



68th IFLA Council and General Conference

August 18-24, 2002

Code Number: 008-122-S
Division Number: IV
Professional Group: Classification and Indexing
Joint Meeting with: -
Meeting Number: 122
Simultaneous Interpretation: -

Asegurando la interoperabilidad entre vocabularios de materia y esquemas de organización del conocimiento: Un análisis metodológico

Lois Mai Chan,

School of Library and Information Science, University of Kentucky
USA

Marcia Lei Zeng

School of Library and Information Science, Kent State University
Kent, USA

Resumen

El heterogéneo entorno de recuperación de información de la World Wide Web ha hecho surgir la necesidad de la interoperabilidad entre diversos sistemas. En la recuperación por materias, los usuarios se encuentran no sólo con diferentes vocabularios y esquemas, sino también con diferentes idiomas. Como resultado de esto, en los últimos años han surgido numerosos proyectos cuyo objetivo es mejorar la interoperabilidad entre vocabularios de materia y esquemas de organización del conocimiento, algunos de los cuales se centran en diferentes vocabularios y otros en diferentes lenguas o idiomas. Esta ponencia intenta analizar los métodos que se han utilizado en estos proyectos. Comienza ésta con una breve aproximación al tema para después hacer un análisis particularizado de los diferentes enfoques y métodos empleados en los esfuerzos más recientes.

1 INTRODUCCIÓN

En el entorno abierto de Internet y de la Web, los recursos de información son heterogéneos y han sido indizados con diferentes vocabularios y organizados de acuerdo a diferentes esquemas. El hecho de

lograr los mejores resultados de recuperación en el proceso de búsqueda de información en varias disciplinas ha representado un reto para el profesional de la información. En la recuperación de información, los usuarios no están, ni deberían estar, al cuidado de los mecanismos que les permiten realizar búsquedas mediante los vocabularios empleados por varios sistemas. La mejor solución sería facilitar una búsqueda única y homogénea en lugar de hacer que el usuario busque de forma separada en bases de datos o colecciones individuales. Para hacer posible esa solución es importante adaptar los diferentes sistemas de organización, tales como vocabularios controlados y sistemas de clasificación y hacerlos interoperables con un sencillo dispositivo de búsqueda.

2 REVISIÓN DE PROYECTOS Y EJEMPLOS DE VOCABULARIOS INTEROPERABLES

Antes de comenzar a examinar sus métodos, vamos a revisar sucintamente algunos esfuerzos recientes para lograr la interoperabilidad entre diferentes vocabularios de materia (tanto controlados como no controlados) y sistemas de organización del conocimiento. Estos incluyen esfuerzos tanto para establecer interoperabilidad entre vocabularios en el mismo idioma o en diferentes idiomas, entre diferentes esquemas de clasificación, y entre vocabularios controlados y esquemas de clasificación. Estos esfuerzos han conducido a la equiparación e integración de sistemas de organización ya existentes, o a la creación de otros nuevos, para compartir información en un entorno interconectado. Los proyectos que presentamos, organizados por similitud de objetivos, son los siguientes:

2.1. Entre vocabularios controlados en la misma lengua

1. Entre los *Encabezamientos de Materia de la Biblioteca del Congreso* (LCSH) y los *Encabezamientos de Materia de Medicina* (MeSH) - Northwestern University (Olson, 2001).
2. Entre diferentes vocabularios controlados - H.W. Wilson Company (Kuhr, 2001)
3. Entre diferentes tesauros alemanes usados para indizar contenidos especializados en Matemáticas y Física, así como en Ciencias Sociales- CARMEN (Content Analysis, Retrieval, Metadata: Effective Networking) (CARMEN WP12, 2000)

