensaje documental para sistematizar la información desde categorías lingüísticas, apareciendo los primeros intentos de alcanzar lenguajes para dialogar con los documentos.

Este planteamiento introdujo los lenguajes documentales en la gestión biblioteconómica. Las clasificaciones decimales fueron una opción válida para organizar la información y recuperarla con eficacia. Pese a que, igual que las nomenclaturas científicas, los sistemas de clasificación se volviesen limitativos, andando el tiempo, para el conjunto de operaciones deseadas por los usuarios.

4. REVOLUCIÓN TECNOLÓGICA Y LENGUAJES COMBINATORIOS

La enorme acumulación de documentos científico-técnicos durante la II Guerra Mundial superó ampliamente los métodos de transmisión y acceso a los resultados de las investigaciones. Vannevar Bush, director del *Office for Scientific Research and Development* durante la presidencia de Roosevelt, reflejó su

práctica en el artículo *As we may a think* que anticipó las lejanas condiciones de la *Sociedad de la Información*. Analizó los métodos e instrumentos que servían para controlar y organizar la información y propuso mejorarlos. Destacó cómo la estructura secuencial de los documentos, en reflejo de la secuencialidad del habla, era la causante de que las clasificaciones alfanuméricas fuesen incapaces de procesar adecuadamente grandes cantidades de información al relacionar las materias por hiponimia, cuando “la mente humana no trabaja de esa manera, sino que opera por medio de asociaciones” (Bush 1945).

Los procesos de selección de recursos y de recuperación de información son fundamentos y principios indefectibles a la hora de crear conocimiento. Para comunicarlo, los científicos se valen de una compleja operación asociativa con sus pares desde el lenguaje. Bush consideró que los sistemas de indización para recuperar los documentos eran muy artificiales, fuesen sistemas de clasificación o categóricos por materias. Pues seguían una jerarquía lógica de tanta rigidez que se situaban en una perspectiva antagónica al funcionamiento asociativo de la mente reflejado en los textos. El problema principal estaba en la manera incorrecta de guardar y organizar la información, por lo que había que dar con un sistema que lo resolviese. Para conseguirlo propuso el sistema *Memex* en el que anticipó hipotéticamente la recuperación mediante combinaciones lógicas de códigos conformes con las materias de los documentos. Su apertura a que actuasen las asociaciones de conceptos superaba con claridad la capacidad de los sistemas jerárquicos a la hora de que circulase la información. Con esta propuesta se anticipaban los índices asociativos en lenguajes de estructura combinatoria junto a la implantación de los sistemas de bases de datos, hipertextos e hipermedia (Robert 1984). De la teoría pasó a la práctica con una máquina homónima de información que imitaba el funcionamiento asociativo del cerebro humano. La máquina almacenaba mucha información que, además, combinaba en hipertexto mediante el empleo de un “índice por asociación (Bush 1991) que ofrecía a los usuarios la posibilidad de seguir trayectorias variadas en procesamiento hipertextual, que supuso una nueva forma de textualidad con nuevas posibilidades de lectura y de escritura. El *Memex* aportaba flexibilidad ante las dificultades que para la efectiva circulación de los contenidos suponían las categorías clasificatorias, que nunca llegaron a tratar con suficiencia el contenido de los documentos.

La creciente especialización del conocimiento y la necesidad de acceder con relevancia a los documentos que satisficiesen las necesidades de las búsquedas hizo que, mediado el siglo XX, se planteasen unos lenguajes docu-

mentales organizados hacia los conceptos. Se necesitaba guiar a los usuarios hacia los significados de las representaciones de información. Se manifestó con mucha claridad en el periodo previo a la Ciencia de la Información cuando la recuperación de la Información se situaba ya como su hecho fundamental (Shera, Cleveland 1977). Fue entonces cuando la automatización documental del análisis de contenido que permitían los computadores limitó el empleo de los sistemas de clasificación a la organización de las colecciones bibliotecarias. Para gestionar la información se integró en bases de datos en la que su recuperación se haría con el empleo de lenguajes naturales controlados por lo que hubo que sistematizar sus estructuras y estándares.

