



TRATAMIENTO DE LOS METADATOS DE CONTENIDO EN LA WEB SEMÁNTICA O CUANDO LA LÓGICA SE HACE EVIDENTE

Mariàngels GRANADOS

*Biblioteca de Catalunya. C/ Hospital, 56 . BARCELONA 08001. ESPAÑA
(HTmgranados2@gmail.com);*

RESUMEN

Se propone un sistema en el uso de los metadatos de contenido susceptible de aplicarse en la Web Semántica que ayude a aumentar la precisión y la coherencia temáticas, ideado en base a las experiencias obtenidas en dos centros de temática multidisciplinar. El sistema de indización y recuperación propuesto se basa en la adopción de los encabezamientos de materia (LCSH) usados como descriptores, rompiendo las cadenas precoordinadas y combinándolos con la CDU. Se justifica el uso de esta clasificación así como también se dan ejemplos comparativos entre el uso de los LCSH y este sistema, enmarcando la propuesta con las características que ofrecen los lenguajes RDF y Dublín Core como elementos estructurales de la Web Semántica.

ABSTRACT

This article propose a new application of content metadata to increase the relevante and precision in web searching. The controled vocabularies used is based in LCSH for indexing but used as descriptors, breaking the precoordinated strings and establishing their equivalence with the UDC. Many reasons are exposed to justify their use. Also there are many exemples to explain the LCSH as they are usually used and the way that this system works. This proposal is based in the characteristics of RDF and Dublin Core languages as structures elements of Semantic Web.

PALABRAS CLAVE

Web Semántica, Clasificación Decimal Universal, Library of Congres Subject Headings, Recuperación de la información, Acceso temático, Metadatos de contenido, Indización.



Frente a la excesiva cantidad de documentos recuperados en las búsquedas en Internet, la falta de calidad y contexto, ha nacido el proyecto de la Web Semántica. Se trata pues de una web dotada de contenido donde además sea posible la interoperabilidad entre los diferentes Sistemas de Organización del Conocimiento existentes.

Por ejemplo, actualmente en una búsqueda en Google, el ordenador procesa la cadena de caracteres que le da el usuario en forma de palabras clave y la compara con el título y contenido de las páginas web que hay en la red. En el momento que encuentra esos mismos caracteres en diferentes registros, ofrece estos como resultado. Pero se trata de una forma de actuar no contextualizada. Es decir, si buscamos la palabra “banco”, los resultados obtenidos tanto pueden hacer referencia a cofradías de pescadores de bancos de atún, como a los diferentes aspectos económicos de entidades financieras. Obtendríamos una información excesiva en documentos de los cuales muy pocos o ninguno son relevantes.

Una solución a este problema de ruido documental consistiría representar la información de manera que los ordenadores puedan discriminar y ayudar. Esto implicaría que en vez de usar palabras clave en las búsquedas, se buscasen conceptos o bien que estas palabras estuviesen relacionadas con un dominio determinado aumentando además la precisión.

Todas estas ideas las asume el proyecto de Web Semántica ideado por Tim Berners Lee.

La Web Semántica se basa pues en un etiquetado sintáctico y semántico a nivel de vocabulario de los documentos. Es decir, se pretende que el contenido de la web no sólo sea recuperable a través de palabras sino también en función de su significado, teniendo que estructurarse en base a una lógica (Slavic, 2005).

Nuestro objetivo en el presente trabajo es proponer una lógica en los metadatos. Por ello nos basaremos en el uso de las clasificaciones –en concreto la CDU- como herramientas lógicas que contextualizan y en los encabezamientos de materia usados como descriptores siguiendo también una pauta lógica en su asignación para representar el contenido..

1. AMBITO DE ACCIÓN.

Para dar a entender nuestro proyecto, antes que nada hay que precisar que la Web Semántica se basa en un modelo multicapa donde figura el XML como parte sintáctica y metalenguaje que se usa para etiquetar y caracterizar los distintos bloques de texto de un documento, y por otro lado el RDF Schema o RDF(S). Es este último el que establece un modelo de datos y la forma en que se estructura la información para que los ordenadores puedan entenderla para darle una función semántica. En concreto nos referimos a la segunda parte estructurada de los ficheros o documentos RDF que es el



llamado cuerpo del fichero. Es aquí donde se pueden definir las propiedades de los recursos.

