



68th IFLA Council and General Conference

August 18-24, 2002

Code Number: 006-122-S
Division Number: IV
Professional Group: Classification and Indexing
Joint Meeting with: -
Meeting Number: 122
Simultaneous Interpretation: -

Interoperabilidad entre materias: acerca del Proyecto “Tesoro de Alto Nivel” (HILT)

Dennis Nicholson
Strathclyde University
Glasgow, UK

Introducción: De HILT Fase I a HILT Fase II

Los aspectos de interoperabilidad entre materias que se tratan en esta ponencia surgen de dos proyectos, llamados ahora HILT Fase I y HILT Fase II. Estos proyectos fueron fundados conjuntamente por el Joint Information Systems Committee (JISC) y el Research Support Libraries Programme (RSLP) y duró cerca de un año. La constitución de HILT Fase II por parte del JISC se espera que esté confirmada en abril de 2002. La Fase II durará 12 meses y retomará el trabajo de HILT Fase I, y las herramientas y experiencia del equipo que la llevó a cabo para lograr un consenso en la comunidad profesional como el logrado en HILT Fase I, creando un servicio piloto de equiparación terminológica o mapa de ruta con un enfoque específico al desarrollo del Distributed National Electronic Resource (DNER), incluyendo – pero no necesariamente limitado a – la Enseñanza Superior (ES) en lo concerniente a terminologías de materia para descripciones de nivel de colección en el plan Information Environment (IE) del JISC. La evaluación del usuario y el análisis del coste-beneficio de varios niveles de servicio serán también aspectos a tratar en el proyecto.

El problema

Asegurar que los usuarios de ES del IE puedan encontrar a través de las materias recursos de aprendizaje, investigación e información es uno de los mayores retos del JISC, el DNER, el Resource Discovery Network (RDN), y los diversos proveedores de servicios de información y enseñanza de archivos, bibliotecas, museos, y el ámbito de servicios electrónicos. Como descubrió HILT Fase I, los distintos proveedores de servicios usan una gran variedad de esquemas de materias (desde los esquemas generales

como LCSH, UNESCO, DDC, y AAT, a los esquemas específicos como MeSH) para describir de manera adecuada y consistente sus recursos de cara a una recuperación fiable. Si queremos que las búsquedas cruzadas y el *browsing* funcionen coherentemente para los usuarios del IE, estos esquemas deben ser equiparados a otro, quizá usando un “modelo” común, como la DDC, con una aplicación internacional y multilingüe así como con el potencial para facilitar su diálogo en red máquina a máquina (M2M). Más importante es quizá, el que las terminologías que tienen en mente diferentes tipos de usuarios de ES sean desprovistas de ambigüedad¹, y traducidas después a aquellos términos asignados por el servicio que los usuarios necesitan para realizar búsquedas cruzadas o hacer un *browsing* de los servicios relevantes a su petición de información. El objetivo de HILT Fase II es construir y evaluar un servicio piloto que hará posible este proceso como un “Servicio compartido” del DNER en el Information Environment.

HILT Fase 1

HILT Fase 1 encontró que:

- En los servicios de información de Gran Bretaña se utilizan esquemas de materia muy diferentes. Se piensa que la búsqueda a través de esos servicios es válida tanto para los usuarios como para el personal técnico.
- Había un fuerte consenso en el ámbito de los archivos, servicios electrónicos, bibliotecas y museos a favor de un proyecto piloto de seguimiento práctico que desarrollaría y determinaría con precisión los costes y beneficios de un mapa de ruta interactivo y en red de las terminologías usadas por estos ámbitos y las relaciones existentes entre esos ámbitos (previamente denominado en HILT como un 'servicio piloto de equiparación' – véase el diagrama del Mapa de Ruta de Terminologías (TeRM) en el Apéndice A para una mejor descripción de qué es y de cómo funcionaría)
- Se requería una mayor investigación para valorar la efectividad, nivel y naturaleza de la necesidad del usuario, práctica, requerimientos de diseño, y costes y beneficios de tal sistema antes de que un compromiso a largo plazo con un servicio caro pudiera estar justificado. Se determinó que esto podría hacerse mejor con un proyecto piloto que examinara estos y otros aspectos relacionados.

Más detalles de HILT Fase 1 pueden encontrarse en la página web de HILT² en general y en el Informe Final HILT³, en particular.

