
El uso de vocabularios controlados en los sistemas de información jurídica: evolución y tendencias actuales de representación

Use of controlled vocabularies in legal information systems: evolution and current trends

María Luisa ALVITE DíEZ

Facultad de Filosofía y Letras, Universidad de León (España)
luisa.alvite@unileon.es

Resumen

La semántica documental ha conformado tradicionalmente un pilar esencial en la organización y recuperación de la información. La irrupción de la Web y su evolución hacia una Web Semántica y de datos enlazados exige la adaptación de estas herramientas y vocabularios documentales hacia nuevos estándares de representación. Se indaga en el estado actual de la gestión de los procesos documentales y el desarrollo de las técnicas semánticas implementadas en los diarios oficiales españoles. Se observan esfuerzos claros por llevar a cabo un análisis documental del contenido normativo, resultando relevante la tendencia hacia el empleo del tesoro EuroVoc o adaptaciones de EuroVoc.

Palabras clave: Estándares semánticos. Lenguajes documentales. Representación de la información. Sistemas de información jurídica.

Abstract

Semantics has had an essential role in the organization and retrieval of information. The emergence of the Web and its evolution towards a Semantic Web and linked data requires the adaptation of these tools and vocabularies to new standards of representation. The current state of document process management and development of semantic techniques implemented in the Spanish official journals are explored. There are efforts to carry out a documentary analysis of legal contents. The significant employment of the EuroVoc thesaurus is highlighted.

Keywords: Indexing languages. Information representation. Legal information systems. Semantic standards.

1. El papel de los lenguajes documentales en la organización y recuperación de la información jurídica

El ámbito jurídico ha sido pionero en la creación de sistemas de información en línea para el acceso al texto completo de documentos legales. El primer sistema implementado para la recuperación de una gran colección legislativa fue presentado en el año 1960 por el *Health Law Center* de la Universidad de Pittsburg y utilizaba ya una especie de tesoro concebido como ayuda en el proceso de búsqueda. Lancaster (1995, p. 184) se refiere a este tesoro como una simple compilación de palabras con significados similares, con una estructura semejante a la del *Roget's Thesaurus* que, sin embargo, suponía una estimable ayuda en la búsqueda evitando al usuario pensar en todas las palabras que pudieran expresar una idea concreta. Se trataba, por tanto, de un tesoro simple que en palabras de Lancaster *carecía de un significativo grado de "estructura"*.

Desde entonces, se han llevado a cabo innumerables avances en líneas de trabajo que consi-

deramos complementarias y que han conducido a la necesaria evolución de los sistemas jurídicos: uso del hipertexto, procesamiento en lenguaje natural, indización automática, inteligencia artificial, aplicaciones derivadas de los distintos lenguajes de marcado y, de modo especialmente relevante, implementaciones de XML (eXtensible Markup Language) en el tratamiento y recuperación semántica de la información legal.

En nuestro país las bases de datos jurídicas cuentan, generalmente, con instrumentos de control terminológico, en muchos casos asociados a un campo que recibe la denominación de "Voz" o "Voces". Ya los repertorios de textos legales del siglo XIX recurrían a índices de "voces" para organizar el complejo lenguaje tratado en la documentación jurídica, esto es, listas autorizadas de términos preferentes que hacen referencia a un marginal que conduce al documento.

Los sistemas tratan de adaptarse al modo de trabajo de los profesionales jurídicos y evitar las dificultades de la búsqueda en lenguaje natural sobre documentos de la extensión considerable

que alcanzan muchos de los registros normativos o jurisprudenciales almacenados en este tipo de bases de datos. Estas series de entradas autorizadas se presentan hoy de modo muy distinto a como lo hacían las obras impresas; así, listas desplegables, opciones de autocompletado o nubes de etiquetas, trasladando el modelo exitoso de la web social, son algunas de las opciones que aparecen en las interfaces de las actuales bases de datos.

Los sistemas jurídicos disponen de varios campos para acceder a información temática y, no obstante, la gran mayoría cuenta con un campo distintivo que sirve de punto de acceso específico al contenido mediante el uso de vocabulario controlado. Este lenguaje controlado elaborado para la indización y recuperación de la documentación jurídica se enfrenta a la dificultad que supone confeccionar un léxico que incluya lenguaje común, lenguaje especializado o técnico de las distintas ramas de conocimiento y lenguaje estrictamente jurídico.

En un estudio previo (Alvite, 2004) en el que abordamos el análisis de instrumentos terminológicos integrados en las interfaces de bases de datos jurídicas comerciales españolas, se observaron diferencias sustanciales en las herramientas empleadas en las bases de datos legislativas y en las jurisprudenciales. Se evidenció igualmente una escasa normalización de la terminología jurídica en los entornos objeto de estudio, así como la priorización de criterios apegados a la práctica jurídica frente al empleo de herramientas consistentes para el control del vocabulario.

Existe abundante literatura científica acerca de las ventajas e inconvenientes que conlleva el uso del lenguaje libre -lenguaje natural- y el controlado, asignado a partir de un lenguaje documental. La combinación de ambos, elegida por muchos sistemas de información, sirve para neutralizar las deficiencias de estos modelos. De un lado, el lenguaje natural permite una mayor especificidad en la recuperación y el sistema contará con un vocabulario ilimitado y completamente actualizado. De otro, los vocabularios controlados reducen la ambigüedad semántica, mejoran la consistencia en la representación de un concepto dado, mediante el control de los sinónimos o cuasi sinónimos, y favorecen la realización de búsquedas más amplias gracias a la estructura que enlaza los términos relacionados semánticamente.

Consideramos que una indización intelectual consistente garantiza mayor eficacia en la recuperación de la información y resulta un valor añadido en la identificación de los documentos

más relevantes en detrimento de los menos relevantes cuando el usuario efectúa búsquedas conceptuales. Nos unimos a la amplia lista de trabajos (Aitchison, Bawden y Gilchrist, 2000; Gil Leiva, 2008; Gil Urdiciain, 1998) que estiman una clara complementariedad entre el lenguaje libre y el controlado, dado que el lenguaje controlado contrarresta las carencias del lenguaje libre y viceversa.

2. La semántica documental

Aludimos en este epígrafe al área de trabajo dedicada a la representación de documentos, tanto a la identificación normalizada de las características externas de los documentos como a la representación de su contenido.