De las características del cuerpo del fichero sabemos de la posibilidad de añadir la información en un determinado orden. Esta característica es la que aprovecharíamos en nuestro sistema de asignación de metadatos de contenido, atendiendo a lo que nosotros llamamos ORDEN NUCLEAR.

Al ser RDF un vocabulario que se expresa con sintaxis XML, proponemos el vocabulario Dublín Core como vocabulario XML para representar la información, ya que reúne unas características que comentaremos más adelante.

2. SISTEMA DE INDIZACIÓN O ASIGNACIÓN DE METADATOS DE CONTENIDO EN LA WEB SEMÁNTICA.

Tim Berners Lee, en su propuesta de Web Semántica, propone añadir lógica a la web, pero anteriormente a él ya encontramos referencias de un aspecto necesario en los Sistemas de Organización del Conocimiento: El orden (Zeng, 2004).

El “Research Seminar on Compatibility and Integration Order Systems” organizado por la Internacional for Knowlwdge Organization (ISKO) y la Society for Professional Information defienden que en Internet hay dos infraestructuras paralelas que son: la tecnología de la red, en constante desarrollo, y la conceptual que teje la estructura del conocimiento:

“The first detailed recomendations from this seminar include the following statement:

“Further investigations should be carried out into the principles and methodology of establishing concordances between order systems with particular attention to order systems of different structure” (ISKO, 1995)”.

Un determinado orden lógico, nuclear, en los lenguajes alfabéticos es lo que se propone, por un lado el presente sistema y, por otro, su transliteración expresada mediante el uso de la CDU.

Proponemos un nuevo sistema de indización que usa el lenguaje de los Library of Congress Subject Headings (LCSH) como descriptores, rompiendo las cadenas precoordinadas y la Clasificación Decimal Universal (CDU), como ontología universal.

En concreto la metodología propuesta se basa en la siguiente premisa:

“El funcionamiento del sistema se basa en la equivalencia o transliteración de la CDU con un sistema de descriptores primarios asignados en un orden lógico (Nuclear) para representar el principal contenido del documento”



Los descriptores secundarios se expresaran en el mismo orden pero sin correspondencia con la CDU.

Es decir, se pretende dar una ordenación lógica a los términos de indización procedentes de los lenguajes documentales precoordinados (LCSH), por cuyo motivo creemos que será necesaria una adaptación de los subencabezamientos, así como fijar una nueva sintaxis lógica y normas que apuntarán a una simplificación del lenguaje en si: facetas que designan forma, geográficos y cronológicos.

Con este sistema se consigue un alto grado de especificidad dada la correspondencia entre un conjunto de descriptores y su CDU. Ningún conjunto de descriptores será mas general o más específico que su notación de CDU.

En lo que se refiere a la visualización de la descripción del documento, ésta sigue un orden específico e incluye descriptores e identificadores en el mismo campo. No se trata del orden en que aparecen en el documento, sino el orden de los principales conceptos del documento. Los descriptores secundarios se asignaran en un campo diferente definido por el sistema informático, y aparecerán en el mismo orden según el mismo criterio. Consideramos muy importante la posibilidad de ponderar los términos de indización para así poder remarcar qué temáticas son tratadas principalmente y cuales en un segundo término, así como qué puntos de vista u otras facetas predominan, ya que con ello se aumenta la precisión.

También se prevé que en una primera fase de implementación se podría habilitar un nuevo campo descriptivo de contenido para la indización de los registros que no fuera de términos controlados de búsqueda o de acceso complementarios para cada recurso. En el momento de asignar los metadatos que deben ir asociados a cada uno de ellos se realizaría utilizando directamente la teoría del orden nuclear de los distintos elementos relacionados y haciendo que se correspondiera con la notación clasificatoria.