Si las bases de datos especializadas querían responder a las necesidades crecientes de los grupos de especialistas debían de atenderlas empleando índices coordinados (Taube 1951). Cada concepto de la base de datos se representaba en una tarjeta que Taube propuso controlar mediante un glosario de unitérminos (1953). Se daban los primeros pasos de la recuperación de la información por computador y la indización se alejaba de la costumbre perdurable durante casi dos mil años y se acercaba al empleo de descriptores cuyo primer tesoro operativo aparecía en coincidencia (Castello, Wall 1959). Otro de los pioneros en adentrarse a la Ciencia de la Información, Calvin Mooers se propuso localizar mediante los computadores la información pertinente en medio de la gran cantidad de bases de datos existentes (1963). Su aportación se entiende si consideramos que precisó el significado de los términos *recuperación de la información e indización por descriptores*, en torno a los que situó los procesos principales y los sistemas apropiados de Ciencia de la Información. Su diseño del *Zato-coding* empleaba unos descriptores normalizados que se convertirían en una de las características básicas de la recuperación y de la Ciencia de la información desde sus inicios y que realizó las indizaciones desde conceptos lingüísticos. Mooers fijó su actividad en el cruce de la Ciencia de la información con la informática, la gestión empresarial y la lingüística. Con esta se relaciona tanto desde la finalidad de tener ambas una perspectiva comunicativa, como desde su común aportación a la comprensión de los procesos de transferencia de información (Hayes 1969).

Hans Peter Luhn fue uno de los innovadores que aplicaron las capacidades de los computadores fuera del ámbito estadístico. Muchas fueron las empresas documentales en las que intervino como los sistemas de difusión selectiva de la información (SDI), las indizaciones KWIC y KWOC, los tesauros de descriptores, la autocodificación y el autoresumen, los perfiles de usuario, ... Todas ellas

contribuciones fundamentales que capacitaron a los computadores para realizar análisis de contenido textual (Cuadra 1964). El problema fundamental de las transformaciones del conocimiento a través de la indización y el resumen reside en explicar las reducciones homomórficas, de tal manera que el contenido del documento se pueda reestructurar desde su serie de descriptores y, al aplicarlos, se localicen informaciones relevantes para las necesidades de los usuarios.

Las aplicaciones lingüísticas a la Documentación se mostraron como el agente más conveniente y directo en la mediación informativa automatizada. Para hacerlo con propiedad, los computadores estaban obligados a distinguir las estructuras gramaticales de una lengua, solucionar las situaciones de sinonimia y polisemia, hacer inferencias desde el conocimiento de mundo y diferenciar la disposición y orden de los textos. Ya que el procesamiento automático de estos no se puede efectuar sin conocer sus estructuras y sin entender las múltiples operaciones que intervienen en su situación contextual, conformación y comprensión. Y en esta intermediación comunicativa y científica se encuentra el lugar de los lenguajes documentales, vocabularios controlados que se creaban empleando sólo una parte de los lenguajes naturales. Desde sus primeras pruebas los computadores hicieron aplicaciones lingüísticas que requerían normalizar sus estructuras. Los descriptores evitaban todas las ambigüedades del lenguaje natural y se establecieron como términos unívocos y preferentes para describir un concepto. La garantía de su normalización provenía de su presencia en el tesoro (Sánchez et al 2012). Los descriptores son términos de alta significación que representan a los conceptos de un dominio. En torno a ellos y a las relaciones semánticas establecidas con otros conceptos, se determina la elaboración de los lenguajes documentales.

Funcionalmente, el tesoro documental es un vocabulario destinado a indizar y recuperar información. La matriz de su vocabulario se basa en reducir el número de términos a emplear mediante la elección de los términos preferentes o descriptores. Mientras que, por el lado estructural, se manejan tres herramientas: un corpus de términos representativos del dominio de aplicación; una organización macroestructural que fija los esquemas y la segmentación del corpus con una macroestructura global dependiente del dominio de aplicación, unas macroestructuras secundarias cuyos macrodescriptores encabezan cada una de las subdivisiones del tesoro y el desarrollo en submacrodescriptores que clasifican las familias de términos; y unas estructuras de superficie, los propios descriptores y sus relaciones (Laipelt 2015). El procedimiento inductivo en los tesauros se efectúa mediante las relaciones semánticas con las que se representa el conoci-

miento y que establecen asociaciones entre los términos, lo que no evita que los usuarios deban de ser especialistas del dominio representado, pues la asignación relevante de significado es siempre mental.

