

ADAPTACIÓN DE LOS VOCABULARIOS DOCUMENTALES AL AMBIENTE DIGITAL EN RED: léxico, significado y relaciones semánticas

José Antonio Moreira*

RESUMEN Se exploran los conceptos en torno a la adaptación de los vocabularios a las exigencias del entorno Web y de la interoperabilidad. El objetivo es identificar las nuevas vías de consistencia terminológica y de ampliación de asociaciones entre los conceptos. Se aborda desde la perspectiva del léxico utilizado y de acuerdo con la creciente carga semántica de las relaciones entre sus elementos. Los vocabularios se diferencian por su conformación con palabras del lenguaje natural, con términos de una especialidad o desde intercambios conceptuales. Se contraponen las propiedades de los lenguajes documentales de origen positivista a las características de los vocabularios enlazados. Se valoran las redes semánticas por su aportación a la visualización en pantallas de los vínculos entre conceptos de tesauros, taxonomías y ontologías, así como por superar el uso privativo de los sustantivos y poder expresar con verbos, en tipología ilimitada, las relaciones asociativas. Finalmente, se justifica el paralelismo entre las declaraciones en lenguaje natural y las hechas en RDF para representar la información en modelos basados en el conocimiento.

Palabras clave: Sistemas de organización del conocimiento. Vocabularios. Palabras. Términos. Conceptos complejos.

* Doutorado pela Universidad Nacional de Educación à Distancia, Espanha. Pós-doutorado pela Universidade de São Paulo, Brasil. Diplomado em professorado de E.G.B. em 20/06/1974. Licenciado em Filosofia y Letras pela Universidad Nacional de Educación à Distancia, Espanha. Professor Catedrático de Biblioteconomia e Documentação da Universidad Carlos III de Madrid, Espanha. E-mail: jamore@bib.uc3m.es.

I PLANTEAMIENTO

Algunos de los actuales Sistemas de organización del conocimiento (SOC) son anteriores a la aparición del ciberespacio, por lo que han tenido que pasar, y aún continúan, por un proceso de reformulación y reinterpretación. En paralelo, conviven con otros nuevos ideados para responder a la nueva situación.

“Vocabularios” denomina hoy a los elementos y los enlaces que representan y organizan la información en Internet. Han desplazado al precursor “Lenguajes documentales” a los que, sin duda, incluyen y muchos de cuyos requisitos de control terminológico mantienen, pero que superan

con amplitud. Desde la perspectiva de su uso, los lenguajes documentales son un sistema de signos que producen y facilitan el intercambio de información. Permiten realizar, pues, actos de habla (COSERIU, 1986, p.17). Pero los más extendidos se componían sólo de palabras, y no de cualquier tipo. De modo que, en comparación con los vocabularios, son menos completos, pues se limitaban a listar expresiones que se combinan, en el mejor de los casos, mediante reglas sintácticas. Pero cuyos enlaces son sugeridos por el conocimiento y la experiencia de los intervinientes, sin hacerlos explícitos con oraciones ni esquemas. Si pensamos en SOC combinatorios, los vocabularios ofrecen mayor amplitud morfológica y sintáctica:

- Sus elementos componentes no son solo sustantivos.
- Ofrecen un número superior de relaciones entre esos elementos.
- En las relaciones, revelan los enlaces que existen entre los conceptos, de modo que los enlaces son también conceptos.

“Vocabularios” manifiesta también el necesario entendimiento con usuarios y ordenadores. Así, al ajustarse a los requisitos de aquellos, pueden simplificar sus elementos y estructuras, como en las folksonomías. Mientras que cuando se ajusta a los del software se complica la lógica de formalización, como en las ontologías. Si las primeras normas de los tesauros, como arquetipo de los SOC, se aplicaban a lenguajes difundidos en papel (ISO 2788: 1986), los tesauros actuales son herramientas electrónicas interoperables, que se relacionan con otros vocabularios y que permiten el intercambio de formatos y protocolos para recuperar información de todo tipo de recursos de información. Incluso hoy, la identidad de los tesauros se determina desde el nombre de vocabularios por la propia norma ISO 25964-1 (2011).

El primer objetivo consiste en atender, dentro del entorno Web y de la interoperabilidad, a los cambios que han sufrido los conceptos y aplicaciones de los SOC preexistentes en su adaptación a las nuevas necesidades. Desde los tesauros, que se han vuelto más conceptuales y se visualizan como redes semánticas, al lenguaje libre renovado con las folksonomías. A la vez que analizar las aportaciones de los vocabularios más recientes, ontologías y taxonomías, para comprobar las nuevas vías de estabilidad terminológica y de ampliación de las asociaciones por significado.