Coincidimos con Codina y Pedraza (2011) en que la semántica documental forma parte del ADN de las ciencias de la documentación como pocas otras cosas. Esta dedicación, en nuestra opinión, no puede desligarse en modo alguno del gran esfuerzo normalizador llevado a cabo a nivel internacional en este campo por parte de la comunidad bibliotecaria.

Tradicionalmente la semántica documental ha venido distinguiendo dos grupos de herramientas: los modelos de metadatos -dirigidos a identificar el objeto documental- y los lenguajes documentales o lenguajes de indización que facilitan la representación formalizada del contenido de los documentos.

Así, en el caso de los sistemas de información legal, dejando a un lado el análisis jurídico imprescindible, una disposición normativa en una base de datos legislativa se representa, habitualmente, mediante dos conjuntos de datos: un grupo de metadatos que incluye componentes como el título, el rango, el órgano emisor, la fecha de publicación, el número de la disposición, etc., más un grupo de palabras clave o descriptores extraídos de una taxonomía o de un tesoro que representan los conceptos tratados en la disposición y permiten, por tanto, un acceso temático a la información.

Atendiendo a Gil Urdiciain (2004), el lenguaje documental es un sistema artificial de signos normalizados que facilita la representación formalizada del contenido de los documentos para permitir la recuperación de información solicitada por los usuarios. Tiene que ser unívoco y estar dotado de una sencilla organización. Subrayamos en este punto la imbricación indisoluble entre recuperación de información y organización o representación del conocimiento.

Si bien existen diversos criterios de sistematización de los lenguajes documentales, tal y como

recoge detalladamente Rodríguez Bravo (2011), son lenguajes controlados principalmente, las clasificaciones, taxonomías, las listas de encabezamientos de materia y los tesauros. Presentan un vocabulario previamente elaborado y admiten un limitado número de modificaciones en el momento de su utilización.

Las clasificaciones son un conjunto ordenado de conceptos que se presentan distribuidos sistemáticamente en clases conformando una estructura (Gil Urdiciain, 2004). Se trata de una estructura jerárquica formada por clases y subclases sin incluir relaciones formales entre los términos. Cada concepto se halla representado por una notación. Existen clasificaciones enciclopédicas como la *Clasificación Decimal Universal (CDU)* o la *Clasificación de la Biblioteca del Congreso (LCC)* y clasificaciones especializadas como la *Clasificación Internacional de Patentes (CIP)*.

Las taxonomías son una categorización de descriptores de forma jerárquica que establece una relación de los objetos de generalización-especialización. Consta de una estructura clasificatoria jerárquica con determinados niveles de profundidad, etiquetas que dan nombre a los conceptos y se agrupan en bloques temáticos y, además, admiten facetas no jerárquicas con lo que obtiene dinamismo y pluralidad de enfoques en los asuntos (Moreiro, 2009). Para Codina y Pedraza (2011) en rigor, una taxonomía es una clasificación. Añaden además, que por una deriva semántica que posiblemente obedece a influencias de prestigio del término, ya que procede de la biología, es utilizado como sinónimo de lenguaje controlado en general, sobre todo en determinadas áreas como en arquitectura de la información y también en sistemas de información no bibliotecarios como los portales corporativos. Las clasificaciones tienen su origen en la comunidad bibliotecaria, presentan relaciones entre su vocabulario más ricas que las de las taxonomías, son empleadas para clasificar y ubicar colecciones de temáticas definidas de antemano y se vinculan, fundamentalmente a la documentación tangible. Por su parte, las taxonomías se crean para categorizar recursos digitales con vistas a su navegación, prestando más atención a los usuarios que a los contenidos.

Las listas de encabezamientos de materia son un tipo de lenguaje precoordinado de estructura asociativa o combinatoria que consta de una relación alfabética de conceptos que sirven para representar los temas de los que trata un documento. Se formaliza por medio de encabezamientos y subencabezamientos que se combinan en el momento de la indización para expresar el tema del documento de forma encadena-

da. Se utilizan principalmente en los catálogos automatizados. Son sistemas más rígidos que los tesauros y, como señala Moreiro (2004), manifiestan confusión entre las relaciones jerárquicas y asociativas, unas relaciones semánticamente más superficiales que las que existen en los tesauros. Entre los ejemplos de Listas de materias, podemos citar los *Encabezamientos de materia de la Biblioteca del Congreso (LCSH)*, *Lista de encabezamientos de materia de la red de bibliotecas del Consejo Superior de Investigaciones Científicas (CSIC)*, *Lista de encabezamientos de la Biblioteca Universitaria de Sevilla*, etc.

Respecto a los tesauros, la Norma ISO 2788-1986 equivalente a la española UNE 50106-1990 los entiende desde dos perspectivas. Según su función, la norma define tesoro como *un instrumento de control terminológico que traduce a un lenguaje documental el lenguaje natural empleado en los documentos y por los usuarios*. Atendiendo a su estructura, el tesoro es *un vocabulario controlado y dinámico de términos relacionados semántica y jerárquicamente, que se aplica a un campo específico de conocimiento*. Un tesoro es un lenguaje postcoordinado de estructura combinatoria formado por un conjunto de términos preferidos (descriptores) y no preferidos (no descriptores). Los términos se vinculan entre sí para establecer relaciones de equivalencia, jerarquía y asociación.

Los sistemas de información jurídicos, como ya hemos señalado en ocasiones anteriores (Alvite, 2004), incorporan distintos modelos de herramientas terminológicas escasamente normalizadas en el sentido documental, mostrando una clara preferencia por las listas de palabras clave, listas autorizadas de descriptores y tesauros, que podemos calificar de asistemáticos en su construcción y alejados de los principios normalizadores en el caso de los sistemas comerciales disponibles. Las tendencias más recientes se dirigen a soluciones automatizadas de expansión semántica y sugerencias automáticas en las que se tiene en cuenta la relevancia, atendiendo tanto a los contenidos como al rango de la disposición normativa o del tribunal, en el caso de las bases de datos de jurisprudencia.

3. Evolución en los sistemas de organización del conocimiento: nuevos modelos de representación

Los orígenes del concepto moderno de lenguaje documental se retrotraen, según López Huertas (1991), a finales del siglo XIX con las aportaciones de Charles Cutter y Melvin Dewey. El mo-

vimiento innovador surge en Estados Unidos donde ven la luz las primeras obras teóricas sobre lenguajes documentales en su sentido actual. A lo largo de la primera mitad del siglo XX aparecen nuevos sistemas de clasificación bibliográfica, nuevas listas de encabezamientos de materia y los primeros lenguajes documentales especializados. Tras la segunda guerra mundial, como apunta Rodríguez Bravo (2011), los sistemas utilizados para la gestión y recuperación de información, en su mayoría enciclopédicos, comienzan a dar muestras de ineficacia para el tratamiento de las publicaciones en centros especializados y se buscan formas alternativas de expresión en la lingüística documental.