La practicidad y utilidad de los tesauros les ha proporcionado una abundante aplicación, debido a que su empleo muestra evidentes ventajas, pues son de creación y gestión sencillas, su eficacia ha sido bien comprobada, muestran considerable coherencia y ofrecen un buen arranque para crear ontologías, por lo que no sorprende que su número sea tan abundante (Monreal, Gil 2010). Sin que se haya podido evitar la falta de pertinencia en la recuperación por la presencia de silencio o ruido. A lo que se vino a sumar el problema del número de documentos y la variedad de soportes que circulan por la Web. Entre los inconvenientes que presenta su uso se puede mencionar que (Morato, Sánchez-Cuadrado 2009), al estar pensados para dominios restringidos no han contado con mecanismos para compartir información en la Web hasta que surgió *Simple Knowledge Organisation System* (SKOS). Tampoco cuentan con recursos que relacionen su vocabulario con los objetos de información que representan. En las jerarquías incluyen instancias, atributos y meronimias por lo que las relaciones son heterogéneas. Sus relaciones están muy acotadas, por lo que es difícil añadir nuevas relaciones. No fueron pensados para estar automatizados y en red. Por lo que el coste de su elaboración, mantenimiento y ejecución es elevado.

5. VOCABULARIOS SEMÁNTICOS PARA AMBIENTES DIGITALES Y EN RED

Desde que Internet permitiese compartir información a través de la web, los Sistemas de organización del conocimiento (*Knowledge Organization Systems*-KOS) preexistentes han tenido que pasar, y en ello siguen, por un proceso de reformulación y reinterpretación. A la vez que han llegado nuevos sistemas. Todos ellos se acogen bajo la denominación de Vocabularios, que se califican de semánticos cuando sus elementos constituyentes son conceptos enlazados. Vocabularios y Vocabularios semánticos formados por los objetos que representan y organizan el conocimiento en Internet, además de sus elementos de enlace. Los lenguajes documentales precedentes están incluidos en ellos, lo que ha servido para prolongar algunas de sus exigencias de control terminológico, pero muchas otras las han sobrepasado ampliamente (Moreiro 2018).

De la Web partió la exigencia de ampliar las estructuras de organización del conocimiento, sobre todo con la actuación de las ontologías y las taxonomías. En el caso de la Web semántica, ontología es sinónimo de vocabulario,

mientras que en la Web social lo sería folksonomía. Por su parte, la representación de los objetos de información del mundo empresarial e institucional se hace en las taxonomías. Y en los estándares de metadatos, vocabulario es el conjunto de elementos descriptivos y sus propiedades. De modo que también la Web ha motivado a adaptarse a los vocabularios preexistentes, como en el caso de los tesauros a los que ha llevado hacia una mayor conceptualización y a representar su visualización como redes semánticas, o al lenguaje libre de las palabras-clave que ahora se aprovecha para conformar las folksonomías. Este panorama múltiple nos avisa de que la representación de los objetos del conocimiento ha rebasado la usual atención a la comunicación de información científico-técnica y se extiende hacia otros aspectos de la vida administrativa, comercial, de ocio y laboral, desde luego siempre en conexión por la red.