2 LOS VOCABULARIOS EN ORGANIZACIÓN DEL CONOCIMIENTO

Saussure fijó en su *Teoría del signo* el axioma que vincula los términos con los conceptos en la comunicación verbal (1945). De forma que los signos surgen de la asociación dual de un significante con un mensaje. Este requisito llega también a los vocabularios de recuperación, cuya naturaleza se encarrila por

la cercanía a uno de los topes, bien sea el de la expresión o el del contenido, siempre graduados por la profundidad y exactitud a conseguir. El planteamiento de Saussure puede extenderse a un triple enunciado que acoja todas las condiciones léxicas de los vocabularios:

Tabla 1: Origen de la representación de los signos en los diferentes vocabularios

	Significante	Palabra
SIGNO		Término
	Significado	Concepto

Fuente: Elaboración propia a partir del axioma del signo de Saussure (1945).

Así pues, el léxico tiene como fin semántico alcanzar los conceptos. Este propósito fija una primera diferenciación entre los vocabularios según se conformen:

- Con palabras sin normalizar del lenguaje natural, tanto extraídas como asignadas.
- Con términos propios de una especialidad del conocimiento.
- Desde intercambios semánticos a partir de los contenidos de los documentos.

Tabla 2: Relación entre los componentes y el tipo de vocabulario

Componentes de representación	Tipo de vocabulario
Palabras	Vocabularios libres
Términos	Vocabularios controlados
Conceptos	Redes semánticas y Ontologías

Fuente: Elaboración propia.

Todos estos vocabularios son SOC y sirven para indizar y recuperar recursos en la Web.

Tabla 3: Elementos componentes y relaciones semánticas en los SOC

Tipos de SOC	Listas	Taxonomías	Tesauros	Redes semánticas	Ontologías
Elementos	Palabras clave → Folksonomías Términos → Diccionarios, glosarios, léxicos... Conceptos → Anillos de sinónimos	Términos	Términos	Conceptos	Conceptos
Relaciones	Equivalencia en ciertos casos.	Equivalencia. Jerarquía.	Equivalencia. Jerarquía. Asociación.	Asociación verbal.	Equivalencia. Jerarquía formalizada. Asociación formalizada. Reglas de inferencia.

Fuente: Elaboración propia a partir de la tipología de vocabularios de la norma Z 39.19-2005.

2.1 Vocabularios de palabras

Alargando esta correlación, la Web social se apoya en las palabras para indizar y recuperar, pues desde la expresión libre no se obliga a depender de vocabularios que intermedien. Se busca en texto completo o en etiquetas indizadas de forma colaborativa y se recuperan los documentos que incluyan o tengan asignadas las mismas expresiones de búsqueda utilizadas. A la facilidad de empleo de las palabras se opone una recuperación disgregada e incoherente y, en ocasiones, incongruente, pues no se evitan los problemas causados por la sinonimia y la polisemia, ni la combinación libre de palabras en sintagmas no deseados. De forma que cualquier búsqueda se dilata al obtenerse resultados inciertos y, casi siempre, incompletos.

Hacer las búsquedas en corrientes de signos, no por conceptos, ha sido un obstáculo para explorar las capacidades de la Web. Las palabras-clave han reaparecido con fuerza al ser empleadas en las etiquetas de las folksonomías. Este uso no es muy diferente del que tuvieron al inicio de la indización automática. Entonces, se querían identificar las palabras más representativas de los textos para extraerlas y

elaborar los índices permutados (LUHN, 1957). Su logro no evitaba una extracción de los datos limitada al título, ni que se seleccionasen los sinónimos, o que se aceptasen como entradas distintas los morfemas flexivos de la misma palabra e incluso que se separasen los términos compuestos. Algunos de estos problemas aún suceden ahora cuando hacemos búsquedas libres en la web, aunque la evolución de las folksonomías hacia procesos semánticos está siendo rápida (SEGUNDO, 2017).

2.2 Vocabularios terminológicos

La concreción terminológica fue un gran afán de los científicos entre los siglos XVI y XX. Querían acordar metalenguajes emancipados de las lenguas naturales que fuesen instrumentos simbólicos y conceptuales al mismo nivel que el resto de coordenadas y unidades de medida universales. Consiguieron dentro de cada dominio unos lenguajes muy formalizados, lógicamente coherentes y de intercambio desambiguado (GUTIÉRREZ, 1998). Estos rasgos pasaron a los lenguajes controlados que se componían con términos unívocos normalizados sobre los que después actuaban criterios

clasificatorios, jerárquicos y asociativos para conformar su estructura.

Con idea de evitar las imprecisiones del lenguaje libre y de poder intercambiar información entre las bases de datos y los usuarios, la primera oleada informática respaldó la creación de vocabularios terminológicos (FELBER; PITCH, 1984). Esta tendencia se consolidó con la norma ISO 2788 (1986) que se ocupó de representar conceptos mediante términos normalizados y de conectarlos por relaciones semánticas. Desde luego, la función de las representaciones es comunicar conceptos, pero su manifestación, por dominios de referencia, se hacía mediante grupos de términos controlados, los descriptores, de significado estable por la univocidad nominal. En el tesoro cada descriptor se dispone dentro de la clase correspondiente, pues su organización no puede

entenderse lejos del orden taxonómico propio de los esquemas jerárquicos de origen positivista. Los tesauros se convirtieron en el arquetipo de los vocabularios terminológicos por su función reguladora y controladora para garantizar indicaciones y recuperaciones pertinentes. Otra de las ventajas viene del uso combinado de descriptores en las búsquedas, o de contar con relaciones de asociación para navegar por sus listas de descriptores. Pese a lo cual, hasta la llegada de los SKOS (2009), presentaban muchos problemas para mostrar con precisión las relaciones entre los términos y para compartir información en la Web.

