

La interoperabilidad entre sistemas de organización del conocimiento: un nuevo desafío

Ana M. Martínez Tamayo^{1,2}, Julia C. Valdez², Edgardo A. Stubbs²,
Yanina González Terán², María Inés Kessler²

¹Instituto de Investigaciones en Humanidades y Ciencias Sociales (UNLP-CONICET). ²Departamento de Bibliotecología, Facultad de Humanidades y Ciencias de la Educación, Universidad Nacional de La Plata. 48 e/6 y 7, 1900 La Plata, Argentina. E-mail: ammarti@speedy.com.ar

Resumen: La interoperabilidad entre distintos sistemas de organización del conocimiento (SOC) ha cobrado gran importancia en los últimos tiempos, con el propósito de facilitar la búsqueda simultánea en varias bases de datos o fusionar distintas bases de datos en una sola. Las nuevas normas para el diseño y desarrollo de SOC, la estadounidense Z39.19:2005 y la británica BS 8723-4:2007, incluyen recomendaciones detalladas para la interoperabilidad. También se encuentra en preparación una nueva norma ISO 25964-1 sobre tesauros e interoperabilidad que se agregará a las anteriores. La tecnología disponible proporciona herramientas para este fin, como son los formatos y requisitos funcionales de autoridades y las herramientas de la Web Semántica RDF/OWL, SKOS Core y XML. Actualmente es difícil diseñar y desarrollar nuevos SOC debido a los problemas económicos, de modo que la interoperabilidad hace posible aprovechar los SOC existentes. En este trabajo se revisan los conceptos, modelos y métodos recomendados por las normas, así como numerosas experiencias de interoperabilidad entre SOC que han sido documentadas.

Introducción

Los sistemas de información disponibles en la Web, entre ellos las bibliotecas virtuales, muchas veces gestionan y brindan acceso a dos o más bases de datos bibliográficas o factuales propias, adquiridas o compartidas, que suelen estar indizadas y clasificadas con distintos sistemas de organización del conocimiento (SOC). Si bien existen modalidades como las nuevas herramientas de descubrimiento NextGen, que permiten a los usuarios consultar varias bases y sitios Web al mismo tiempo y con los mismos términos (Schneider, 2006), éstos por lo general se limitan a la lengua natural, de modo que el usuario enfrenta

problemas como la pérdida de información por el uso de sinónimos, homónimos o términos polisémicos, en diferentes dominios o en distintos idiomas.

Por otro lado, hay servicios de información que requieren intercambiar los registros bibliográficos e incluso fusionar bases de datos desarrolladas previamente con total independencia, por ejemplo para implementar un catálogo colectivo. Surge así el problema de la diferencia de los SOC usados en las bases de datos individuales. Así, muchos servicios de información en línea necesitan que los SOC de sus bases de datos sean compatibles conceptual y terminológicamente (Aitchison et al., 1997; Chang y Zeng, 2002; Lancaster, 2002).

Esta necesidad ha llevado al tratamiento detallado del tema en la nueva norma estadounidense ANSI/NISO Z39.19:2005 para el diseño y desarrollo de SOC (National Information Standards Organization, 2005) y en la norma británica BS-8723, específicamente su parte 4 (BSI Group, 2007). Actualmente se encuentra en desarrollo la norma ISO 25964 sobre el diseño de tesauros y su interoperabilidad con otros SOC. Por otro lado, hay numerosas experiencias publicadas en revistas y congresos de la especialidad. En consecuencia, el propósito de este trabajo es presentar el estado del arte de este tema.

Interoperabilidad entre SOC

Existen varias definiciones de interoperabilidad. La norma estadounidense ANSI/NISO Z39.19:2005 y la norma británica BS8723-4:2007, citadas anteriormente, coinciden en definir la interoperabilidad como la capacidad que tienen dos o más sistemas o componentes de intercambiar información y usar esa información que se ha intercambiado.

La demanda de interoperabilidad entre SOC ha surgido debido a dos tendencias: la tecnología disponible en la World Wide Web que permite el intercambio de datos en una escala nunca antes imaginada y la presión económica que dictamina u obliga a que los recursos de información preparados para un determinado contexto o aplicación deban ser accesibles para otras aplicaciones y/o contextos. Esta demanda se presenta principalmente en tres escenarios: las organizaciones globales que operan alrededor del mundo y necesitan

acceder a la información y al conocimiento que generan en su propia lengua o en otras, el sector público cuya información necesita ser accesible para una variada audiencia, o bien las colecciones de datos indizados en épocas pasadas con diferentes SOC (BSI Group, 2007).