Aparece así un nuevo lenguaje documental, el tesoro. Sostiene Moreiro (2007) que la creciente especialización del conocimiento, hacia la mitad del siglo XX, empezó a impulsar la idea del concepto como núcleo organizador de los lenguajes documentales. Se impuso a principios de los años sesenta del pasado siglo la idea de descriptor asociada a la de concepto. Los tesauros heredan los planteamientos determinados por la aparición en 1852 del *Thesaurus of English Words and Phrases* de Roget, un tesoro lingüístico que, con respecto a los diccionarios comunes, contextualizaba el sentido en que cada término debía ser comprendido. Paulatinamente el tesoro fue perfeccionándose y, como puntualiza el propio Moreiro, durante la década de los setenta, ante la aparición de las bases de datos, el tesoro pasó a tener un gran protagonismo en la recuperación de la información.

Los lenguajes documentales han demostrado su eficacia durante décadas, pero el contexto actual derivado del desarrollo de Internet y de la Web como plataforma de edición electrónica universal exige nuevos modelos de representación de la información.

Las limitaciones de la World Wide Web, como subraya García Marco (2009), tanto en lo que se refiere a la implementación del hipertexto como en su capacidad de gestión de datos son precisamente el trampolín del que parte la Web semántica. Una Web entendida como un espacio compartido para el intercambio de datos altamente estructurados. Ahora bien, como el propio García Marco señala, la Web semántica aporta la construcción de niveles de sucesiva abstracción y control: el documento estructurado (XML), la definición del esquema documental (DTD, XML Schema), el nivel de relaciones entre metadatos (RDF Schema) y, por último, el modelado de dominios de conocimientos (OWL).

La evolución actual de las herramientas terminológicas para la organización del conocimiento ha de entenderse dentro de este marco contextual de la Web semántica, un ámbito que aglutina un conjunto de tecnologías con un grado de abstracción cada vez mayor y en cuya evolución, como apunta Pastor Sánchez (2011, p. 42) “son fundamentales las constantes sinergias entre cada uno de los niveles, producto de la aplicación y evaluación de las diferentes tecnologías”.

En primer lugar, en lo que atañe a la estructuración del documento, hemos de hablar del triunfo rotundo del marcado en XML, un modelo de datos no propietario, independiente de plataforma y que garantiza la compatibilidad entre sistemas. En el caso que nos ocupa en este trabajo, los documentos legales por su carácter fuertemente estructurado y por las complejas conexiones e interrelaciones que requieren expresar, se adaptan de modo inmejorable a esta tecnología (Andrés Rivero y Hernández Marín, 2009; Martínez, Fuente y Derniame, 2003)

En segundo lugar, el nivel de control asentado en los esquemas de metadatos y en RDF (Resource Description Framework), metamodelo de metadatos, cuenta con un más que notable desarrollo en el ámbito de las colecciones digitales. Formarían parte de los esquemas de metadatos tanto las tradicionales reglas de catalogación —Anglo-American Cataloguing Rules (AACR2), Resource Description and Access (RDA)— como los sistemas más recientes: Dublin Core, MARCXML, EAD (Encoded Archival Description), VRA Core, MODS (Metadata Object Description Standard), METS (Metadata Encoding & Transmission Standard), MIX (NISO Metadata for Images in XML), PREMIS o ESE (Europeana Semantic Elements) son algunos de los más ampliamente extendidos.

En este punto, resulta imprescindible introducir el concepto de interoperabilidad, entendido de modo genérico como el compromiso para maximizar la posibilidad de intercambio y reutilización de la información. En el caso español, el Esquema Nacional de Interoperabilidad se establece en el apartado 1 del artículo 42 de la Ley 11/2007, de 22 de junio, de acceso electrónico de los ciudadanos a los Servicios Públicos (BOE núm. 150 de 23/6/2007). Su finalidad es la creación de las condiciones necesarias para garantizar el adecuado nivel de interoperabilidad técnica, semántica y organizativa de los sistemas y aplicaciones empleados por las Administraciones públicas, que permitan el ejercicio de derechos y el cumplimiento de deberes a través del acceso electrónico a los servicios públicos, a la vez que redundan en beneficio de la eficacia y la eficiencia.

Pues bien, uno de los cimientos básicos en el marco de las colecciones digitales para lograr esta interoperabilidad entre los sistemas reside en los metadatos. Se trata de “datos sobre datos”, aunque como señala acertadamente Pastor Sánchez (2011, p. 21) sería más correcto afirmar que los metadatos son elementos que describen un determinado objeto siguiendo algún modelo o conjunto de reglas. Se trata, por tanto, de descripciones estructuradas y codificadas que describen características y propiedades de objetos y recursos para facilitar su localización, recuperación, valoración, administración, persistencia e interoperabilidad.

El uso de meta-información estructurada permite integrar más fácilmente colecciones distintas al establecer un marco común de trabajo y, en consecuencia, mejora la recuperación de los objetos digitales y de su contenido. Esta preocupación por la recuperación de información se ha fijado igualmente en la recuperación temática y así, Blocks, Cunliffe y Tudhope (2006) subrayan la renovada atención prestada en los últimos años a los tesauros, cuyo origen sitúan en el interés por los metadatos en la Web; estándares como Dublin Core recomiendan que la materia de un recurso electrónico sea tomada de un vocabulario controlado.

Las evidentes deficiencias en la recuperación de información en Internet parecen exigir vocabularios interoperables y multilingües. Esta demanda ha conducido a un esfuerzo a nivel internacional por desarrollar normativas que adapten las tradicionales herramientas terminológicas a las estructuras y soportes de la información digital. El trabajo de García Marco *et al.* (2007) relata este proceso así como las nuevas tendencias internacionales, lideradas por la iniciativa americana ANSI/NISO Z39.19-2005 y la propuesta por el British Standard Institute, BS 8723: Structured Vocabularies for Information Retrieval. Se trata, apostilla el estudio, de establecer mecanismos de interconexión entre los distintos lenguajes documentales y conectarlos con los desarrollos que se están llevando a cabo en el ámbito de las ontologías. Las novedades para Gil Leiva (2008, p. 154) están relacionadas con la inclusión de las taxonomías y ontologías en la norma, así como aspectos encaminados a crear mejores sistemas de almacenamiento y recuperación por medio de protocolos y formatos que faciliten la interoperatividad entre vocabularios controlados.