Para poder entender mejor el sistema, debemos hacer mención de los índices permutados de descriptores/CDU y CDU/descriptores. Se trata de unos índices generados por el sistema que tiene que ver con cómo se ordenan y procesan los datos a nivel interno. Estos índices facilitan la comparación entre los términos y las notaciones dentro del mismo concepto según su proximidad, la cual representa la frecuencia de su uso. Con esta base podemos establecer si a un conjunto de descriptores le corresponde la misma notación de CDU para representar el mismo concepto. Los indizadores pueden así seguir un modelo cuando necesiten representar el mismo concepto, o en el caso de un indizador aislado, puede éste asegurarse de actuar consistentemente si el concepto aparece más de una vez. Y aquí cabe mencionar otra de las propiedades y ventajas de este sistema, que es la coherencia o consistencia, pues es uno de los mayores problemas de la indización. La siguiente figura permite visualizar además cuanta más o menos relevancia tiene cada palabra o término empleado en función del orden que el descriptor ocupa:

- Aritmética ; Educación primaria ; Didáctica 511:373.43
- Aritmética ; Educación primaria ; Innovación curricular 511:371.014.5
- Aritmética ; Educación primaria ; Innovación curricular 511:371.014.5
- Aritmética ; Educación primaria ; Innovación curricular 511:371.014.5
- Educación musical ; Educación primaria 78:373.3.02

Figura 1.- Ejemplo de Índice de descriptores/CDU en un catálogo de biblioteca.

En el contexto de Internet estos índices pueden contribuir a la eficacia de los motores de búsqueda ya que para dibujar estos clusters se basan en algoritmos.

Tim Berners Lee, en su planteamiento de Web Semántica apunta la idea de coherencia al igual que el proyecto Dublin Core (Hillmann):

“La comunidad del Dublín Core reconoce que la coherencia en la creación de metadatos es una clave importante para conseguir una recuperación completa y una visualización inteligible en medio de fuentes dispares de registros descriptivos. Los metadatos incoherentes esconden efectivamente los registros deseados y producen unos resultados en la búsqueda que son incompletos, imprevisibles o desiguales”

En efecto, como estándar y módulo complementario del RDF, ha nacido el Dublin Core. Concretamente en el elemento “materia” sugiere unos determinados vocabularios controlados juntamente con palabras clave. Los vocabularios recomendados son: LCSH, los encabezamientos de Materias Medicas, la Clasificación de la LC, la Clasificación de Dewey y la Clasificación Decimal Universal. El Subcommittee de la American Library Association presentó en 1999 un informe con recomendaciones en la línea de considerar para el Dublin Core y, en general, el uso de texto –preferiblemente de lenguajes controlados- y datos de clasificación. También la inclusión en el elemento temático de descriptores junto con palabras libres para una mayor precisión y añadir un poco de flexibilidad (elementos “dc:title”, “dc:abstract”) que se suma a las posibilidades que ofrecen los lenguajes controlados para aumentar los resultados y permitir el browsing.

Dado que el Dublin Core es extensible, es decir, que reconoce la importancia de proporcionar un mecanismo para ampliar el conjunto de elementos, creemos que los descriptores secundarios, en nuestro caso podrían darse en otro elemento “dc:subject” diseñado para tal fin. Esto permitiría conseguir un aumento de la precisión.

3. POR QUÉ EL USO DE LA CDU?

Tal como hacíamos referencia en la introducción de la presente comunicación, Internet precisa herramientas que contextualicen la información, que ayuden a desambiguar el problema de los homónimos. Además, en el entorno digital hace falta un sistema que



permita agrupar las distintas unidades de contenido u objetos en clusters que ofrezcan, a su vez, un sistema de búsqueda por navegación eficaz, y donde las diferentes notaciones y conceptos estén vinculados y relacionados jerárquica y sistemáticamente con un carácter universal como las ontologías.

Es en este aspecto donde toma importancia el papel de las clasificaciones, especialmente las facetadas y sintéticas como es el caso de la Clasificación Decimal Universal.