Antes, los sistemas de indización no se habían preocupado de representar los vínculos semánticos que asocian un concepto con otro. Hasta que lo hicieron las redes semánticas, lo que ha tenido mucha repercusión en el paso de los términos a los conceptos, en definitiva, a alcanzar una representación que pueda llamarse semántica. Las redes no solo muestran la existencia de una relación, si no que facilitan su visualización al identificarla, la sitúan en su contexto y explican la trayectoria de su distribución. Asociar conceptos fue una meta desde que Bush (1945) lo señalase como el trayecto adecuado para representar el conocimiento. Las relaciones que los tesauros atendían con tanta pasividad se ven ahora reavivadas desde las redes semánticas, campo de cruce entre la representación, recuperación y organización del conocimiento con los lejanos mapas de conceptos, herramienta didáctica en su origen y con los sociogramas que muestran cómo se establecen los vínculos entre los integrantes de grupos y organizaciones. Así como, desde luego, con las redes matemáticas de elementos que convergen en un espacio de definiciones formales y las informáticas basadas en la teoría de grafos, pues han proporcionado herramientas para visibilizar y representar grandes volúmenes de datos. Como consecuencia los tesauros se dinamizaron con un número mayor de relaciones, además identificadas, para mejorar la representación y aumentar la precisión y la eficacia

El tesoro en papel, modelo arquetípico de los lenguajes documentales fijaba sus elementos y relaciones de acuerdo con lo establecido en la norma ISO 2788 (1986). Ahora se ha convertido en una herramienta interoperable que se relaciona con otros vocabularios y que permite el intercambio de formatos y la recuperación de información de cualquier recurso. Precisamente se le llama vocabulario en la norma ISO 25964-1 (2011). A lo que se añade la facilidad

de visualización ofrecida por las pantallas que ha propiciado la representación de los vocabularios mediante redes semánticas de forma utilizando grafos de conexión que identifican y nombran las relaciones semánticas que se dan entre los términos: equivalencia, jerarquía o asociación. Este hecho fue determinante para que la asociación de conceptos se aproximase mucho a la que se establece entre los términos del lenguaje natural. El desarrollo de los tesauros como redes de conceptos ha tenido consecuencias inmediatas sobre el léxico empleado para indizar y, más aún, para sus relaciones, pues la identificación verbal de asociaciones funcionales ha sido una de las novedades más alentadoras.

La potente comunicación realizada con los documentos multimedia fue otra causa que obligó a los vocabularios a un cambio rápido, sobre todo a la hora de indizar imágenes y documentación software. Los descriptores expresados exclusivamente mediante sustantivos quedaban despegados de cualquier posibilidad de describir las acciones consustanciales a sus contenidos. Para hacerlo había que emplear verbos, pues expresan acciones, frente a los sustantivos semánticamente más estáticos y sin capacidad de representar las condiciones de admisibilidad de los hechos y las tareas en movimiento. De modo que se pasó a designar mediante verbos los hechos en ejecución, los procesos o las relaciones entre entidades, que hasta ahora solo se habían representado mediante sustantivos estáticos en los vocabularios combinatorios.

La facilidad de almacenamiento y difusión de los recursos digitales ha impulsado a los sistemas de información a ofrecer nuevos servicios. El continuo aumento en la facilidad de acceso a la información les obliga, cada vez más, a una recuperación pertinente y una representación de los contenidos relevante. Para lo que tienen que disponer de métodos de interrogación capaces de realizar:

- Búsquedas en colecciones internas y externas.
- Búsquedas cruzadas (interoperabilidad).
- Búsquedas por navegación (visualización e interfaces basados en KOS).
- Indización y clasificación automáticas.
- Extracción de términos.
- Apoyo a la recuperación en diferentes lenguas mediante “vocabularios” interconectados.

Otro gran reto para la organización del conocimiento se deriva de la comunicación científica propugnada por la Ciencia abierta (*Open Science*) que, por su carácter colaborativo y sin barreras en el acceso a la investigación,

implica cambios notables en la manera de concebir tanto la investigación como la difusión de sus resultados. En sus planteamientos subyacen las características de rapidez de comunicación, apertura y reproducibilidad consustanciales a los objetos de información digitales sobre los que se difunde la ciencia actual (Rodrigues 2004). Procurando así facilitar el acceso a los resultados de una investigación que suele haberse financiado con fondos públicos. Y que va destinado a todos los integrantes de la comunidad científica, sin olvidar al resto de la sociedad que los apoya.