En agosto de 2011 se ha publicado la norma *ISO 25964-1:2011 Information and documentation - Thesauri and interoperability with other vocabularies - Part 1: Thesauri for information*

retrieval. La segunda parte de la norma está aún en desarrollo.

La ISO 25964 se asienta en la norma BS 8723 y ha estructurado su contenido en dos partes diferenciadas, la primera trata los aspectos relacionados con los tesauros monolingües y plurilingües e incluye un modelo de datos y formatos/protocolos para el intercambio de datos. La segunda parte está previsto que cubra los aspectos relacionados con la interoperabilidad entre tesauros y otros tipos de vocabularios, como clasificaciones, taxonomías y ontologías.

A este necesario esfuerzo normalizador se ha unido el impulso de la Web Semántica en la que necesariamente han de incorporarse los sistemas de organización del conocimiento como tesauros, vocabularios controlados, topic maps, esquemas de clasificación y otras herramientas conceptuales. Ahora bien, esta Web Semántica requiere que los datos sean legibles y comprensibles por los agentes software, para lo cual son necesarios estándares que permitan la representación de información en formatos interoperables. En este marco, parece evidente que la integración en el nivel de conocimiento debe venir de la mano de la utilización combinada de ambos instrumentos (herramientas conceptuales y estándares de representación), e incluso que su uso debería ser suficiente para garantizarla (Martínez González *et al.*, 2009).

Con este fin surgen los distintos estándares de representación, comenzando con XML, y siguiendo con los que se apoyan en él. RDF permite representar metadatos. RDF Schema añade la posibilidad de representar el vocabulario utilizado en grafos RDF y relacionarlo mediante estructuras clasificatorias sencillas como taxonomías. Sirven de muestra los ejemplos siguientes:

- Zthes (s. d.), un modelo de representación de tesauros, pero también glosarios, diccionarios, clasificaciones, taxonomías, ontologías simples, ficheros de autoridades, etc., con el objetivo de facilitar la interoperabilidad de tesauros construidos con ISO 2788 y ANSI/NISO Z39.19. El modelo es una aplicación de XML y el acceso puede realizarse a través de los protocolos SRU/SRW y Z39.50.
- Existen otros conjuntos de metadatos con los que se puede representar un tesoro, si bien no han sido creados para este fin:

MADS (Metadata Authority Description Standard, Biblioteca del Congreso de Estados Unidos). Formato XML compatible con MARC 21. Con algunas de sus etiquetas se

puede representar una lista de encabezamientos de materia o un tesoro MADS.

Topic MAPs (ISO/IEC 13250). Para Eíto (2003) un topic map equivaldría a un índice creado para una colección de recursos disponibles en formato electrónico. Similar a los mapas conceptuales, no obstante Topic Maps es un estándar. En este índice, señala el mismo autor, contaríamos con una serie de topics (temas o materias), que se utilizarán para describir el contenido de los recursos y facilitar su recuperación así como relaciones que se establecen entre estos topics (una serie de recursos que serán indizados o descritos mediante los topics anteriores). Las aplicaciones disponibles para la gestión de topic maps resuelven importantes problemas relacionados con su visualización, la navegación de estos espacios informativos y su transferencia y validación.

Por su parte, SKOS (Simple Knowledge Organization System) es un "modelo de datos diseñado para compartir sistemas de organización de conocimiento en la Web" (W3C, 2009). Se trata de una iniciativa del Consorcio Web (W3C) que proporciona un modelo para expresar la estructura básica y el contenido de tesauros, clasificaciones, listas de encabezamiento de materia, taxonomías o folksonomías.

SKOS se puede implementar sobre RDF, utilizarse aisladamente, o combinarlo con OWL. Esta capacidad de interacción con otras propuestas creemos que dota al modelo de una enorme potencialidad. Según Pastor Sánchez (2011, p. 99) la importancia de SKOS se debe a que constituye un estándar sencillo para publicar sistemas de organización del conocimiento en la Web semántica. En la práctica, añade, SKOS es la especificación del W3C utilizada para realizar esta tarea de modo eficiente, lo cual, unido al carácter extensible de RDF, permite ampliar los vocabularios modelados en SKOS con ontologías OWL. SKOS se ha consolidado como un estándar de *facto*, tal y como se recoge en el Consorcio Web (World Wide Web Consortium, 2012). Es el modelo utilizado para la representación de vocabularios del Congreso de Estados Unidos, tesauros como AGROVOC o EuroVoc, los encabezamientos de materia de las bibliotecas nacionales de Francia o Alemania, etc.

Un vocabulario documental representado en SKOS continúa siendo un tesoro, una clasificación o una lista de materias, no obstante, su disposición en un lenguaje de alto nivel lógico como RDF favorece la interconexión entre los vocabularios especificados mediante SKOS y

puede contribuir a mejorar la calidad de los lenguajes documentales de origen.

Asimismo, la representación de herramientas terminológicas en modelos vinculados a la Web Semántica guarda una estrecha relación con lo que se viene denominando Linked Data y Open Linked data, una iniciativa que pretende facilitar el acceso a los datos, su publicación y reutilización a través de la Web y la construcción de aplicaciones en torno a los datos así expuestos. Como recogen Peset, Ferrer-Sapena y Subirats-Coll (2011), a escala internacional, el W3C identificó la comunidad bibliotecaria como uno de sus focos de interés a través de sus grupos dinámicos o *Incubator groups*, en concreto, creó en 2010 el Library Linked Data Incubator Group (World Wide Web Consortium, 2011) orientado a la publicación de datos estructurados en *Linked data* procedentes de bibliotecas, conformando específicamente las Listas de Autoridades y los vocabularios, dos de los *clusters* de iniciativas establecidas. La propia Biblioteca Nacional de España ha anunciado a finales de 2011 su participación en esta iniciativa. Además, los principios de esta Web de datos están siendo aplicados con éxito en el ámbito de las administraciones públicas. En España, gobiernos autonómicos como los del País Vasco, Cataluña o Asturias, corporaciones locales como los Ayuntamientos de Zaragoza o Gijón, y agencias estatales como el Instituto Geográfico Nacional, están comenzando a publicar sus datos de forma libre y gratuita.