Entre muchas de las aportaciones y aplicaciones que la CDU nos ofrece destacaríamos:

- Los códigos de clasificación pueden ser usados como puente para salvar los problemas que presentan los registros en diferentes lenguas.
- Puede ayudar a organizar los recursos y a construir una arquitectura de la información semántica de forma global e integral. Cómo la organización de conceptos está basada solo en relaciones temáticas y no alfabéticas, los temas o conceptos que están dentro de una disciplina no están dispersos (como ocurre con las clasificaciones alfabéticas). Por lo tanto, cada término expresado en forma de expresión alfabética va a estar ubicado dentro de una jerarquía y así pues dentro de una red semántica, ayudando a solventar problemas de homonimia, al estar estos conceptos relacionados con la disciplina en cuestión, evitando también cualquier ambigüedad. En definitiva, se trata de la polijerarquía que en el contexto de los lenguajes controlados es la relación jerárquica que se establece entre un concepto específico y dos o más que le son genéricos.
- A partir de la confección inicial de una notación, ésta permite ser usada muchas veces y, por lo tanto, resulta a largo plazo económica.
- Permite la construcción de notaciones equivalentes a una cadena de descriptores.
- Es flexible al permitir la combinación de elementos (estructura sintética) y adaptar la ordenación de subdivisiones según las necesidades.
- Permite la simplificación en la representación de conceptos ya que la sintaxis se puede aplicar para relacionar unos términos con otros.
- Engloba todas las disciplinas del saber.
- Es escalable, es decir, permite decidir el nivel de especificidad que se quiera alcanzar en el momento de la clasificación a partir de diferentes criterios. Al tratarse de una clasificación numérica ordenada por el principio de los números decimales, cabe la posibilidad de que un número determinado sea susceptible de

un aumento indefinido, y que pueda ser subdividido indefinidamente por sucesiva agregación de cifras.

- La CDU no solo se puede utilizar para organizar y visualizar recursos a nivel de interficie sino también como metadatos en el momento de indicarlos. Entendemos así que el proceso de recuperación más eficiente es el que se lleva a cabo sobre la descripción de los documentos y no sobre los documentos mismos a texto completo, tal y como funcionan la mayoría de los buscadores de la web actual.

Uno de los primeros proyectos internacionales destinado a organizar el control de autoridades de materia combinándolo con la CDU fue Oxford University's HUMBUL project (2002). Humbul fue uno de los mayores dominios temáticos de la web para la investigación en humanidades..

Antes del paso a llamarse Intute el 2006, en la visualización de los registros obtenidos de una búsqueda en Humbul se marcaba una distinción tipográfica y de disposición entre los descriptores con dos objetivos:

- ◆ Se marcaban en negrita las categorías de primer nivel, seguidas respectivamente de las de segundo nivel.
- ◆ La disposición de los distintos elementos podría seguir un orden de mayor a menor relevancia en función del contenido y la ocurrencia de cada término en el registro.
- ◆ La notación del código de clasificación y el valor alfabético se daban mezclados.

Figura 2.- Ejemplo de recurso web de la pasarela temática Humbul (2006)

Aberdeen art gallery and museums

- This Web site encourages exploration and understanding of and participation around the collection of Aberdeen Art Gallery and Museums. A selection of images have been digitized to represent the scope and depth of the collections and arranged in virtual tours, including some of relevance for art history, maritime history, the history of science and industry, archaeology and numismatics. ... ([View Full Record](#))
- [Museums, Libraries, Archives](#) [Archaeology](#) [antiquities / analysis of artifacts](#) [museums / art galleries](#)
[History](#) [numismatics](#) [history of science](#) [maritime / naval history](#) [art / architecture history](#) [museums / art galleries](#)
- <http://www.aagm.co.uk/code/emuseum.asp>
- Catalogued by : Alun Edwards on 2003-02-27

Dentro del proyecto Intute: Arts and Humanities, se explicita y registra en diversos campos los distintos valores de contenido; eso sí, siempre en lenguaje verbal, es decir que en el campo de clasificación se da la expresión alfabética y, por lo tanto, el código numérico de la CDU es implícito. Además, viendo la descripción del mismo recurso vemos que se añade el elemento “Keywords-uncontrolled” y en el de clasificación aparecen los términos ordenados jerárquicamente.