2.2. Entre vocabularios de materia múltiples en diferentes lenguas y sistemas de clasificación

1. Entre tesauros, sistemas de clasificación, sistemas codificados, y listas de términos controlados en el campo de la Biomedicina – UMLS (Unified Medical Language System) Metathesaurus (National Library of Medicine, 2001)
2. Entre servicios distribuidos con diferentes vocabularios de indización utilizados por varias comunidades, tales como archivos, el sector de la Enseñanza Superior, bibliotecas, museos, la National Grid for Learning, y el Resource Discovery Network, etc. - HILT (High-Level Thesaurus Project). (HILT, 2000; Nicholson, Wake y Currier, 2001a)
3. Entre “Vocabularios de entrada” usados por sistemas (por ejemplo, índices a *BIOSIS Concept Codes*, *INSPEC Thesaurus*, *U.S. Patent and Trademark Office Patent Classification*, etc.) con objeto de equipararlos a “Vocabularios de interrogación” con entrada en una búsqueda – Proyecto DARPA Unfamiliar Metadata, Universidad de California Berkeley (Buckland et al., 1999).
4. Entre esquemas de clasificación locales a un esquema común (DDC (*Clasificación Decimal de Dewey*) – Proyecto Renardus (Koch, Neuroth, and Day, 2001)
5. Entre cuatro vocabularios controlados y esquemas: *Clasificación Temática Polaca* (PTC), descriptores basados en el *Thesaurus of Common Topics* (TCT), *Clasificación Decimal Universal* (CDU), y *Lenguaje de Encabezamientos de Materia* (SHL) de la Biblioteca Nacional en Varsovia – Polish Project (Scibor and Tomasik-Beck, 1994).

6. Entre vocabularios controlados utilizados por cuatro catálogos de bibliotecas nacionales en tres idiomas: inglés, francés y alemán - MACS (Multilingual Access to Subjects) (Freyre and Naudi, 2001).
7. Entre vocabularios para una base de datos multilingüe acerca del patrimonio francés – Merimee (Véase el informe estadístico en Doerr, 2001.)

2.3. Entre un vocabulario controlado y un sistema de clasificación universal

1. Entre LCSH y LCC (*Clasificación de la Biblioteca del Congreso*) –*Classification Plus* (un producto en CD-ROM) y *Classification Web* (un interfaz basado en la Web que está desarrollándose), Biblioteca del Congreso
2. Entre LCSH y DDC (Vizine-Goetz, 1996)
3. Entre CDU y GFSH (*General Finish Subject headings*) (Himanka y Vesa, 1992)

2.4 Entre sistemas de clasificación

1. Entre MSC (la Sociedad Matemática Americana (AMS) *Mathematics Subject Classification*) y la Clase 510 de DDC – Universidad del Estado de Nueva York en Albany, Nueva York. (Iyer and Giguere, 1995).
2. Entre SAB (*Klassifikationssystem för svenska bibliotek*) y DDC – Biblioteca Real Sueca (IFLA, 2001:34)

2.5 Nuevo sistema para diferentes idiomas

El Proyecto HEREIN (la red europea de información sobre las políticas de patrimonio cultural) realizó un tesoro compuesto de términos extraídos de informes sobre políticas de patrimonio cultural). Se creó sin ninguna referencia directa a los términos o a la estructura de un tesoro ya existente. - The HEREIN Project (<http://www.european-heritage.net/en/index.html>, click Thesaurus)

3 MÉTODOS USADOS PARA LOGRAR Y MEJORAR LA INTER-OPERATIVIDAD

La búsqueda de la compatibilidad del vocabulario no es nueva. Mucho antes de la llegada de la era electrónica, los profesionales de las bibliotecas y de la información habían explorado y empleado varios métodos para reducir el conflicto entre diferentes vocabularios usados en el mismo sistema. Los primeros métodos se basaron casi completamente en esfuerzos de tipo intelectual. Conforme avanzó el proceso de automatización, surgieron nuevos métodos para lograr y mejorar la interoperatividad, y la tecnología informática comenzó a ser empleada a fondo en el entorno de la interconexión. La siguiente sección recoge tanto métodos convencionales como nuevos que son ampliamente aceptados.