Por último encontramos las ontologías, definidas por Gruber (1993) como representación formal y explícita de la estructura conceptual de un campo del conocimiento. Las relaciones conceptuales que son representadas en una ontología son muy variadas y dependientes del campo del conocimiento a estructurar. Una ontología se construye con la finalidad de compartir y reutilizar la información almacenada, la cual, por estar formalizada, puede ser interpretada tanto por personas como por programas informáticos. OWL (W3C, 2004) aporta el vocabulario estándar para representar ontologías, las herramientas conceptuales más potentes de las que disponemos para expresar semántica al más alto nivel. Las ontologías han adquirido un papel relevante como herramientas que soportan las búsquedas conceptuales y otra serie de funcionalidades. García, Pareja y Pradana (2008) resumen los beneficios adicionales que aporta la formalización del conocimiento como ontología y no como tesoro.

Coincidimos con Codina y Pedraza (2011) para quienes las ontologías, constituyendo evidentemente una de las tecnologías más prometido-

ras para el futuro de los sistemas de información no conforman en la actualidad un modelo de éxito en lo que atañe a su implementación en sistemas de información; aún son soluciones, en este sentido, inmaduras.

Esta situación puede constatarse en los casos de aplicación práctica recogidos por el propio W3C (World Wide Web Consortium, 1994-2012). De los 45 enumerados, dos son de temática jurídica:

- iLaw - Intelligent Legislation Support System. Pyung Kim *et al.* KISTI and MOJ, Korea. Aplicación de tecnologías semánticas, en concreto OWL y SPARQL para mejorar la búsqueda, la integración de datos y la generación de contenidos dinámicos.
- Helping New Judges Answer Complex Legal Questions. Pompeu Casanovas. Institute of Law and Technology de la Universidad Autónoma de Barcelona y Consejo General del Poder Judicial. Aplicación de vocabularios RDF y vocabularios *ad hoc* para mejorar la búsqueda de información jurídica.

Existe un debate abierto que trata de contraponer tesauros y ontologías, si bien parece más adecuado considerar que existe un *continuum* con las clasificaciones y taxonomías en un extremo y las ontologías en el otro, y los tesauros en algún lugar intermedio del mismo y como un posible punto de partida para la construcción de ontologías (Fensel, D. *et al.*, 2005; Codina y Pedraza, 2011). Por su parte, García Jiménez (2004) habla de un proceso de retroalimentación entre ambos instrumentos, señala la dificultad para entender el tesoro como un tipo de ontología, no sólo por sus diferentes trayectorias históricas sino también por sus diferentes vinculaciones operativas y teóricas, si bien, añade, los tesauros encajan en aquellas ontologías que se han calificado de ligeras.

La literatura coincide en calificar las ontologías como una tecnología semántica prometedora y con enormes posibilidades, de eficacia contrastada en campos concretos de la inteligencia artificial y la lingüística, pero en su aplicación a los sistemas de información aún no constituye una solución suficientemente robusta. Además, en muchos casos se diseñan para propósitos específicos y no son reutilizadas a posteriori, limitándose su funcionalidad al sistema local para el que se elaboraron.

4. El tratamiento semántico de la información en los diarios oficiales

Como una primera aproximación al estado actual de la gestión de los procesos documentales

y al desarrollo de la semántica documental en los sistemas legales de las administraciones públicas nos hemos fijado en los diarios y boletines.

Para conocer la situación de desarrollo actual de técnicas semánticas implementadas en los diarios oficiales de nuestro país se procedió al envío, en el mes de octubre de 2011, de un cuestionario abierto a los servicios de documentación responsables del Boletín Oficial del Estado, de los Boletines de las diecisiete comunidades autónomas y de las ciudades autónomas de Ceuta y Melilla indagando en el uso de lenguajes controlados, el empleo de XML y el uso o modelado de algún esquema de metadatos. El cuestionario completo figura como anexo a este texto.

La tasa de respuesta a nuestra solicitud de información no ha sido tan exhaustiva como hubiésemos deseado, pero creemos que nos permite extraer ciertas conclusiones. Doce Diarios han respondido a nuestra solicitud de información: Boletín Oficial de la Junta de Andalucía (BOJA), Boletín Oficial de Aragón (BOA), Boletín Oficial del Principado de Asturias (BOPA), Boletín Oficial de Islas Baleares (BOCAIB), Boletín Oficial de Canarias (BOC), Boletín Oficial de Cantabria (BOC), Diario Oficial de Castilla-La Mancha (DOCM), Boletín Oficial de la Junta de Castilla y León (BOCyL), Diari Oficial de la Generalitat de Catalunya (DOGC), Diario Oficial de Extremadura (DOE), Diario Oficial de Galicia (DOG), Boletín Oficial de La Rioja (BOR), Boletín Oficial de la Comunidad de Madrid (BOCM), Boletín Oficial de la Región de Murcia (BORM), Boletín Oficial de Navarra (BON), Boletín Oficial del País Vasco (BOPV), Diari Oficial de la Comunitat Valenciana (DOCV), Boletín Oficial de la Ciudad Autónoma de Ceuta (BOCCE) y Boletín Oficial de la Ciudad Autónoma de Melilla (BOCME).

Según se aprecia en la tabla I, en la página siguiente, el sistema de gestión de mayor implementación es Oracle y seis de las instituciones señalan el empleo específico de una herramienta para el control del vocabulario en la aplicación empleada, observándose la coincidencia de Oracle Text en el Boletín Oficial de Castilla y León, el Diari Oficial de la Generalitat de Catalunya y el Boletín Oficial de la Región de Murcia.

En el BOE se utiliza una herramienta propia desarrollada a partir de Oracle adaptada específicamente a este organismo. Esta herramienta no incluye gestión del Tesoro, sino que las materias se seleccionan a partir de un desplegable. Para suplir la carencia de gestor de Tesoro, se utilizan más o menos materias generales o específicas dependiendo del caso. No

existe control de Autoridades; para facilitar la gestión de Gaceta, se elaboró un proyecto de control de Autoridades para aquellos boletines anteriores a 1959, ya que en ellos sólo es posible la búsqueda a partir del título de la disposición y no del texto completo de la misma, pero aún no se ha logrado su puesta en marcha.

