
Representación de tesauros en SKOS: tendencias y desafíos para su integración en la Web Semántica

Thesauri representation in SKOS: trends and challenges for their integration in the Semantic Web

María-Luisa ALVITE-DÍEZ (1), M. Mercedes MARTÍNEZ-GONZÁLEZ (2)

(1) Facultad de Filosofía y Letras, Universidad de León, Campus de Vegazana s/n, 24071 León, luisa.alvite@unileon.es.
(2) Departamento de Informática, Universidad de Valladolid, Campus Miguel Delibes s/n, 47011 Valladolid, mercedes@infor.uva.es

Resumen

Los tesauros se han adaptado progresivamente a los entornos digitales demostrando su capacidad de integración con las tecnologías de la Web Semántica. No obstante, la convergencia de tesauros y Web Semántica no es tan directa como aparentemente pudiera parecer. Este trabajo analiza y contrasta los constructos y reglas de integridad de ISO 25964 y SKOS (Simple Knowledge Organization System). Se examinan igualmente las representaciones en SKOS de los tesauros AGROVOC, EuroVoc y Unesco con el fin de estudiar las soluciones llevadas a cabo. Por último, se abordan los retos percibidos, considerando particularmente la integración de los tesauros con los datos enlazados y el desarrollo de ontologías.

Palabras clave: Datos enlazados. ISO 25964. Ontologías. SKOS. Tesauros. Web Semántica.

1. Introducción

La publicación en agosto de 2011 de ISO 25964-1:2011 (*UNE-ISO 25964-1:2014 Información y documentación. Tesauros e interoperabilidad con otros vocabularios. Parte 1: Tesauros para la recuperación de la información*) y de la segunda Parte en marzo de 2013, ISO 25964-2:2013 (*UNE-ISO 25964-2:2016. Parte 2: Interoperabilidad con otros vocabularios*) ha supuesto la superación de la aproximación léxica a la organización del conocimiento sustentada en el estándar ISO 2788-1986. El nuevo estándar de tesauros supone un gran salto que permite conectar estos vocabularios con otros Sistemas de Organización del Conocimiento (SOC) y hace posible expresarlos de forma compatible con otros sistemas electrónicos (García Marco, 2016). Sin embargo, la nueva ISO 25964 vio la luz cuando SKOS se había consolidado ya como estándar de la Web semántica (Miles y Bechhofer, 2009), proporcionando un modo de representación de vocabularios controlados mediante RDF (Resource Description Framework).

La convergencia de los tesauros y la Web Semántica no es tan inmediata como en principio

Abstract

Thesauri have progressively adapted to digital environments, demonstrating their capacity for integration with Semantic Web technologies. However, the convergence of thesauri and Semantic Web is not as straightforward as it might seem. This paper analyses and contrasts the constructs and integrity rules of ISO 25964 and SKOS (Simple Knowledge Organization System). The SKOS representations of the thesauri AGROVOC, EuroVoc and Unesco are also examined in order to study the solutions implemented. Finally, the perceived challenges are addressed, in particular the integration of thesauri with Linked Data and the development of ontologies.

Keywords: ISO 25964. Linked data. Ontologies. Semantic Web. SKOS. Thesauri.

podría parecer. El enfoque restrictivo que guía la definición de los estándares de tesauros diverge con el enfoque abierto de los estándares de la Web Semántica, cuyas recomendaciones se inclinan a aceptar un conjunto más amplio de SOC, como folksonomías, glosarios, listas de encabezamientos de materia, clasificaciones, taxonomías, etc., incluso aquellos sistemas que, siendo “incompletos” (sirve de ejemplo el hecho de no representar/establecer todas las relaciones recíprocas) cumplen con un conjunto de reglas mínimas de integridad que los hacen reconocibles como SOC, de modo que pueden completarse aplicando alguna regla de inferencia.

Este estudio explora la fase actual de integración de los tesauros con la Web Semántica y plantea retos futuros para avanzar en la adaptación de estos Sistemas de Organización del Conocimiento con los datos enlazados (*Linked Data*) y el desarrollo de ontologías.

2. Metodología

El trabajo está conectado con estudios previos (Martínez González y Alvite Díez, 2019a, 2019b, 2020). Para analizar las similitudes y diferencias

entre ISO 25964 y SKOS, y conocer los patrones de representación de tesauros en SKOS se han seguido los pasos siguientes:

1. Comparación y correspondencia de los constructos de ISO 25964 y SKOS.
2. Estudio de las restricciones de integridad en ambos estándares.
3. Análisis de las soluciones de representación de tesauros en RDF/SKOS. En este punto se han examinado los tesauros AGROVOC (1), EuroVoc (2) y Unesco (3).

