

Gestión de fuentes de información en ciencia y tecnología: desarrollo del programa CUMULUS

**Blázquez
Ochando,
Manuel**

Departamento de Biblioteconomía y Documentación, Facultad de
Ciencias de la Documentación, UCM.
C/Santísima Trinidad, 37. CP 28010, Madrid
Correo electrónico:
Sitio web personal: <http://www.mblazquez.es/>

Resumen:

Las fuentes de información especializadas en ciencia y tecnología carecen de sistemas adecuados para su gestión, catalogación y representación automatizada. Esto se debe a que los gestores bibliográficos priman un enfoque de descripción bibliográfica por encima de la consideración de fuente de información institucional y documental. Con el objetivo de experimentar un nuevo modelo de gestión de las fuentes de información basada en su estudio analítico, se ha diseñado el programa CUMULUS que aporta nuevas soluciones en el tratamiento documental de las fuentes y su relación semántica con terceros recursos. El sistema resultante es capaz de organizar las fuentes de información con una descripción y control exhaustivo de sus autoridades así como articular su recuperación semántica, lo que lo convierte en una nueva herramienta para tratar las fuentes y asistir en la enseñanza y formación de futuros documentalistas.

Palabras clave:

Fuentes de información, ciencia y tecnología, automatización, web semántica.

INTRODUCCIÓN

Las fuentes de información son canales y medios por los que se transmite información y documentación, relativas al desempeño de las actividades y funciones de un cierto sujeto productor. En el caso de las fuentes especializadas en ciencia y tecnología, las vías de transmisión de la información científica tienen su origen en los centros de investigación, instituciones universitarias e incluso entidades públicas y privadas (FRANCIS, A.T., 2008). Es necesario matizar que estos productores se basan a su vez en fuentes personales que transmiten sus conocimientos por medio de la documentación. De esta forma la institución proporciona los medios de comunicación necesarios para su difusión, configurándose todo un espectro de recursos que difícilmente puede ser controlado y gestionado en su totalidad por los principales centros de documentación científica.

El objeto de estudio es el desarrollo de un sistema automatizado para la gestión de fuentes de información en ciencia y tecnología, que ha dado como resultado el programa CUMULUS. La finalidad es ayudar a controlar mejor documentalmente las instituciones y centros de investigación, sus sujetos productores, los medios de comunicación científica, así como sus fuentes personales. Tal sistema ha sido creado para identificar, describir y relacionar dichas fuentes y facilitar su posterior representación y recuperación.

En relación a precedentes de aplicaciones similares, cabe destacar que no se encuentran sistemas diseñados específicamente para tal propósito. Sí que es cierto que existen sistemas de gestión bibliográfica (ALONSO ARÉVALO, J., 2007), que permiten la elaboración de bibliografías y repertorios a partir de las consultas efectuadas en catálogos automatizados. A pesar de que la documentación que se gestiona es una fuente de información en sí misma, ésta lo es a un bajo nivel en relación a las fuentes personales e institucionales que lo generan, pero además hay que considerar que también son fuentes aquellos servicios que proporcionan acceso a tales documentos. En cuanto a los CMS (Content Management System) (COYSH, S. and Jeffery, B., 2008), si bien permiten la introducción de directorios de enlaces, ello por sí mismo no configura un sistema que describa de forma exhaustiva y completa las fuentes de información en la web a partir de sus campos fundamentales.

METODOLOGÍA

Para llevar a cabo la investigación se han seguido los siguientes pasos; 1) Marco de gestión, 2) Requerimientos del sistema, 3) Diseño del sistema CUMULUS, 4) Pruebas y evaluación del sistema, 5) Experiencia aplicada a la enseñanza.