Los modelos y métodos de interoperabilidad entre SOC

Se debe distinguir entre el *SOC fuente* (source) que sirve como punto de partida cuando se busca un término correspondiente en otro SOC y *SOC blanco* (target) en el que se busca un término que corresponde a otro término existente en el SOC fuente (BSI Group, 2007).

La norma estadounidense describe los siguientes métodos para lograr la interoperabilidad entre SOC (National Information Standards Organization, 2005):

- Derivación o modelado: consiste en seleccionar un SOC fuente para desarrollar o derivar otro SOC blanco. Ejemplo es el proyecto BiblioPhil de Rumania, que ha derivado un tesoro trilingüe a partir de la Clasificación Decimal Universal (CDU) (Frâncu y Sabo, 2010).

- Traducción o adaptación: algunos SOC blanco se han creado mediante la traducción o adaptación de los términos de un SOC fuente en otra lengua. Un ejemplo es Descriptores en Ciencias de la Salud (DeCS) que es la traducción al español y al portugués de Medical Subject Headings (MeSH), pero que además ha hecho una adaptación del área de salud pública al ambiente latinoamericano.

- SOC satélites: el SOC fuente funciona como una superestructura (macrotesauro) para SOC más específicos (microtesauros), que pueden usarse independientemente.

- Nodo: diferentes nodos de la estructura jerárquica de un SOC fuente se vinculan a términos de un SOC más específicos que se aplican a un único nodo de la jerarquía.

- Mapeo directo: es el proceso de establecer equivalencias y otras relaciones entre los términos de distintos SOC. Un ejemplo es el proyecto CAT-AGROVOC, en el que se mapearon el Tesoro Chino de Agricultura (CAT) y el tesoro AGROVOC de la Organización de las Naciones Unidas para la Agricultura y la Alimentación (Liang et al., 2005).

- Mapeo por coocurrencia: este método usa la coocurrencia de términos en distintas bases de datos para encontrar familias de términos. Los términos se seleccionan de los metadatos a partir de palabras claves y términos controlados que se han asignado a los contenidos, o bien de las palabras seleccionadas del texto completo.

- Conexión: se utiliza un SOC de conexión que contiene los puntos focales o notaciones a través de los cuales se agrupan los términos de varios SOC. Ejemplo de esta modalidad es el proyecto MSAC, en el que las bibliotecas nacionales de la República Checa, Croacia, Letonia, Lituania, Macedonia, Eslovaquia y Eslovenia han creado una lista de autoridades de materia multilingüe (una autoridad por idioma), conectadas por notaciones de CDU (Balikova, 2005).

- Vinculación a través de una lista de temporal unificada: diferentes SOC se vinculan mediante mapeo de términos, no solo aquellos conceptualmente equivalentes, sino también otros que pueden estar relacionados. El mapeo se realiza de forma dinámica en respuesta a la consulta de un usuario específico y el resultado se muestra en una lista temporal, es decir no se genera un nuevo SOC.

- Vinculación a través de servidores de SOC: involucra la transmisión de una consulta a uno o más servidores de SOC. Cada SOC individual se registra en un servidor y éste acepta consultas contra cualquier SOC especificado en el protocolo del servidor. Las consultas de los usuarios se envían al servidor central que pasa la consulta a los SOC registrados, recoge los distintos resultados y los presenta al usuario.

Por su parte, la norma británica presenta tres modelos básicos (BSI Group, 2007):

1) Unidad estructural: todos los SOC participantes comparten exactamente la misma estructura jerárquica y las relaciones asociativas entre conceptos. Es equivalente a los métodos de derivación y traducción de la norma estadounidense.

2) Pares no equivalentes: vincula dos SOC que no comparten la misma estructura, por ejemplo alguna combinación de tesauros, esquemas de clasificación, encabezamientos de materia, taxonomías, listas de autoridades, etc. Es equivalente al mapeo directo de la norma estadounidense.

3) Estructura de columna vertebral o espina: en los casos en que se requiere vincular tres o más SOC, dada la complejidad de esta situación, se hace necesario seleccionar uno de los SOC como columna vertebral o estructura básica. Es equivalente al método de conexión de la norma estadounidense.