Finalmente se abordan los retos futuros, considerando particularmente la integración de los tesauros con los datos enlazados y el desarrollo de ontologías.

3. Afinidades y divergencias en los constructos previstos en ISO 25964 y en SKOS

Existen similitudes entre los tesauros y SKOS que facilitan el mapeo de ciertos constructos del tesoro a los de SKOS. Un tesoro se representa como un esquema de conceptos SKOS, los conceptos se definen de forma similar en ambos estándares. Ambos estándares disponen de etiquetas preferidas y no preferidas y notas. Los conceptos pueden relacionarse mediante relaciones jerárquicas (TG/TE) y asociativas (TR). Además, los términos en ISO 25964 pueden representarse con etiquetas SKOS-XL, y en ambas normas es posible establecer relaciones entre ellos.

Constructo ISO 25964	Constructo SKOS
Thesaurus	skos:ConceptScheme
ThesaurusConcept	skos:Concept
ThesaurusTerm	skosxl:Label
ConceptGroup (subject category, domain, microthesaurus, theme)	(Subclass of) skos:Collection skos:ConceptScheme
Note	skos:note
ThesaurusArray	(subclass of) skos:Collection
CompoundEquivalence	No equivalente en SKOS

Tabla 1. Constructos en ISO 25964 y en SKOS

Las dificultades se presentan cuando ciertas construcciones de ISO 25964 no tienen correspondencia directa en las construcciones de SKOS. Como se observa en la Tabla 1, no existe una posibilidad única de representar grupos de conceptos en SKOS y la equivalencia compuesta es específica de los tesauros. Además, la posibi-

lidad de introducir tipos en las relaciones jerárquicas y la polijerarquía, son otras características inherentes a los tesauros que no cuentan con un equivalente en SKOS. Esto añade un nivel de complejidad que dificulta la representación de los tesauros con los estándares de la Web Semántica.

La representación de grupos de conceptos (ConceptGroup) en SKOS se ha resuelto con soluciones heterogéneas. Los tipos de construcciones SKOS que se pueden utilizar para ello son los esquemas de conceptos y las colecciones, pero ninguno de ellos representa exactamente la naturaleza que tienen los grupos de conceptos en la norma ISO 25964. Por ello, las ontologías que amplían SKOS proponen nuevas clases para los grupos de conceptos, distinguiendo incluso entre dominios y microtesauros. Esta es la forma de preservar la idiosincrasia del tesoro: estas clases representan exactamente ese tipo de constructo. Un ejemplo reseñable es el Tesoro EuroVoc, en el que los dominios han sido etiquetados como miembros de las clases skos:Concept y eu:Domain. Esta solución muestra el deseo de preservar la naturaleza de estas construcciones de tesoro y el modo de lograrlo es a través de ontologías *ad-hoc* capaces de proporcionar la semántica que no se encuentra en SKOS, sirven de ejemplo Agrontology (AGROVOC), EuVoc (EuroVoc) e ISO-THES (Unesco).

Las tablas II, III, IV y V muestran ejemplos de los tres tesauros objeto de examen. Los ejemplos que figuran se han extraído de los archivos RDF descargados de los sitios web oficiales de cada uno de los tesauros. En cuanto a las celdas, cada una contiene el enunciado en lenguaje natural, seguido tras dos puntos de su expresión en RDF (cuando hay dos posibilidades, con SKOS y SKOS-XL, se indican ambas).

A pesar de que el *anidamiento entre grupos de conceptos* ha recibido una atención dedicada en ISO 25964, *hasSubGroup/hasSuperGroup*, no existe ninguna equivalencia en SKOS. Las soluciones para el anidamiento de grupos de conceptos utilizadas en los tesauros examinados y en ISO-THES (4) consisten en crear nuevas propiedades para relacionar los grupos de conceptos. Cada ontología propone sus propias propiedades para este fin. ISO-THES propone añadir una relación "subgrupo/supergrupo". Sin embargo, otros tesauros, como EuroVoc, proponen sus propias propiedades para este objetivo. Además, esto guarda relación con la falta de una equivalencia clara para los grupos de conceptos en SKOS.

