



SISTEMA DE ETIQUETADO AUTOMÁTICO DE CONTENIDOS EDUCATIVOS PARA FAVORECER SU REUSABILIDAD E INTEROPERABILIDAD

*AUTOMATIC LABELING SYSTEM FOR EDUCATIONAL CONTENT TO
PROMOTE THEIR REUSABILITY AND INTEROPERABILITY*

Jordi Juárez Mecías jordi.juarez@udl.cat
Óscar Flores Alarcia oscar.flores@udl.cat
Universidad de Lleida

RESUMEN

En este trabajo presentamos un proyecto de etiquetaje automático de objetos de aprendizaje. Destacando la importancia del etiquetado de materiales para favorecer su reusabilidad, se realiza un análisis de los problemas que nos encontramos para añadir metadatos a los recursos digitales. También se hace un estudio de alternativas al etiquetado manual de recursos así como las ventajas de un buen etiquetado de materiales. Como aplicación práctica se hace una aproximación a lo que sería un sistema de etiquetado automático de materiales, dando pautas sobre cómo montar un sistema autónomo de recopilación, etiquetado y organización de objetos de aprendizaje.

PALABRAS CLAVE: E-learning, metadatos, web semántica, buscador facetas, etiquetado automático.

ABSTRACT

The purpose of this research is to present a project of automatic labeling for learning objects. Focusing on the importance of labeling materials to promote their reusability, we have analyzed the problems we have encountered to add metadata to digital resources. We have also carried out a study of alternatives to manual labeling and of the advantages of adding correct metadata to our resources. As a practical application, we have developed some ideas about what an automatic labeling system would be like. We aim to give guidance about how to set up an autonomous system for the collection, labeling and organization of learning objects.

KEY WORDS: Elearning, metadata, semantic web, Faceted search, automatic metadata generation

1. INTRODUCCIÓN

En el libro “Las condiciones del aprendizaje”, Gagné (1987:2) define el aprendizaje como un cambio en la disposición o capacidad humana perdurable en el tiempo y que no es debido simplemente a los procesos de crecimiento biológico. Podemos decir que un proceso de enseñanza-aprendizaje es una situación en que el alumno interactúa con una serie de estímulos que le provocan un cambio en sus capacidades; en definitiva, que facilitan su aprendizaje.

Estos estímulos con los que interactúa el alumno pueden presentarse de diferentes maneras: en forma de explicaciones, de actividades, de situaciones concretas, de problemas, etc. Ya sea leyendo un texto o resolviendo un problema, el alumno siempre se enfrenta a un material de aprendizaje que le implica desarrollar alguna clase de actividad.

Hoy en día la web nos ofrece una fuente de recursos inagotables para el aprendizaje. Cualquier concepto, idea, procedimiento, lo encontraremos en Internet. Además, este recurso lo podemos buscar en diferentes formatos: textual, imagen fija, vídeos, animaciones, presentaciones... Entre otras cosas, este hecho aporta un gran ahorro económico ya que los costes de crear materiales desde cero son más elevados que reusar materiales ya desarrollados (Juárez y Flores, 2008:3). Son los denominados objetos de aprendizaje (Pire et al., 2012:2).

No existe consenso sobre la noción de objeto de aprendizaje. En un sentido amplio, las definiciones van desde aquellas que incluyen personas, objetivos, organizaciones y acontecimientos, hasta los que lo consideran como cualquier entidad digital que puede ser usada o referenciada durante un proceso de enseñanza-aprendizaje apoyado a través de las TIC (Morales y Agüera, 2002:26).

Existen muchas definiciones de OA, entre las cuales podemos destacar:

“Cualquier entidad, digital o no digital, que se puede utilizar, reutilizar o referirse durante un proceso de aprendizaje a través de las tecnologías” (IEEE Learning Technology Standards Committee, 2001:1).

“Cualquier recurso digital que se pueda reutilizar por apoyar el aprendizaje” (Wiley, 2000:7).