HILT Fase II: Objetivos

HILT Fase II se sitúa en el proceso de confección del proyecto piloto, centrándose – tal como recomendaba HILT Fase I – en los requerimientos de la terminología y de los tesauros de nivel de colección, pero también teniendo presente la necesidad de extender esto a las necesidades de recuperación. Utilizará el trabajo de HILT Fase I, y las herramientas y experiencia del equipo que lo llevó a cabo, para configurar un mapa de ruta de terminologías o servicio TeRM, similar al que se proponía en HILT Fase I, con los siguientes objetivos:

- a. Ofrecer un foco práctico experimental en el cual investigar y establecer los requerimientos de un servicio de terminología de materias para el JISC Information Environment, con particular referencia a DNER, RDN, usuario, nivel de colección, compatibilidad internacional, y local, regional y consideraciones de acceso en Gran Bretaña.

¹ El proceso de determinar si el usuario que teclea 'lotus' está buscando información sobre el coche, el programa informático, la flor, o alguno de los muchos significados adicionales de este término.

² <http://hilt.cdli.strath.ac.uk/>

³ <http://hilt.cdli.strath.ac.uk/Reports/FinalReport.html>

- b. Hacer recomendaciones concernientes a un posible servicio futuro, teniendo en cuenta una serie de factores, incluyendo el nivel y naturaleza de la necesidad del usuario, práctica, requerimientos de diseño, efectividad, funcionalidad existente en paquetes informáticos comerciales como oposición a desarrollos informáticos originales, y finalmente los costes y beneficios que para los usuarios de ES supone un servicio completo de terminologías centrado fundamentalmente en necesidades de nivel de colección.

HILT Fase II: Participantes

HILT Fase II tendrá una duración de 12 meses, y contará con aproximadamente la misma combinación de participantes que HILT Fase I, pero con la adición de una mayor implicación de representantes del DNER, el RDN, y ES. Concretamente:

- El Centro para la Investigación de la Biblioteca Digital (CDLR) en la Universidad de Strathclyde– líder;
- Representante de DNER
- mda (antes Asociación de Documentación de Museos);
- Consejo Nacional de Archivos (NCA);
- National Grid for Learning (NGfL) Escocia;
- Online Computer Library Center (OCLC);
- Representante de RDN
- Representante de ES
- Biblioteca Escocesa y Consejo de Información (SLIC);
- Universidad Escocesa de Industria (SufI);
- UK Office for Library and Information Networking (UKOLN).
- Expertos en terminología, Alan Gilchrist y Leonard Will (evaluador externo)

Mediante su compromiso en el proyecto CAIRNS⁴ (que utilizó el reforzamiento de la colección para configurar mini-grupos), los proyectos SCONE y SEED⁵ que intentaban construir una base de datos de colecciones de sectores cruzados⁶, y HILT⁷ Fase I, el sitio líder - Centro para la Investigación de la Biblioteca Digital (CDLR) en la Universidad de Strathclyde⁸ - tiene amplia experiencia en el uso de descripciones de nivel de colección en un entorno distribuido dinámicamente, y en los problemas de terminología asociados. También tiene disponible un rico entorno de información distribuida en el que estudiar la operación del proyecto piloto y sus interacciones con usuarios y servicios. Esto incluye el catálogo distribuido de CAIRNS, la Biblioteca Nacional de Escocia (NLS), NGfL, SLAINTE, y las bases de datos de la Biblioteca Digital de Glasgow (GDL), un mecanismo de reforzamiento de las posibilidades de la colección basado en materias, la base de datos de colecciones llamada SCONE, un servidor de impresos electrónicos (Iniciativa de Archivos Abiertos (OAI)), el Fondo de Nuevas Oportunidades (NOF) y otras bases de datos de proyectos de digitalización, y el potencial para instalar otras bases de datos Z39.50. Otros participantes – concretamente UKOLN, mda, NCA, el RDN, el DNER, y los expertos en terminología HILT, añaden profundidad y amplitud de miras al equipo. Además, la OCLC ha acordado prestar ayuda a la investigación ofreciendo acceso a una equiparación legible por máquina de LCSH a DDC y acceso a expertos. El CDLR también trabaja estrechamente con diez colegios de enseñanza superior dentro del proyecto RSLP GDL.