Por ejemplo:

Virtual Training Suite	URL	http://www.aagm.co.uk/ [English]
Support materials	Keywords - controlled	Aberdeenshire ; Aberdeen Art Gallery ; Aberdeen Art Gallery ; Aberdeen Maritime Museum ; The Tolbooth (Aberdeen) ; Provost Skenes House (Aberdeen) ; Hall Russel (Firm) ; art galleries ; galleries ; museums ; events ; Aberdeen-Scotland-United Kingdom ; exhibitions
Intute events	Keywords - uncontrolled	North Sea ; whaling ; oceanographic instruments
Features	Type	Datasets; Images; Other organisations
AHRC Projects	Classification	History > Museums / Art Galleries Archaeology > Museums / Art Galleries Cross-Disciplinary > Collections and Exhibitions > Museums and Galleries > UK Local > Scotland Museum, Library, and Archive Studies > Museums / Art Galleries History > Art / Architecture History History > Maritime / Naval History History > History of Science History > Antiquities
Blog	Responsibility	Aberdeen City Council, Aberdeen Art Gallery and Museums (Copyright holder)
Conferences & Events		
Critical Thinking		
Electronic journals		
Limelight		
MyIntute		
NW Film Archive		
On This Date		
People Index		
Timelines		

Figura 3.-Ejemplo de recurso web de la pasarela temática Intute (2008)

Las opciones de búsqueda que ofrece este sistema de recuperación de la información son múltiples y entre ellos hay la opción de navegar por un listado alfabético de las subclases temáticas de segundo nivel; por ejemplo, las correspondientes a la clase “Archeology”:

Conferences & Events	<ul style="list-style-type: none"> Academic Libraries / Archives / Museums Aegean PreHistory Agriculture / Horticulture Americas : Archaeology Antiquities / Analysis of Artifacts Archaeological Conservation Archaeological Drawing Archaeological Practice Archaeological Recording / Documentation Archaeological Sciences Archaeological Surveying / Prospection Archives Archives Collections Finding Aids Art History Asia : Archaeology Audiovisual / Specific Media Collections Australasia : Archaeology British Isles : Archaeology Byzantine Studies Computing and Archaeology Cultural / Political / Historical Surveys Cultural Heritage Digital Repositories Egyptology Environmental Archaeology Epigraphy Etruscan Studies 	<ul style="list-style-type: none"> Industrial Archaeology Islamic Studies Jewish Studies Landscape Archaeology Libraries Library Collections Finding Aids Listed Buildings Lithics Marine Archaeology Mesopotamian Studies Metallurgy Military Studies / Warfare and Defence Museum Collections Finding Aids Museums / Art Galleries Museums, Libraries, Archives National Libraries / Archives / Museums Numismatics Onomastics / Place-names Other African Archaeology Other European Archaeology Other Middle Eastern Archaeology Palaeography Papyrology Persian Studies Phoenician and Punic Studies Preservation of Monuments Provincial / Colonial Administration Religion / Ritual / Funerary Rescue Archaeology Roman Britain Rome
Critical Thinking		
Electronic journals		
Limelight		
MyIntute		
NW Film Archive		
On This Date		
People Index		
Timelines		

Figura 4.- Ejemplo de subclases vinculadas al campo arqueología (Arts & Intitute, 2008)

Esta pasarela temática es un ejemplo de recurso digital especializado que utiliza como puntos de acceso de contenido, a la vez como metadatos, descriptores, identificadores y un sistema de clasificación jerárquico. Además permite la navegación por la estructura así como también consultar otros documentos indizados con los mismos términos. Los distintos puntos de vista o disciplinas relacionadas desde las cuales es tratada la filosofía quedan reflejadas en las subclases de segundo nivel que así lo explicitan.

El principal problema que nos encontramos actualmente en el uso de la CDU, es la falta de un formato válido para gestionarla (Slavic, 2005, 2007) y conseguir una aplicación de software que sea robusta, flexible y escalable, que permitan optimizar y explotar todas sus posibilidades.

4. POR QUÉ EL USO DE LOS LCSH (Library of Congress Subject Headings) USADOS COMO DESCRIPTORES?

Si bien por un lado disponemos de herramientas que ayudan a organizar y contextualizar la información, como es el caso de la CDU, que salva las barreras lingüísticas y que engloba todo el conjunto de la ciencia humana mediante símbolos numéricos, necesitamos también la expresión alfabética de estos conceptos.

