UNESKOS
Publicación como Linked Open Data de la Nomenclatura Internacional de Ciencia y Tecnología y del Tesauro UNESCO.

JUAN-ANTONIO PASTOR-SÁNCHEZ
Universidad de Murcia
pastor@um.es

FRANCISCO-JAVIER MARTÍNEZ-MÉNDEZ
Universidad de Murcia
javima@um.es

ROSANA LÓPEZ-CARREÑO
Universidad de Murcia
rosanalc@um.es

JOSÉ-VICENTE RODRÍGUEZ-MUÑOZ
Universidad de Murcia
jovi@um.es

Resumen El objetivo del proyecto UNESKOS es la representación de la Nomenclatura Internacional Normalizada de Ciencia y Tecnología y del Tesauro de la UNESCO mediante tecnologías propias de la Web Semántica. Se pretende ofrecer dichos vocabularios como conjuntos de datos RDF con una estructura que facilite su consulta y reutilización según los principios Linked Open Data. La representación de ambos vocabularios se ha desarrollado aplicando SKOS. Si bien el modelado de la Nomenclatura ha resultado sencillo, el del Tesauro ha sido algo más complejo debido a su estructura de dominios de conocimiento y micro-tesauros. La consulta, tanto del Tesauro como de la Nomenclatura, puede realizarse a través de una web multilingüe que permite la búsqueda (por etiquetas preferentes y alternativas) y la navegación por la estructura de relaciones semánticas. Cada categoría y concepto se identifica mediante una URI derreferenciable que permite la negociación de contenido en múltiples formatos. También se ofrece un SPARQL Endpoint para la ejecución de consultas y la recuperación de datos RDF. SKOS se ha demostrado muy adecuado para representar vocabularios sencillos como es el de la Nomenclatura, y algunos más complejos como el Tesauro de la UNESCO. En este último se ha identificado la necesidad de definir elementos que complementen una futura versión de SKOS. El modelado obtenido es compatible con una interfaz de consulta clara y sencilla. La publicación bajo los principios de Linked Open Data permite una reutilización sencilla y ágil de los datos RDF y posibilitan el desarrollo de proyectos futuros, como la alineación con otros vocabularios mediante relaciones de mapeado.

Palabras-clave SKOS. Linked Open Data. UNESCO. Vocabularios Controlados. RDF.

Abstract The goal of UNESKOS project is to represent the International Standard Nomenclature for fields of Science and Technology and the UNESCO Thesaurus. These vocabularies are provided as RDF datasets with an appropriate structure for access and reuse the contained data, applying the Linked Open Data principles. The representation of these vocabularies has been developed using SKOS. The modeling of the Nomenclature is simple. However, the Thesaurus needs a more complex structure because it is organized into several knowledge domains and micro-thesauri. The queries to the Thesaurus and the Nomenclature can be done through a multilingual website that allows searching (for preferred and alternative labels) and browsing the structure of semantic relationships. Categories and concepts are identified by a derreferenciable URI with content negotiation in multiple formats. There is a SPARQL endpoint to query and retrieval RDF data. SKOS

has proven suitable for representing simple vocabularies like the nomenclature, and in more complex cases such as the UNESCO Thesaurus. In this case has been identified the need to define elements that complement SKOS in a future versión of the Thesaurus. The obtained modelling is easily compatible with the development of a clear and simple query interface. The publication, under the principles of Linked Open Data, provides an simple and flexible reuse of RDF data and enables the development of future projects like the alignment with other vocabularies using mapping relationships.

Keywords SKOS. Linked Open Data. UNESCO. Controlled Vocabularies. RDF.

Introducción

En en el ámbito de la Web Semántica resulta esencial disponer de vocabularios controlados adecuados a la descripción de contenidos y conjuntos de datos. Esta importancia deriva de la capacidad de los vocabularios controlados para la desambiguación terminológica y la organización semántica de conceptos (Montalvo, 2011). De hecho, los vocabularios controlados, los tesauros en particular, constituyen una fuente para el desarrollo de ontologías (García-Torres, et al, 2008).

Desde su introducción, SKOS se ha convertido en el instrumento más aplicado para representar en forma de conjuntos de datos RDF lenguajes documentales del tipo tesauros, clasificaciones, encabezamientos de materia, glosarios, etc. (Pastor et al., 2009), (Pastor et al., 2012). De hecho, se ha convertido en uno de los componentes clave en el despliegue de la Web Semántica, puesto que permite aplicar técnicas y principios propios de la organización del conocimiento. Además de la representación, mediante SKOS es preciso considerar determinados aspectos relacionados con el acceso y la reutilización. En este sentido Linked Open Data va más allá de la mera aplicación de SKOS para modelar un vocabulario, puesto que propone una serie de criterios y técnicas para facilitar tanto su acceso como su interoperabilidad.

Con este panorama podemos encontrar diferentes sistemas de organización del conocimiento (KOS)[1] que si bien están publicados en Internet, su reutilización deja bastante que desear. Una de las líneas de trabajo de nuestro grupo de investigación se ha centrado tanto en la Nomenclatura Internacional Normalizada de Ciencia y Tecnología y en el Tesauro de la UNESCO, vocabularios que han sido objeto de nuestro interés por la relevancia que poseen ambos en la gestión de información científica relacionada con procesos y proyectos de investigación. La Nomenclatura es ampliamente utilizada en la gestión y realización de proyectos de investigación, tesis doctorales, documentos científicos de todo tipo (informes, artículos, monografías). Por su parte, el tesauro de la UNESCO es profusamente utilizado como vocabulario de ámbito general para la indización de colecciones y repositorios digitales, estudios informétricos y cienciométricos, así como fuente para el desarrollo de nuevos vocabularios y ontologías. Este tesauro dispone de un sitio web para su consulta, pero no dispone de funciones para la reutilización o acceso a términos específicos de un modo sencillo y transparente.

Las tareas desarrolladas en el seno del proyecto UNESKOS tenían como objetivo la representación, publicación y acceso mediante tecnologías de la Web Semántica de la Nomenclatura y del Tesauro. El presente trabajo expone cómo se han llevado a cabo los distintos pasos del proyecto, los problemas encontrados, las decisiones tomadas, y las soluciones adoptadas. Se detalla la metodología utilizada, abordando además un breve análisis de ambos vocabularios destacando sus principales características. En la sección de resultados se muestra detalladamente el modelado escogido para su representación mediante SKOS junto el uso de diferentes herramientas para su disponibilidad, como Linked Open Data a partir de la publicación de conjuntos de datos RDF, derreferenciación de URIs[2] y puesta en marcha de un SPARQL Endpoint. A continuación se debaten algunas de las decisiones tomadas en relación al modelado SKOS, la posible extensión de esta recomendación para incorporar ciertos puntos de la norma de tesauros ISO-25964 y las

[1] Se utilizará en adelante la abreviatura KOS cuyo origen está en el término en Inglés Knowledge Organization System para hacer referencia a los sistemas de organización del conocimiento desde una aproximación general.

[2] Una URI se utiliza para identificar de forma unívoca un recurso en entorno web. Una URI derreferenciable permite además su localización para poder acceder a dicho recurso.

repercusiones que esto tendría en el proyecto. Para finalizar, se aporta una serie de conclusiones en las que también se apuntan posibles ampliaciones de UNESKOS en el futuro.

1 Metodología

A continuación se detalla la metodología seguida para la publicación como Linked Open Data tanto de la Nomenclatura como del Tesauro. Para comprender mejor los distintos pasos seguidos es imprescindible realizar una descripción global de la estructura de ambos vocabularios.