El punto crítico de la interoperabilidad entre distintos SOC es la relación de equivalencia, pues no es posible establecer jerarquías y asociaciones si no se han establecido primero las equivalencias entre términos. Un mismo concepto puede estar representado por un término preferente distinto en cada SOC (por ejemplo *maní* y *cacahuate*), ya sea que usen la misma lengua o no (*maní* y *peanut*). A veces un concepto se puede representar por un término en un SOC, pero requiere dos o más términos en el otro (*administración de bibliotecas* y *administración AND bibliotecas*). Un término puede representar un concepto dentro de un dominio y un concepto diferente en otro dominio (*mercurio* en astronomía, metalurgia y mitología). En consecuencia, se pueden encontrar diferentes grados de equivalencia:

- Equivalencia exacta: Es la situación ideal. El SOC blanco contiene un concepto idéntico al concepto del SOC fuente, por ejemplo: el epígrafe *catalogación* de la Lista de Encabezamientos de Materia para Bibliotecas (LEMB), el descriptor *catalogación* del Tesouro de la Unesco y la clase *025.3* de la CDU.

- Equivalencia inexacta: los conceptos en el SOC fuente y en el SOC blanco se superponen. Puede establecerse una relación de equivalencia, dependiendo del grado de superposición y del contexto en el que será utilizado el mapeo. Por ejemplo, en el Tesouro de Unesco, el descriptor *país en desarrollo* es un término específico de *país menos desarrollado*, mientras que el Tesouro Spines los considera sinónimos.

- Equivalencia parcial: El concepto en uno de los SOC tiene un alcance más general que el concepto en otro SOC. En contextos en los que el mapeo jerárquico no es posible, la mejor solución puede ser establecer una relación de equivalencia. Por ejemplo, en epígrafe de LEMB *bibliobuses* y la clase *021.65* de CDU que es *Bibliotecas móviles. Bibliotecas ambulantes*. Alternativamente, se pueden combinar dos o más términos preferentes identificados en el SOC blanco que, juntos, representen un concepto del SOC fuente. Por

ejemplo, el encabezamiento de materia *publicidad en bibliotecas* de LEMB equivale a dos descriptores del Tesauro de Unesco: *publicidad* y *biblioteca*.

- No equivalencia: el SOC blanco no contiene un concepto que coincida con el concepto del SOC fuente, ni de manera parcial o inexacta.

Experiencias de interoperabilidad entre SOC

El establecimiento de equivalencias entre dos o más SOC tiene una larga historia (Aitchison et al., 1997; Lancaster, 2002). En 1995 el tema había adquirido tal importancia, que la International Society for Knowledge Organization (ISKO) organizó en Varsovia el Seminario sobre Compatibilidad e Integración de Sistemas de Ordenación que culminó con una serie de recomendaciones tanto para la investigación, la educación y el intercambio de información, como para el desarrollo de sistemas y esfuerzos cooperativos nacionales e internacionales (International Society of Knowledge Organization, 1996).

En la última década se han desarrollado nuevos proyectos de interoperabilidad entre SOC, que pueden agruparse como sigue:

- Entre listas de encabezamientos de materia (precoordinadas): un ejemplo es el proyecto MACS de la Conference of European National Libraries (CENL), que ha vinculado tres listas de epígrafes: LCSH en inglés, Répertoire d'autorité-matière encyclopédique et alphabétique unifié (Rameau) en francés y Schlagwortnormdatei (SWD) en alemán (Landry, 2000).

- Entre tesauros (poscoordinados): entre ellos el proyecto ASEG del Centro de Ecosistemas Acuáticos de Cataluña, que ha vinculado cinco tesauros que abarcan las ciencias acuáticas, pesca, agricultura, alimentación, medio ambiente y economía (Reverté y Sebastiá, 2009).

- Entre listas de encabezamientos de materia y tesauros: el proyecto SWD-IZ es una experiencia alemana que relaciona los conceptos indizados en dos bases de datos, una que utiliza el Thesaurus Sozialwissenschaften (IZ) de ciencias sociales y la otra que usa la lista de epígrafes SWD (Zhang, 2006).

- Entre sistemas de clasificación: Library of Congress ha establecido equivalencias entre las notaciones de Library of Congress Classification (LCC) y las del Sistema de Clasificación Decimal de Dewey (SCDD) (Library of Congress, 2004).

- Entre listas de epígrafes, tesauros y sistemas de clasificación: un ejemplo es el proyecto KoMoHe del Bundesministerium für Bildung und Forschung de Alemania, iniciado en 2004 y concluido en 2007. En total se han vinculado 25 SOC (16 tesauros, 4 listas de descriptores, 2 listas de encabezamientos de materia y 3 sistemas de clasificación, incluyendo SCDD), de 11 disciplinas, en alemán, inglés y ruso (Mayr, 2008, 2009).

Los ejemplos anteriores son solo ilustrativos, ya que existen numerosas experiencias.

Herramientas para lograr la interoperabilidad entre SOC

Muchas experiencias de interoperabilidad entre SOC han recurrido a herramientas ya existentes y de fácil acceso, entre ellas las listas de autoridades y las herramientas de la Web Semántica (Library of Congress, 2010a,b; Willer, 2009; International Federation of Library Associations and Institutions, 2000; International Federation of Library Associations and Institutions, 2010; World Wide Web Consortium, 2009; Guzmán Luna et al., 2006; Cavieres Abarca et al., 2010).

Agradecimiento

Este trabajo es parte del proyecto acreditado 11/H526 del Programa de Incentivos del Ministerio de Educación de la Nación.

Bibliografía

- Aitchison, Jean; Alan Gilchrist y David Bawden. 1997. Thesaurus construction and use: a practical manual. 3rd ed. London: Aslib. xvi, 212 p.
- Balikova, Marie. 2005. Multilingual subject access to catalogues of national libraries (MSAC): Czech Republic's collaboration with Slovakia, Slovenia, Croatia, Macedonia, Lithuania and Latvia. En World Library and Information Congress, IFLA General Conference and Council. (71th : 2005 : Oslo) The Hague: IFLA <http://archive.ifla.org/IV/ifla71/papers/044e-Balikova.pdf>. [Consulta 30 Sep 2010].
- BSI Group. 2007. Structured vocabularies for information retrieval: guide: Part 4: Interoperability. London: The Group, 55 p. (BS 8723-4:2007).
- Cavieres Abarca, Alonso; Sergio Fredes Mena y Arturo Ramírez Novoa. 2010. Tesauros y Web semántica: diseño metodológico para estructurar contenidos Web mediante SKOS Core. En Serie bibliotecología y gestión de información. N° 57, p. 1-64. http://eprints.rclis.org/18169/1/Serie_N%C2%BA_57_Mayo_2010_Tesauros_y_Web_Sem%C3%A1ntica.pdf [Consulta 30 Sep 2010].
- Chan, Lois Mai y Marcia Lei Zeng. 2002. Asegurando la interoperabilidad entre vocabularios de materia y esquemas de organización del conocimiento. En IFLA Council and General Conference (68th : 2002; Glasgow). The Hague: IFLA. <http://archive.ifla.org/IV/ifla68/papers/008-122s.pdf>. [Consulta 30 Sep 2010].
- Frâncu, Victoria y Sabo, Cosmin-Nicolae. 2010. Implementation of a UDC-based multilingual thesaurus in a library catalogue: the case of BiblioPhil. En Knowledge organization. Vol. 37, n° 3, 209-215.
- Guzmán Luna, Jaime A.; Durley Torres Pardo y Alba Nubia López García. 2006. Desarrollo de una ontología en el contexto de la Web semántica a partir de un tesauro documental tradicional. Revista interamericana de bibliotecología Medellín. Vol. 29, n° 2, 79-94.

- International Federation of Library Associations and Institutions. 2000. Concise UNIMARC classification format . 2a ed. The Hague: The Federation. <http://archive.ifla.org/VI/3/p1996-1/concise.htm>. [Consulta 30 Sep 2010].
- International Federation of Library Associations and Institutions. 2010. Functional requirements for subject authority data. The Hague: The Federation. <http://nkos.slis.kent.edu/FRSAR/FRSAD-Report.pdf>. [Consulta 30 Sep 2010].
- International Organization for Standardization. 2010. Information and documentation: thesauri and interoperability with other vocabularies: Part 1. Thesauri for information retrieval. Geneva: The Organization. 134 p. (ISO/DIS 25964-1:2010).
- International Society for Knowledge Organization. 1996. Compatibility and integration of order systems: research seminar proceedings of the TIP/ISKO Meeting (1995 : Warsaw). Warsaw: Stowarzyszenie Bibliotekarzy Polskich.
- Lancaster, Frederick W. 2002 [1986]. El control del vocabulario en la recuperación de información. 2a ed. Valencia: Universitat de Valencia. 286 p.
- Liang, Anita C. et al. 2005. The mapping schema from Chinese Agricultural Thesaurus to AGROVOC. En Agricultural Ontology Service Workshop on Ontologies: the More Practical Issues and Experience. (6th : 2005 : Vila Real, Portugal). Rome: Food and Agriculture Organization. <ftp://ftfao.org/docrep/fao/008/af241e/af241e00.pdf>. [Consulta 30 Sep 2010].
- Library of Congress. 2004. Classification Web: quick start tutorial. Washington DC: The Library. <http://classificationweb.net/tutorial/1intro.html>. [Consulta 30 Sep 2010].
- Library of Congress. 2010a. MARC 21 format for authority data. Washington DC: The Library. <http://www.loc.gov/marc/authority/ecadhome.html>. [Consulta 30 Sep 2010].
- Library of Congress. 2010b. MARC 21 format for classification data. Washington DC: The Library. <http://www.loc.gov/marc/classification/eccdhome.html>. [Consulta 30 Sep 2010].
- McCulloch, Emma y George Macgregor. 2008. Analysis of equivalence mapping for terminology services. En Journal of information science. Vol. 34, nº 1, 70-92.

- Mayr, Philipp y Vivien Petras. 2009. Cross-concordances: terminology mapping and its effectiveness for information retrieval. En *International cataloging and bibliographic control*. Vol. 38, n° 3, 43-52.
- National Information Standards Organization. 2005. Guidelines for the construction, format and management of monolingual thesauri. 4ª ed. Bethesda MD: The Organization. 172 p. http://www.niso.org/kst/reports/standards/kfile_download?id%3Austring%3Aiso-8859-1=Z39-19-2005.pdf&pt=RkGKiXzW643YeUaYUqZ1BFwDhIG4-24RJbcZBWg8uE4vWdpZsJDs4RjLz0t90_d5_ymGsj_IKVAGZww13HuDIYn5U74YdfA-3TffjxYQ25QrtR8PONuJLqxvo-l0NIr5 [Consulta 30 Sep 2010].
- Reverté Reverté, Carmen y Montserrat Sebastiá Salat. 2009. Aquatic science subject Gateway Project as a model of interoperability. En *ISKO-UK 2009 Conference*. (2009 : London). London: ISKO-UK Chapter. http://www.iskouk.org/conf2009/papers/reverte_ISKOUK2009.pdf. [Consulta 30 Sep 2010].
- Schneider, Karen G. 2006. How OPACs suck, part 2: the checklist of shame [en línea]. En: ALA TechSource. Chicago: American Library Association <http://www.alatechsource.org/blog/2006/04/how-opacs-suck-part-2-the-checklist-of-shame.html>. [Consulta 30 Sep 2010].
- Wang, Jun. 2003. A knowledge network constructed by integrating classification, thesaurus, and metadata in digital library. En *International information & library research*. Vol. 35, n° 2/4, 383-397.
- Willer, Mirna, ed. 2009. UNIMARC manual authorities format. 3rd ed. Munich: K.G. Saur. 309 p. (IFLA Series on bibliographic control; 38).
- World Wide Web Consortium. 2009. Manual de SKOS (simple knowledge organization system: nota del Grupo de Trabajo del W3C de 18 de agosto de 2009. [s.L.]: El Consorcio. <http://skos.um.es/TR/skos-primer/> [Consulta 30 Sep 2010].
- Zeng, Marcia y Maja Zummer. 2009. Introducción de FRSAD y correspondencia con SKOS y otros modelos. En *World Library and Information Congress, IFLA General*

I Jornada de Intercambio y Reflexión acerca de la Investigación en Bibliotecología,
La Plata, 6-7 de diciembre de 2010. La Plata: Facultad de Humanidades y Ciencias de la Educación de la
Universidad Nacional de La Plata.

Conference and Council. (75th : 2009 : Milan). The Hague:
IFLA. <http://www.ifla.org/files/hq/papers/ifla75/200-zeng-es.pdf> [Consulta 30 Sep 2010].

- Zhang, Xueying. Concept integration of document databases using different indexing languages. Information Processing & Management. Vol. 42, n° 1, 121-135, 2006.