Estas propiedades de los lenguajes documentales se correspondían aun con un marco de representación de la información de carácter positivista.

Tabla 4: Características de los Lenguajes documentales de base positivista

Lenguajes normalizados que buscaban una comunicación científica inequívoca
Perseguían la esencia universal en la constitución de los objetos y fenómenos
El léxico se veía como una nomenclatura
El sustantivo era la forma de representación indispensable. El resto de categorías gramaticales se sustantivaban
Contaban con unas formas apriorísticas de representación basadas en categorías generales
En su estructuración predominaban las relaciones jerárquicas
Se planteaban eliminar las imperfecciones de los discursos de las humanidades a las que se quería llevar hacia la exactitud de las ciencias más formalizadas

Fuente: Elaboración propia.

Estas bases constituyeron los tesauros tradicionales, con términos ligados a los conceptos que representan. Por más que su empleo fuese demasiado rígido en la descripción, poco exacto al fijar las relaciones y carente de adaptación respecto al potencial de relación entre los términos. En respuesta a los nuevos requisitos, estas mismas limitaciones forzaron a una renovación de la capacidad de representación. El progreso fue tan notorio que pronto se establecieron otras normas de

indudable transcendencia para los nuevos vocabularios.

2.3 Vocabularios conceptuales complejos

La interactuación de los vocabularios en internet valoró el léxico a partir de su función comunicativa (HARVEY, 1994, p.37), desestimando los intentos de fijar los términos como referentes ideales y homogéneos.

Tabla 5: Características de los vocabularios semánticos

El léxico está considerado a partir de su función comunicativa
El sujeto de la interpretación es el usuario, no el documentalista o el científico
Se emplea un léxico más cercano al lenguaje natural
El valor de los términos está relativizado por el contexto o situación del discurso
Se utiliza toda la riqueza expresiva del lenguaje, rompiendo la limitación impuesta por el empleo exclusivo de sustantivos
Cada documento contiene un modelo léxico
Predominan las relaciones asociativas. Se ha superado la categorización única por jerarquía

Fuente: Elaboración propia.

La norma ISO 25964-1 (2011) supuso un gran avance respecto a la ISO 2788 de construcción de tesauros, al facilitar una mejor comprensión de los conceptos, evidenciar las relaciones interconceptuales y orientar la configuración de los SOC. La actual función comunicativa del léxico ha llevado a admitir el empleo de la sinonimia y la polisemia por razones de funcionalidad, por la flexibilidad de modelos cognitivos en constante modificación y por la diversidad de los procesos de categorización. Situándonos así dentro de la indización inteligente propia de las bases de conocimientos (ROLSTON, 1990). De modo que se limitaba el silencio al asociar términos sinónimos, al valerse de la relación de las abreviaturas con su expresión *in extenso*, al remitir desde los términos canónicos a sus formas flexivas o al enviar desde los términos genéricos a sus específicos. También se evitaba el ruido, al sortear las ambigüedades causadas por la homografía y la polisemia. La navegación se expandía, dentro de un corpus de documentos, mediante los enlaces de un término de búsqueda con sus genéricos o con sus específicos, e incluso por categorías semánticas cercanas. Se superaba así la obligación de univocidad de un término para un concepto y se facilitaba buscarlo por todas las formas posibles, aunque se usase aun la expresión.

Las normas han evolucionado para recoger este nuevo estado de cosas (Z 39.19-2005; BSI Group 2005-2007; ISO 25964-1-2011). Sin perder

el carácter terminológico de los vocabularios controlados, los han rebasado con amplitud. Ciertamente, los nuevos vocabularios conocen una progresión semántica, incluso los que se limitan a trabajar en la vía expresiva. En este sentido, es indudable la contribución de los anillos de sinónimos a la hora de evitar los equívocos en las búsquedas. La presencia de los anillos de sinónimos es un intento de alcanzar los conceptos por las palabras, sin control de los términos. Muestran la ventaja de acomodarse a los principios de la web social, pues al expandirse la búsqueda desde las palabras empleadas hacia sus equivalentes sinónimos, se delimita la significación del concepto buscado mediante el empleo de todas sus posibles denominaciones (MOREIRO, 2013). De modo que se explora un concepto, por más que el acceso a su significado se haga mediante el empleo de los significantes equivalentes. Al unirse todos los significantes de un mismo significado la recuperación se resuelve por la intersección de cada una de las palabras del anillo con todas las demás. Por lo que ninguna de ellas resulta preferente sino concurrente.