	Herramienta de gestión	Control del vocabulario
BOE	Oracle	
BOA	BRS	
BOPA	DOGMA/TAMINO y Framework PA/ORACLE	√
DOCM		
BOCyL	Oracle	Oracle Text
DOGC	Oracle Universal Content Management	Oracle Text Thesaurus
DOE	Microsoft SQL Server	√
DOG	Informix, OpenOffice, Indesing	
BOCM		
BORM	Oracle 11G	Oracle Text
BOPV		
DOCV	Oracle	Basis

Tabla I. Funcionalidad en los boletines

La utilidad de las tecnologías XML para el tratamiento, difusión e intercambio de la documentación legal frente a las bases de datos convencionales se ha demostrado desde hace más de una década. La representación en XML permite reconocer la estructura del documento implícita en su contenido, marcar relaciones semánticas o de referencias cruzadas e incorporar al propio texto legal metainformación. Valoramos en este sentido el uso de XML como un pilar que proporciona beneficios estructurales y de contenido sobre este tipo de documentación y que permite en un segundo nivel incrustar modelos semánticos y vocabularios de metadatos estandarizados. El empleo de XML se acredita en la gestión de seis de los Diarios que han respondido a la encuesta (en la tabla II se puede encontrar el detalle), si bien en algunos casos, se limita a secciones concretas de la documentación tratada.

Actualmente el BOE sólo emplea XML para la Legislación Consolidada. Asimismo, el Boletín

Oficial del País Vasco nos comunica que está prevista su utilización en la próxima actualización de la sede web de consulta al BOPV. Por su parte el BORM emplea XML para la aplicación que da acceso a los contenidos a través de dispositivos móviles.

	XML	Esquema de metadatos
BOE (Legislación Consolidada)	√	
BOA		
BOPA	√	√
DOCM		
BOCyL	√	√
DOGC	√	√
DOE		
DOG		
BOCM	√	√ Proyecto
BORM	√ (Dispositivos móviles)	
BOPV	√ Proyecto	√ Proyecto
DOCV	√	

Tabla II. Esquemas de metadatos en los boletines

En cuanto a la implementación de esquemas de metadatos se han recibido respuestas afirmativas del Boletín Oficial del Principado de Asturias, Boletín Oficial de Castilla y León y Diari Oficial de la Generalitat de Catalunya (Orengaya y Giralt, 2011) y proyectos en desarrollo de los Boletines de la Comunidad de Madrid (BOCM), en este caso para la gestión de anuncios, y en el Boletín Oficial del País Vasco para el tratamiento de datos de carácter personal. En ningún caso se nos ha comunicado el empleo de modelos de metadatos estandarizados o adaptaciones de los mismos.

Tampoco se ha registrado en ninguna de las respuestas la aplicación de tecnologías semánticas como RDF, OWL o SKOS.

En cuanto al control del vocabulario propiamente, de los doce diarios recogidos en la tabla III, siete señalan la elaboración y empleo de un lenguaje controlado para el tratamiento de la información. Se observa un cuidado particular en asegurar la univocidad en la indización y

recuperación de organismos e instituciones, esto es, en el control de autoridades aplicado específicamente a las entidades. En el DOGC con control de autoridades también para “Planes y Programas”. En este Diario podemos hablar de una dedicación concienzuda en este punto; se emplearon como punto de partida tres fuen-

tes de referencia de acreditada solvencia (Catàleg d'autoritats de noms i títols de Catalunya (Cantic), Llista d'encapçalaments de noms i títols de la Biblioteca de Catalunya (Lenoti) y la Base de datos de organismos y responsables de la Generalitat de Catalunya).

	BOE	BOA	BOPA	DOCM	BOCyL	DOGC	DOE	DOG	BOCM	BORM	BOPV	DOCV
Vocabulario ad-hoc	√		√			√	√		√		√	√
Control de entidades	√	√	√			√			√			√
Control de topónimos	√											√
Clasificación	√								√			√
Tesaurus EuroVoc	√ (*)		√			√				√	√	√

Tabla III. Uso de vocabularios en los boletines —Tesaurus Eurovoc incluye uso efectivo y proyecto; (*) adaptación —

Solamente el Boletín Oficial del Estado y el Diari Oficial de la Comunitat Valenciana señalan el empleo específico de control de autoridades para topónimos.

Han creado sistemas clasificatorios el BOCM y el DOCV. En el caso del BOE para asignar los códigos geográficos se ha utilizado la Clasificación Decimal Universal (CDU). Por su parte, el listado de materias está dividido en tres grandes grupos: Materias, Instituciones y Geográficos. Cada descriptor tiene un código numérico compuesto de diez dígitos, el primero indica si pertenece al grupo de Materias (dígito 1), Instituciones (dígito 2) o Geográfico (dígito 3). A continuación, dos bloques de tres dígitos cada uno, que indican la familia a que pertenecen; otro bloque de dos dígitos para indicar mayor especificidad (si es necesario), y por último un dígito que indica si el término es un descriptor válido o no. Para los códigos de Instituciones y Materias se utilizará la CDU o el Tesaurus EuroVoc según convenga.

Existen tesaurus jurídicos en castellano, resultando ejemplos relevantes el Tesaurus de Derecho del IEDCYT (Consejo Superior de Investigaciones Científicas) o el Tesaurus del Centro de Documentación Judicial (Cendoj). Por su parte, EuroVoc es un tesaurus que se puede usar libremente, en el marco de un convenio con la Oficina de Publicaciones de la Unión Europea, que es quien se encarga de su mantenimiento. Es un tesaurus multilingüe y multidisciplinar que abarca la terminología de los ámbitos de actividad de la Unión Europea, especialmente adecuado a las labores parlamentarias. Es la herramienta terminológica oficial que la Oficina de Publicaciones utiliza en sus sistemas de infor-

mación, conformándose como un estándar de referencia entre los tesaurus que se usan en el campo de la información jurídica. Su utilización por parte de un buen número de organismos oficiales en los distintos países de la Unión Europea, como el Congreso y el Senado en España o numerosos parlamentos autonómicos induce su utilización en cuantos sistemas buscan interoperabilidad con cualquiera de ellos y esta ha sido la razón que nos ha llevado a inquirir específicamente sobre su uso en los sistemas objeto de análisis.

Como se observa en la tabla anterior, cinco de los sistemas emplean el Tesaurus Eurovoc: Boletín Oficial del Principado de Asturias, Diari Oficial de la Generalitat de Catalunya, Boletín Oficial de la Región de Murcia, Boletín Oficial del País Vasco y Diari Oficial de la Comunitat Valenciana.