Por los motivos que vamos a detallar a continuación, los LCSH, han sido hasta ahora, en el ámbito de los sistemas tradicionales de recuperación, también utilizados en los



OPACS, el paradigma como sistema más representativo utilizado principalmente en centros de información multidisciplinares de Europa:

- Tienen un rico vocabulario que cubre todas las disciplinas del saber, aunque de manera general.
- Controlan sinónimos y homónimos.
- Tienen el soporte de la Library of Congress.
- Han estado traducidos a muchas lenguas y adoptados por muchas bibliotecas.
- Facilitan el *browsing*.
- Se encuentran en millones de registros bibliográficos.

Pero el rápido aumento de recursos disponibles en la web hace necesaria la existencia de una alternativa de esquema temático de metadatos más simple, que en el momento de la recuperación se adapte a los usuarios, sin necesidad de que éstos conozcan los lenguajes desde el punto de vista técnico, y ni tan a fondo como los profesionales de la información y documentación, puesto que: (Dean),

- La sintaxis de los LCSH no es compatible con muchos otros lenguajes controlados.
- Fuera de los OPACS, la estructura precoordinada de los LCSH no es adecuada para los buscadores que precisan un esquema postcoordinado más simple y adaptable a grandes volúmenes de información y a un entorno automatizado.
- La estructura precoordinada de las cadenas de encabezamientos es costosa de mantener, pues el uso de los LCSH requiere personal altamente cualificado que realice las tareas de control y revisión, por cuyo motivo cual no es susceptible de facilitar la indización automática.
- La sintaxis es compleja y precisa de normas complejas de aplicación.

Por ejemplo:

La subdivisión: “Razas” no se admite bajo el epígrafe: 650 7 Animales domésticos, en cambio si bajo el epígrafe: 650 7 Animales

Pero si sustituyéramos el uso de los encabezamientos por descriptores, contemplando su estricta aplicación, sin un orden, provocaría una falta de precisión en la recuperación de la información. De ahí que se proponga el uso de los LCSH pero usados como descriptores, rompiendo las cadenas precoordinadas y dando un orden que no tiene en cuenta las normas preestablecidas de los lenguajes precoordinados y que corresponde al uso de la lógica. A este orden le llamamos ORDEN NUCLEAR.



5. LOS LCSH (Library Of Congreso Subject Headings) /Vs. ORDEN NUCLEAR.

A continuación detallamos una serie de ejemplos de diferentes documentos indizados según LCSH, en concreto la traducción y adaptación que hace la LEMAC (Llista d'Encapçalaments de Matèria en Català) y clasificados según la 5ª ed. CDU (AENOR) en el catálogo de la Biblioteca de Catalunya y el sistema de indización que se propone en este artículo:

Ejemplo 1:

El documento:

Un cos sense límits : 26 octubre 2007-27 gener 2008. Barcelona : Fundació Joan Miró, 2007. 251 p. : il. col. ; 24 cm

Se trata del catálogo de una exposición de pintura moderna cuyo tema es la figura humana.

Según los LCSH:

Tenemos que utilizar dos entradas:

650 7 Figura humana en el arte|xExposiciones
650 7 Pintura moderna|yS. XXI|xExposiciones

Según el ORDEN NUCLEAR:

Figura humana ; Pintura Moderna ; S. XXI ; Exposiciones CDU:
75..041(061.4)"20"

Ejemplo 2:

El documento:

La Cuina del bosc / Duran, Nuria, [et al.].. 2ª ed. Barcelona : Pòrtic, 2008. 191 p. :il. col. ; 22 cm.(Portic Natura ; 20)

Es una guía de plantas silvestres comestibles que a la vez ofrece recetas de cocina con cada una de estas plantas.

Según los LCSH:

650 7 Plantas silvestres comestibles|xGuías
650 7 Cocina



Según el ORDEN NUCLEAR:

Como los dos conceptos corresponden a temas principales de documento, utilizaríamos:

Plantas silvestres comestibles|xGuías CDU: 633.82(036)
Plantas silvestres comestibles|xCocina CDU: 633.82:641.5(083.2)

Ejemplo 3:

I Congreso internacional sobre sesgo de género y desigualdades en la calidad académica : foros y talleres de trabajo. Bellaterra : Universidad Autónoma de Barcelona, Observatori per la Igualtat, 2008. 143 p., taules ; 23 cm

El documento apunta el problema de la discriminación sexual en cuanto a la evaluación de la calidad en la educación de los profesores universitarios según el sexo.