3 LOS VOCABULARIOS COMO REDES SEMÁNTICAS

Si se quiere conseguir una Web semántica es forzoso operar con conceptos. En su logro ha sido decisiva la consolidación

de las redes semánticas junto a la utilización de esquemas lógicos. Las redes semánticas llegaron de la mano de los inicialmente conocidos como Tesoros conceptuales. El paso desde los tesauros ISO 2788 se dio con el fin de adaptar los índices para su visualización generalizada en las pantallas. La propuesta no consistió tanto en manejar conceptos como en navegar por ellos y visualizarlos mediante grafos de conexión explícitos que, por ello, identifican la relación existente entre términos (ROE; THOMAS, 2004). Al entrecruzar los nodos de conceptos se mapeaba el contexto de significación siguiendo las características relaciones de equivalencia, jerarquía o asociación. Esas relaciones semánticas se significan mediante grafos integrados por los conceptos o nodos y por los arcos que interconectan los conceptos. Era un paso definitivo para que el sistema conceptual asociativo fuese cada vez más cercano al que existe entre los términos del lenguaje natural.

En la representación de los vocabularios mediante redes semánticas ha sido determinante la facilidad de visualización ofrecida por las pantallas. Para mostrar gráficamente el conocimiento han contribuido de manera decisiva los mapas conceptuales. En las redes, los nodos representan los conceptos, mientras que los enlaces simbolizan las relaciones que se dan entre los conceptos, con un indudable impacto visual respecto a las de jerarquía (CAÑAS; NOVAK, 2009). Pero también ha sido patente el influjo de los sociogramas que reflejan los vínculos existentes entre los integrantes de grupos y organizaciones y muestran la **influencia y las relaciones de preferencia** que se dan dentro de ellos (VAZ, 2011). Por fin, ha sido decisiva la aportación de las redes matemáticas e informáticas basadas en la teoría de grafos, pues han proporcionado herramientas para visibilizar y representar grandes volúmenes de datos (TUTTE, 2001, p.31).

Hasta ahora, los sistemas de indización desatendían los vínculos semánticos existentes entre los conceptos de un texto. Con el empleo de redes semánticas para representarlos se ha progresado mucho en la transición de los términos a los conceptos. Esas redes declaran la reciprocidad por la que se relacionan los conceptos, e incluso alcanzan a mostrar cómo se vinculan sus nodos y la trayectoria seguida en

su distribución, muy cercana a la propuesta de esquemas basados en marcos según el modelo de Entidad/Relación (SHASTRI, 1988). Las redes o marcos representan bien el conocimiento complejo, sobre todo cuando se expresa mediante el lenguaje natural. Al sindicar nodos, las redes han ampliado los propósitos seguidos por los mapas conceptuales en educación. Aunque la asociación de conceptos siempre fue una aspiración en la representación del conocimiento desde que Bush (1945) propusiera que los SOC debían de actuar enlazando conceptos, a imitación de la articulación mental. Ahora se busca reflejar las asociaciones mediante vínculos que conectan los conceptos en un sentido determinado, definiendo así su navegabilidad. Toda asociación queda caracterizada mediante la función que juegan en ella cada uno de los conceptos localizados en los límites del arco. Sobre esta revitalización de las relaciones se ha podido superar la pasividad del léxico de los tesauros al adaptar su significación a la situación concreta y explicarla enunciativamente.

El dinamismo de esos enlaces entre nodos permitió asociaciones más funcionales cuya determinación pasó a hacerse mediante identificación verbal. Se sitúa aquí uno de los mayores avances respecto a los tesauros desarrollados según la norma ISO 2788 (1986) que obligaba a la representación nominal de los descriptores. Los sustantivos refieren los eventos desde una perspectiva pasiva, pues no tienen capacidad de representar las condiciones de admisibilidad de las acciones. Además, los tesauros lo hacían con bastante artificialidad, ya que cualquier acción se representaba con nombres. Las acciones, componente sustancial de los mensajes audiovisuales y de los documentos software, tienen que describirse mediante verbos, precisamente porque los verbos expresan acciones. Para simbolizar esta aserción sirva de ejemplo que comer es verbo y que este hecho no puede expresarse mediante ningún sustantivo sin que se desnaturalice seriamente su significación ¿Cómo representar a quien come con el sustantivo comida? Cuando elegimos esta alternativa expresamos la acción por quien la padece. Siendo así, los estados de hechos, los procesos o las relaciones entre entidades solo se pueden significar si empleamos verbos para hacerlo. Hasta ahora, esto quedaba sin atender en los lenguajes combinatorios.

Si consideramos una fotografía en la que aparece en plano medio una mujer que realiza la acción de comer una manzana, a la hora de indizarla con un tesoro ISO 2788 su representación se limitaría a sustantivos, siendo lo más notorio la reducción pasiva de la acción:

Mujer comida tomate

De esta forma el enlace verbal se sustantivaba, mientras que dichas normas abrieron la posibilidad de que ese enlace pudiese mostrarse como acción y funcionase de conector relacional entre el sujeto y el objeto:

Mujer come tomate

Mujer cultiva tomate

La variación en las relaciones entre sujeto y objeto se hace explícita en el empleo verbal asociativo, cuyo número se extiende tanto como posibilidades ofrece el lenguaje

Mujer compra tomate natural:

Este modelo se replica en las redes semánticas, donde el conocimiento se representa mediante frases simples dispuestas en triplas. En ellas el nodo conceptual sujeto se asocia con el nodo objeto mediante un arco de enlace, que declara las propiedades o predicados, dando lugar al grafo de representación:

Sujeto → Verbo → Objeto

La posibilidad de expresar con verbos las relaciones existentes entre nodos de conceptos establece una de las mayores diferencias entre los SOC tradicionales y su paso a SKOS. El origen se sitúa en la necesidad de declarar las relaciones mediante el empleo de verbos, de modo que los grafos de representación acabaron con la exclusividad de uso de los sustantivos (MOREIRO et al. 2004). Además, la utilización verbal tuvo otra consecuencia, pues los tipos de relaciones asociativas acotados por las normativas precedentes se abrieron sin límite a cuantas pueden establecerse al emplear el lenguaje natural. Lo que tenía la ventaja añadida de precisar el contexto de significación de los nodos de acuerdo con las relaciones existentes entre ellos.

Esta forma de formular las redes semánticas se ha incorporado a las propuestas de simbolización lógica imprescindibles a la hora de intercambiar con los ordenadores la estructura de los conceptos. La representación del conocimiento mediante redes semánticas ha tenido una correspondencia indudable en la creación de declaraciones con RDF. Lo que puede parecer obvio, pues no es otra cosa que formalizar el lenguaje natural según la **lógica de los predicados**. De manera consecuente, las partes de una declaración o sentencia se concretan en la terna sujeto, verbo y objeto.

Tabla 6: Terna de una declaración en lenguaje natural

Sujeto →	Verbo →	Objeto
Recurso o cosa sobre la que versa la declaración	Propiedad o característica del sujeto que se expresa mediante esa declaración	Valor de la propiedad a la que se refiere el predicado

Fuente: Elaboración propia.

Por su parte, las declaraciones con RDF se expresan en un grafo unidireccional donde los nodos aluden tanto al sujeto

como al objeto. Mientras que el arco de relación se refiere a los predicados o propiedades.

Tabla 7: Representación del conocimiento por frases estructuradas

<i>Representación del conocimiento</i>	<i>Frases estructuradas en grafos</i>
Redes semánticas	concepto → relación → concepto nodo agente → arco → nodo objeto sujeto → verbo → predicado
Declaraciones con RDF	

Fuente: Elaboración propia.

Se muestra así el paralelismo que mantienen las estructuras gramaticales básicas con el procesamiento lógico elemental de los esquemas para elaborar ontologías. Lo que ha traído progresión para los vocabularios semánticos. Esta triple concurrencia no supone sorpresa alguna. Son hechos concomitantes que coinciden en la misma propuesta al buscar procesar el significado mediante la lógica gramatical.

4 REPRESENTACIÓN DE INFORMACIÓN EN MODELOS BASADOS EN EL CONOCIMIENTO

Sigue siendo cierto que la sistematización jerárquico-taxonómica ofrece muchas ventajas para organizar el conocimiento. Así se justifica al observar que está presente en todos los esquemas de clasificación, en los vocabularios controlados, en los modelos conceptuales e incluso en las ontologías (CAMPOS et al, 2011). No obstante, los vocabularios conceptuales y ontológicos son mucho más que el orden dado a sus elementos componentes por unas listas con relaciones taxonómicas, ya que su auténtica finalidad es dotar de significado a los objetos informativos para que las máquinas los puedan entender. Si se trata de obtener comprensión semántica hay que facilitar los términos con univocidad y precisión a los ordenadores, lograr que sean interoperables con los de otros vocabularios y originar inteligibilidad acerca de las posibles relaciones que puedan darse entre sus conceptos.

No puede haber Web semántica sin una auténtica recuperación semántica. La búsqueda tiene que efectuarse por significados, por conceptos, por ideas. Lo que obliga al software a distinguir entre ideas y términos. A conseguirlo contribuye la lógica formal, pero también el propio léxico cuyas posibilidades de tratamiento han aumentado notablemente con los estándares recientes. En definitiva, los vocabularios especifican, en un modelo de mundo concreto, cual es la naturaleza y los atributos de los conceptos que contienen y qué relaciones les conectan con otros conceptos (SÁNCHEZ et al, 2007). Hemos de ver la representación ontológica del conocimiento como una descripción explícita y formal de las clases (conceptos) dentro de un marco de conocimiento compartido, así como de las propiedades que delimitan las características y atributos de esas clases, junto a las restricciones que introducen en los atributos (BREITMAN et al, 2007).

Frente a los vocabularios anteriores, las ontologías no cuentan con una disposición de relaciones acotada y preestablecida para representar la información. Requieren un mayor desarrollo de asociación semántica entre los conceptos y también un nivel de descripción de los términos más alto. Por lo que identifican expresamente los sinónimos, suprimen los homónimos y declaran las relaciones semánticas que se dan entre los conceptos que comprenden. Incluso facilitan la incorporación a su estructura de toda relación semántica con tal de que esté bien definida, cuando así lo requiera la representación activa del conocimiento. Son vocabularios enlazados con la red cuya adaptación al espacio conceptual se produ-

ce a través del lenguaje XML. De forma que se configuran como corporación de términos en una red semántica neuronal, donde cada nodo es un concepto al que se asocian otros conceptos. Estos son elementos precisos, con relaciones concretas en esa red, por lo que las ontologías actúan como vocabularios capaces de organizar la información de cualquier objeto.