El Boletín Oficial del Estado emplea una lista normalizada de materias exclusivamente para lo publicado en la base de datos Iberlex, es decir las disposiciones de carácter general de ámbito estatal, autonómico y europeo desde 1960. A finales de la década de los 90 se elaboró un Tesaurus común con el Senado y el Congreso, que tenía como punto de partida el Tesaurus EuroVoc adaptado a la legislación española. Actualmente, el BOE utiliza su propio Tesaurus, resultado de ese trabajo conjunto, pero cuyas materias y estructura de familias son básicamente las establecidas por EuroVoc.

En el BOPA no se emplea EuroVoc, sin embargo, sí se utiliza en la base de datos legislativa INVESDOC en la que se hace un vaciado de las disposiciones normativas de interés publicadas en el Boletín Oficial del Principado de Asturias,

EuroVoc se complementa con un listado de materias elaborado *ad hoc*.

Merece una especial atención la gestión documental llevada a cabo en el Diari Oficial de la Generalitat de Catalunya y recogida con detalle en el artículo de Orenga-Gaya y Giralt (2011). Entre los años 2009 y 2010 se realizó una adaptación rigurosa de los descriptores originales del DOGC a EuroVoc. Los descriptores originales del DOGC se incluyeron como términos equivalentes (no-descriptores) del descriptor aceptado de EuroVoc. En esta fase se identificaron también los términos del DOGC sin correspondencia en EuroVoc. Posteriormente se realizó la traducción de los antiguos descriptores temáticos del DOGC a sus descriptores equivalentes de EuroVoc y se ha concluido el proceso elaborando pautas de indización y formación de usuarios.

5. Reflexiones finales

Los portales legislativos institucionales han pasado progresivamente de meros repositorios de documentos normativos a sistemas de recuperación de información, con mejores prestaciones gracias a un mayor cuidado en el tratamiento documental y jurídico.

Se observan esfuerzos claros por llevar a cabo un análisis documental del contenido normativo que redunde en un mayor rendimiento de estos sistemas legales y una optimización de la recuperación de información por parte de los usuarios finales. El empleo efectivo o en fase de implementación de XML y los proyectos de esquemas de metadatos parecen avanzar en la línea necesaria de la descripción semántica de la documentación.

En cuanto al análisis conceptual, valoramos muy positivamente el empleo de vocabularios para el control de entradas temáticas, de entidades y topónimos. En nuestra opinión resulta relevante la tendencia observada que nos permite hablar de un empleo mayoritario del tesoro EuroVoc o adaptaciones de EuroVoc. Abogamos por el empleo de una herramienta consolidada en entornos semejantes y que podría resultar una alternativa viable como vocabulario estándar y multilingüe en sistemas de información legal como los que nos ocupan, que permitiría el intercambio de información entre distintas Administraciones.

No se observa en los Diarios oficiales aquí tratados ninguna iniciativa para la representación de lenguajes mediante estándares semánticos, pero la decisión de la Oficina de Publicaciones de la UE de representar EuroVoc en SKOS

puede marcar la pauta para que sistemas como los aquí observados caminen hacia la necesaria interoperabilidad de la información gestionada por los mismos.

Creemos que se necesitan políticas de acercamiento y esfuerzos conjuntos para gestionar de un modo más eficaz estos sistemas de información en aras de prestar servicios de mayor calidad que permitan reutilizar la información. En este sentido, cabe destacar el Grupo de Trabajo creado entre el Parlamento de Cataluña y la Entidad Autónoma Diari Oficial i de Publicacions (Orenga-Gaya y Giralt, 2011) que persigue, entre otros, la coordinación de la solicitud de alta de nuevos descriptores a EuroVoc, el establecimiento de un protocolo de indización conjunto para los documentos del Parlament que se envíen a publicar al DOGC así como la elaboración conjunta del microtesoro de la Administración catalana.

Agradecimientos

Agradecemos la colaboración de los responsables de los servicios de documentación de los Boletines y Diarios Oficiales que han atendido amablemente nuestra solicitud de información.

Referencias

- Aitchison, J.; Bawden, D.; Gilchrist, A. (2000). *Thesaurus Construction and Use: A Practical Manual*. Londres: Aslib, 2000.
- Alvite Díez, M. L. (2004). Las herramientas terminológicas en los sistemas de información jurídica. // *Scire: Representación y organización del conocimiento*. 10:1 (2004) 77-90.
- Andrés Rivero, J. de; Hernández Marín, R. (2009). CRONOLEX: sistema para la representación dinámica de cuerpos legales. // *Scire: Representación y organización del conocimiento*. 15:1 (2009) 133-147.
- ANSI/NISO (2005). *ANSI/NISO Z39.19-2005. Guidelines for the construction, format, and management of monolingual controlled vocabularies*. Bethesda: NISO Press, 2005.
- Blocks, D.; Cunliffe, D.; Tudhope, D. (2006). A reference model for user-system interaction in thesaurus-based searching. // *Journal of the American Society for Information Science and Technology*. 57:12 (2006) 1655-1665.
- Codina, L.; Pedraza-Jiménez, R. (2011) Tesauros y ontologías en sistemas de información documental. // *El Profesional de la información*. 20:5 (2011) 555-563.
- Eito Brun, R. (2003). Topics maps y la indización de recursos electrónicos en la web. // *El profesional de la información*. 12:2 (2003) 141-148.
- Fensel, D.; Hendler, J. A.; Lieberman, H.; Wahlster, W. (Eds.) (2005). *Spinning the semantic web: Bringing the World Wide Web to its full potential*. Cambridge: The MIT Press, 2005.
- García Jiménez, A. (2004). Instrumentos de representación del conocimiento: tesauros versus ontologías. // *Anales de Documentación*. 7 (2004) 79-95. <http://revistas.um.es/analesdoc/article/view/1691/1741> (2012-06-21)