“Objeto de aprendizaje es una nueva manera de pensamiento del contenido de aprendizaje. Tradicionalmente, el contenido viene a través de un bloque de varias horas denominado curso. Los objetos de aprendizaje son unidades de aprendizaje muy pequeñas, con un extensión, por ejemplo, de entre 2 y 15 minutos” (Wisconsin Online Resource Center, [en línea]: 1).

“Un objeto de aprendizaje se define como la experiencia educacional independiente más pequeña que contiene un objetivo, una actividad de aprendizaje y una evaluación (...):

1. *Objetivo: componente estructural del OA que describe el resultado previsto de la actividad que se aprende.*

2. *Actividad de aprendizaje: componente estructural del OA que enseña el objetivo.*

3. *Evaluación: componente estructural del OA que determina si se ha logrado el objetivo” (L’Allier, 1997).*

Morales y Agüera (2002:26) proponen una serie de atributos esenciales que distinguen un OA de simples piezas de información:

- Debe ser un objeto educativo.
- Debe proporcionar una cantidad de conocimientos o habilidades relativamente pequeña.
- Debe ser autocontenido.
- Debe ser útil en más de una secuencia de instrucción.
- Debe ser fácil de identificar y por lo tanto de buscar.
- Debe ser independiente de un sistema administrador de aprendizaje.
- Debe ser accesible desde una gran variedad de plataformas.

Por su parte, Naharro (2007:4) expone las características que deberán cumplir los objetos de aprendizaje para asegurar la calidad en la creación de los mismos:

- Formato digital: puede ser actualizado y/o modificado constantemente; es utilizable desde Internet y accesible a diferentes personas simultáneamente desde distintos lugares.
- Propósito pedagógico: el objeto de aprendizaje incluye no sólo contenidos sino que también guía el propio proceso de aprendizaje del estudiante.
- Contenido interactivo: implica la participación activa de cada individuo (profesor-alumno/s) en el intercambio de información. Para ello es necesario que el objeto incluya actividades (ejercicios, simulaciones, cuestionarios, diagramas, gráficos, diapositivas, tablas, exámenes, experimentos, etc.) que permitan facilitar el proceso de asimilación y el seguimiento del progreso de cada alumno.
- Es indivisible e independiente de otros objetos de aprendizaje, por lo que debe tener sentido en sí mismo y ser autocontenido, y no puede descomponerse en partes más pequeñas.
- Es reutilizable en contextos educativos distintos a aquel para el que fue creado. Esta característica es uno de los principios que fundamentan el concepto de objeto de aprendizaje.

Desde un punto de vista formativo, las posibilidades que nos ofrecen los OAs son importantes, sobre todo por el hecho de disponer de bases de datos de diferentes tipos de OAs, los llamados repositorios, ya existentes en diferentes plataformas (Roy, Sarkar y Ghose, 2010:108):

- ARIADNE (<http://www.ariadne-eu.org>).
- National Science, Mathematics, Engineering, and Technology Education Digital Library (NSDL) (<http://www.smete.org>).
- Multimedia Educational Resources for Learning and Online Teaching (<http://www.merlot.org>).
- Health Education Assets Library (<http://www.healcentral.org>).
- Education Network Australia (<http://www.edna.edu.au/>).
- iLumina (<http://www.ilumina-dlib.org>)
- LearnAlberta Online Curriculum Repository (<http://www.learnalberta.ca/>).
- Campus Alberta repository of educational objects (<http://www.ucalgary.ca/commons/careo/>).
- LydiaLearn (<http://www.lydialearn.com/>).

Pero uno de los principales problemas que nos encontramos cuando queremos utilizar OAs es localizar los recursos que mejor se adapten a nuestras necesidades. Buscar un recurso de un tema determinado puede parecer fácil utilizando motores de búsqueda como Google, Bing o Yahoo. Pero si queremos hacer una búsqueda más detallada (idioma, dificultad, destinatario del recurso) nos encontramos con más dificultades, ya que estas herramientas no devuelven la información deseada o devuelven una gran cantidad de enlaces. Generalmente, el usuario se centra en los primeros resultados, que no siempre sirven si la búsqueda se hace considerando sólo las palabras clave (Pire et al., 2012:2).