RDF ha impulsado la reutilización para que las distintas aplicaciones Web puedan intercambiar su información. Cuando se quiere localizar un término con equivalencia exacta entre varias ontologías se parte de una ontología fuente y se alcanza otra de destino (NOY; MCGUINNNESS, 2000). Pero, además, al elaborar ontologías se pueden aprovechar SOC anteriores sobre dominios cercanos, en especial tesauros en formato electrónico lo que hace más fácil su importación a otro sistema. Para lograrlo, a la vez que mantener la carga semántica original, hay todo un desarrollo metodológico. Como en *Methontology* que permite aplicar en un contexto imprevisto el conocimiento inicialmente organizado con otra intención (CORCHO et al, 2005). Se usa así un concepto similar o relacionado con los que se manejan, aunque provenga de un vocabulario ajeno al que se construye. Este intercambio automático de información se consigue al conjugar la interoperabilidad de los documentos con sintaxis XML con la interoperabilidad semántica, derivada de tener los conceptos y sus propiedades un significado perfectamente constituido. Pero el primer paso para que un sistema de información comparta datos procedentes de otros sistemas y alcance una representación compatible se da mediante la interoperabilidad sintáctica. Lo posibilitan modelos de codificación y esquemas de metadatos estandarizados que se aplican a documentos estructurados. Después, para que el intercambio de información sea factible, los términos contenidos tanto en los vocabularios de metadatos como en los controlados y en las ontologías deben de tener equivalencia de significado entre sí y con los documentos manejados (DEXTRE; ZENG, 2012). Por tanto, que sus conceptos puedan interoperar depende de que haya un marco de consistencia en el que se representen con precisión los contenidos según unos esquemas semánticos que incluyen vocabularios controlados como tesauros, listados de encabezamiento de materias, taxonomías, ontologías, anillos de sinónimos, etc.

Skos (2009) incluye un esquema para codificar vocabularios controlados en XML y luego migrarlos al entorno de la web semántica. De forma que facilita publicar vocabularios, pero, sobre todo representar las relaciones existentes entre esquemas conceptuales distintos, para luego unirlos a partir de su semejanza y de las relaciones reales que se dan entre las entidades que los compusieron y administran. Con mayor detalle, a la interoperabilidad le corresponde hallar los términos de un recurso que sean válidos en otro o capturar vocabulario de un recurso y lograr aprovecharlo para un uso apropiado al contexto de otro (MÉNDEZ; GREENBERG, 2012). Sin duda, su aplicación más representativa se propone agrupar diferentes recursos en otro mayor para hacer más comprensibles los mapas conceptuales y terminológicos. La adaptación de los vocabularios más tradicionales para organizar la información de la web requiere un tratamiento y codificación diferentes que conocemos como NKOS (*Networked Knowledge Organization Systems/Services/Structures*). Su tipificación es fundamental para los estándares de la Web Semántica, pues busca aunar los distintos recursos existentes en diferentes sistemas de información digital en la Web, como diccionarios, tesauros, ontologías, esquemas de clasificación, etc. (EUZENAT; TRONCY, 2004).

No cabe duda de que en el logro de unos vocabularios con mejor descripción de los recursos de información han influido mucho los modelos de metadatos mediante los que se pueden expresar sus propiedades bibliográficas: autor, materia, editorial, fecha, etc. Precisamente los esquemas (*schemas*) son formatos de metadatos que, al estar codificados en un lenguaje de marcado normalizado y compartido como XML, son legibles por ordenador. Se basan en la información que el propio recurso proporciona y cuyas propiedades se incluyen en un esquema. Sin embargo, los esquemas no son vocabularios como pueden ser los descriptores de una taxonomía, de una lista de encabezamientos o de un tesoro, los elementos de una ontología o las codificaciones de una clasificación decimal. Al lado de los esquemas suelen aparecer *schemes* que representan las materias de un campo del conocimiento. Desde los criterios usuales en la organización y representación de la información, podemos considerarlos como sinónimos de vocabularios (CAPLAN, 2003: 5-6), pues cumplen la función de indizar y recuperar por ma-

terias. De forma que los *schemes* son metadatos en RDF/XML que se establecen a partir de las materias o del contenido de los recursos.

5 APRECIACIÓN FINAL

Con idea de situar el marco de conceptos de referencia se han considerado los fundamentos de los diferentes tipos de vocabulario empleados para analizar, organizar y representar la información. El punto de partida lo marcó la norma Z 39.19-2005, seguida después por otras que modificaron anteriores propuestas de organización, sobre todo con agregación de clases y oferta de características orientadas siempre hacia una mayor complejidad de tratamiento:

- Palabras-clave independientes en las folksonomías, con rápida evolución hacia los esquemas.
- Listados de términos individuales.
- Sistemas de categorización y clasificación de términos: taxonomías, ...
- Grupos o esquemas enlazados de relaciones: tesauros, redes semánticas y ontologías.