1. Derivación/Modelaje – Se desarrolla un vocabulario especializado o más sencillo con un vocabulario existente y más completo, que se toma como punto de partida o modelo.
2. Traducción/Adaptación – Se desarrolla un vocabulario controlado que consta de términos traducidos de otro en un idioma diferente con o sin modificación.
3. Equiparación (intelectual) – Se desarrolla un sistema equiparado que consta básicamente de términos establecidos como equivalentes entre diferentes vocabularios controlados o entre términos y números de clasificación. Tal equiparación requiere generalmente una gran dedicación y un enorme esfuerzo intelectual.
4. Equiparación (asistida por ordenador) – Se desarrolla un sistema equiparado basándose parcial o totalmente en la tecnología informática.
5. Vinculación – Se desarrolla una lista de términos que se vinculan a otros términos que no son equivalentes conceptualmente pero que están muy relacionados lingüísticamente. Tales vínculos o enlaces se realizan con objeto de perfeccionar los resultados de la recuperación.
6. Conversión – Se desarrolla un lenguaje o esquema que sirve como un intermediario para moverse entre términos equivalentes en diferentes vocabularios.

4 MÉTODOS UTILIZADOS EN EL ALMACENAMIENTO Y GESTIÓN DE LOS VÍNCULOS

Una vez se ha establecido la equiparación, se necesita un dispositivo para almacenar y mantener los vínculos que soportan el gran número de términos de indización y las complejas relaciones que se generan entre ellos. Para este propósito, se han investigado y usado diferentes opciones:

1. Registros de autoridad – Campos especiales de los formatos de autoridad pueden ser empleados para almacenar los vínculos.
2. Concordancias – La elaboración de concordancias requiere la distinción de un vocabulario/esquema matriz y de uno o más vocabularios/esquemas finales.
3. Red semántica – Una red semántica, también llamada Web semántica, compuesta de un estructura organizada que sirve como “hilo” o espina dorsal. Cada unidad en la red representa un concepto alrededor del cual un grupo de términos equivalentes de diferentes vocabularios se identifican y almacenan.

5. DISCUSIÓN

Una serie de aspectos comunes han aparecido en nuestro análisis de los métodos utilizados en los numerosos proyectos discutidos anteriormente.

5.1 Aspectos generales en la equiparación

5.1.1 Equiparación de vocabularios multilingües.

En el corazón del vocabulario de materias multilingüe está la equiparación o establecimiento de equivalencias. Las relaciones uno a uno entre términos de diferentes vocabularios y diferentes idiomas son las equiparaciones ideales, pero con frecuencia resultan difíciles de conseguir. Diferentes expresiones lingüísticas para el mismo concepto, diferentes grados de especificidad, y términos polisémicos son algunas de las dificultades con las que se tienen que enfrentar los que intentan equiparar vocabularios y los que crean vocabularios multilingües o multidisciplinares. Los complejos requisitos y procesos de la equiparación de términos, que son con frecuencia imprecisos, tienen repercusión en los siguientes aspectos de la equiparación de vocabularios (Koch, Neuroth, y Day, 2001): *browsing*, visualización, exhaustividad y la interrelación entre consistencia, precisión y pertinencia. Varios niveles de equiparación/vinculación pueden coexistir en el mismo proyecto, tales como los que han sido identificados por el proyecto MACS: nivel terminológico (encabezamiento de materia), nivel semántico (registro de autoridad), y nivel sintáctico (aplicación) (Freyre y Naudi, 2001).

5.1.2 Integración de perspectivas de diferentes culturas.

Bajo el punto de partida de que todos los idiomas son iguales en concordancia, existe una cuestión sobre si las perspectivas de una cultura en concreto que está expresada mediante un vocabulario controlado o una clasificación pueden ser transferidos apropiadamente, en el proceso de equiparación, a una lengua diferente. Hudon (1997) señalaba los siguientes problemas asociados con sistemas multilingües:

- 1) los que resultan de adaptar una lengua para acomodarla a una estructura conceptual extranjera hasta el punto de que pueda ser reconocible por sus propios hablantes;
- 2) los que resultan de transferir una estructura conceptual completa de una cultura a otra, ya sea apropiada o no; y
- 3) los que resultan de traducir literalmente términos de la lengua fuente a expresiones sin sentido en el idioma final, etc.