1) Marco de gestión

Las fuentes de información en ciencia y tecnología (ALONSO ARÉVALO, J., 2005) constituyen un activo que puede ser abordado en su gestión documental desde múltiples puntos de vista. La catalogación tradicional es una perspectiva que si bien tiene en cuenta los asientos bibliográficos fundamentales de cualquier fuente de información, no siempre resulta satisfactoria la descripción que de esta se obtiene. La razón elemental de este hecho reside en que las fuentes de información también pueden estar constituidas por las propias instituciones e investigadores que generan la documentación científica. En consecuencia la catalogación debe ser flexible y adaptarse a las necesidades de cada tipo de fuente.

El marco de gestión debe permitir la identificación de cada fuente según su tipología y determinar su orden jerárquico con respecto al resto de fuentes. Esto implica que existe un orden para analizar las fuentes de información en ciencia y tecnología, tal y como se muestra en la *figura 1*.

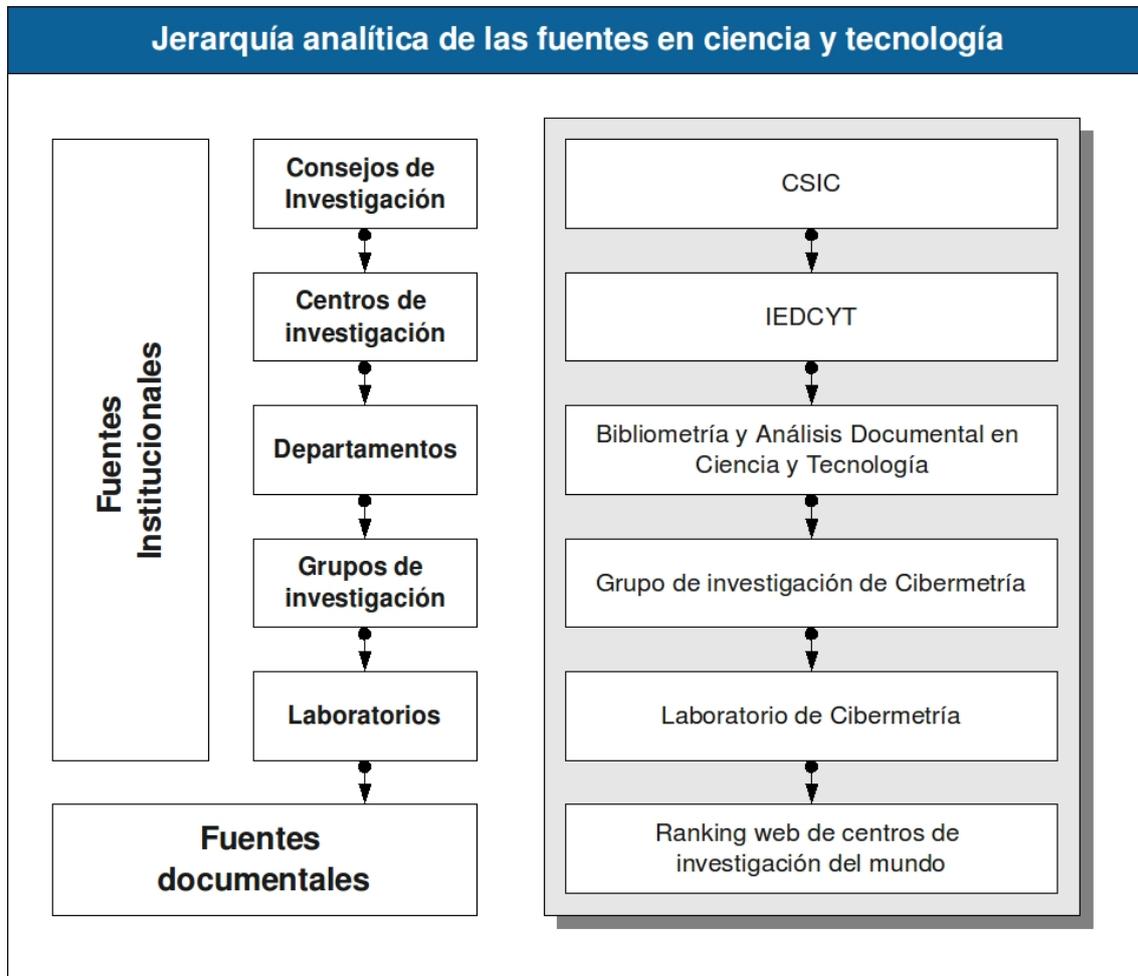


Figura 1: Ejemplo de jerarquía analítica del CSIC

Tal y como se muestra en la *figura 1* las fuentes documentales dependen directamente del trabajo desarrollado en laboratorios, de las investigaciones de los grupos de investigación, que conforman departamentos y centros de investigación, gestionados por consejos o consorcios de investigación nacionales o internacionales. Por este motivo se propone un marco de gestión en el que se pone de relieve las relaciones que existen entre las distintas fuentes, a fin de conocer su dependencia y facilitar posteriores trabajos de ordenación y recuperación semántica.

2) Requerimientos del sistema

El sistema CUMULUS es una aplicación web programada en PHP y soportada por la base de datos MySQL, constituyendo de esta forma el grueso del programa. A parte incorpora otras tecnologías basadas en lenguajes de marcado XML, RDF (RDF/XML Syntax Specification, 2004), XSL (STAYTON, B., 2007) y XPath (XML Path Language XPath W3C Consortium, 1999) que actúan en el plano de la representación y recuperación de las fuentes de información y sus contenidos. La elección de estas tecnologías está justificada con el objetivo de reducir al mínimo los requerimientos de instalación de CUMULUS en cualquier servidor web Apache.