- García Marco, F. J. (2009). Perspectivas sobre el uso de la Web Semántica en el tratamiento de información y documentación legislativa. // *Scire: Representación y organización del conocimiento*. 15:1 (2009) 59-86
- García Marco, F. J. (Coord.); Agustín Lacruz, M. C.; Caro Castro, C.; Martínez Usero, J. A.; San Segundo Manuel, R. (2007). Proyectos internacionales de reforma y ampliación de las normas sobre tesauros para su adaptación a los nuevos contextos de integración e interoperabilidad en el entorno digital. // VIII Congreso ISKO-España. León: Universidad de León, 2007. 389-398.
- 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 (2008) 8-21.
- Gil Leiva, I. (2008). Manual de indización. Gijón: Trea, 2008
- Gil Urdiciain, B. (1998). Evaluación del rendimiento de tesauros españoles en sistemas de recuperación de información. // *Revista Española de Documentación Científica*. 21:3 (1998) 286-302.
- Gil Urdiciain, B. (2004). Manual de lenguajes documentales. 2ªed. revisada y ampliada. Gijón: Trea, 2004.
- Gruber, T. R. (1993). A translation approach to portable ontologies. // *Knowledge acquisition*. 5:2 (1993) 199-220.
- Hernández-Pérez, T.; Rodríguez-Mateos, D.; Martín-Galán, B.; García-Moreno, M. A. (2009). El uso de metadatos en la administración electrónica española: los retos de la interoperabilidad. // *Revista Española de Documentación Científica*. 32:4 (2009) 67-91.
- ISO (2011). ISO 25964-1:2011 Information and documentation. Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval. http://www.iso.org/iso/catalogue_detail.htm?csnumber=53657 (2012-03-21)
- Lancaster, F. W. (1995). El control del vocabulario en la recuperación de información. Valencia: Universitat de València, 1995.
- López-Huertas Pérez, M. J. (1991). Lenguajes documentales: aproximación a la evolución histórica de un concepto. // *Boletín de la ANABAD*. XLI:1 (1991) 61-70.
- Martínez González, M. M.; Vicente Blanco, D.J.; Fuente Redondo, P. de la; Adiego Rodríguez, J.; Pisabarro Marrón, A. M.; Sánchez Felipe, J. M. (2009). Estructura, semántica, extracción de información y XML legislativo: experiencias en la Universidad de Valladolid. // *Scire: Representación y Organización del Conocimiento*. 15:1 (2009) 173-186.
- Martínez, M. M.; Fuente, P. de la; Derniame, J.C. (2003). XML as a means to support information extraction from legal documents. // *International Journal of Computer Systems Science and Engineering*. 18:5 (2003) 263-277.
- Méndez, E. (2010). Tendencias en recuperación de información: principios y retos para una nueva década de datos enlazados. // *Anuario ThinkEPI*. 4 (2010) 231-239.
- Moreiro González, J. A. (2004). El Contenido de los documentos textuales: su análisis y representación mediante el lenguaje natural. Gijón: Trea, 2004.
- Moreiro González, J. A. (2007). Evolución paralela de los lenguajes documentales y la terminología // VIII Congreso ISKO-España. León: Universidad de León, 2007. 389-398.
- Moreiro González, J. A. (2009). Taxonomía. // Díaz Nafria, J.M.; Salto Alemany, F.; Pérez Montoro, M. (coords.). *Glosario Bitrum*. <http://glossarium.bitrum.unileon.es/Home/taxonomia> (2012-03-15)
- Orenga-Gaya, L.; Giralt, O. (2011). El Diari Oficial de la Generalitat de Catalunya: génesis de un diario digital. // *El profesional de la información*. 20:3 (2011) 340-344.
- Pastor Sánchez, J.A. (2011). Tecnologías de la web semántica. Barcelona: Editorial UOC, 2011.
- Peset, F.; Ferrer-Sapena, A.; Subirats-Coll, I. (2011). Open data y Linked open data: su impacto en el área de bibliotecas y documentación. // *El profesional de la información*. 20:2 (2011) 165-173.
- Rodríguez Bravo, B. (2011). Apuntes sobre representación y organización de la información. Gijón: Trea, 2011.
- Rowley, J. (1994). The controlled versus natural indexing languages debate revisited: a perspective on information retrieval practice and research. // *Journal of Information Science*. 20:2 (1994) 108-119.
- Sánchez-Jiménez, R.; Gil Urdiciain, B. (2007). Lenguajes documentales y ontologías. // *El Profesional de la Información*. 16:6 (2007) 551-560.
- Tenopir, C.; Ro, J.S. (1990). Full text databases. New York: Greenwood Press, 1990.
- UNE (1990): 50106-1990, Documentación. Directrices para el establecimiento y desarrollo de tesauros monolingües. Madrid: AENOR, 1990.
- UNE (1997): 50125-1997, Documentación. Directrices para la creación y desarrollo de tesauros multilingües. Madrid: AENOR, 1997.
- World Wide Web Consortium (1994-2012). Semantic Web. Case Studies and Use Cases. <http://www.w3.org/2001/sw/sweo/public/UseCases/>.
- World Wide Web Consortium (W3C). (2004). OWL Web Ontology Language guide. W3C Recommendation 10 Feb. 2004, 2004. <http://www.w3.org/TR/owl-guide>. (2012-03-10).
- World Wide Web Consortium (W3C). (2009). SKOS Simple Knowledge Organization System Reference. W3C Recommendation 18 August 2009, 2009. <http://www.w3.org/TR/2009/REC-skos-reference-20090818/> (2012-03-10).
- World Wide Web Consortium, Library Linked Data Incubator Group (2011). Use case / Case studies. <http://www.w3.org/2005/Incubator/ld/wiki/UseCases>.
- World Wide Web Consortium (2012). W3C. Semantic Web. SKOS/Datasets: <http://www.w3.org/2001/sw/wiki/SKOS/Datasets>.
- Zthes (s. d.). Zthes: the Zthes specifications for thesaurus representation, access and navigation. <http://zthes.z3950.org/>.

Anexo: Cuestionario

- ¿Qué herramienta tecnológica emplean para la gestión del Diario?
 - ¿Incluye un módulo o herramienta concreta para la gestión del Tesoro o del Control de Autoridades?
- ¿Para el tratamiento de los textos emplean XML?
- ¿Han elaborado algún esquema de metadatos propio?
- ¿Emplean algún vocabulario controlado o lista normalizada de control de autoridades para la indización y recuperación de información en su Diario Oficial?
 - ¿Desde qué fecha?
 - ¿Se trata de un vocabulario ad hoc o una adaptación de un sistema ya desarrollado?
- En el caso de que no lo utilicen, ¿se han planteado el empleo de EuroVoc o su adaptación?

Enviado: 2012-04-26. Versión corregida: 2012-08-21.

Aceptado: 2012-08-21.