Esta autora resume estos aspectos como: problemas de gestión, problemas lingüísticos/semánticos, y problemas relacionados con la tecnología.

5.1.3 Equiparación de sistemas con diferentes estructuras

Existen diferencias básicas en los términos de las macro-estructuras de vocabularios controlados y de los sistemas de clasificación. Los tesauros construidos siguiendo la ISO 2788 y otros estándares nacionales aseguran que la estructura y “gramática” de tales vocabularios sean consistentes o compatibles. La construcción de encabezamientos de materia y esquemas de clasificación, por otra parte, se ha guiado por patrones o ejemplos ya existentes. Si hay diez sistemas universales diferentes, entonces habrá diez directrices diferentes. En consecuencia, los sistemas de organización del conocimiento difieren unos de otros en su estructura y características semánticas, léxicas y de notación o entrada (Iyer y Giguere, 1995). Por ejemplo, éstos pueden cubrir diferentes dominios temáticos, o pueden tener diferente alcance o cobertura; pueden tener diferencias semánticas que son causadas por variaciones en la estructura conceptual; sus niveles de especificidad y el uso de la terminología puede variar; y las características sintácticas, tales como el orden de las palabras de los términos y la elección del encabezamiento, pueden ser distintas.

Estas incompatibilidades han supuesto, desde el principio, un problema para cualquier esfuerzo de equiparación. Por ejemplo, el establecimiento de concordancia o traducción entre un tesoro y una clasificación o entre varios sistemas resulta algunas veces extremadamente difícil, sino imposible. Esto resulta especialmente cierto cuando el sistema final presenta un nivel de especificidad más alto que el sistema fuente u otros sistemas.

5.2 Opciones metodológicas

Para los proyectos que pretenden conseguir una interoperabilidad entre sistemas de organización del conocimiento con objeto de satisfacer las nuevas exigencias de los usuarios en entornos de red, una de las decisiones importantes que hay que tomar es la elección de un método apropiado. La primera pregunta compleja que hay que contestar es: ¿integrar, equiparar, o crear un sistema nuevo? Las opciones son similares a las que ha sugerido Riesthuis (2001) con respecto a las diferentes formas de crear tesauros multilingües:

1. traducción
2. fusión
3. crearlo partiendo de cero

Dentro de cada una de estas opciones hay múltiples posibilidades, tal como han sugerido los responsables del proyecto HILT en un sistema de referencia por coordenadas (Nicholson, Wake y Currier, 2001b).

Estos investigadores proponen tres opciones básicas:

- Usar o crear un esquema único (LCSH, UNESCO, basado en DDC, CDU, completamente nuevo);
- Equiparar esquemas ya existentes (LCSH, UNESCO, basado en DDC, CDU);
- Equiparar esquemas existentes a corto plazo que conducirán a la larga a un esquema sencillo.

Basándose en las opciones anteriores, pueden consignarse las siguientes consideraciones:

- Estructura de tesoro adicional;
- Nuevo micro-tesoro de materias específicas;
- Equiparación entre micro-tesoros ya existentes;
- Capacidad multilingüe;
- Control de la comunidad;

- Métodos asistidos por máquina;
- Métodos asistidos por Inteligencia Artificial;
- Formación del usuario;
- Facilidades para ayudar a los usuarios;
- Mapas mentales para usuarios;
- Consistencia en la aplicación de los términos garantizada por la formación y asesoramiento;
- Bibliotecarios formados para ayudar a mejorar las recuperaciones de los usuarios

La elección del método básico con la combinación de alguna de las consideraciones que hemos realizado puede conducir a varios productos finales y requiere una dedicación de tiempo y recursos diferentes. Cualquier método y su combinación con otro proceso tiene sus pros y contras. Es necesario realizar una investigación exhaustiva e identificar problemas potenciales cuando se emplea un método concreto.