Olvidándonos del gasto que para las universidades supone acceder mediante pago a unas publicaciones que, en la mayoría de los casos, se han generado en su seno, se procura un desbloqueo no solo de las publicaciones (acceso abierto), sino también de los datos, las metodologías y los procesos seguidos en las investigaciones. Lo que requiere de unas TIC bien dotadas para manejar los numerosos datos que de ellos se derivan. Aparecen, pues, muchas tareas para los sistemas de información y Documentación, responsables de preservar los extensos repositorios de datos generados en la actividad investigadora, pero asimismo de buscarlos, recuperarlos y permitir su reutilización desde la interoperabilidad entre infraestructuras. Por lo que entra en juego aquí, con toda razón, la utilización de los vocabularios semánticos.

Sin duda, como manifiesta la propia denominación, el fin perseguido por los Vocabularios semánticos es alcanzar la búsqueda y recuperación por conceptos. Al efecto cuentan con vocabularios de base lógica como las ontologías y los tesauros conceptuales (Sales; Café 2008), junto a vocabularios en los que permanece la base léxica o terminológica como son la red semántica *WordNet* y los vocabularios incluidos en SKOS.

Uno de los rasgos de la Web social consiste en aportar nuevos servicios compartidos gracias a la participación colectiva. Este tipo de organización ha servido para mejorar las búsquedas refinadas como la indización por reputación, la estima de la opinión de los participantes o la agregación por asignación de etiquetas que caracteriza a la indización social. Procesos en los que, para la organizar el conocimiento, se respetan las ideas y opiniones de los usuarios. El vocabulario empleado es el lenguaje libre, al representar la información con etiquetas de palabras-clave que constituyen las folksonomías. Las folksonomías han mejorado la indización por materias tradicional, desde un planteamiento que es, a la vez, gratuito y especializado (Mai 2011). Aunque sus logros se supeditan al sector en el que se manejan, pues reflejan la estabilidad terminológica dependiendo de la información manejada, que puede ser general, cultural o incluso

científica. Se ha percibido que las folksonomías incrementan la pertinencia en la indización cuando quienes indizan son competentes en un campo cognitivo o técnico (Moreiro; Bolaños 2018). Se percibe por esa vía un avance en la hibridación de las folksonomías al conjuntarse las ventajas del lenguaje libre con las del controlado, de forma que se aproximan a las relaciones semánticas de los vocabularios controlados. Este hecho supone que las etiquetas se van sometiendo a una formalización no forzada que se deriva del establecimiento de relaciones entre los conceptos por intervención de programas de co-wording (Rose 2016). Al agruparse automáticamente las etiquetas se pueden formar jerarquías e identificar su dependencia del contexto. Se acercan así al conocimiento semántico-léxico que requieren las relaciones jerárquicas y asociativas. Incluso, desde las folksonomías se han ido abriendo líneas de negocio, al aprovechar la organización que, para almacenar y difundir los objetos de información, se realiza en la actividad corporativa e institucional.

Por su parte, la Web semántica, con sus esquemas y patrones, constituye la base para organizar, representar y recuperar el conocimiento digital desde la contribución de varias tecnologías que agregan semántica comprensible para las máquinas respecto al contenido de los documentos. Estas herramientas inteligentes de representación del conocimiento buscan generar significados concretos desde discursos y documentos extensos. Tienen la finalidad de proporcionar un acceso inteligente a la información tan heterogénea que está distribuida por toda la Web, posibilitando que los agentes de software intermedien entre las necesidades de los usuarios y los objetos de información. Sin olvidarse de crear una Web de datos con significado (conocimiento), de forma que el software pueda aprender respecto a lo que quieren decir los datos, como acerca de la información necesaria para procesarlos. Esta Web emplea vocabularios normalizados con sintaxis uniforme y semántica no ambigua, a partir de lenguajes para la expresión de ontologías, KOS y vocabularios de metadatos (Marcondes, Campos 2008).