1.1 Sobre la Nomenclatura

UNESCO comenzó los trabajos de desarrollo de un sistema normalizado para la clasificación de los campos de Ciencia y Tecnología en 1966. Dicha nomenclatura fue propuesta por las divisiones de Política Científica y de Estadística de Ciencia y Tecnología de dicho organismo entre 1973 y 1974. En 1988 se propuso oficialmente por la UNESCO aunque en la práctica había comenzado a utilizarse antes de esta fecha[3].

Su objetivo inicial era la clasificación de artículos científicos y tesis doctorales. Sin embargo, actualmente es utilizada por numerosos organismos y entidades para la descripción y clasificación de recursos relacionados con la investigación científica, proyectos, grupos y líneas de investigación, patentes, informes de evaluación científica, etc. (Martínez-Frias, Hochberg, 2007). La versión de 1988 no ha sido revisada posteriormente por la UNESCO y aunque han surgido numerosas tecnologías y disciplinas científicas desde entonces, la nomenclatura es ampliamente utilizada e incluso ha sido revisada por algunos organismos[4].

Aunque suele afirmarse que la Nomenclatura tiene tres versiones, en realidad se trata de distintos niveles de detalle jerárquico. La organización básica sigue una estructura clasificatoria de tres niveles en el que cada categoría tiene asociada una notación de 2 a 6 dígitos. El primer nivel utiliza dos dígitos para identificar los campos principales de la clasificación. En el segundo nivel se identifican las disciplinas asociadas a los diferentes campos mediante 4 dígitos. El tercer nivel utiliza 6 dígitos para hacer lo propio con las subdisciplinas. Los códigos son acumulativos, por ejemplo: 22 Física, 2202 Electromagnetismo, 2202.05 Rayos gamma. Algunas categorías tienen uno o varios reenvíos que permiten identificar categorías relacionadas entre sí, por ejemplo: 2202.09 Propagación de ondas electromagnéticas (véase 2105 Radiastronomía). La Nomenclatura se ofrece en tres ediciones: Inglés, Francés y Español.

1.2 Sobre el Tesauro de la UNESCO

UNESCO hizo pública la primera edición de este tesauro en 1977 (Aitchison y Clarke, 2004). Inicialmente, su principal aplicación fue ayudar en la consulta de bases de datos de dicha organización, si bien su uso se ha extendido a otros contextos: la desambigüación terminológica, como fuente para la elaboración de otros tesauros (Dunsire, 2011), para la docencia o para la descripción de recursos educativos (García y Jaroszczuk, 2009).

Desde el año 2000 existe una versión web para su consulta, primero en Inglés y luego, de forma progresiva, se han incorporado progresivamente otros idiomas (Francés, Español y Ruso). Se

[3] En 1983 por Resolución de 23 de septiembre de 1983 (BOE 14 de octubre) pasa a ser la clasificación utilizada por el Ministerio de Ciencia y Tecnología del Gobierno de España.
[4] La Universidad del País Vasco dispone de una versión de la edición en Español que incluye actualizaciones en algunos campos o disciplinas. Más información en: http://www.et.bs.ehu.es/varios/unesco.htm.

trata de un tesauro multidisplinar, monojerárquico[5], multilingüe que cumple las normas ISO-2788 e ISO-5964 (Ewketu, 2011). Está estructurado en términos entre los que se establecen relaciones de equivalencia (sinonimia o cuasi-sinonimia), jerárquicas y asociativas, variando el número de términos en función del idioma.

Los términos se dividen en preferentes y no-preferentes a partir de las relaciones de equivalencia. De este modo, sobre un término no-preferente se define una relación de sinonimia "USE" (usar, úsese) con un término preferente. El Tesauro de la UNESCO asocia los distintos términos a uno o varios micro-tesauros que a su vez están asociados a siete áreas o temas principales. Uno de los aspectos más interesantes de este tesauro es que su dominio de conocimiento posee una cobertura general. También hay que hacer especial mención al hecho que desde la puesta en marcha del servicio de consulta a través de la Web, el tesauro se está actualizando constantemente.

1.3 Planificación

En este apartado procedemos a describir los pasos realizados durante el proyecto, algunos de los cuales se ampliarán más adelante. En primer lugar se procedió a trabajar con la Nomenclatura. Tras analizar el trabajo realizado y recibir sugerencias por parte de usuarios y especialistas para mejorar el resultado, se procedió con el Tesauro.

La secuenciación de dichos pasos para cada uno de los vocabularios fue la siguiente:

1. Obtención de una versión digital del vocabulario;
2. Elaboración de ficheros de texto normalizados en cada idioma;
3. Obtención de una serialización RDF/Turtle y RDF/XML;
4. Volcado de la serialización en un triplestore;
5. Configuración del servidor para la derreferenciación de URIs;
6. Desarrollo de la interfaz de navegación y búsqueda;
7. Implementación de un SPARQL Endpoint.

El desarrollo del primer paso ha variado sustancialmente para cada uno de los vocabularios. En el caso de la Nomenclatura, no pudimos encontrar fuentes digitales procesables. Únicamente se localizaron archivos PDF con los correspondientes documentos originales escaneados6. Se utilizó el documento de la Universidad del País Vasco 7 como versión base puesto que permitía su recuperación y procesamiento para la elaboración de un fichero de texto normalizado en el paso posterior. En este mismo paso se realizó la traducción de dicho fichero para obtener las versiones en Inglés y Francés utilizando los documentos originales escaneados. Los datos relativos al Tesauro se obtuvieron del propio sitio web del Tesauro de la UNESCO8. Se aplicaron técnicas de 'web scraping'[9] para el envío automático de formularios y análisis mediante expresiones regulares de los

[5] Si bien el Tesauro de la UNESCO es mono-jerárquico en su mayor parte no sucede así con los términos que hacen referencia a descriptores geográficos.
[6] Los documentos originales en los tres idiomas se encuentran en: http://unesdoc.unesco.org/Ulis/cgi-bin/ulis.pl?catno=82946
[7] Disponible en: http://www.et.bs.ehu.es/varios/unesco.htm
[8] Disponible en: http://databases.unesco.org/thessp/
[9] Extracción de datos de las páginas web sin estructura para darles un formato.

resultados obtenidos (Schrenk, 2012 p.49-75). A partir de dichos datos se construyeron los ficheros de texto normalizados en cada idioma.

Este conjunto de ficheros fueron procesados conjuntamente en el paso 3 para obtener las representaciones en SKOS. Como SKOS es una ontología OWL que sigue el modelo de descripción RDF, esta representación se serializó en RDF/Turtle y RDF/XML. El procesamiento de los ficheros se realizó mediante un sencillo 'script'[10] PHP. Para la Nomenclatura y el Tesauro se utilizaron scripts distintos, ya que la estructura de los ficheros de texto era diferente en cada caso.

En el paso 4 se procedió a volcar en un 'triplestore'[11] dichas serializaciones. En UNESKOS se utilizó el triplestore suministrado por ARC2, un conjunto de librerías que facilita el desarrollo de aplicaciones PHP que operan con datos RDF. En este caso, el triplestore hace uso de una base de datos MySQL. Por lo tanto, los requisitos del servidor utilizado en el proyecto son muy básicos y se identifican con la oferta estándar de cualquier proveedor de alojamiento web, lo que favorecería una potencial portabilidad y reutilización del proyecto.