Tanto los nuevos vocabularios como las revisiones profundas de los preexistentes ofrecen una gran complejidad disciplinar y de uso que obliga a abordar su comprensión desde fundamentos teóricos provenientes de diversos campos, como la Lógica formal, la Estadística, la Informática, la Lingüística, la Semiótica o la Lexicografía que en esta reflexión aparecen mezcla-

dos. Aunque en todos los casos ayudan a mejorar la eficiencia de los servicios de información y posibilitan categorizar los recursos mediante esquemas de organización determinados para facilitar su recuperación consecutiva. Como su representación es heterogénea, la forma con que estos recursos están estructurados puede responder a diferentes esquemas de organización.

Asimismo, proporcionan la interpretación de las estructuras de conocimiento de manera lógica y organizada para los usuarios y las máquinas. Lo que supone una progresión hacia la comprensión abstracta de los contenidos, de los significados y de los contextos.

La atención preferente hacia las necesidades de los usuarios hace que se maneje la terminología específica de las instituciones, puesto que la mayor parte de la información que se utiliza es de origen interno. Por esa misma razón, emplean diferentes vocabularios, algunos de ellos controlados, otros no, incluso dentro del mismo dominio.

Si calificamos a la Web como semántica es porque los procesos de búsqueda y recuperación que se efectúan en ella son semánticos. De modo que se tienen que realizar por conceptos, por los significados de las palabras, de los enlaces y de los términos de cualquier recurso documental que contenga. Precisamente, el software y las formalizaciones lógicas de las bases de conocimientos tienen que ser capaces de distinguir entre ideas y términos. En la captación de los conceptos interviene la lógica formal, pero también el propio léxico cuyas posibilidades de tratamiento han aumentado notablemente tal como reflejan las últimas normas internacionales.

Artigo recebido em 08/01/2018 e aceito para publicação em 05/02/2018

ADAPTAÇÃO DOS VOCABULÁRIOS DOCUMENTAIS AO AMBIENTE DIGITAL ONLINE: *léxico, significado e relações semânticas*

RESUMO

São explorados os conceitos relativos a adaptação dos vocabulários às demandas do ambiente da Web e da interoperabilidade. O objetivo é identificar as novas vias de consistência terminológica e expansão de associações entre conceitos. É abordado da perspectiva do léxico usado e segundo a carga semântica crescente das relações entre seus elementos. Os vocabulários são diferenciados pela sua conformação com palavras da linguagem natural, com termos de uma especialidade ou de intercâmbios conceituais. As propriedades das linguagens documentárias, de origem positivista, são contrastadas com as características dos vocabulários enlaçados. As redes semânticas são valorizadas pela sua contribuição para a visualização em tela dos links entre conceitos de tesauros, taxonomias e ontologias, bem como por superar o uso exclusivo de substantivos e por poder expressar com verbos, em tipologia ilimitada, as relações associativas. Finalmente, é justificado o paralelismo entre as declarações em linguagem natural e as feitas em RDF para representar a informação em modelos baseados no conhecimento.

Palavras-chave: Sistemas de organização do conhecimento. Vocabulários. Palavras. Termos. Conceitos complexos.

SEMANTIC VOCABULARIES ADAPTATION TO THE NETWORKED DIGITAL ENVIRONMENT: *lexicon, meaning and semantic relationships*

ABSTRACT Are explored the concepts around vocabularies adjustment to the requirements of the online environment and interoperability. The aim is to identify new ways of terminological consistency and expanding relationships between concepts. It is approached from the perspective of the used lexicon and in accordance with the growing semantic load of the relations between its elements. Vocabularies are differentiated by their composition with natural language words, with subject domain terms or from conceptual exchanges. Controlled languages properties of positivist origins are contrasted with the characteristics of linked vocabularies. Semantic networks are valued for their contribution to the visualization in screens of the links between thesaurus, taxonomies and ontologies concepts, as well as for overcoming the exclusive use of nouns and being able to express associative relations with verbs, in unlimited typology. Finally, the parallelism between declarations in natural language and those made in RDF is justified to represent information in knowledge-based models.

Key-words: Knowledge organization systems; Vocabularies; Words; Terms; Complex concepts; Semantic networks; Ontologies; Relationships between concepts.