3) Diseño del sistema CUMULUS

El diseño del sistema se basa en funciones y módulos. Las funciones son las operaciones principales del sistema, mientras que los módulos son programas que asisten a las funciones para desempeñar una determinada tarea. En la *tabla1* se exponen las funciones y módulos más importantes del sistema, su clasificación y descripción.

<i>Tabla1: Funciones y módulos esenciales del sistema CUMULUS</i>			
Nombre	Tipo	Clasificación	Descripción
Insertar func.insert.php	Función	Catalogación	Permite catalogar y registrar cualquier fuente de información institucional y documental.
Editar func.edit.php	Función	Edición	Posibilita la recuperación y edición de las fuentes de información ya catalogadas y registradas.
Buscar func.search.php	Función	Recuperación	Permite la consulta interna para la recuperación de las fuentes.
Sindicación func.sync.php	Función	Representación	Función de control de la salida de canales de sindicación con el repertorio completo de fuentes catalogadas.
Web semántica func.rdfgen.php	Función	Representación y Recuperación	Función que construye la web semántica de las fuentes de información y sus relaciones con terceros recursos y autoridades.
Directorios func.makedir.php	Función	Representación y Recuperación	Función que construye un directorio web con interfaz de recuperación basada en la clasificación por tesoro, tipificaciones, autoridades y relaciones.
Visor de registros func.viewer.php	Función	Representación	Función para la visualización completa de una fuente de información o recurso individualmente.
Administradores func.gestadmins.php	Función	Administración	Función para la gestión de los administradores de la aplicación.
Usuarios func.gestusers.php	Función	Administración	Función para la gestión de los usuarios de la aplicación.