Uno de los objetivos principales de UNESKOS era la publicación Linked Open Data de ambos vocabularios, por ello, cada vocabulario cuenta con su propio espacio de nombres XML. Además se procedió a la adecuada configuración del servidor para la negociación de contenido y la consiguiente derreferenciación de las URIs correspondientes a los diferentes elementos del Tesauro y la Nomenclatura.

Con posterioridad, se desarrolló la interfaz de navegación y búsqueda para cada uno de los vocabularios. También se dotó de contenidos adicionales: estadísticas, créditos, descarga de serializaciones RDF, etc. El desarrollo se realizó en PHP y ARC2, desarrollando dos pequeños programas que se adaptaban a las peculiaridades tanto del Tesauro como de la Nomenclatura. Estos programas hacen uso de una clase creada para ex profeso para el proyecto UNESKOS. Esta clase define una capa de abstracción con respecto a ARC2, definiendo una serie de propiedades y métodos para el recorrido y la consulta de conjuntos de datos RDF que hacen uso de SKOS.

El último paso realizado consistió en la implementación de un SPARQL Endpoint, desde el que se pueden realizar consultas para recuperar datos concretos de un determinado vocabulario.

2 Resultados

Para la presentación de los resultados, en primer lugar vamos a llevar a cabo una breve descripción de SKOS indicando todo lo referente al trabajo de modelado y a la arquitectura del sistema para cumplir con los requisitos Linked Open Data. Como producto final del proyecto UNESKOS se va a disponer de un sitio web[12] en el que se pueden consultar tanto la Nomenclatura como el Tesauro. Por otro lado, el modelado realizado para ambos vocabularios forma parte de los resultados del proyecto, puesto que se tomaron ciertas decisiones con el objeto de alcanzar una representación adecuada.

[10]En Informática es un guión, archivo de órdenes o archivo de procesamiento por lotes, vulgarmente referidos con el barbarismo script (del latín *scriptum*, escrito).
[11]Un triplestore es una base de datos que permite el almacenamiento de sentencias RDF.
[12]La Nomenclatura y el Tesauro están disponibles en <http://skos.um.es/unesco6> y <http_//skos.um.es/unescothes> respectivamente.

2.1 Modelado mediante SKOS

SKOS es una ontología OWL-Full desarrollada para la representación de KOS de estructura relativamente sencilla. Como toda ontología OWL se basa en el RDF, el modelo de datos sobre el que se desarrolla la Web Semántica. En RDF la información se representa mediante tripletas del tipo sujeto-predicado-objeto. El sujeto es un recurso web identificado con una URI. El predicado es una propiedad o relación que se declara sobre el sujeto. El objeto es el valor de la propiedad u otro recurso con el que se relaciona el sujeto.

En SKOS los sujetos pueden ser de tres clases: conceptos, esquemas de conceptos y colecciones. Los elementos fundamentales son los conceptos que tienen asignadas etiquetas en uno o varios idiomas. Existen diferentes tipos de etiquetas:

- preferentes (skos:prefLabel): términos utilizados en la indización. Su función es idéntica a la de los términos descriptores de los tesauros;
- alternativas (skos:altLabel): utilizadas para representar términos sinónimos o cuasi-sinónimos de los preferentes. Permiten enriquecer la diversidad léxica de un KOS y ofrecer múltiples puntos de acceso a un concepto que puede representarte con diferentes términos;
- Ocultas (skos:hiddenLabel): no visibles para los usuarios, pero si para las aplicaciones informáticas. Son útiles para el control de variantes terminológicas con errores ortográficos, diversas formas de acrónimos y abreviaturas, etc.

Existen una serie de propiedades que permiten vincular conceptos a través de la definición de relaciones semánticas:

- Jerárquica específica (skos:narrower): relación que indica que existe un concepto cuyo significado es más específico que el concepto sobre el que se define la relación (por ejemplo: *Botánica* con *Genética Vegetal*);
- Jerárquica genérica (skos:broader): es la relación inversa de la anterior;
- Asociativa (skos:related): relación que indica que dos conceptos están relacionados semánticamente. (Por ejemplo: *Botánica* con *Biología*).

Los conceptos se asocian a un esquema de conceptos (skos:ConceptScheme) que generalmente identifica un único KOS. También es posible agrupar los conceptos en colecciones simples (skos:Collection) u ordenadas (skos:OrderedCollection).

SKOS también dispone de un conjunto de propiedades que permiten configurar redes entre diferentes KOS mediante relaciones de mapeado a establecer entre conceptos de diferentes esquemas:

- Equivalencia jerárquica específica (skos:narrowMatch): se utiliza cuando se desea indicar que un concepto de un esquema tiene un significado más específico que un concepto de otro esquema.
- Equivalencia jerárquica genérica (skos:broadMatch): es la relación inversa de la anterior.

- Equivalencia asociativa (skos:relatedMatch): cuando dos conceptos de diferentes esquemas están relacionados semánticamente.
- Equivalencia exacta (skos:exactMatch): cuando ambos conceptos tienen exactamente el mismo significado.
- Equivalencia cercana o próxima (skos:closeMatch): cuando ambos conceptos tienen un significado aproximado sin llegar a ser exacto.

Adicionalmente, SKOS define una serie de condiciones de consistencia y reglas -especialmente referidas a la transitividad de las relaciones semánticas- que delimitan el ámbito de aplicación de los vocabularios en procesos lógicos de inferencia (Miles y Bechhofer, 2009).

2.2 Modelado de la Nomenclatura

Como se ha indicado anteriormente, la Nomenclatura Internacional Normalizada de Ciencia y Tecnología tiene una estructura jerárquica con determinados reenvíos de unas categorías a otras. La terminología se encuentra en tres idiomas: Español, Inglés y Francés. Este vocabulario ha precisado de un modelado sencillo para el que SKOS se ha adaptado perfectamente. Se definió un único esquema de conceptos, las categorías se representaron mediante conceptos asociados al esquema de conceptos, sobre los que se definieron una etiqueta preferente en cada idioma. A cada concepto se asoció una propiedad skos:notation para representar la notación de dígitos correspondiente. Las relaciones semántica jerárquicas genéricas y específicas, así como las relaciones asociativas, se expresaron con las propiedades skos:broader, skos:narrower y skos:related respectivamente. El modelado expresado en RDF/Turtle de la categoría “2105 Radioastronomía” sería:

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
<http://skos.um.es/unesco6/2105>
  rdf:type                skos:Concept ;
  skos:prefLabel          "Radio-astronomie"@fr ,
                          "Radioastronomy"@en ,
                          "Radioastronomía "@es ;
  skos:inScheme           <http://skos.um.es/unesco6/00> ;
  skos:narrower           <http://skos.um.es/unesco6/210599> ,
                          <http://skos.um.es/unesco6/210502> ,
                          <http://skos.um.es/unesco6/210501> ;
  skos:notation           "2105" ;
  skos:broader            <http://skos.um.es/unesco6/21> ;
  skos:related            <http://skos.um.es/unesco6/220209> .
```

Los campos principales de la Nomenclatura se han identificado como conceptos cabecera del esquema de conceptos mediante la relación skos:topConceptOf y skos:hasTopConcept. El conjunto de datos de la Nomenclatura está formado por un total de 20.820 tripletas RDF[13]. La tabla 2 muestra los resultados finales cuantitativos más destacados:

Entidad SKOS	Tipo	Número
skos:ConceptScheme	Clase	1
skos:Concept	Clase	2504
skos:prefLabel	Propiedad	7515
skos:broader	Propiedad	2480
skos:narrower	Propiedad	2480
skos:related	Propiedad	780
skos:hasTopConcept	Propiedad	24
skos:topConceptOf	Propiedad	24
skos:inScheme	Propiedad	2504

Tabla 1: Detalle de las estadísticas del conjunto de datos de la Nomenclatura desglosados por entidades SKOS.

La Nomenclatura sirvió como una primera toma de contacto del proyecto con un caso real de modelado con SKOS. El hecho más destacable del procedimiento tiene que ver con la integridad terminológica de la propia nomenclatura. Durante el proceso de modelado se verificó que existían más de 200 categorías con la misma etiqueta preferente en el mismo idioma (por ejemplo, la disciplina 1105 y la subdisciplina 610701 tienen la misma etiqueta preferente: "Metodología" o las subdisciplinas 220303 y 330705 tienen ambas como etiqueta preferente, "Válvulas electrónicas"). A este respecto, la Guía de Referencia de SKOS establece en su condición de integridad S14 que las propiedades skos:prefLabel, skos:altLabel y skos:hiddenLabel son disjuntas entre sí. Esto significa que en un mismo idioma no deberían existir dos etiquetas iguales, sean del tipo que sean. El incumplimiento de esta condición de integridad significa que los procesos de búsqueda y recuperación de conceptos, basados en el texto de las etiquetas, se ven distorsionados al ser recuperados varios conceptos con idénticas etiquetas a partir de un mismo mismo término.

[13] El conjunto de datos de la Nomenclatura puede descargarse libremente en: http://skos.um.es/unesco6/downloads.php

2.3 Modelado del Tesauro

El Tesauro de la UNESCO posee un ámbito multidisciplinar (macro-tesauro) y se estructura en 7 áreas de conocimiento y más de 88 micro-tesauros que actualmente contienen más de 8.600 términos en Español y Francés, más de 7.100 términos en Inglés y casi 7.000 en Ruso[14].

Las equivalencias entre los términos de diferentes idiomas permitieron la definición de conceptos etiquetados mediante skos:prefLabel (descriptores) y skos:altLabel (no-descriptores). Las propiedades SKOS para la representación de las relaciones jerárquicas y asociativas se utilizaron para modelar los términos específicos, genéricos y relacionados de cada término del Tesauro. Aquellos conceptos sin relaciones jerárquicas genéricas se establecieron como conceptos cabecera.

La principal dificultad del proceso de modelado estribó en la representación de la estructura de áreas de conocimiento y micro-tesauros. En principio, SKOS no ofrece ningún elemento para representar este tipo de estructuras de forma directa[15]. Las alternativas analizadas fueron las siguientes:

[a] Definir el Tesauro de la UNESCO, cada área de conocimiento y cada micro-tesauro como esquemas de conceptos y utilizar la propiedad in:Scheme para jerarquizarlos. Suponiendo que <S> representara el Tesauro de la UNESCO en su totalidad, <S1> el área de conocimiento de "Educación" y <S105> el micro-tesauro "Ciencias de la educación y ambiente educacional" se representaría del siguiente modo:

<S>	rdf:type	skos:ConceptScheme
<S1>	rdf:type	skos:ConceptScheme
<S105>	rdf:type	skos:ConceptScheme
<S1>	skos:inScheme	<S>
<S105>	skos:inScheme	<S1>

[b] Definir algún tipo de propiedad personalizada y con una semántica diferente de skos:inScheme para definir jerarquías de esquemas de conceptos, del tipo: ex:subScheme y ex:subSchemeOf. Siguiendo el ejemplo anterior[16]:

<S>	rdf:type	skos:ConceptScheme
<S1>	rdf:type	skos:ConceptScheme
<S105>	rdf:type	skos:ConceptScheme

[14] La organización general del tesauro se ha descrito previamente en la sección 2.2.

[15] A este respecto resulta interesante consultar el debate que tuvo lugar en la lista public-esw-thes del W3C: http://lists.w3.org/Archives/Public/public-esw-thes/2012Dec/0006.html

[16] Se utiliza el prefijo "ex" para hacer mención de las propiedades que deberían crearse con esta opción.

<S>	ex:subScheme	<S1>
<S1>	ex:subScheme	<S105>
<S1>	ex:subSchemeOf	<S>
<S105>	ex:subSchemeOf	<S1>

[c] Aplicar la correspondencia entre ISO 25964 y SKOS/SKOS-XL[17] (ISO TC46/SC9/WG8 e Isaac, 2012) que en la práctica supone una ampliación SKOS. Utilizando esta propuesta en el ejemplo anterior, el Tesauro se definiría como un esquema de conceptos y las áreas y los micro-tesauros como grupos de conceptos (iso-thes:ConceptGroup). Se subordinarían las áreas al esquema de conceptos mediante iso-thes:microThesaurusOf y los micro-tesauros a - las áreas utilizando las propiedades iso-thes:subGroupOf / iso-thes:superGroupOf.

<S>	rdf:type	skos:ConceptScheme
<S1>	rdf:type	iso-thes:ConceptGroup
<S105>	rdf:type	iso-thes:ConceptGroup
<S1>	iso-thes:microThesaurusOf	<S>
<S1>	iso-thes:subGroup	<S105>
<S105>	iso-thes:subGroupOf	<S1>

- Definir el Tesauro como un esquema de conceptos y las áreas y micro-tesauros como colecciones. Las áreas se asociarían al esquema con la propiedad skos:inScheme. Mediante skos:member se indicaría la pertenencia de un micro-tesauro a una área de conocimiento. Continuando con el ejemplo anterior:

<S>	rdf:type	skos:ConceptScheme
<S1>	rdf:type	skos:Collection
<S105>	rdf:type	skos:Collection
<S1>	skos:inScheme	<S>

[17] Se trata de una propuesta que todavía está pendiente de su publicación definitiva y por lo tanto todavía no se ha definido un espacio de nombres para las diferentes elementos. En dicha propuesta se utiliza el prefijo iso-thes para los diferentes elementos, por lo que en este trabajo se hará lo mismo.

<S1> skos:member <S105>

Las cuatro alternativas mostradas conllevan la vinculación de los conceptos del Tesauro al esquema de conceptos que lo representa aplicando la propiedad skos:inScheme. En el caso de la opción [a] dicha vinculación se realizaría con el esquema de conceptos del micro-tesauro al que pertenezca el concepto. En las alternativas [b] [c] y [d] los conceptos se asociarían a las entidades que representan los micro-tesauros mediante la propiedad skos:member. Es decir, el caso de un concepto <C> asociado al micro-tesauro <S105> se expresaría del siguiente modo:

<C> rdf:type skos:Concept

<S105> skos:member <C>

La opción [a] se descartó debido a que en la práctica habitual de modelado con SKOS las entidades del tipo skos:ConceptScheme se utilizan para modelar KOS independientes. Como los micro-tesauros que nos ocupan no tienen dicha naturaleza, se descartó esta opción. El modelado con la propuesta [b] se rechazó debido a que en un principio el equipo del proyecto UNESKOS quería evitar la definición de un vocabulario ad hoc, buscando en todo momento el uso de las clases y propiedades ofrecidas de forma nativa por SKOS. También se descartó la opción [c], puesto que en el momento de realizar el modelado la propuesta ISO-THES todavía estaba pendiente de perfilar (Pastor, 2013) e incluso todavía no disponía de un espacio de nombres normalizado que permitiera hacer referencia a los diferentes elementos.

Finalmente se optó por aplicar la alternativa [d] al no precisar de elementos diferentes de los que incorpora SKOS y permitir el modelado de los micro-tesauros dentro de su pertenencia a un esquema de conceptos que representa al Tesauro en su totalidad[18], obteniéndose como resultado final un conjunto de datos formado por 69.776 tripletas[19]. En la tabla 2 se detallan los datos cuantitativos de los elementos más relevantes.

Entidad SKOS	**Tipo**	**Número**
skos:ConceptScheme	Clase	1
skos:Concept	Clase	4408
skos:Collection	Clase	96
skos:prefLabel	Propiedad	17980
skos:altLabel	Propiedad	13868
skos:broader	Propiedad	4244

[18] Se puede descargar el conjunto de datos libremente en: http://skos.um.es/unescothes/downloads.php.
[19] El conjunto de datos del Tesauro puede descargarse libremente en: http://skos.um.es/unescothes/downloads.php.

skos:narrower	Propiedad	4244
skos:related	Propiedad	12196
skos:hasTopConcept	Propiedad	583
skos:topConceptOf	Propiedad	583
skos:inScheme	Propiedad	4415
skos:member	Propiedad	4496

Tabla 2: Detalle de las estadísticas del conjunto de datos del Tesauro desglosados por entidades SKOS.

2.4 Aplicación de los principios Linked Open Data.

Además del modelado de vocabularios mediante SKOS, nuestro proyecto también tenía como objetivo la aplicación de los principios Linked Open Data para la publicación de dichos conjuntos de datos. La expresión Linked Open Data se refiere a la publicación de datos estructurados en la Web. Los conjuntos de datos pueden ser reutilizados, recuperados e incluso enlazados desde otros conjuntos de datos. Este planteamiento conlleva la apertura de datos para su uso público, partiendo de la interoperabilidad que aportan los formatos y protocolos abiertos desarrollados por el W3C. Se utiliza el protocolo HTTP para que una aplicación informática pueda seguir los enlaces establecidos entre diferentes conjuntos de datos. Es un enfoque diferente al utilizado en los sistemas basados en procesos de agregación, como ocurre con el protocolo OAI-PMH[20] que conlleva la creación de catálogos centralizados cuyo contenido se obtiene a partir de agregadores que "recolectan" los metadatos de repositorios externos.

A partir de la propuesta original realizada por Berners-Lee (2006) y algunos otros autores (Haslhofer y Schandl, 2010) hemos definido una serie de principios Linked Open Data para este proyecto:

1. Utilizar URIs derreferenciables.
2. Implementar mecanismos para la negociación de contenido.
3. Aplicar estándares abiertos con respecto a los formatos empleados para suministrar los datos.
4. Permitir la recuperación selectiva mediante SPARQL.
5. Incluir enlaces a URIs de recursos externos que permita descubrir nuevos objetos.
6. Existencia de una licencia adecuada para la libre reutilización de los datos.

[20] Open Archives Initiative-Protocol Metadata Harvesting es un protocolo utilizado para la recolección de metadatos que describen objetos de información. Se basa en el uso de una serie de comandos para solicitar conjuntos de metadatos expresados mediante Dublin Core a servicios web de proveedores de datos. Se utiliza XML, tanto para representar los mensajes de solicitud como para hacer lo propio con los datos obtenidos.

El uso de URIs derreferenciables es el mecanismo esencial para la identificación y el acceso a objetos o recursos específicos de un conjunto de datos. Se ha dotado a cada objeto de los vocabularios de su propia URI individual. Para ello se han definido dos espacios de nombres:

- Nomenclatura: <http://skos.um.es/unesco6/>
- Tesauro: <http://skos.um.es/unescothes/>

Estos espacios de nombres se utilizan como prefijos de los objetos de los elementos respectivos. Por ejemplo, la URI para el concepto del Tesuro de la UNESCO etiquetado con el término en Español "Almacenamiento de agua" es <http://skos.um.es/unescothes/C04305>. Este criterio se ha aplicado a todos los tipos de objetos: conceptos, esquemas y colecciones.