La trascendencia que tienen las ontologías para la ciencia se encuentra en las posibilidades que ofrecen de guardar y organizar el conocimiento para que pueda entenderse y utilizarse por los computadores, de forma que se realice su gestión desde la inteligencia artificial. Las ontologías aparecieron como aplicaciones de esta para almacenar el conocimiento y actuar sobre su información y datos buscando imitar los procesos mentales humanos para aprovechar sus significados (Sánchez Cuadrado et al. 2007). De suerte que en el concepto de ontología cabe cualquier sistema de organización del conocimiento que maneje

conceptos para describir y representar un dominio cognitivo. Por lo que su amplio espectro acoge desde la noción más simple en sus relaciones, las taxonomías, como los tesauros y modelos de conceptos, hasta llegar a las teorías lógicas que representan la noción más compleja.

Las ontologías son lenguajes multicapa que reutilizan recursos a partir de nuevos modelos de procesamiento desde los que se establecen relaciones de confianza y pruebas de autenticidad y que garantizan a los usuarios la disponibilidad de los recursos (Sánchez Cuadrado et al. 2007). Para organizar el conocimiento en la web a partir de los sistemas terminológicos preexistentes, NKOS (*Network Knowledge Organization Systems/Services*) logró determinar el significado de las unidades lingüísticas y sus relaciones, al aplicar unos esquemas de codificación de vocabularios mediante metadatos orientados al contenido y utilizables por personas y por computadores (RSS). De modo que los primeros recursos empleados para alcanzar las condiciones de procesamiento ontológico son los vocabularios de metadatos (SKOS: *Simple Knowledge Organization System*). Además, se apoya en una sintaxis común, XML; en la expresión del conocimiento en estructuras simples predefinidas (p.e. RDF); y maneja URIs para identificar los recursos publicados en la Web, pues se aprovecha el “http” de la URI para que estos recursos se puedan localizar y consultar. En virtud de ello, la Web semántica se sirve de unos vocabularios basados en ontologías para representar el conocimiento, desde los que se puede localizar y acceder a los recursos.

6. CONSIDERACIONES FINALES

Cada momento comunicativo va unido a unos determinados rasgos sociales y tecnológicos y procura solucionar los problemas que plantean el crecimiento, el almacenamiento, la difusión y el acceso a la información. Para hacerlo se vale de una filosofía de actuación heredada en la comunicación de la información y de unos procesos precedentes cuya función y respuestas se busca mejorar para atender a la evolución del conocimiento.

En la actualidad, el conocimiento muestra una fuerte propensión a continuar organizándose en taxonomías, como prueba el hecho de que están presentes en los esquemas de clasificación, los tesauros y hasta en las ontologías, los vocabularios semánticos, que son conceptuales y ontológicos, están muy lejos de limitarse a cualquier listado organizado por jerarquía y clase, aunque tampoco lo rehúyan. Su preocupación principal está en dotar de significado a los objetos de información para que puedan ser comprendidos por los computadores. El camino pasa por formalizar los términos con toda exactitud, facilitar los enlaces

a vocabularios existentes y entender las relaciones que pueden darse entre sus elementos. Esto supone unos vocabularios más interoperables que especifiquen cómo cada concepto se relaciona con otros conceptos de su contexto.

Parece una tautología, pero no existirá Web semántica sin contar con una recuperación que sea semántica, que permita efectuar la búsqueda por conceptos, por significados, por ideas. De modo que el software está obligado a diferenciar una idea de un término. La percepción de los conceptos se hace desde la lógica formal, pero también desde el propio léxico cuya posibilidad de tratamiento ha aumentado notablemente en los últimos estándares.

A la hora de representar y organizar el conocimiento es cada vez más común generar ontologías que relacionen los conceptos individuales pertenecientes a una clase, de forma que esta les integra en alguna base de conocimiento. Cada concepto se describe explícita y formalmente dentro de un marco de conocimiento participado, así como las propiedades que especifican sus características y atributos. A lo que se añade que la reutilización se realiza sin pérdida alguna de significado, de suerte que el conocimiento organizado originalmente para un entorno puede aplicarse en otro distinto. Los sistemas tienen que poder intercambiar automáticamente su información, pues todas las propiedades tienen un significado bien establecido (interoperabilidad semántica) con una sintaxis XML para los objetos de información más habituales (interoperabilidad sintáctica). De manera que cualquier aplicación podrá usar un concepto relacionado con los que maneja, aunque no estuviese incluido en su ontología.