Tesaurus mod.category.php	Módulo	Clasificación	Módulo para la edición de tesauros electrónicos integrados que asiste en el proceso de catalogación de las fuentes de información.
Autoridad personal mod.personal.php	Módulo	Gestión de autoridades	Módulo para la gestión de las autoridades personales y su aplicación en el proceso de catalogación.
Autoridad corporativa mod.corporate.php	Módulo	Gestión de autoridades	Módulo para la gestión de las autoridades corporativas, entidades e instituciones.
Autoridad editorial mod.publisher.php	Módulo	Gestión de autoridades	Módulo para la gestión de las autoridades editoriales.
Autoridad geográfica mod.place.php	Módulo	Gestión de autoridades	Módulo para la gestión de las autoridades geográficas.
Tipología web mod.typeref1.php	Módulo	Tipificación	Módulo para la gestión de tipologías web que sirven para tipificar la fuente según su representación en la red.
Tipología origen mod.typeref2.php	Módulo	Tipificación	Módulo para la gestión de tipologías según el origen de las fuentes de información, institucionales, documentales, etc.
Tipología nivel mod.typeref3.php	Módulo	Tipificación	Módulo para la gestión de tipologías según el nivel de la fuente, primaria, secundaria, terciaria, etc.
Tipología contenido mod.typeref4.php	Módulo	Tipificación	Módulo para la gestión de tipologías según el contenido, permite definir el tipo documental.
Relaciones mod.linkrel.php	Módulo	Relación	Módulo para el establecimiento y configuración de relaciones entre las distintas fuentes de información.

La estructura de CUMULUS, tal como demuestra la *tabla1*, contempla todas las operaciones y funciones necesarias para la catalogación, clasificación, control de autoridades, interrelación con terceras fuentes y recursos, recuperación y representación de forma tal, que cumple el marco de gestión propuesto anteriormente. A continuación se detallan todas las funciones y su relación con cada módulo;

Funciones de edición: insertar y modificar fuentes de información

Constituyen funciones de edición que posibilitan la catalogación completa de una fuente de información. Para ello cuenta con la asistencia de los módulos de autoridades personales, corporativas, editoriales, geográficas, clasificación por tesauro y configuración de relaciones con terceras fuentes de información. La pantalla de edición está diseñada para mantener visible casi toda la información vital de la catalogación, por lo que todos los módulos son visibles y editables en la misma pantalla por medio de un visor frontal que presenta en cada momento la información demandada.

The screenshot shows the Cumulus system's editing interface. The top header includes the Cumulus logo and the text 'Sistema de Gestión de Fuentes de Información'. The user is identified as 'Administrador: Manuel Blázquez' with a 'Cerrar sesión' link. The interface is divided into several sections:

- Identificación:** Fields for Título, Subtítulo, Sitio web (htmlUrl), and Canal de sindicación (xmlUrl). There are 'X Borrar' buttons next to the title and URL fields.
- Tipología web:** A dropdown menu.
- Tipología según el origen:** A dropdown menu with an 'X Borrar' button.
- Tipología según el nivel:** A dropdown menu with an 'X Borrar' button.
- Tipología según el contenido:** A dropdown menu with an 'X Borrar' button.
- Resumen (255 palabras):** A text area.
- Autoridades:** Fields for Idioma and Derechos.
- Fecha de publicación:** Fields for Día, Mes, and Año.
- Gestión:** Fields for Fecha de catalogación del recurso o fuente de información and usuario. There are 'Guardar registro' and 'Borrar datos' buttons.
- Contenidos:** A section with a 'relación de contenidos' and a 'Página1 - Página2 - Página3 - Página4 - Página5' navigation.
- Visor Frontal:** A section with a '+ Mostrar Visor Frontal' button and several links for navigation.
- Tipologías según el contenido:** A table with columns for denominación, descripción, and actions (asignar, modificar, eliminar). The table lists various document types like 'Actas de congresos', 'Artículos científicos', 'Artículos de divulgación', and 'Índices bibliográficos'.

Figura 2: Impresión de pantalla del módulo de edición

En cuanto al esquema de catalogación, está basado en parte en las principales áreas comunes entre las Normas MARC y las reglas de catalogación, a saber área de título (otras formas del título, subtítulo) y mención de responsabilidad (autoridades personales y corporativas), área de publicación (fecha de publicación, editorial y lugar). Además contempla la introducción de áreas de descripción por páginas atendiendo a los asuntos que se determinen, en este caso los valores por defecto permiten elegir entre descripción de servicios de información y documentación, estructura y organización institucional, publicaciones científicas, revistas y series, centros de investigación dependientes, bases de datos, etc.