<i>Constructo ISO 25964</i>	<i>Constructo SKOS</i>
<i>Thesaurus</i>	<i>skos:ConceptScheme</i>
SKOS AGROVOC	AGROVOC ejemplo
skos:ConceptScheme	Agrovoc is a Thesaurus: <http://aims.fao.org/aos/agrovoc> rdf:type skos:ConceptScheme.
SKOS EuroVoc	EuroVoc ejemplo
skos:ConceptScheme eu:Thesaurus	EuroVoc is a Thesaurus: <http://eurovoc.europa.eu/100141> rdf:type skos:ConceptScheme, eu:Thesaurus.
SKOS Unesco	Unesco ejemplo
skos:ConceptScheme	Unesco is a Thesaurus: <http://vocabularies.unesco.org/t hesaurus> rdf:type skos:ConceptScheme.
<i>ThesaurusConcept</i>	<i>skos:Concept</i>
SKOS AGROVOC	AGROVOC ejemplo
skos:Concept	Climate is a ThesaurusConcept: agrovoc:c_1665 rdf:type skos:Concept.
SKOS EuroVoc	EuroVoc ejemplo
skos:Concept eu:ThesaurusConcept	Climate is a ThesaurusConcept: eurovoc:6011 rdf:type skos:Concept, eu:ThesaurusConcept
SKOS Unesco	Unesco ejemplo
skos:Concept	Climate is a ThesaurusConcept: thesaurus:concept5434 rdf:type skos:Concept.
<i>ThesaurusTerm</i>	<i>skosxl:Label</i>
SKOS AGROVOC	AGROVOC ejemplo
skosxl:Label skos-xl:literalForm	'Climate' is a ThesaurusTerm: agrovoc:xl_en_1299512695556 rdf:type skosxl:Label; skosxl:literalForm "climate"@en.
SKOS EuroVoc	EuroVoc ejemplo
skosxl:Label eu:SimplePreferredTerm	'Climate' is a ThesaurusTerm: eurovoc:218442 rdf:type eu:SimplePreferredTerm, skosxl:Label.
SKOS Unesco	Unesco ejemplo
skosxl:Label	'Climate' is a ThesaurusTerm: thesaurus:concept5434-xl- d1e27930 rdf:type skosxl:Label.

Tabla II. Constructos en ISO 25964 y en SKOS: Conceptos y Términos

Las colecciones de SKOS se pueden anidar, pero no hay ninguna propiedad que relacione dos grupos de conceptos en SKOS, que parecería el candidato ideal para la anidación de grupos de conceptos de ISO 25964. Así pues, existen varios enfoques para resolver el problema de representar con SKOS la existencia de varios niveles de grupos de conceptos en los tesauros, así como su anidamiento.

<i>Constructo ISO 25964</i>	<i>Constructo SKOS</i>
<i>ConceptGroup</i>	(subclass of) <i>skos:Collection</i> , <i>skos:ConceptScheme</i>
SKOS AGROVOC	AGROVOC ejemplo
—	—
SKOS EuroVoc	EuroVoc ejemplo
skos:ConceptScheme, eu:MicroThesaurus, skos:Concept, eu:Domain	Natural environment is a Microthesaurus: eurovoc:100243 rdf:type skos:ConceptScheme, eu:MicroThesaurus.
	Environment is a Domain: eurovoc:100155 rdf:type skos:Concept, eu:Domain.
SKOS Unesco	Unesco ejemplo
isothes:ConceptGroup, unes:MicroThesaurus, unes:Domain, skos:Collection	Meteorology is a Microthesaurus: thesaurus:mt2.45 rdf:type isothes:ConceptGroup, unes:MicroThesaurus, skos:Collection.
	Science is a Domain: thesaurus:domain2 rdf:type isothes:ConceptGroup, unes:Domain, skos:Collection.

Tabla III. Grupos de conceptos (dominios, microtesauros...) en ISO 25964 y en SKOS

<i>Constructo ISO 25964</i>	<i>Constructo SKOS</i>
isPartOf / contains (relates ConceptGroup, Thesaurus)	skos:inScheme
SKOS AGROVOC	AGROVOC ejemplo
----	----
SKOS EuroVoc	EuroVoc ejemplo
dct:isPartOf	Natural environment isPartOf EuroVoc: eurovoc:100243 dct:isPartOf eurovoc:100141.
SKOS Unesco	Unesco ejemplo

skos:inScheme	Meteorology isPartOf Unesco Thesaurus: thesaurus:mt2.45 skos:inScheme <http://vocabularies.unesco.org/thesaurus>.
hasSubGroup / hasSuperGroup (relates ConceptGroup, ConceptGroup).	skos:member
SKOS AGROVOC	AGROVOC ejemplo
—	—
SKOS EuroVoc	EuroVoc ejemplo
eu:domain	Natural environment hasSuperGroup Environment: eurovoc:100243 eu:domain eurovoc:100155.
SKOS Unesco	Unesco ejemplo
isothes:superGroup	Meteorology hasSuperGroup Science: thesaurus:mt2.45 isothes:superGroup thesaurus:domain2.
isMemberOfGroup / hasAsMember (relates ThesaurusConcept, ConceptGroup).	skos:member, skos:inScheme
SKOS AGROVOC	AGROVOC ejemplo
—	—
SKOS EuroVoc	EuroVoc ejemplo
skos:inScheme	Climate isMemberOfGroup Natural environment: eurovoc:6011 skos:inScheme eurovoc:100243. Climatic zone isMemberOfGroup Natural Environment: eurovoc:4786 skos:inScheme eurovoc:100243.
SKOS Unesco	Unesco ejemplo
skos:member	Climate isMemberOfGroup Meteorology: thesaurus:mt2.45 skos:member thesaurus:concept5434.