Construyendo el TeRM

⁴ Véase <http://cairns.lib.strath.ac.uk/> -el catálogo Z39.50 incluye universidades, NLS, NGfL, y otros

⁵ Véase <http://scone.strath.ac.uk/> y <http://seed.cdrl.strath.ac.uk/>

⁶ Véase <http://scone.strath.ac.uk/service/index.cfm>

⁷ Véase <http://hilt.cdrl.strath.ac.uk/>

⁸ Véase <http://cdlr.strath.ac.uk/>

Para el propósito de este proyecto, el piloto TeRM sería construido usando el software Wordmap⁹, disponible comercialmente. Se espera de éste (a través de la experiencia de HILT Fase I) que ofrezca una buena ilustración inicial del tipo de facilidades necesarias para el piloto. Esto no implica una preferencia por este software, ni siquiera de un software comercial en contra de uno “casero” o más abierto. El proyecto intentaría desarrollar una especificación completa de requerimientos mediante actividades evaluadoras dirigidas por grupos de usuarios y servicios, y expertos externos. Se compararían entonces *todos* los paquetes relevantes disponibles, habiendo realizado con anterioridad una investigación profunda de todas las soluciones, comerciales o no. WordMap se encontraría entre aquellos capaces de ofrecer software que podría resolver una parte significativa de la especificación, pero no estaría entre los preferidos. La cuestión de si una solución abierta es preferible a una solución comercial también se evaluaría.

Existen buenas razones para usar una parte específica de software comercial en esta etapa de desarrollo. La experiencia de HILT Fase I sugiere que los participantes del proyecto encuentran más fácil analizar los requerimientos de ese servicio dándoles un ejemplo ilustrativo real en el que centrarse. Se considera esencial, sin embargo, dotar pronto al proyecto de un piloto ilustrativo con objeto de captar el interés y atención de los usuarios y otros participantes y darles un entorno práctico en el cual puedan considerar y estudiar el problema. Wordmap está siendo empleado porque queremos tener una demostración real de trabajo en una etapa inicial para que los usuarios y proveedores de servicio puedan interactuar con ella. Intentar elaborar el requerimiento completo *antes* de implementar un piloto ilustrativo daría como resultado, según se piensa, una investigación pobre, ya que los usuarios y proveedores de servicio no habrían sido suficientemente estimulados para operar en un contexto que permita surgir una especificación completa. Esta aproximación, que se ve como pragmática, nos permitirá evaluar en un tiempo razonable los usos reales y los diferentes aspectos a tener en cuenta, al mismo tiempo que evitará el derroche y riesgos potenciales que supondría partir de cero en el desarrollo, antes de que se haya establecido un estudio completo de las necesidades.

Terminologías y Aspectos relacionados con la Terminología

El TeRM ilustrado inicialmente se basaría en terminologías de RDN¹⁰, en terminologías disponibles en el apartado de taxonomía de Wordmap, el cual incluye, en particular, un conjunto de términos que son usados por usuarios generales de Internet, y en partes seleccionadas de LCSH, DDC, UNESCO, y AAT. OCLC ofrecerá una equiparación LCSH – DDC, y posiblemente también otra equiparación de DDC a Encabezamientos de Materia Conspectus¹¹. El Tesauro de la UNESCO está disponible en línea¹² y acudiremos a fuentes de información manuales para obtener selecciones AAT. El objetivo sería una equiparación *selectiva* suficiente para el propósito del piloto en una primera etapa – por ejemplo, no un mapa terminológico exhaustivo. Se tendrá en cuenta también los diferentes aspectos apuntados por el evaluador de HILT Fase I, Leonard Will (Informe Final HILT, Sección 10), y dos cuestiones adicionales:

1. La cuestión de si el TeRM necesita o no un “hilo” central

Un elemento clave en el desarrollo de dicho piloto será “traducir” la propia terminología del usuario a las diferentes terminologías empleadas en el entorno distribuido, y hacerlo de manera inteligente y útil. Esto requerirá una cierta cantidad de interacción usuario-TeRM para eliminar la ambigüedad del término o términos empleados por el usuario (por ejemplo, ¿qué se quiere buscar cuando se busca

⁹ Véase www.wordmap.com

¹⁰ Véase, a modo de ejemplo, la lista creada por Andy Powell en <http://www.rdn.ac.uk/cgi-bin/browse>

¹¹ Los encabezamientos de materia Conspectus se usan en la base de datos CAIRNS.