Módulos de autoridades

Los módulos de autoridades permiten la gestión de personas, entidades, instituciones, editores, editoriales y lugares por medio de un esquema de descripción común que tiene en cuenta su denominación completa, tipo de autoridad, descripción, ocupación y referencia web.

Módulo de tesauro o categoría temática

Su principal función es la edición de un tesauro o grupo de categorías temáticas jerárquicas de hasta tres niveles. Este instrumento permite clasificar automáticamente las fuentes de información mediante la asignación de descriptores que heredan la estructura jerárquica de origen. Su funcionamiento ha sido integrado en las funciones de edición y catalogación del sistema CUMULUS.

Módulos de tipificaciones

Su función es tipificar las fuentes de información, a partir de su clasificación según el tipo de recurso electrónico, tipo de fuente según el origen (institucional, personal, documental), tipo de fuente según el nivel (primaria, secundaria, terciaria, complementaria) y tipo de fuente según su contenido (monografías, actas, revistas, artículos, catálogos, bases de datos, directorios, repertorios, etc.). Su funcionamiento al igual que el resto de módulos ha sido integrado directamente en las funciones de edición.

Módulo de relaciones

Permite configurar hasta 10 relaciones de una fuente de información con respecto a otras fuentes y recursos. El módulo de relaciones contempla diversos tipos que son utilizados para elaborar la web semántica y el directorio de fuentes. Los tipos de relaciones incorporados son los presentes en la *tabla2*.

Relación	Descripción
Relación de subordinación (relSub)	La relación de subordinación permite expresar qué recursos dependen jerárquicamente del que se está describiendo y catalogando, como por ejemplo el instituto IEDCYT subordinado al consejo de investigación CSIC.
Relación entre iguales (relEqu)	La relación entre iguales permite conocer que otras fuentes y recursos se encuentran al mismo nivel en cuanto a contenidos y tipología. Un ejemplo de ello pueden ser los distintos consejos de investigación europeos, como el CNRS Francés, el CSIC español, el CNR italiano, etc.
Relación alternativa (relAlt)	La relación alternativa está diseñada específicamente para contenidos y fuentes alternativas para abordar una temática o área de conocimiento. Por ejemplo el Centro de Investigación de Biología Molecular Severo Ochoa puede completarse con la fuente del ISIB CNR instituto de ingeniería biomédica italiano.
Relación es parte de (isPartOf)	La relación es parte de permite definir que una fuente o recurso es una parte de un todo formado por terceros recursos. Un ejemplo de ello es la relación que puede existir entre diferentes ediciones de un congreso científico.
Relación es definido por (isDefinedBy)	Hace posible la definición de fuentes de información a partir de terceros recursos y fuentes. Utilizado en recursos que abordan o estudian la propia fuente que se está describiendo. Un ejemplo de esta relación es la que se encuentra en el sitio web personal de un investigador y su currículum, como documento o recurso que lo define.
Relación es servicio de (isServiceOf)	Relaciona servicios de tipo informativo, documental, científico o transaccional propios de una fuente

	institucional.
Relación es recurso de (isResourceOf)	Relaciona habitualmente un recurso o fuente documental perteneciente a una fuente institucional. Por ejemplo la base de datos ICYT perteneciente al IEDCYT.

Funciones de representación: directorio, visor y canal de sindicación

Las funciones de representación han sido diseñadas para facilitar el acceso a la información a cualquier usuario registrado o no en la aplicación CUMULUS, para ello se dispone de un directorio que ordena y representa todos los contenidos catalogados y clasificados por su valor Pagerank y Alexa, así como por sus fechas de actualización. También pueden utilizarse filtros de categorías del tesoro y sus autoridades. En relación al visor, consiste en un sencillo programa que permite visualizar de forma individual y completa toda la información de las fuentes en las que se desea profundizar sobre sus contenidos, mostrando una ficha exhaustiva. En cuanto al canal de sindicación se refiere, permite al usuario conocer las últimas fuentes y recursos descritos en el sistema.