La existencia de URIs derreferenciables permite utilizar el protocolo HTTP en la negociación de contenido, es decir, la transmisión de datos entre cliente y servidor, pudiendo indicar el primero el formato en el que éste último debe suministrar los datos. Existen varias opciones para realizar dicha negociación de contenido (Saumermann et al., 2008), (Heath y Bizer, 2011). UNESKOS implementa una negociación de contenido mediante la existencia de URIs neutras y específicas (según el formato) y el uso de reenvíos 303. Las URIs neutras se corresponden con URIs derreferenciables que apuntan a un objeto específico de un vocabulario, como en el ejemplo anterior. Para atender peticiones de un objeto en formatos específicos existen URIs diferenciadas, derivadas mediante un sufijo a partir de la URI neutra21. Los formatos disponibles son HTML, RDF/XML, RDF/Turtle, JSON y JSON-LD. Siguiendo con el ejemplo anterior de "Almacenamiento de agua" en el Tesauro de la UNESCO se tendría la siguiente tabla:

URI	Descripción
<http://skos.um.es/unescothes/C04305>	URI neutra
<http://skos.um.es/unescothes/C04305/html>	Versión HTML
<http://skos.um.es/unescothes/C04305/rdfxml>	Versión RDF/XML
<http://skos.um.es/unescothes/C04305/turtle>	Versión RDF/Turtle
<http://skos.um.es/unescothes/C04305/json>	Versión JSON
<http://skos.um.es/unescothes/C04305/jsonld>	Versión JSON-LD

Tabla 3: Esquema utilizado para la derreferenciación de URIs.

Si un cliente web solicita un objeto en un formato determinado a través de la URI neutra, el servidor de UNESKOS indica que debe acceder a otra dirección devolviendo el código de estado HTTP "303 See others" junto con la URI donde se encuentran los datos en el formato solicitado. Si el cliente se conecta directamente a una URI correspondiente a un formato concreto, recupera los datos en dicho formato.

En todo momento se han utilizado estándares abiertos con respecto a los formatos empleados. La versión HTML está optimizada para la consulta y navegación a través de un agente de

[21] Una URI proporciona un identificador genérico para un recurso web, con independencia de que exista una dirección diferente y específica para cada formato en el que se proporcione la información sobre dicho recurso.

usuario convencional. Esta versión utiliza RDFa 1.1 para el marcado semántico, de forma que es posible extraer información RDF mediante el análisis del documento HTML, permitiendo que una misma URI sirva tanto para la consulta por personas como para su procesamiento automático por una aplicación informática. El formato RDF/XML aunque resulta muy verboso puede ser reutilizado fácilmente para su conversión en otros formatos a través de XSLT. RDF/Turtle es un formato compacto, de fácil lectura por personas y procesamiento directo por máquina. JSON y JSON-LD permiten la recuperación y procesamiento de datos mediante Javascript en el lado del cliente por parte de prácticamente cualquier agente de usuario.

UNESKOS dispone de un SPARQL Endpoint[22] que permite realizar consultas en este lenguaje para seleccionar datos concretos. Se trata de un punto de acceso unificado, tanto a la Nomenclatura, como al Tesauro. Es posible realizar consultas mediante los comandos SELECT, CONSTRUCT, ASK y DESCRIBE, así como visualizar y obtener ficheros de los datos recuperados en múltiples formatos. Se piensa conservar esta infraestructura para otros conjuntos de datos que se puedan publicar en el futuro, ampliándola con nuevos conjuntos de datos y espacios de nombre correspondiente. La ilustración 1 muestra un ejemplo de consulta para recuperar las URIs y las etiquetas preferentes en Español de todos los conceptos que están definidos como conceptos cabecera de la Nomenclatura. El SPARQL Endpoint de UNESKOS está registrado en el Status Endpoint Status de Mondeca Labs[23].

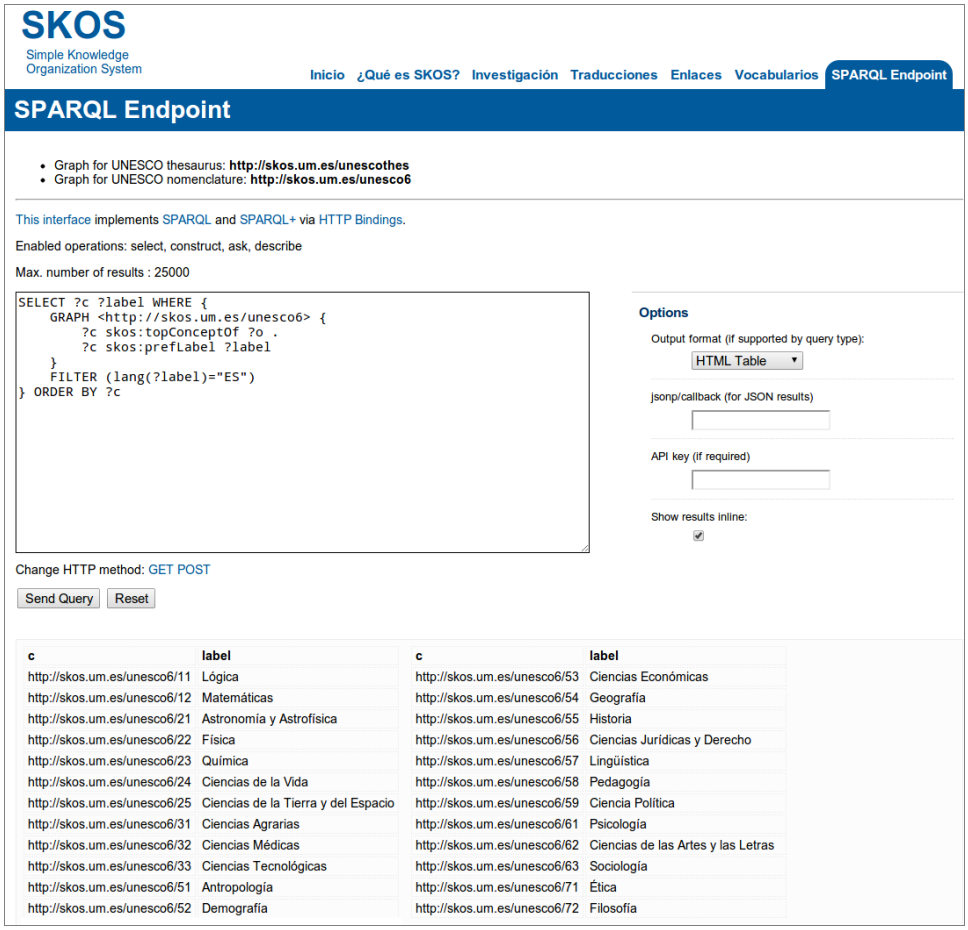

Ilustración 1: Ejemplo de consulta en el SPARQL Endpoint del proyecto UNESKOS y resultados obtenidos.

[22] Un *SPARQL Endpoint* es una *dirección web* que permite recuperar datos RDF utilizando el lenguaje de consulta SPARQL.
[23] Más información en: http://labs.mondeca.com/sparqlEndpointsStatus/

En el contexto del proyecto UNESKOS se están desarrollando actualmente una serie de tareas cuyo objetivo es la alineación entre sí de la Nomenclatura y del Tesauro, así como con otros vocabularios externos. Esto implica el uso de las relaciones semánticas de equivalencia (mapeado) con otros vocabularios como EUROVOC o AGROVOC (por ejemplo). Uno de los primeros resultados ha sido la obtención un conjunto de datos experimental a través de un proceso automático de alineación mediante Apache SOLR. El análisis de dicho conjunto de datos se está llevando a cabo actualmente. También está por definir si los resultados de estas operaciones se ofrecerán como conjuntos de datos separados o integrados en los ya existentes, con independencia de visualizarlos de forma unificada en la interfaz web de visualización y navegación. Con respecto a la licencia, en principio los conjuntos de datos pueden descargarse y utilizarse libremente. Sin embargo, la nota genérica de copyright de UNESCO[24] no permite el uso comercial de sus contenidos. Por este motivo, hemos decidido que la licencia más adecuada para el conjunto de datos era Creative Commons Atribución, No Comercial, Compartir Igual 3.0 Unported (CC BY-NC-SA 3.0)[25].

Principio Linked Open Data	**Presencia en UNESKOS**
Derreferenciación de URIs	Cada concepto, colección o esquema de conceptos tiene una URI propia para su identificación y acceso
Negociación de contenido	Actualmente soporta la negociación de contenido HTML con RDFa embebido, RDF/XML, Turtle, N3, JSON y JSON-LD directamente o con referencia mediante URL neutra.
Uso de estándares abiertos	Los formatos utilizados se basan en estándares abiertos.
Enlaces a recursos externos	Se está trabajando en la alineación tanto de la Nomenclatura como del Tesauro con otros vocabularios. De forma experimental se han definido relaciones de mapeado entre ambos conjuntos de datos.
Disponibilidad de un SPARQL Endpoint	Se ofrece un SPARQL Endopoint unificado.
Licencia libre de los datos	Actualmente los conjuntos de datos están disponibles para su descarga y uso libres. Licencia CC BY-NC-SA compatible con el copyright de la UNESCO que no permite el uso comercial sin autorización expresa.

Tabla 4: Resumen de características Linked Open Data del proyecto UNESKOS

[24] Más información en: http://www.unesco.org/new/es/terms-of-use/terms-of-use/copyright
[25] Más información en: http://creativecommons.org/licenses/by-nc-sa/3.0/deed.es

Ambos conjuntos se encuentran registrados en el catálogo *the Data Hub* en el que se han descrito los datos, servicios y formatos utilizados[26].

2.5 Arquitectura funcional y dinámica del sistema

El diseño del sistema se ha realizado siguiendo los principios indicados en la sección anterior. A través de una interfaz web cualquier usuario puede consultar el Tesauro y la Nomenclatura, navegando a través de la estructura semántica de los vocabularios o buscando en las etiquetas preferentes y alternativas. La navegación puede iniciarse mediante una búsqueda o bien consultando las áreas de conocimiento, micro-tesauros y conceptos. La dinámica de navegación entre elementos se realiza de un modo muy sencillo e intuitivo a través de enlaces de hipertexto. Se ha utilizado una presentación alfabética para la visualización de los términos y las relaciones de cada elemento del tesauro.

Ilustración 2: Ejemplo de consulta de la interfaz web del Tesauro de la UNESCO.

[26] Dichas descripciones se encuentran en http://datahub.io/en/dataset/unesco6 y en http://datahub.io/en/dataset/unescothes para la Nomenclatura y el Tesauro respectivamente.

Ilustración 3: Ejemplo de búsqueda en la interfaz web del Tesauro de la UNESCO. Puede observarse como una misma búsqueda localiza términos descriptores y no-descriptores, indicando entre paréntesis la abreviatura del idioma del término.

La redirección HTML únicamente se ha utilizado para agentes de usuario que soliciten los datos en dicho formato conectándose a la URI neutra de un elemento del vocabulario. Tal y como se ha indicado anteriormente, también es posible que otro tipo de clientes soliciten una URI neutra en un formato diferente a HTML. En tal caso, el servidor devuelve un código 303 "See others" junto con la dirección correcta a la que debe conectarse el cliente para obtener los datos en el formato solicitado.

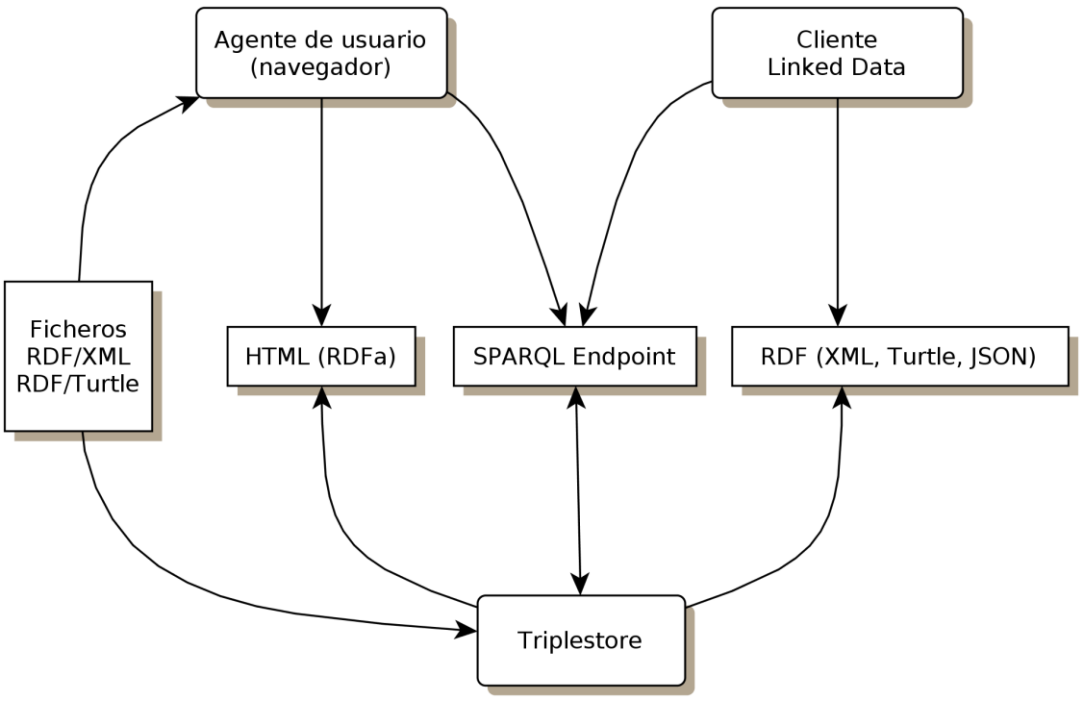

Ilustración 4: Esquema funcional de consulta y acceso a los datos del proyecto UNESKOS

Tanto el SPARQL Endpoint, como el envío de los datos en cualquier formato hacen uso directamente del triplestore utilizado. La única excepción es la descarga completa de los conjuntos de datos a través de los correspondientes ficheros RDF/XML y RDF/Turtle.

3 Discusión

Los principales aspectos de debate abiertos en torno al proyecto UNESKOS giran en torno a dos puntos fundamentales: el modelado del tesauro y la incorporación de ciertos elementos tecnológicos que permitan una mayor eficiencia, escalabilidad y actualización del sistema de acceso a los datos.

Con respecto al primer punto, el modelado, hay que indicar SKOS se mostró más que suficiente para representar la Nomenclatura, pero algo limitado para el modelado del Tesauro. Los principales problemas tienen que ver con la propia naturaleza del Tesauro de la UNESCO: su división en áreas de conocimiento y micro-tesauros no conlleva una aplicación inmediata y única del modelo de SKOS para la representación de la estructura. La ausencia de elementos específicos de SKOS para realizar esta tarea implica la inexistencia de una normalización para representar una organización basada en micro-tesauros. No es un caso único, el tesauro EUROVOC (cuya estructura de micro-tesauros es similar) utiliza una ontología ad hoc para modelar este tipo de agrupaciones[27]. Es evidente que sería preferible un enfoque más cercano a la normalización, como ISO-THES, en lugar de diseñar una nueva ontología. Esto es lo que hemos intentado durante el desarrollo del proyecto UNESKOS aunque, tras realizar el modelado del Tesauro de la UNESCO se han identificado una serie de carencias en la ontología SKOS con vistas a la publicación de KOS como Linked Open Data. Estas carencias se refieren a la capacidad de descubrir datos a partir de los enlaces que conectan los elementos del vocabulario, más concretamente: en SKOS no hay ninguna propiedad que permita modelar un enlace desde un Concepto hacia la Colección a la que pertenece.

En el caso del Tesauro de la UNESCO esto significa que únicamente a partir de los datos RDF de un determinado concepto no se puede determinar a qué micro-tesauro pertenece. El sistema diseñado en el proyecto UNESKOS obtiene esta información a partir de una consulta SPARQL. Sin embargo, entendemos que una aproximación puramente basada en Linked Open Data precisa que la relación desde un concepto hacia la colección a la que pertenece se declare explícitamente. Otro punto interesante se relaciona con la definición de los conceptos cabecera de un micro-tesauro, este aspecto resulta de gran importancia ya que los conceptos cabecera constituyen el punto de inicio para el recorrido de una estructura jerárquica. En el modelado actual, esta información se obtiene a partir de la intersección de dos conjuntos: el formado por todos los conceptos del tesauro definidos como conceptos cabecera y el formado por todos los conceptos perteneciente a una colección. Partiendo únicamente de los datos de una colección y utilizando exclusivamente SKOS es imposible conocer directamente aquellos conceptos que constituyen el inicio de la estructura jerárquica de un micro-tesauro. Al igual que en el caso anterior, este inconveniente se ha salvado mediante consultas SPARQL. Por ejemplo, la consulta para obtener los conceptos cabecera del micro-tesauro de "Política Educativa" del Tesauro de la UNESCO sería:

```
SELECT ?concepto {
  FROM <http://skos.um.es/unescothes/> {
        ?concepto skos:topConceptOf <http://skos.um.es/unescothes/CS001>.
        <http://skos.um.es/unescothes/COL110> skos:member ?concepto .
  }}
```

[27]Más información en: http://eurovoc.europa.eu/drupal/?q=ontology

Sin embargo, un software que actuara a modo de rastreador RDF y cuyo cometido fuera el descubrimiento de conceptos SKOS, basándose exclusivamente en los enlaces entre recursos, no podría obtener dicha información. Una posible solución sería el uso de un pequeño número de elementos que permitiría solucionar esta asimetría de SKOS. Esto permitiría modelar un vocabulario definiendo relaciones en ambos sentidos entre elementos de distinto tipo, algo que tampoco contempla actualmente la propuesta de ISO-THES[28]. Actualmente estamos trabajando en el modelado del Tesauro de la UNESCO utilizando como complemento de SKOS las propiedades de la Tabla 5.

Elemento	Descripción	Inverso	Dominio	Rango
uneskos:contains	Permite relacionar un esquema de conceptos con cualquier elemento de SKOS.	skos:inScheme	skos:ConceptScheme	
uneskos:isMemberOf	Permite relacionar un concepto con una colección	skos:member	skos:Concept	skos:Collection
uneskos:hasMainConcept	Identifica los conceptos que son puntos de acceso a una jerarquía de conceptos de una colección. Subpropiedad de skos:member	uneskos:mainConceptOf	skos:Collection	skos:Concept
uneskos:mainConceptOf	Relaciona un concepto a una colección como punto de acceso a una jerarquía. Subpropiedad de uneskos:memberOf	uneskos:hasMainConcept	skos:Concept	skos:Collection

Tabla 5: Elementos propuestos por el proyecto UNESKOS para la ampliación de SKOS. Se ha utilizado "uneskos" como prefijo de dichos elementos.

Con respecto a la plataforma, se está trabajando en la incorporación de un buscador basado en Apache SOLR que solucione determinados inconvenientes que plantea el uso de variantes ortográficas y léxicas de los términos utilizados en la consulta. Esto también permitiría optimizar los procesos de búsqueda en las etiquetas preferentes y no preferentes, que actualmente se realizan mediante consultas SPARQL relativamente complejas.

Conclusiones y futuros trabajos

Los resultados de modelado e implementación de la Nomenclatura Internacional Normalizada de Ciencia y Tecnología y del Tesauro de la UNESCO han resultado altamente satisfactorios a nivel general. La ontología SKOS se ha mostrado adecuada en su mayor parte para satisfacer los objetivos del trabajo aunque los principios de publicación Linked Open Data exigen una revisión de ciertos

[28] Más información en: http://www.niso.org/lists/25964info/archive/subject?list_name=25964info&monthdir=201302

aspectos de SKOS, principios que la propuesta de integración con ISO-25964 (ISO-THES) tampoco tiene en cuenta.

Por lo tanto, para futuros trabajos en el seno del proyecto UNESKOS se deberá llevar a cabo un modelado más adecuado del Tesauro de la UNESCO para su reutilización como Linked Open Data. Se utilizarán para ello los elementos apuntados anteriormente en la discusión de resultados de manera que se pueda modelar adecuadamente la estructura de áreas de conocimiento y micro-tesauros de este vocabulario. Uno de los retos de dicho modelado se centrará en introducir el mínimo de cambios imprescindibles en relación al conjunto de datos actual.

La alineación de vocabularios constituye un campo muy interesante, en especial mediante la aplicación de procesos automatizados basados en técnicas avanzadas de recuperación de información, tales como los vectores semánticos, clasificación mediante VSM o indización aleatoria. Estas técnicas, que permiten operar simultáneamente con conceptos y términos, se están demostrando muy eficacaces en el desarrollo de nuevas técnicas de recuperación de información.

La integración entre servicios web basados en SPARQL Endpoint y herramientas como Apache SOLR[29] resulta de gran interés, no solamente en la tarea de alineación de vocabularios sino también para simplificar los procesos de búsqueda, mejorando los resultados y como un medio de reutilización inmediato de conjuntos de datos RDF.

Finalmente, indicar que con la disponibilidad de ambos vocabularios, en especial de la Nomenclatura, se abren nuevas posibilidades para la publicación e integración de otros conjuntos de datos relacionados en los que se utiliza este vocabulario para la descripción del contenido de recursos relacionados con el ámbito académico y de investigación.

Referencias

AITCHISON, J., & CLARKE, S. D. (2004). The thesaurus: a historical viewpoint, with a look to the future. *Cataloging & Classification Quarterly*, 37(3-4), 5-21.

BERNERS-LEE, T. (2006). *Linked Data: Design Issues*. Recuperado el 2 de junio de 2010, de <http://www.w3.org/DesignIssues/LinkedData.html>.

DUNSIRE, G. (2011). Enhancing Information Services Using Machine-to-Machine Terminology Services. *Subject Access: Preparing for the Future*, *42*, 111.

EWKETU, M. (2011). *The UNESCO Thesaurus*. UN-LINKS Meeting, 28-30 Nov. Recuperado el 1 de mayo de 2013 de <http://www.unesco.org/library/PDF/The%20UNESCO%20Thesaurus.pdf>.

GARCÍA, N. E., JAROSZCZUK, S. E., & DE BIBLIOTECOLOGÍA, C. (2009). Objetos digitales: una experiencia de representación con metadatos Dublin Core. *I Encuentro Nacional de Catalogadores: experiencias en la organización y tratamiento de la información en bibliotecas argentinas*. Buenos Aires: Biblioteca Nacional, 193-206.

GARCÍA-TORRES, A., PAREJA-LORA, A. & PRADANA-LÓPEZ, D. (2008). Reutilización de tesauros: el documentalista frente al reto de la web semántica. *El profesional de la información*, 17(1), 8-21.

[29] Apache Solr es una plataforma de código abierto para el desarrollo de motores de búsqueda web que está basado en el software de recuperación de información Apache Lucene. Más información en: http://lucene.apache.org/Solr/

HEATH, T. & BIZER, C, (2001). Linked Data: Evolving the Web into a Global Data Space. *Synthesis Lectures on the Semantic Web: Theory and Technology,* 1(1), 1-136. Recuperado el 1 de mayo de 2013 de <http://linkeddatabook.com/book>.

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

ISO/DIS 25964-1:2013 (2013). *Thesauri and interoperability with other vocabularies*. Part 2: Interoperability with other vocabularies. ISO, 2013.

SAUERMANN, L. & CYGANIAK, R. (2011*). Cool URIs for the semantic web. W3C Interest Group Note* 03 December 2008. Recuperado el 1 de mayo de 2013 de <http://www.w3.org/TR/cooluris/>.

HASLHOFER, B. & SCHANDL, B. (2010). Interweaving OAI-PMH data sources with the linked data cloud. *International Journal of Metadata, Semantics and Ontologies archive*, 5(1), 17-31. Recuperado el 1 de mayo de 2013 de: <http://dx.doi.org/10.1504/IJMSO.2010.032648>.

ISO TC46/SC9/WG8 & Isaac, A. (2012). *Correspondence between ISO 25964 and SKOS/SKOS-XL Models.* Recuperado el 1 de mayo de 2013 de <http://www.niso.org/schemas/iso25964/correspondencesSKOS/>.

MARTÍNEZ-FRÍAS, J. & HOCHBERG, D. (2007). Classifying science and technology: Two problems with the UNESCO system. *Interdisciplinary Science Reviews*, 32(4), 315-319.

MILES, A., & BECHHOFER, S. (2009). *SKOS simple knowledge organization system reference*. W3C Recommendation 18 August 2009. Recuperado el 1 de mayo de 2013 de <http://www.w3.org/TR/skos-reference>.

MONTALVO-MONTALVO, M. (2011). LCSH, FAST y DELICIOUS: vocabularios normalizados y nuevas formas de catalogación temática. *Anales de Documentación,* 14(1). Disponible en línea, recuperado el 1 de mayo de 2013, de: <http://revistas.um.es/analesdoc/article/view/120141>.

PASTOR-SÁNCHEZ, J. A., MARTÍNEZ-MÉNDEZ, F. J., & RODRÍGUEZ-MUÑOZ, J. V. (2012). Aplicación de SKOS para la interoperabilidad de vocabularios controlados en el entorno de linked open data. *El profesional de la información*, 21(3), 245-253.

PASTOR SÁNCHEZ, J. A., MARTÍNEZ MÉNDEZ, F. J., & RODRÍGUEZ MUÑOZ, J. V. (2009). Advantages of thesaurus representation using the Simple Knowledge Organization System (SKOS) compared with proposed alternatives.*Information Research*, 14(4), paper 422. Disponible en: <http://InformationR.net/ir/14-4/paper422.html>.

PASTOR SÁNCHEZ, J. A. (2013). ISO-THES: Ampliando SKOS a partir de la norma de tesauros ISO-25964. *Anuario ThinkEPI*, 7, 189-193.

SCHRENK, M. (2012). *Webbots, spiders, and screen scrapers: A guide to developing Internet agents with PHP/CURL*. San Francisco, No Starch Press, 63-75.