Igualmente, hay que destacar la notoria participación de los vocabularios semánticos en la gestión estratégica de los objetos de información dentro de empresas y otras organizaciones. De modo que han rebasado ampliamente la anterior atención bastante limitada al ámbito científico y técnico. Han pasado a actuar como estructuras que organizan y recuperan la información y los datos, por lo que se han establecido como un factor diferenciador en los negocios, en las administraciones y en otros ambientes institucionales. Principalmente, administran y ordenan los recursos de información digitales que, como sistemas complejos, alojan en sus servidores Web, buscando categorizarlos y navegar por ellos. A lo que se suma que conforman una red semántica de conceptos interrelacionados que no solo cubre las necesidades empresariales, pues alcanza incluso al modo como los trabajadores se relacionan con la información. Incluso conectan con la experiencia y características del personal propio de cada institución.

REFERENCIAS

ARISTÓTELES. (2001). **Tratados de Lógica: el Organon**. México: Porrúa.

AYUSO GARCÍA, M^a D.; MOREIRO GONZÁLEZ, J. A. (col.). (1998). **Conceptos fundamentales de la teoría de la Documentación y Estudio terminológico del Tratado de Documentación de Paul Otlet**. Murcia: Diego Marín Librero.

BACON, F. (1975). **Instauratio Magna. Novum Organum**. Nueva Atlántida. México: Porrúa.

BEARE, G. (2007). *Past, present and future*. **The Indexer**, v. 25, n° 4, pp. 257-264.

BURY, R. de. (1972). **El Philobiblion**. Barcelona: Zeus.

BUSH, V. (1945). As we may think, **Atlantic Monthly**, n° 176, pp. 101-108.

BUSH, V. (1991). Memex revisited. En, **From Memex to hypertext**. Academic Press Professional, pp. 197-216.

CASSON, L. (2001). **Libraries in the Ancient World**. New Haven: Yale University Press.

CASTELLO, J. S.; WALL, E. (1959). Recents improvements in techniques for storing and retrieving information, en Taube, M. et al. **Emerging solutions for mechanizing the Storage and retrieval of information**. Washington: Documentation Inc., pp. 141-169.

CUADRA, C. A. (1964). Identifying key contributions to information science. **American Documentation**, v. 15, n° 4, pp. 289-295.

CUTTER, Ch. (1876). *Rules for a dictionary catalog*. Washington: Government Printing Office.

DEWEY, M. (1876). **A classification and subject index for cataloguing and arranging the books and pamphlets of a Library**. Amherst, Mass.

KRIEGER, M. G.; BOCORNY, M. J. (2004). *Introdução à terminologia: teoria e pratica*. São Paulo: Editora Contexto.

GOMES, H. E. (2009). Tendências da pesquisa em organização do conhecimento. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 2, nº 1, pp. 60-88.

HAYES, R. M. (1969). Information Science in Librarianship. **Libri**, v. 19, pp. 216-236.

ISO 25964-1. (2011). **Information and documentation: Thesauri and interoperability with other vocabularies - Part 1: Thesauri for information retrieval**. Genève: ISO.

ISO 2788: (1986). **Information and Documentation – Guidelines for the establishment and development of monolingual thesauri: international standard ISO 2788**. Genève: ISO.

ISIDORO DE SEVILLA. (1951). De librariis et eorum instrumentis. **En sus Etimologías**. Madrid: Biblioteca de autores cristianos. Libro VI.

IZQUIERDO ARROYO, J. M.; MORENO FERNÁNDEZ, L. M. (1994). Listas de encabezamiento de materia y Thesauri en perspectiva comparada. **Documentación de las Ciencias de la Información**, v. 17, pp. 287-309

LAIPELT, R.C.F. (2015). **Metodologia para seleção de termos equivalentes e descritores de tesouros**. Tese doutoral. São Leopoldo: Universidade do Vale do Rio dos Sinos.

MAI, J. E. (2011). Folksonomies and the new order: authority in the digital disorder. **Knowledge Organization**, v. 38, nº 2, pp. 114-122.