Tabla IV. Relaciones para grupos de conceptos en ISO 25964 y en SKOS

Las diferentes formas de representar los grupos de conceptos con SKOS también dan lugar a diferencias a la hora de tratar los conceptos superiores. Los tesauros que utilizan esquemas de conceptos para representar microtesauros, como EuroVoc, utilizan la propiedad skos:hasTopConcept para este fin. Sin embargo, cuando se utilizan colecciones para representar microtesauros,

como hace ISO-THES, es el caso del Tesoro Unesco, no es posible utilizar esta propiedad de SKOS. En esta situación, el hecho de que un concepto sea un concepto superior no puede expresarse explícitamente con una declaración. Así, afirmar que un concepto del tesoro es un concepto superior de algunos microtesauros (sirve de ejemplo EuroVoc), no puede representarse con la propiedad skos:topConceptOf, que parece el primer enfoque que intentaría un diseñador. Esto se ha dejado a la inferencia (Isaac, 2013). A pesar de ser absolutamente correcto desde una perspectiva técnica, es cierto que confiar en la inferencia aumenta la dependencia del software encargado de la inferencia (razonadores), y que la información de que un concepto es un concepto cabecera no se comparte explícitamente. Una vez que se sabe que un concepto es un concepto superior, sería bueno poder compartir este conocimiento, evitando así tener que inferirlo de nuevo. Por ello, nos parece interesante disponer de un medio para compartir este conocimiento, aunque sea con propiedades ad-hoc creadas específicamente para este fin.

Constructo ISO 25964	Constructo SKOS
hasTopConcept/isTopConceptOf (relates ThesaurusConcept, ConceptGroup)	skos:topConceptOf / skos:hasTopConcept
SKOS AGROVOC	AGROVOC ejemplo
skos:topConceptOf	Activities isTopConceptOf Agrovoc: agrovoc:c_330834 skos:topConceptOf <http://aims.fao.org/aos/agrovoc>.
SKOS EuroVoc	EuroVoc ejemplo
skos:hasTopconcept	climate isTopConceptOf natural environment: eurovoc:6011 skos:topConceptOf eurovoc:100243.
SKOS Unesco	Unesco ejemplo
skos:topConceptOf	Climatology isTopConceptOf UNESCOThesaurus: thesaurus:concept183 skos:topConceptOf <http://vocabularies.unesco.org/thesaurus>.

Tabla V. Relaciones para conceptos cabecera (top) en ISO 25964 y en SKOS

En cuanto a las implicaciones que tiene la introducción de tipos de relaciones Término Genérico/Término específico (TG/TE), conviene tener en cuenta el hecho de que la forma de representar la diversidad de jerarquías asociadas

a los tipos de relaciones TG/TE en SKOS no tiene aún una solución definitiva —véanse las propuestas realizadas por Isaac (2013) y Alexiev, Isaac, y Lindenthal (2016)—. Además, la introducción de tipos afecta a la transitividad, esta propiedad se aplica cuando las relaciones jerárquicas utilizadas para el razonamiento son todas genéricas (tipo de), pero no si se mezclan con relaciones partitivas (parte de).

En lo que se refiere a la *polijerarquía*, ésta no es un tipo de constructo, sino una situación en la que la estructura de un tesoro se vuelve más compleja que un árbol, así, un concepto determinado puede darse en más de una posición en la estructura jerárquica del tesoro. El modelado, y el razonamiento, es más fácil y robusto cuando la jerarquía resultante es arborescente (las posibilidades de obtener conclusiones incoherentes son menores que con otros gráficos), por eso se recomienda controlar la polijerarquía. En el caso en el que exista polijerarquía en un tesoro, ésta puede representarse con SKOS, ya que éste no restringe el uso de la polijerarquía. Sin embargo, no dispone de medios específicos para representar y controlar la polijerarquía. Así pues, la polijerarquía es, de hecho, otra característica específica de los tesauros que se controla mediante soluciones *ad hoc*, como la que se encuentra en EuroVoc, en cuya representación se emplea el atributo propio: `eu:hasPolyhierarchy` para indicar que un microtesoro tiene polijerarquía.

4. Afinidades y divergencias en las reglas de integridad entre ISO 25964 y SKOS

Los axiomas de integridad de la Recomendación SKOS garantizan un buen conjunto de restricciones de integridad. Tanto en ISO 25964 como en SKOS se encuentran restricciones sobre el número de términos preferentes para un concepto dado, la disyunción de conceptos con otras construcciones utilizadas para agruparlos, la restricción de que sólo los conceptos pueden participar en relaciones semánticas genéricas/específicas y asociativas (TG/TE, TR) o que las relaciones jerárquicas son incompatibles con las relaciones asociativas (TR).

Sin embargo, tal y como se observa en la Tabla VI, las reglas de integridad del tesoro son más restrictivas que las de SKOS. La falta de una correspondencia directa entre los grupos de conceptos y las construcciones SKOS ha dado lugar a soluciones heterogéneas para representar los grupos de conceptos con SKOS.

Esto significa que los axiomas SKOS que se aplican dependen del constructo SKOS utilizado, lo

que en última instancia significa que no es posible fijar una equivalencia directa entre algunas reglas de integridad de los tesauros y los axiomas SKOS.

Esto es lo que sucede, por ejemplo, cuando la restricción que garantiza que un constructo no puede ser simultáneamente un grupo de conceptos y un concepto (Condición 1d) tiene que ser mapeada a SKOS. Los axiomas S9 y S37 expresan restricciones similares para las construcciones SKOS. La solución en EuroVoc se asienta en la utilización de diferentes clases de SKOS para los dominios y los microtesauros, que son grupos de conceptos (Véase Tabla III). La clase SKOS asignada a cada dominio es `skos:Concept`, mientras que cada microtesoro se ha representado como un `skos:conceptScheme`. En casos como éste, los axiomas S9 y S37 no garantizan la Condición 1d.

ISO 25964	SKOS
Condición 1a: Solo un término preferente por cada concepto	Axioma S14
Condición 1b: Los grupos de conceptos no pueden ser repetidos	No
Condición 1c: Los conceptos no pueden ser repetidos	No
Condición 1d: Conceptos y grupos de conceptos son conjuntos disjuntos	Axiomas S9 y S37
Condición 1e: Términos preferentes, Términos no preferentes, ... conjuntos disjuntos	Axioma S13
Condición 2: Solo los conceptos participan en relaciones jerárquicas y asociativas	Axiomas S19 y S20
Condición 3: Incompatible con relaciones semánticas	Axioma S27
Condición 4: Prohibidos ciclos	No
Condición 5: Relaciones inversas	No

Tabla VI. Comparación de las restricciones de integridad en ISO 25964 y SKOS

También hay diferencias en las restricciones para los ciclos y en las relaciones recíprocas. Los ciclos que implican relaciones jerárquicas están prohibidos en los tesauros. Algo similar ocurre con los conceptos cabecera (top). No existe ningún axioma de SKOS que exprese una restricción similar a la Condición 4. Hay que subrayar a este respecto, como se señalaba con anterioridad, que SKOS está concebido para un conjunto de SOC más amplio que los tesauros. En consecuencia, las restricciones que son exclusivas de los tesauros deben ser garantizadas por los propios editores de tesauros, además de la integridad proporcionada por SKOS.

Las relaciones recíprocas se exigen en la norma ISO 25964. Si un concepto A está relacionado con un concepto B en algunos tesauros mediante una relación como TG, TE o TR, las relaciones recíprocas también existen. Por lo tanto, ambas relaciones deben estar presentes en la representación RDF del tesoro. Sin embargo, SKOS no requiere la existencia de relaciones recíprocas, pueden estar o no (permitiendo así economizar el número de tripletas RDF almacenadas). Esta es una flexibilidad de SKOS que no comparte con los tesauros.

5. Integración de los tesauros en la Web Semántica

Como se ha señalado, en los tres tesauros examinados en este estudio, se han utilizado ontologías para ampliar SKOS con semántica específica de los tesauros que no está disponible en SKOS. Esto es también lo que hace la ontología ISO-THES.

Se observan ciertos desafíos en cuanto a la incorporación de los tesauros en la Web Semántica. El primero es poder encontrar una forma común de representar grupos de conceptos. Por ahora, hay varias propuestas para ampliar SKOS con clases específicas para ellos. Sería un paso adelante encontrar una solución común, o al menos, poder expresar la equivalencia entre las diferentes soluciones disponibles.