Según los LCSH:

650 7 Discriminación sexual en la educación universitaria|xCongresos
650 7 Profesores universitarios|xEvaluación|xCongresos
650 7 Educación universitaria|xControl de calidad|xCongresos
650 7 Igualdad entre sexos|Congresos

Según el ORDEN NUCLEAR:

Profesores universitarios ; Discriminación sexual ; Evaluación ; Congresos
CDU:378.12:159.922.1(061.3)
Educación universitaria ; Control de calidad ; Congresos
CDU:378.14(061.3)

Como se puede observar a veces en los LCSH, la precoordinación se hace en un mismo encabezamiento y a veces en más de uno, lo cual provoca confusión, ya que fácilmente, -aunque estos ejemplos no lo ilustren- se puede pensar que el documento habla de dos temas dispares. Esto cobra especial relevancia en casos donde los documentos tienen multitud de encabezamientos, pues además, al no haber ponderación (es decir al no establecer campos diferenciados entre conceptos primarios y conceptos secundarios que ofrecería precisión), la confusión aumenta por la falta de un orden.



6. EL SISTEMA DE RECUPERACIÓN EN QUE SE BASA EL PROYECTO DE METADATOS

En Internet la propuesta estaría basada en optar por una interficie de búsqueda simple fácil de usar y con diversas opciones adecuadamente indicadas.

Estableciendo los términos de búsqueda, se recurriría a operadores de proximidad o delimitación que contemplasen el orden lógico en que los términos obtenidos aparecieran, a fin de aportar a los resultados mayor o menor relevancia, teniendo así en cuenta la cadena lógica de términos. Es decir, se priorizaría según el orden nuclear de izquierda a derecha en lo que hace referencia al lenguaje controlado, teniendo en cuenta si se tratase de un término aceptado o no, y también según el campo donde el término apareciera en el recurso (descriptores primarios, secundarios, CDU, título, resumen). Por ello entendemos que habría una ponderación no sólo respecto a las premisas de calidad y precisión que ofrecen los descriptores primarios y secundarios, sino que además se pretende enfatizar la complementariedad de los lenguajes controlados con las palabras del lenguaje natural. Eso si situados estratégicamente, y con un filtro de calidad en algunos casos como por ejemplo el resumen. La notación decimal correspondiente no aparecería en la visualización del recurso, aunque sería una de las premisas principales de organización interna de la información.

En sistemas de información cerrados, como opción alternativa, se establecería un formulario de búsqueda avanzada que permitiría distinguir las materias primarias de las secundarias mediante el acceso por campos, y con la ayuda de un enlace que permitiría consultar en línea el tesoro o lenguaje empleado, con las notas de aplicación y alcance correspondientes. Este sistema ofrecería la ayuda a través de desplegables sucesivos de selección a tres niveles de jerarquía con todas las clases y subclases que tuvieran representación en los mismos, a modo de menú (expresión alfabética de las notaciones de la CDU). Al Igual que Humbul, en este sentido sería interesante poder navegar a partir de las grandes categorías o traducción alfabética de las notaciones que a manera de clusters agrupasen todos los recursos pertinentes utilizando los operadores booleanos. Algunos buscadores soportan y admiten la búsqueda por campos.

A manera de índice A-Z del entorno web se presentarán listados de enlaces entre los códigos de clasificación y los términos de indización a partir de los índices permutados generados automáticamente por el sistema así como también la posibilidad de consulta de las tablas de correspondencias entre ambos. Sería interesante que el programa ofreciera la posibilidad de ordenar los resultados alfabéticamente por materias, clases, grado de relevancia según el metadato de origen o grado de ponderación en los encabezamientos y orden nuclear de elementos.

Independientemente al ámbito en el que se efectúe la búsqueda, más o menos amplio, las redes de relaciones de los lenguajes documentales debe favorecer, por un lado, la consulta hipertextual a partir del conjunto de metadatos asignados a cada uno de ellos y



que remitan a los índices permutados, por ejemplo, y generar sugerencias o recomendar estrategias alternativas de recuperación en función de los términos introducidos: por ejemplo, "también encontraríais información en: ..." o "si su ámbito de interés es ... en relación a ... consulten; si es consulten ..." (de utilidad para descriptores que puedan formar parte de diferentes jerarquías o disciplinas: "Conocimiento, Teoría del (Religión)" o "Conocimiento, Teoría del (Sociología)").

7. CONCLUSIONES

La propuesta de esta nueva metodología pretende favorecer la relevancia y precisión en las búsquedas y beneficiarse de las múltiples funcionalidades que la clasificación y indización aportan.

Si bien la postcoordinación se asocia más a la recuperación a partir de palabras no controladas o del lenguaje natural, más cercano a los usuarios, mientras que los lenguajes controlados se asocian más como la recuperación de términos precoordinados, por medio del procedimiento descrito se garantiza la posibilidad de una mayor flexibilidad y sintaxis respecto la búsqueda postcoordinada.

Se trata pues de aprovechar las propiedades para establecer un orden y ponderación que ofrece el entorno de la Web semántica para conseguir los objetivos anteriormente propuestos.

Un inconveniente sería el hecho que en la recuperación en Internet implicaría la existencia de un tesoro multilingüe o una determinada adaptación del LCSH. Así mismo se hace patente la necesidad de un formato adecuado para la gestión de la CDU, además de lograr una aplicación adecuada de software.

Con todo ello se podría conseguir que Internet funcionase de manera automática a partir del filtro de todos los datos relativos a los índices permutados con que previamente se habría alimentado el sistema, conjuntamente con las remisiones del lenguaje natural no aceptadas.

Creemos que este sistema podría ser también aplicable en el entorno de las bibliotecas tradicionales y también en buscadores de Internet.



BIBLIOGRAFIA

1. BERNERS-LEE, Tim; HENDLER, James and LASSILA, Ora. "The Semantic Web"
En: *Scientific American Magazine*, May, 17, 2001.
2. DEAN, Rebecca J. *FAST: Development of Simplified Headings for Metadata*. OCLC.
[consulta : 2 de noviembre de 2008]. Disponible en Web: <<http://www.oclc.org/research/projects/fast>>.
3. FRANKLIN, Rosemary Aud. "Re-Inventing Subject Access for the Semantic Web"
En: *Online Information Review*, 2003, vol. 27, no. 2. pp. 94-101.
4. GRANADOS, Mariàngels; and NICOLAU, Anna. *Improving Subject Searching in Databases through a Combination of Descriptors and UDC*. Zadar: Bobcatsss, 01/29, 2008. [consulta: 3 de noviembre de 2008]. Disponible en Web: <<http://edoc.hu-berlin.de/conferences/bobcatsss2008/granados-mariangels-151/PDF/granados.pdf>>
5. HILLMAN, Diane. *Utilitzar El Dublin Core*. [Consulta 2 de noviembre de 2008].
Disponible en Web: <http://www.bnc.es/sprof/normalitzacio/dc/dc_guide0.php>.
6. MCILWAINE, I. C. *Guía Para El Uso De La CDU*. Madrid: AENOR (Asociación Española de Normalización y Certificación), 2003. ISBN 84-8143-332-2.
7. SLAVIC, Aida. "On the Nature and Typology of Documentary Classifications and their use in a Networked Environment" En: *El Profesional De La Información*, 2007, vol. 16, no. 6. pp. 580-589.
8. SLAVIC, Aida. "Knowledge Organization Systems, Network Standards and Semantic Web" En: *Informacijske Znanosti u Procesu Promjena*, 2005. pp. p.5-22.
9. ZENG, Marcia Lei; and CHAN, Lois Mai. "Trends and Issues in Establishing Interoperability among Knowledge Organization Systems" En: *Journal of the American Society for Information Science and Technology*, 2004, vol. 5, no. 55. pp. 377-395.
10. O'NEIL, Ed; and CHAN, Lois m. *Faceted Application of Subject Terminology: A Faceted LCSH Based Subject Vocabulary*. New Orleans ed. ALA, June 24, 2006.
[consulta 3 de noviembre 2008]. Disponible en Web: <<http://www.oclc.org/research/projects/fast/ala06-5f.ppt#9>>