6. CONCLUSIÓN

1. ¿Qué hemos aprendido de los proyectos?
2. ¿Qué puntos quedan pendientes?
3. ¿Qué es necesario, desde el punto de vista intelectual y técnico, para hacer avanzar este campo de estudio?

A partir de los ejemplos incluidos en este artículo, podemos resumir las siguientes tendencias que conforman la corriente dominante:

1. La necesidad de interoperabilidad entre sistemas de organización del conocimiento es una cuestión inevitable en el entorno interconectado de hoy día.
2. Varios métodos se han utilizado para conseguir esta interoperabilidad. Puede o no que sea necesario un sistema convertidor. Puede ser o no necesario construir una concordancia entre los diferentes vocabularios involucrados sea la situación ideal. O, igualmente, puede ser que la interoperabilidad se consiga de forma más efectiva mediante los registros de autoridad de materia de varios sistemas en línea.
3. Mientras que la equiparación de vocabularios constituye todavía un enorme esfuerzo intelectual, la tecnología informática se ha aplicado para ayudar a la gestión de grandes ficheros de datos de materia y a la gestión de vínculos o enlaces. Los niveles más altos de equiparación computerizada de sistemas han sido también objeto de experimentación o evaluación. Tanto la equiparación humana como asistida por ordenador co-existirán durante los próximos años.
4. Se han iniciado numerosos proyectos de equiparación de lenguas y estructuras. Estos proyectos han identificado y experimentado con variedad de métodos. Es seguro predecir que habrá más productos y servicios multilingües, y muchos de ellos incluirán sistemas de estructuras múltiples tales como tesauros, clasificaciones, encabezamientos de materia, e índices de términos asignados a registros de bases de datos.

La necesidad de compatibilizar diferentes vocabularios de materia en el entorno de red es indiscutible. Los resultados de los recientes esfuerzos por conseguir interoperabilidad entre vocabularios de diferentes tipos y en diferentes idiomas son esperanzadores. La pregunta permanece: ¿Hemos explotado completamente las posibilidades de la tecnología en nuestros esfuerzos por mejorar el acceso por materias a las miríadas de recursos ahora disponibles en el entorno en red?

AGRADECIMIENTOS

Agradecemos a los principales investigadores de los proyectos de interoperabilidad incluidos en este artículo -Traugott Koch (Suecia y Dinamarca), Patricia Kuhr (Estado Unidos), Martin Kunz (Alemania), Max Naudi (Francia), Dennis Nicholson (Gran Bretaña), and Tony Olson (Estado Unidos)-, quienes generosamente dedicaron su tiempo a contestar a nuestras preguntas, a dar detalles relativos a sus proyectos, o a revisar este artículo.