REFERENCIAS

- BRASCHER, M.; CAFÉ, L. Organização da Informação ou organização do conhecimento? In: **IX ENANCIB. Diversidade cultural e políticas de Informação.** São Paulo: USP, 2008. 1 CD-ROM.
- BREITMAN, K. K., CASANOVA, M. A.; TRUSZKOWSKI, W. **Semantic Web: Concepts, Technologies and Applications.** London: Springer-Verlag, 2007.
- BSI Group. BS 8723/1-4. **Structured vocabularies for information retrieval: guide.** London: BSI, 2005-2007.
- BUSH, V. As we may think. **Atlantic Monthly**, n. 176, p. 101108, 1945.
- CAMPOS, M^a. L. de A.; SILVA, J. da. M. A Representação de Domínios de Conhecimento e uma Teoria de Representação: a ontologia de fundamentação. **Informação & Informação**, v. 16, n. 2, p. 140-164, 2011.
- CAÑAS, A.; NOVAK, J. ¿Qué es un Mapa Conceptual? Pensacola: Institute for Human and Machine Cognition, 2009. Disponible: <<http://cmap.ihmc.us/docs/mapaconceptual.html>> Acceso en: 08 ago. 2017.
- CAPLAN, P. **Metadata Fundamentals for all Librarians.** Chicago: American Library Association, 2003.
- CORCHO, O.; FERNÁNDEZ-LÓPEZ, M.; GÓMEZ-PÉREZ, A.; LÓPEZ-CIMA, A. Building legal ontologies with METHONTOLOGY and WebODE. In BENJAMINS, R.; SELIC, B.; GANGEMI, A. (ed.). **Law and the semantic web: legal ontologies, methodologies, legal information retrieval, and applications.** Berlin: Springer, 2005, p. 142-157.
- COSERIU, E. **Introducción a la lingüística.** Madrid: Gredos, 1986.
- DEXTRE CLARKE, S. G.; ZENG, M. L. From ISO 2788 to ISO 25964: The evolution of thesaurus standards towards interoperability and data modelling. **Information Standards Quarterly (ISQ)**, v. 24, n. 1, p. 20-26, 2012.
- EUZENAT, J.; TRONCY, R. Web sémantique et pratiques documentaires. LE MOAL, J. C.; HIDOINE, B. Y CALDERAN, L. (eds). **Publier sur internet.** Paris: ADBS, 2004, p. 157-188.

- GUTIÉRREZ RODILLA, B. M. **La ciencia empieza en la palabra.** Análisis e historia del lenguaje científico. Barcelona: Península, 1988.
- HARVEY, D. **The Condition of Postmodernity.** An Inquire into the Origins of Cultural Change. Great Britain: Cambridge University Press, 1994.
- ISO 25964-1. **Information and documentation: Thesauri and interoperability with other vocabularies - Part 1: Thesauri for information retrieval.** Genève: ISO, 2011. Disponible:<<http://www.niso.org/schemas/iso25964/schema-intro/#about>>. Acceso en: 08 ago. 2017.
- ISO 2788: 1986. Information and Documentation - **Guidelines for the establishment and development of monolingual thesauri: international standard ISO 2788 / ISO.** 2nd ed. 1986-11-15. Genève: ISO, 1986.
- MÉNDEZ, E.; GREENBERG, J. Linked data for Open Vocabularies and HIVE's global framework. **El profesional de la Información**, v. 21, n. 3, p. 236-244, 2012.
- MOREIRO-GONZÁLEZ, J. A. Hacia la primacía de los conceptos sobre los términos en los vocabularios para la web semántica. **Anuario ThinkEPI**, v. 7, p. 173-177, 2013.
- NOY, N. F.; MCGUINNESS, D. L. **Ontology development 101: a guide to creating your first ontology.** Palo Alto: Stanford University, 2000. Disponible: <<http://protege.stanford.edu/ontology101-noy-mcguinness.html>>. Acceso en: 17 may. 2017.
- ROLSTON, D. W. **Principios de inteligencia artificial y sistemas expertos.** Bogotá: McGraw-Hill, 1990.
- SÁNCHEZ-CUADRADO, S.; MORATO, J.; PALACIOS, V., LLORENS, J.; MOREIRO, J. A. De repente, ¿todos hablamos de ontologías? **El Profesional de la Información**, v. 16, n. 6, p. 562-568, 2007.
- SEGUNDO, J. E. Santarem. O uso de elementos semânticos no processo de recuperação da informação em ambientes digitais. **Texto Digital**, v. 13, n. 2, p. 93-111, 2017.
- SHASTRI, L. **Semantic Networks: An Evidential Formalization and Its Connectionist Realization.** Los Altos, California: Morgan Kaufmann, 1988.
- SKOS Simple Knowledge Organization System Reference. **W3C Recommendation 18 August 2009.** Disponible:< <http://www.w3.org/TR/2009/REC-skos-reference-20090818/>>. Acceso en: 09 dic. 2017.
- ROE, S. K.; THOMAS, A. R. (eds) **The THESAURUS: Review, Renaissance and Revision.** Binghamton: Haworth Information Press, 2004.
- TUTTE, W. T. **Graph Theory.** Cambridge: Cambridge University Press, 2002.
- VAZ, GLAUBER J. A construção dos sociogramas e a teoria dos grafos. **Revista Brasileira de Psicodrama**, v. 17, n. 2, p. 67-78, 2009.
- Z 39.19-2005: NISO (National Information Standard Organization). **Guidelines for the construction, format and management of monolingual controlled Vocabularies.** Bethesda, Maryland: NISO Press, 2005: Disponible:< <http://www.niso.org/standards/resources/Z39-19-2005.pdf>>. Acceso en: 11 ene. 2017.