Un segundo reto proviene de la diversidad de tipos de relaciones jerárquicas que introduce la norma ISO 25964. A pesar de que hay algunas propuestas valiosas, como la de Alexiev, Isaac y Lindenthal (2016), todavía no existe una solución única para estas especificaciones recogida en un estándar, bien del W3C, bien ISO, o comúnmente aceptada por la comunidad de tesauros. Pero lo que está claro es que SKOS no prevé tipos en las relaciones jerárquicas. Las consecuencias de usar tipos en relaciones jerárquicas van más allá de ampliar SKOS con nuevas clases y propiedades, dado que afecta a la inferencia. Es cierto que SKOS no asocia la transitividad a las propiedades `skos:broader` y `skos:narrower`, sino a las relaciones `skos:transitivenarrower` y `skos:transitivebroader`, de modo que el razonamiento (5) sobre la transitividad sólo se hará si estas últimas propiedades están presentes. Pero, de todos modos, debería proporcionarse algún medio para limitar la inferencia a las situaciones en las que puede llevarse a cabo.

El último desafío es encontrar un medio para expresar y compartir restricciones específicas de los tesauros que no están en SKOS, como las que se presentan para los ciclos y las relaciones recíprocas. Mientras estos conocimientos no se

compartan de alguna manera, las herramientas de gestión de tesauros se encargan de implementarlas como funcionalidad extra. No obstante, si se pudiesen compartir, podrían ser reutilizados por cualquier herramienta capaz de entender este conocimiento.

Las ontologías pueden ayudar a llenar el vacío existente entre los tesauros y SKOS, ampliando SKOS con clases, propiedades y restricciones específicas de los tesauros, que no ofrece SKOS. Esto es lo que han hecho ontologías como ISO-THES y las ontologías asociadas a los tesauros observados en este trabajo.

La web semántica ofrece medios para expresar la equivalencia entre clases de diferentes ontologías. Esto podría ser útil, ya que el software podría reconocerlas como equivalentes y aplicar tratamientos basados en ello. Así, poder expresar, y compartir estas equivalencias contribuiría a una mejor explotación de las relaciones entre tesauros. Esto también podría ayudar a expresar la equivalencia entre las diferentes clases y/o propiedades utilizadas en cada tesoro para representar grupos de conceptos y anidamiento de grupos de conceptos. Creemos que una ontología común sería la mejor solución para abordar esta cuestión. Podría ser entendida y gestionada por cualquier herramienta de gestión de tesauros, proporcionando así a los diseñadores de tesauros una solución directa para los grupos de conceptos. ISO-THES podría ser la base de una ontología de este tipo, pero serían necesarios algunos cambios para superar la limitación actual que tiene con la representación de los Top Concepts.

La respuesta a la cuestión de compartir la semántica entre ontologías asociadas a diferentes tesauros procede del propio ámbito de las ontologías (Euzenat y Shvaiko, 2013). Estos alineamientos deben estar disponibles para las herramientas que han de ser capaces de procesarlos. OWL (Web Ontology Language) proporciona los cimientos para expresar alineaciones, en concreto, la propiedad `owl:same` se utiliza para expresar equivalencias entre clases ontológicas. Otra cuestión es cómo se pueden realizar estos alineamientos, ya sea manualmente, o bien mediante procesos automáticos. Conviene recordar que la alineación de ontologías puede ser un escollo importante, en el que obtener un buen porcentaje de alineaciones confiables puede ser complicado. Creemos que podría resultar útil contar con alguna ontología general, no específica de un determinado tesoro, que idealmente surgiera de un consenso de la comunidad de tesauros y pudiera utilizarse como referencia para los alineamientos. Este esfuerzo requeriría la colaboración de los desarrolladores de herramientas de gestión de tesauros para que ofrecieran

soporte para dicha ontología por defecto en sus herramientas. La ontología ISO-THES u otras similares podrían desempeñar ese papel.

Otro de los retos a tener en cuenta es poder expresar las restricciones de los tesauros y compartirlas. Esta capacidad de expresar restricciones es otra cuestión estrechamente relacionada con las ontologías. En el contexto de la Web Semántica existen varios lenguajes creados para expresar este tipo de conocimiento. OWL ofrece algunas posibilidades para expresar restricciones, pero también hay otras propuestas realizadas en este sentido, por ejemplo, SHACL (6), un lenguaje para validar grafos RDF contra un conjunto de condiciones.

Los tesauros se han beneficiado de las tecnologías de la Web Semántica de forma sustancial. Las herramientas de gestión de tesauros han adoptado RDF para almacenar, distribuir y compartir tesauros. Se utilizan repositorios semánticos para almacenar las representaciones RDF de los tesauros (por ejemplo, VocBench y Synptica utilizan Ontotext GraphDB, PoolParty utiliza RDF4J y la red de vocabularios empresariales TopQuadrant utiliza Jena).

Los tres tesauros examinados en este estudio permiten la descarga de los archivos RDF desde los URL correspondientes. Estos archivos pueden insertarse en cualquier almacén RDF y utilizarse para aplicaciones específicas. Además, ofrecen la posibilidad de consultar sus datos a través de puntos de acceso SPARQL (SPARQL endpoints), un avance notable en las posibilidades de acceso y de reutilización de sus contenidos. Las consultas se construyen utilizando SPARQL, un lenguaje estandarizado, potente y flexible, para la consulta de grafos RDF.

Por último, parece claro que los tesauros han incorporado con entusiasmo la posibilidad de utilizar las propiedades de SKOS (`skos:exactMatch`, `skos:closeMatch`, etc.) para alinear conceptos de diferentes tesauros. Los tesauros analizados en este trabajo ofrecen sus alineaciones con varios tesauros, que pueden descargarse en RDF. De este modo, se crea una red de conceptos enlazados, que además implica que cualquier dato vinculado con estos conceptos está a su vez enlazado. Esto encaja muy bien con los principios de datos enlazados, y es prometedor como medio para que los tesauros contribuyan a la organización del conocimiento en la Web de datos. En este sentido, subraya Dextre Clarke (2019) que el éxito de los tesauros en la actualidad depende de la interoperabilidad y de las oportunidades abiertas en aplicaciones de datos enlazados.

6. Conclusiones

Los constructos de la norma ISO 25964 se han representado en los tesauros examinados utilizando SKOS, SKOS-XL para los términos, y con soluciones *ad-hoc* para situaciones cuyo mapeo de ISO 25964 a SKOS no era directo. Los tesauros observados utilizan ontologías que amplían la ontología SKOS con semántica adicional para estas situaciones. Esto es lo que ocurre con los grupos de conceptos y el anidamiento de grupos de conceptos, o con semánticas específicas, como las relaciones enriquecidas de AGROVOC. También sería la situación de las relaciones jerárquicas tipificadas (tipo de, instancia de, parte de); sin embargo, no hemos encontrado ejemplos en los tres tesauros utilizados en este estudio, y tampoco ninguna referencia en sus ontologías específicas. Una situación similar sería la de la equivalencia compuesta. En este caso, algunos tesauros como EuroVoc las prevén, su documentación sobre cómo se ha representado EuroVoc con SKOS habla de equivalencias compuestas, aunque no hemos localizado ejemplos de su representación en la versión 4.12 examinada.

Por su parte, ISO-THES es una ontología general que amplía SKOS con clases y propiedades de ISO 25964 que no tienen una representación directa en SKOS (Pastor Sánchez, 2013). Su principal diferencia con respecto a las ontologías que se han creado individualizadamente para un tesoro dado, es que no está asociada a un tesoro específico, por lo que no proporciona clases y/o propiedades para la semántica específica de un tesoro. Esto es, ISO-THES proporciona extensiones para lo que está en ISO 25964 pero no está recogido en SKOS. Tiene clases específicas para grupos de conceptos (`isother:ConceptGroup`) y microtesauros (`isother:MicroThesaurusOf`). Ambas son subclases de `skos:Collection`.

Una ventaja de utilizar `skos:Collection` para los grupos de conceptos es que SKOS tiene propiedades que pueden utilizarse para el anidamiento de colecciones. `skos:member` relaciona dos colecciones, mientras que SKOS no tiene una propiedad que relacione dos esquemas de conceptos. Por lo tanto, el uso de colecciones en lugar de esquemas de conceptos parece estar mejor alineado con el anidamiento de grupos de conceptos. El inconveniente es que las colecciones no pueden tener conceptos superiores en SKOS. La propiedad `skos:topConceptOf` relaciona un concepto y un esquema conceptual, nunca una colección.

De los tres tesauros examinados, solo el de Unesco utiliza activamente las clases ISO-THES. En la representación de este tesoro en RDF hay

declaraciones que indican que una construcción es de un tipo de una de sus clases. ISO-THES también se utiliza para la representación de los tesauros Getty, donde se pueden encontrar algunas construcciones que no se utilizan en el tesoro de la Unesco, por ejemplo, Arrays. Con todo, a pesar de ser una propuesta general, ISO-THES no se utiliza de forma generalizada en la representación de los tesauros.

Los resultados del análisis realizado en este estudio evidencian que los tesauros se encuentran en pleno proceso de integración en la Web Semántica. Asimismo, se constata que tesauros y datos enlazados han encontrado una buena interrelación mutua. Los tesauros aprovechan las tecnologías de los datos enlazados para mostrar la alineación entre ellos, y los datos enlazados han encontrado en los tesauros un gran caso de uso para demostrar su potencial.

Creemos que el futuro exige avanzar en consensos sobre las reglas de integridad, cómo expresarlas y cómo compartirlas. Poder compartir restricciones significa poder simplificar el desarrollo de herramientas de gestión de tesauros y, por tanto, hacer que los tesauros sean más capaces de mostrarse como datos abiertos en la web de datos.

Finalmente, la propuesta de una ontología general, amparada en el consenso de la comunidad competente en estos vocabularios controlados, serviría para realizar alineamientos confiables entre tesauros.

Notas

- (1) <http://www.fao.org/agrovoc/es/access>
- (2) <https://op.europa.eu/en/web/eu-vocabularies/dataset/-/resource?uri=http://publications.europa.eu/resource/dataset/eurovoc>
- (3) <http://vocabularies.unesco.org/browser/thesaurus/es/index>
- (4) <http://purl.org/iso25964/skos-thes>
- (5) Cabe matizar que cuando nos referimos a razonamiento lo hacemos en un sentido amplio del término, como cualquier proceso capaz de inferir nuevo conocimiento a partir de un conjunto de datos. Este proceso puede apoyarse sobre los razonadores asociados a las ontologías, como por ejemplo aquellos que suelen incluirse en los editores de ontologías más populares, o directamente sobre reglas embebidas en el código de las aplicaciones que manipulan los datos. Si bien esta última solución no potencia compartir reglas y procesos con otros miembros de una

comunidad, es cierto que es la opción utilizada en muchas situaciones en las que soluciones arquitectónicamente más elegantes son menos eficientes en términos de coste computacional.

- (6) <https://www.w3.org/TR/shacl/>

Referencias

- Alexiev, Vladimir; Isaac, Antoine; Lindenthal, Jutta (2016). On the composition of ISO 25964 hierarchical relations (BTG, BTP, BTI). // *International Journal on Digital Libraries*. 17:1 (2016) 39-48.
- Dextre Clarke, Stella G. (2019). The information retrieval thesaurus. // *Knowledge Organization*. 46:6 (2019) 439-459.
- Euzenat, Jérôme; Shvaiko, Pavel (2013). *Ontology Matching*. Berlin, Heidelberg: Springer, 2013.
- García Marco, F. J. (2016). Evolution of Thesauri and the History of Knowledge Organization. // *Brazilian Journal of Information Science*. 10:1 (2016) 1-11.
- Isaac, Antoine (2013). Correspondence between ISO 25964 and SKOS/SKOS-XL models. 2013. https://groups.niso.org/apps/group_public/download.php/12351/Correspondence%20ISO25964-SKOSXL-MADS-2013-12-11.pdf
- Miles, Alistair; Bechhofer, Sean (2009). SKOS Simple Knowledge Organization System Reference. W3C Recommendation 18 August 2009. World Wide Web Consortium. <https://www.w3.org/TR/2009/REC-skos-reference-20090818/>
- Martínez González, María Mercedes; Alvite Díez, María Luisa (2019a). The support of constructs in thesaurus tools from a Semantic Web perspective: Framework to assess standard conformance. // *Computer Standards & Interfaces*. 65 (2019) 79-91.
- Martínez González, María Mercedes; Alvite Díez, María Luisa (2019b). Thesauri and Semantic Web: Discussion of the Evolution of Thesauri Toward Their Integration With the Semantic Web. // *IEEE Access*. 7 (2019) 153151-153170.
- Martínez González, María Mercedes; Alvite Díez, María Luisa (2020). A semantic web methodological framework to evaluate the support of integrity in thesaurus tools. // *Journal of Information Science*. 46:3 (2020) 378-391.
- Pastor-Sánchez, J. A. (2013). ISO-THES: ampliando SKOS a partir de la norma de tesauros ISO 25964. // *Anuario ThinkEPI*. 7 (2013) 189-193.
- Suominen Osmo; Mader Christian (2014). Assessing and improving the quality of SKOS vocabularies. // *Journal of Data Semantics*. 3:1 (2014) 47-73.
- UNE-ISO 25964-1:2014 Información y documentación. Tesauros e interoperabilidad con otros vocabularios. Parte 1: Tesauros para la recuperación de la información. Madrid: AENOR, 2014.
- UNE-ISO 25964-2:2016 Información y documentación. Tesauros e interoperabilidad con otros vocabularios. Parte 2: Interoperabilidad con otros vocabularios. Madrid: AENOR, 2016.

Enviado: 2021-03-26. Segunda versión: 2021-07-10.
Aceptado: 2021-10-05.