REFERENCES

- Buckland, M., et al. (1999). Mapping entry vocabulary to unfamiliar metadata vocabularies, *D-Lib Magazine*, 5(1). <http://www.dlib.org/dlib/january99/buckland/01buckland.html>, (consulta Feb. 5, 2002).
- CARMEN. WP12: Cross concordances of classifications and thesauri. <http://www.bibliothek.uni-regensburg.de/projects/carmen12/index.html.en> (consulta Feb. 5, 2002)
- Doerr, Martin. (2001) Semantic problems of thesaurus mapping. *Journal of Digital information*, 1 (8). <http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Doerr/#Nr.52>
- Freyre, Elisabeth and Max Naudi. (2001) MACS: Subject access across languages and networks. In *Subject Retrieval in a Networked Environment: Papers Presented at an IFLA Satellite Meeting sponsored by the IFLA Section on Classification and Indexing & IFLA Section on Information Technology, OCLC, Dublin, Ohio, USA, 14-16 August 2001*. Dublin, OH: OCLC.
- HILT. (2000) *HILT: High-Level Thesaurus Project Proposal*. <http://hilt.cdlr.strath.ac.uk/AboutHILT/proposal.html>. (consulta Feb.5, 2002)
- Himanka, Janne and Kautto Vesa. (1992) Translation of the Finish Abridged Edition of UDC into General Finish Subject Headings. *International Classification* 19(3):131-134.
- Hudon, Michele. (1997) Multilingual thesaurus construction: integrating the views of different cultures in one gateway to knowledge concepts. *Knowledge Organization* 24(2): 84-91.
- IFLA Section on Classification and Indexing. (2001) *Newsletter* Nr.24, December 2001.
- Iyer, Hermalata and Mark Giguere. (1995). Towards designing an expert system to map mathematics classificatory structures. *Knowledge Organization* 22(3/4):141-147.
- Koch, Traugott, Heike Neuroth, and Michael Day. (2001) Renardus: cross-browsing european subject gateways via a common classification system (DDC). In *Subject Retrieval in a Networked Environment: Papers Presented at an IFLA Satellite Meeting sponsored by the IFLA Section on Classification and Indexing & IFLA Section on Information Technology, OCLC, Dublin, Ohio, USA, 14-16 August 2001*. Dublin, OH: OCLC. <http://www.lub.lu.se/~traugott/drafts/preifla-final.html> (consulta Feb.5, 2002)
- Kuhr, Patricia S. (2001) Putting the world back together: mapping multiple vocabularies into a single thesaurus. In *Subject Retrieval in a Networked Environment: Papers Presented at an IFLA*

Satellite Meeting sponsored by the IFLA Section on Classification and Indexing & IFLA Section on Information Technology, OCLC, Dublin, Ohio, USA, 14-16 August 2001. Dublin, OH: OCLC.

National Library of Medicine. (2001) *Fact Sheet: UMLS ® Metathesaurus ®* Last updated 2001. <http://www.nlm.nih.gov/pubs/factsheets/umlsmeta.html>. (consulta Feb. 5, 2002)

Nicholson, Dennis and Susannah Wake. (2001a) HILT: Subject retrieval in a distributed environment. In *Subject Retrieval in a Networked Environment: Papers Presented at an IFLA Satellite Meeting sponsored by the IFLA Section on Classification and Indexing & IFLA Section on Information Technology, OCLC, Dublin, Ohio, USA, 14-16 August 2001. Dublin, OH: OCLC.*

Nicholson, Dennis, Susannah Wake, and Sarah Currier. (2001b) High-Level Thesaurus Project: investigating the problem of subject cross-searching and browsing between communities. In *Global Digital Library Development in the New Millemnnium: fertile ground for distributed cross-disciplinary collaboration*, edited by Ching-Chih Chen. Beijing: Tsinghua University Press, 2001.

Olson, Tony. (2001) Integrating LCSH and MeSH in information systems. In *Subject Retrieval in a Networked Environment: Papers Presented at an IFLA Satellite Meeting sponsored by the IFLA Section on Classification and Indexing & IFLA Section on Information Technology, OCLC, Dublin, Ohio, USA, 14-16 August 2001. Dublin, OH: OCLC.*

Riesthuis, Gerhard J.A. (2001) Information languages and multilingual subject access. In *Subject Retrieval in a Networked Environment: Papers Presented at an IFLA Satellite Meeting sponsored by the IFLA Section on Classification and Indexing & IFLA Section on Information Technology, OCLC, Dublin, Ohio, USA, 14-16 August 2001. Dublin, OH: OCLC.*

Scibor, Eugeniusz and Joanna Tomasik-Beck. (1994) On the establishment of concordances between indexing languages of universal or interdisciplinary scope (Polish Experiences). *Knowledge Organization* 21(4):203-212.

Vizine-Goetz, Diane. (1996) Classification Research at OCLC. *Annual Review of OCLC Research*, pp. 27-33.

Traducción: Jesús Jiménez Pelayo