MARCONDES, C. H.; CAMPOS, M. L. A. de. (2008). Ontologia e Web Semântica: o espaço da pesquisa em ciência da informação. **Ponto de Acesso**, v. 2, nº 1, pp. 107-136.

MILLARES CARLO, A. (1971). **Introducción a la Historia del libro y de las Bibliotecas**. México: FCE.

MILLARES CARLO, A. (1973). **Técnica de la investigación bibliográfica**. Caracas: Universidad Católica Andrés Bello, Institutos Humanísticos de Investigación.

MONREAL SOLER, C.; GIL LEIVA, I. (2010). Posibilidades y límites de los tesauros frente a otros sistemas de organización del conocimiento: folksonomías, taxonomías y ontologías. **Revista Interamericana de Bibliotecología**, v. 33, nº 2, pp. 361-377.

MOOERS, C. (1963). The indexing language of an Information Retrieval System, en Simonton, W. **Information retrieval today**. Minneapolis: University of Minnesota, pp. 21-36.

MORATO, J., SÁNCHEZ-CUADRADO. (2008). **Sistemas de Conocimiento para la Recuperación de Información**. Universidad Carlos III de Madrid. <http://ocw.uc3m.es/ingenieria-informatica/sistemas-avanzados-de-recuperacion-de-informacion/material-de-clase-1/MC-F-003.pdf>

MOREIRO, J. A. (2018). Adaptación de los vocabularios documentales al ambiente digital en red: léxico, significado y relaciones semánticas. **Informação & Sociedade: Estudos**, v. 28, nº 1, pp. 35-46.

MOREIRO-GONZÁLEZ, J. A.; BOLAÑOS-MEJÍAS, C. (2018). Folksonomy Indexing from the Assignment of Free Tags to Setup Subject: A Search Analysis into the Domain of Legal History. **KO Knowledge Organization**, v. 45, nº 7, pp. 574-585.

MUELLER, S. P. M. (1995). O crescimento da ciência, o comportamento científico e a comunicação científica: algumas reflexões. **Revista da Escola de Biblioteconomia da UFMG**, v. 24, nº2, pp.63-84.

ROBERT, N. (1984). The pre-history of the information retrieval thesaurus, **Journal of Documentation**, v. 40, nº 4, pp. 271-285.

RODRIGUES, E. (2004). Acesso livre ao conhecimento: a mudança do sistema de comunicação da ciência e os profissionais de informação. **Cadernos BAD**, nº 1, pp. 24-35.

ROGET, P. M. (1883). **Thesaurus of english words and phrases**. London: Avenel Books.

ROSE, G. (2016). **Visual methodologies: An introduction to researching with visual materials**. London: Sage.

SALES, R. de; CAFÉ, L. (2008). Semelhanças e diferenças entre tesauros e ontologias. **DataGramZero - Revista de Ciência da Informação**, v. 9, nº 4, artigo 02.

SÁNCHEZ-CUADRADO, S., MORATO, J., PALACIOS, V., LLORENS, J., MOREIRO, J. A. (2007). De repente, ¿todos hablamos de ontologías? **El profesional de la información**, v. 16, nº 6, pp. 562-568.

SÁNCHEZ-CUADRADO, S.; COLMENERO, M. J.; MOREIRO, J. A. (2012). Tesauros: estándares y recomendaciones. **El profesional de la información**, v. 21, nº 3, pp. 229-235.

SHERA, J.H.; CLEVELAND, D.B. (1977). History and foundations of Information Science, **Annual Review of Information Science and Technology**, nº 12. pp. 249-275.

TAUBE, M. (1951). Functional approach to bibliographic organization, en SHERA, J. H., EGAN, M. (ed.). **Bibliographic Organization**. Chicago: Chicago Univ. Press, pp. 53-73.

TAUBE, M. et. al. (1953-1959). **Studies in coordinate indexing**. 5 v. Washington: Documentation Inc.

TERRADAS, M. L. (1983). **La documentación médica como disciplina**. Valencia: Centro de documentación e informática biomédica.

UNESCO. (1977). **Guia para a redação de artigos científicos destinados à publicação**. Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia.