

SKOS y la Gestión de Vocabularios para la Interoperabilidad Semántica

Miona Dimic, Cecilio Angulo Bahón
IDEAI-UPC, Universitat Politècnica de Catalunya

Introducción

En muchos sectores industriales, la terminología suele documentarse en forma de **vocabularios controlados**. Un vocabulario controlado se define como “una lista prescriptiva de términos, encabezamientos o códigos, cada uno de los cuales representa un concepto” [1]. Un ejemplo de vocabulario controlado es cualquier catálogo de términos estandarizados [2]:

Código INE	Provincia
01	Araba/Álava
02	Albacete
03	Alicante/Alacant
04	Almería
...	...

Fuente: Instituto Nacional de Estadística (INE), Lista estándar de provincias.
https://www.ine.es/daco/daco42/codmun/cod_provincia_estandar.htm

Cuando se introducen niveles jerárquicos entre los términos, se obtiene lo que se llama una **taxonomía**, que permite identificar claramente categorías y subcategorías. Además, si vamos más allá de las relaciones jerárquicas (términos más amplios y más específicos) y añadimos relaciones asociativas (términos relacionados), una taxonomía evoluciona hacia un **tesauro** [2]. Los tesauros suelen incluir definiciones de los términos, sinónimos, términos equivalentes en otros idiomas, etc.

AGROVOC [3] es un ejemplo de tesauro que cubre terminología en las áreas de interés de la FAO (Organización de las Naciones Unidas para la Alimentación y la Agricultura) [4]. Se puede buscar por términos específicos, explorar de forma jerárquica o acceder mediante servicios web. Por ejemplo, el término “*mercurio*” y su descripción en AGROVOC pueden consultarse en http://aims.fao.org/aos/agrovoc/c_4751 (o mediante buscador: https://agrovoc.fao.org/browse/agrovoc/en/page/c_4751).

En este caso se observa una jerarquía clara con términos más amplios, definiciones en varios idiomas y una lista de equivalencias exactas con conceptos presentes en otros vocabularios. Esto es muy importante porque apoya la interoperabilidad entre diferentes vocabularios, permitiendo pasar rápidamente de uno a otro y estar seguros de que se está hablando del mismo concepto. En el ejemplo de *mercurio*, en la lista de términos exactamente equivalentes aparece una referencia al tesauro NALT (NAL Agricultural Thesaurus) [5], que también define el *mercurio* como un elemento químico: <https://lod.nal.usda.gov/nalt/44818>

Sin embargo, es importante entender que disponer de la misma etiqueta no siempre significa que el concepto etiquetado sea el mismo. Esto se puede encontrar, por ejemplo, en el vocabulario UAT (Unified Astronomy Thesaurus) [6], que define Mercurio como un planeta (<http://astrothesaurus.org/uat/1024>), y no como un elemento

químico, como ocurre en AGROVOC y NALT.

Esta ambigüedad en el etiqueta conduce a una pregunta importante: si nuestro caso de uso encaja mejor con el concepto definido en un cierto vocabulario específico y cómo podemos reutilizarlo en nuestro sistema indicando claramente a qué vocabulario original pertenece.

La respuesta está en el uso de un **URI (Uniform Resource Identifier)**: un identificador único global que distingue un concepto de cualquier otro en la web.

Si observamos los tres vocabularios citados en el ejemplo y sus respectivos términos para *mercurio*, cada uno tiene su propia dirección web:

- http://aims.fao.org/aos/agrovoc/c_4751
- <https://lod.nal.usda.gov/nalt/44818>
- <http://astrothesaurus.org/uat/1024>

Estas direcciones web corresponden exactamente a los URIs, actuando como identificadores únicos que permiten referirse sin ambigüedad al concepto de cada vocabulario.

¿Por qué es tan potente el uso de identificadores? Porque se puede acceder a mucha información simplemente usando uno de estos URI. Un URI no solo nombra un concepto, sino que lo identifica de forma clara y proporciona información estructurada, como definiciones, sinónimos, relaciones jerárquicas, enlaces a otros vocabularios, traducciones y mucho más.

Todo esto es posible porque estos conceptos suelen expresarse usando **SKOS** (Simple Knowledge Organization System) [7], un estándar del W3C (World Wide Web Consortium) [8] para representar vocabularios controlados de forma legible por máquinas, lo que permite una interpretación consistente y la interoperabilidad entre sistemas.

Representación de vocabularios controlados con SKOS

SKOS es un modelo de datos basado en RDF (Resource Description Framework) [9] para representar tesauros, taxonomías, esquemas de clasificación y sistemas de encabezados de materias en la web [10]. Se identifica mediante el namespace URI: <http://www.w3.org/2004/02/skos/core#> donde se puede encontrar su definición completa.

SKOS proporciona un conjunto de URIs que se crean añadiendo un nombre local (por ejemplo, `prefLabel`) a este namespace. Por ejemplo:

<http://www.w3.org/2004/02/skos/core#Concept> representa un Concepto (en inglés: Concept) y <http://www.w3.org/2004/02/skos/core#ConceptScheme> representa un Esquema de Conceptos (en inglés: Concept Scheme) en SKOS.

En adelante, estos URI se escribirán como: `skos:Concept` y `skos:ConceptScheme` asumiendo que el namespace ha sido asignado al prefijo *skos*.

skos:Concept y skos:ConceptScheme son elementos clave para representar vocabularios controlados. Un skos:Concept puede entenderse como un término, idea o noción [7]. Los conceptos pueden existir por sí solos, pero normalmente se organizan dentro de un skos:ConceptScheme, que define el alcance y los límites de los conceptos dentro de un dominio. Cada concepto se vincula a su esquema mediante la propiedad skos:inScheme.

Para entenderlo mejor, consideremos el vocabulario NALT y su definición del concepto *mercurio*, representado con SKOS: <https://lod.nal.usda.gov/nalt/44818>

```
@prefix agrovoc: <http://aims.fao.org/aos/agrovoc/> .
@prefix cabt: <http://id.cabi.org/cabt/> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix gacs: <http://id.agrisemantics.org/gacs/> .
@prefix lcsh: <http://id.loc.gov/authorities/subjects/> .
@prefix nalt: <https://lod.nal.usda.gov/nalt/> .
@prefix naltv: <https://lod.nal.usda.gov/naltv#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

<https://lod.nal.usda.gov/nalt> a skos:ConceptScheme .

nalt:44818 a skos:Concept,
    naltv:Chemical ;
    dcterms:created "2006-01-19"^^xsd:date ;
    dcterms:modified "2017-10-16"^^xsd:date ;
    skos:broader nalt:3195 ;
    skos:exactMatch agrovoc:c_4751,
        gacs:C3980,
        cabt:71993,
        lcsh:sh85083794 ;
    skos:inScheme <https://lod.nal.usda.gov/nalt>,
        <https://lod.nal.usda.gov/nalt-core> ;
    skos:prefLabel "mercury"@en,
        "mercurio"@es ;
    skos:related nalt:10454,
        nalt:268,
        nalt:310130 ;
    naltv:marc001 "23030" .
```

Se puede ver que NALT define su skos:ConceptScheme de forma única asignándole el URI <https://lod.nal.usda.gov/nalt/> . De la misma manera, un skos:Concept definido para el término *mercurio* tiene el URI <https://lod.nal.usda.gov/nalt/44818> , que se representa como nalt:44818 usando el prefijo *nalt*.

La pertenencia de este concepto a su esquema se declara mediante skos:inScheme, y la conexión con conceptos equivalentes en otros vocabularios se establece con skos:exactMatch. Por ejemplo, agrovoc:c_4751 (http://aims.fao.org/aos/agrovoc/c_4751) es el URI que identifica *mercurio* en

AGROVOC. Este ejemplo también muestra cómo SKOS permite conectar conceptos relacionados mediante `skos:related`, organizar la jerarquía con `skos:broader` y proporcionar etiquetas en español e inglés mediante `skos:prefLabel`. Además, se pueden utilizar otras propiedades, como definiciones, notas de alcance y etiquetas alternativas, para aportar información semántica adicional. Para ver más posibilidades, consulta el SKOS Primer [10].

Ventajas de modelar con SKOS

Modelar vocabularios con SKOS ofrece varias ventajas prácticas para organizar el conocimiento y lograr interoperabilidad. Cada concepto se identifica con un URI, que funciona como un identificador único global. Esto asegura que un concepto pueda ser referenciado de forma clara entre sistemas, incluso cuando diferentes vocabularios usan las mismas etiquetas. Al apuntar a un URI específico, un sistema puede indicar claramente de qué definición de concepto depende y evitar confusiones de significado.

Los conceptos también se convierten en **recursos accesibles digitalmente**. Como los vocabularios SKOS generalmente se publican como Linked Data, cada URI se puede consultar a través de servicios web o APIs para obtener información estructurada como etiquetas, definiciones, relaciones jerárquicas, traducciones y enlaces a conceptos equivalentes en otros vocabularios.

SKOS facilita además la interoperabilidad entre vocabularios. Propiedades como `skos:exactMatch`, `skos:broader`, `skos:narrower` y `skos:related` permiten conectar conceptos entre vocabularios, haciendo posible navegar e integrar conocimiento de distintos sistemas terminológicos.

Además, SKOS puede mejorar la recuperación de información para sistemas de **Inteligencia Artificial** (IA). Los sistemas que usan generación aumentada por recuperación (RAG) pueden recorrer grafos de conocimiento para obtener contexto relevante ante preguntas complejas. Al basarse en conceptos claramente definidos con URIs y relaciones semánticas explícitas, estos sistemas pueden ir más allá de la similitud vectorial y recuperar información según el significado exacto de los conceptos, aumentando la precisión y confiabilidad.

En el contexto de los **Espacios de Datos**, SKOS se recomienda comúnmente como estándar para modelar vocabularios. Tanto el Blueprint DSSC [11] como el Marco de Interoperabilidad Técnica del Centro de Referencia de Espacios de Datos [12] destacan a SKOS como un enfoque preferido para lograr **interoperabilidad semántica**.

Referencias

- [1] Asociación Española de Normalización (UNE), "UNE-ISO 25964-2:2016: Información y documentación. Tesauros e interoperabilidad con otros vocabularios. Parte 2: Interoperabilidad con otros vocabularios," 2016. [En línea]. Disponible: <https://www.une.org/encuentra-tu-norma/busca-tu-norma/norma?c=N0057671>
- [2] J. Talisman, "The Ontology Pipeline." [En línea]. Disponible: <https://jessicatalisman.substack.com/p/the-ontology-pipeline>
- [3] Subirats-Coll, K. Kolshus, A. Turbati, A. Stellato, E. Mietzsch, D. Martini, and M. Zeng, "AGROVOC: The linked data concept hub for food and agriculture," Computers and Electronics in Agriculture, vol. 196, p. 105965, 2022, doi: <https://doi.org/10.1016/j.compag.2020.105965> .
- [4] Food and Agriculture Organization of the United Nations (FAO), "FAO." [En línea]. Disponible: <https://www.fao.org/>
- [5] U.S. Department of Agriculture, National Agricultural Library, "NAL Agricultural Thesaurus (NALT)." [En línea]. Disponible: <https://lod.nal.usda.gov/nalt>
- [6] Unified Astronomy Thesaurus (UAT)."UAT." [En línea]. Disponible: <https://astrothesaurus.org/>
- [7] A. Miles y S. Bechhofer, "SKOS Simple Knowledge Organization System Reference," W3C Recommendation. [En línea]. Disponible: <https://www.w3.org/TR/skos-reference/>
- [8] W3C, "World Wide Web Consortium (W3C)". [En línea]. Disponible: <https://www.w3.org/>
- [9] "RDF - Semantic Web Standards,". [En línea]. Disponible: <https://www.w3.org/RDF/>
- [10] A. Isaac y E. Summers, SKOS Simple Knowledge Organization System Primer, W3C Working Group Note, 18 August 2009. [En línea]. Disponible: <https://www.w3.org/TR/2009/NOTE-skos-primer-20090818/>
- [11] Data Spaces Support Centre (DSSC), "Blueprint 3.0" [En línea]. Disponible: <https://blueprint.dssc.eu/>
- [12] Gobierno de España, "Marco de Interoperabilidad Técnico," Centro de Referencia de Espacios de Datos (CRED). 2025. [En línea]. Disponible: <https://cred.digital.gob.es/content/dam/cred/img/docs/MarcoInteroperabilidadTecnico.pdf>