Funciones de recuperación: búsqueda interna y recuperación semántica

Las funciones de recuperación tienen por objeto recuperar las fuentes y recursos catalogados. Para el usuario registrado, se ofrece una modalidad de recuperación interna basada en la función FULLTEXT de SQL, que emplea operadores booleanos y fichero inverso de la base de datos. Para todos los usuarios existe un sistema de recuperación semántica básico que opera en base a RDF y permite obtener datos relacionados con una búsqueda dada.

4) Pruebas y evaluación del sistema

Con la finalidad de someter a prueba el sistema, se trabaja en la catalogación y clasificación analíticas de una muestra de 50 fuentes de información del Consejo Superior de Investigaciones Científicas CSIC y del Instituto de Estudios Documentales en Ciencia y Tecnología IEDCYT, antiguo CINDOC, tratando de establecer un cuadro de relaciones entre ambos recursos hasta la consecución de las fuentes propiamente documentales. El desarrollo de estos trabajos pueden ser consultados en la web del sistema CUMULUS, disponible en: <http://www.mblazquez.es/testbench/cumulus/>

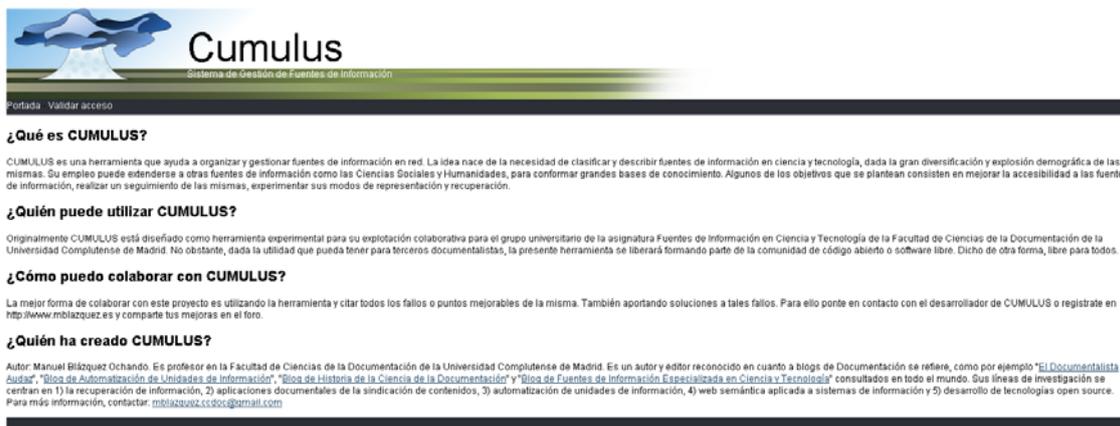


Figura 3: Portada del programa CUMULUS

El resultado obtenido hasta el momento permite incidir en algunos aspectos mejorables del sistema, como la necesidad de perfeccionar los perfiles de catalogación, a fin de que se adapten mejor a la descripción de recursos bibliográficos. Esto sucede fundamentalmente cuando se alcanza un nivel de descripción de monografías o artículos de publicaciones científicas. Ello es debido a que el sistema CUMULUS no ha sido concebido como un programa de gestión bibliográfica. A pesar de todo, el sistema CUMULUS permite una descripción suficiente que sí permite dar acceso a este tipo de documentos máxime cuando se encuentran disponibles en red, debido ello a la introducción de campos de control de URLs web y de canales de sindicación propias de cada recurso.

5) Experiencia aplicada a la enseñanza

El sistema CUMULUS es una aplicación útil para la enseñanza de las fuentes de información especializadas en ciencia y tecnología. Muestra de ello es la experiencia que se está llevando a cabo con los alumnos del grupo de fuentes de información de 2º de diplomatura de la Facultad de Ciencias de la Documentación de la Universidad Complutense, en el presente curso académico.