¹² Véase <http://www.ulcc.ac.uk/unesco/thesaurus.htm>

lotus: la flor, o el coche, o el paquete informático, o qué?). Se plantea entonces una cuestión sobre si es:

- a. Factible
- b. Mejor en términos de buenos resultados de recuperación para el usuario (advirtiéndose que esto incluye una necesidad de recuperar por encima de las barreras lingüísticas)

para que esta interacción tenga lugar entre el usuario y un esquema central sencillo al cual están equiparados en el TeRM todos los demás esquemas del entorno, o entre el usuario y cada esquema individual. Siguiendo esto, si la mejor respuesta es un esquema central que sirve de modelo, se plantea la cuestión sobre cual de los esquemas existentes, si hay alguno, serviría mejor a este propósito, siendo la mejor posibilidad DDC (un esquema bien estructurado, jerárquico que ya ha sido traducido a un número importante de idiomas).

2. La cuestión de si a largo plazo la mejor solución para el problema de las terminologías de materia en un entorno distribuido no sería la identificación y adopción de un esquema simple aceptado como adecuado para cubrir todos los propósitos en todos los dominios – sino un nuevo esquema, o un esquema existente modificado para acomodarlo a los requerimientos de un modelo aceptado.

Este es, en esencia, un añadido al análisis coste-beneficio de la idea de un servicio de equiparación terminológico –un punto de referencia externo que sirva para evaluar el valor de un TeRM contra otras posibles soluciones.

Construyendo el Entorno de Investigación

Este se lograría añadiendo una serie de DNER y otras colecciones -incluyendo colecciones RDN, colecciones de archivos, colecciones de museos, y una colección local OAI-, a una *copia* de la base de datos de colecciones SCONE¹³ para crear un banco de pruebas de bases de datos de colecciones de HILT Fase II y un escenario basado en CLD y en la búsqueda cruzada usando el mecanismo de configuración dinámica y las facilidades de búsqueda de CAIRNS. El objetivo sería utilizar 'esquemas naturales de materia' para las colecciones en el entorno, y usar el piloto TeRM para eliminar la ambigüedad de los términos de los usuarios y resolver diferencias entre esquemas. Serían utilizadas una serie de configuraciones, asociadas aproximadamente a concentradores de materia relacionadas con intereses temáticos, pero que representan una variedad de circunstancias de usuarios, de alcance local, regional, nacional y de toda Gran Bretaña (general y por concentradores temáticos)¹⁴. El objetivo sería enlazar o vincular el TeRM al mecanismo de configuración si es posible (la experiencia de CAIRNS sugiere que podría ser), o simular este aspecto, si no es posible (esto sería menos elegante, pero suficiente para los propósitos de investigación del proyecto)

Compromisos de HILT Fase II

Los compromisos especificados para HILT Fase II son:

- 1.** Mayor comprensión del problema y necesidades de los usuarios de ES respecto a la recuperación por materias en el projectado JISC Information Environment, ambos dentro de JISC, servicios de JISC, y – aparte de las actividades de difusión – en la comunidad como un todo.

¹³ <http://scone.strath.ac.uk/Service/Index.cfm>

¹⁴ Concretamente, una universidad, un colegio ES, entorno ES en Glasgow, entorno ES en Escocia, entornos de nivel ES DNER, entornos ES en un centro de materia RDN

2. Un estudio en profundidad de los requerimientos para la equiparación terminológica en el DNER y servicios asociados de Gran Bretaña, teniendo en cuenta entornos de materia locales, regionales, nacionales, internacionales, de ES, y archivos, bibliotecas, museos, y servicios electrónicos.
3. Un servicio piloto de terminologías para el JISC IE (con funcionalidad limitada y con un servicio completo que posiblemente requiera un cambio de software).
4. Costes de las necesidades, instalación y mantenimiento, y costes contra beneficios, para un servicio futuro, incluyendo tanto al usuario como los requerimientos terminológicos y funcionales máquina a máquina.
5. Informe final sobre el proyecto, junto con recomendaciones adecuadas.

En espera de recibir los fondos necesarios (sólo tenemos una notificación informal en el momento en que escribimos esta ponencia), HILT Fase II comenzará en mayo de 2002.

Traducción: *Jesús Jiménez Pelayo*

***Nota:** Pueden verse ejemplos en www.wordmap.com con www.oingo.com y vivisimo.com