Se ha propuesto un trabajo de curso que consiste en analizar diversos centros de investigación, asignados individualmente a cada alumno a partir del (Ranking Web of Research Centers, 2010) que elabora todos los años el Laboratorio de Cibermetría (CybermetricsLabs: Observatorio de ciencia y tecnología en internet, 2008) del CSIC. Aprovechando el trabajo de identificación básica de estas fuentes, es decir su denominación y sitio web, el alumno efectúa una catalogación completa basada en los campos y opciones del programa CUMULUS, expuestas anteriormente. Su objetivo es profundizar en el conocimiento de la institución, la identificación y catalogación de fuentes relacionadas, hasta obtener las fuentes documentales, bases de datos, repertorios, revistas y medios de comunicación pertinentes, propios al centro o consorcio de investigación analizado. Esta experiencia permite al alumno familiarizarse con el análisis de las fuentes de información, su gestión y organización, poner en práctica sus conocimientos de análisis documental y lenguajes documentales. De cara a la investigación y la Documentación, el trabajo que los alumnos realizan permitirá construir la primera red semántica de fuentes de información especializadas, así como un sistema repertorial completo que puede ser de utilidad para la documentación e investigadores.

CONCLUSIONES

1. El programa CUMULUS dota a los documentalistas de un instrumento nuevo para catalogar y clasificar fuentes de información especializadas, de forma tal que puedan generarse directorios completos con tales contenidos.
2. La gestión de las fuentes de información en ciencia y tecnología se basa en la descripción analítica multinivel, que contempla las fuentes institucionales implicadas, jerárquicamente relacionadas y sus fuentes documentales y personales dependientes.
3. Las relaciones semánticas entre las fuentes de información pueden constituir un buen método de acceso a los recursos descritos, potenciando la recuperación de

la información científica relevante, no prevista en origen por el usuario.

4. La implantación de sistemas automatizados como el CUMULUS, puede asistir la enseñanza en Ciencias de la Documentación. Ello debe plantear un debate en torno al papel de la automatización en el desarrollo de nuevas herramientas documentales, que en el caso de las fuentes de información podrían haber sido desatendidas.

BIBLIOGRAFÍA

ALONSO ARÉVALO, J. 2005. Comunicación científica y edición alternativa. Visibilidad y fuentes de información. *ByD*.

ALONSO ARÉVALO, J. 2007. Gestores de referencias: Endnote, Procite, RefWorks. *In: Biblioteca Virtual en Ciencias de la Salud*. Salamanca.

COYSH, S. and B. JEFFERY. 2008. Decanting the Library within a Course management System CMS. *In: 37th Workshop on Instruction in Library Use*. British Columbia.

CybermetricsLabs: Observatorio de ciencia y tecnología en internet. 2008. [online]. [Accessed 15 Feb 2010]. Available from World Wide Web: < HYPERLINK "<http://internetlab.cindoc.csic.es/>" <http://internetlab.cindoc.csic.es/> >

FRANCIS, A.T. 2008. Information management in science and technology. Macmillan, pp.115-127.

Ranking Web of Research Centers. 2010. [online]. [Accessed 12 Feb 2010]. Available from World Wide Web: < HYPERLINK "<http://research.webometrics.info/>" <http://research.webometrics.info/> >

RDF/XML Syntax Specification. 2004. [online]. [Accessed 10 May 2009]. Available from World Wide Web: < HYPERLINK "<http://www.w3.org/TR/rdf-syntax-grammar/>" <http://www.w3.org/TR/rdf-syntax-grammar/> >

STAYTON, B. 2007. *DocBook XSL: The Complete Guide*. Santa Cruz: Sagehill.

XML Path Language XPath W3C Consortium. 1999. [online]. [Accessed 23 Feb 2007]. Available from World Wide Web: < HYPERLINK "<http://www.w3.org/TR/xpath>" <http://www.w3.org/TR/xpath> >