Una manera de resolver este problema es añadir metadatos a los recursos educativos que se ofrecen en la web, de manera que les demos cierta semántica y facilitemos su almacenamiento, clasificación y reutilización. Los metadatos proporcionan información complementaria sobre un recurso, de forma que es más fácil organizarlos y reutilizarlos, y su finalidad sería obtener una descripción clara de un recurso sin tener que acceder a él. Pero a día de hoy, intentar utilizar metadatos para etiquetar un curso o un elemento digital comporta algunos problemas (Sonntag, 2004):

- Actualmente los buscadores generalistas no soportan búsquedas por metadatos. Existen repositorios de recursos que usan estándares de metadatos pero de una forma limitada. Una alternativa a los buscadores clásicos es el uso de los buscadores por facetas (Juárez, Flores y Gil, 2009), que favorecen la búsqueda por etiquetas, pero que requieren un buen etiquetado de los recursos a indexar.
- Otro problema es que la mayoría de los usuarios no añade metadatos a los recursos que crea. Incluso en repositorios especializados encontramos una falta de metadatos en los recursos. Una solución a este problema sería el desarrollo de herramientas que permitan un etiquetado automático de materiales, aunque en este caso encontraríamos la limitación de existir campos complejos de determinar (por ejemplo, algunos estándares incluyen etiquetas educativas como el nivel de interactividad del recurso, la densidad semántica, la edad del destinatario, etc.).
- Un tercer problema es la existencia de diferentes estándares de metadatos (Juárez y Flores, 2008:7; Roy, Sarkar y Ghose, 2010:105):

- The Dublin Core Metadata Initiative (<http://dublincore.org/>).
 - The IEEE Learning Object Metadata (<http://ltsc.ieee.org/wg12/index.html>).
 - The IMS Global Learning Consortium (<http://www.imsglobal.org/>).
 - The Advance Distributed Learning Initiative (<http://www.adlnet.org>).
 - The CanCore Learning Resource Metadata Initiative (<http://www.cancore.ca/>)
- Cada uno de estos estándares tiene sus características, por lo que se hace más complicado un consenso a la hora de categorizar un recurso. Éste es un problema que se minimizaría si se especificase correctamente el estándar a utilizar.
 - La calidad de los metadatos es otro problema que podemos encontrarnos, ya que actualmente el etiquetado depende del usuario final, por lo que éstos pueden ser incorrectos o poco precisos.

Ante esta problemática, el Área de Soporte a la Innovación Docente y E-learning, una unidad estructural del Instituto de Ciencias de la Educación – Centro de Formación Continua de la Universidad de Lleida, <http://www.ice.udl.cat/aside>, hace tiempo que desarrolla una línea de trabajo relacionada con el etiquetaje de contenidos y el desarrollo de la web semántica.

El proyecto que presentamos se centra en hacer una aproximación a lo que sería un sistema automático de etiquetado de materiales, que, como hemos comentado, es una de los principales problemas que consideramos que se debe resolver para conseguir que la web semántica sea una realidad.

El objetivo final es conseguir un sistema autónomo de etiquetado, de manera que el usuario sólo tenga que suministrar los materiales y el aplicativo los etiquete y organice.

2. ETIQUETADO AUTOMÁTICO DE MATERIALES: PROYECTO OC-E (OpenCalais-Exhibit)

Estamos de acuerdo en que el etiquetado de objetos de aprendizaje facilita su búsqueda y reutilización, pero la mayoría de las veces el etiquetado requiere de un esfuerzo adicional por parte del creador del material y muchas veces el vocabulario del estándar de metadatos hace que no sea fácil que la tarea sea realizada de forma manual. Igualmente muchos creadores de recursos ven el etiquetado de sus materiales como trabajo adicional y costoso, y son reacios a desarrollar esta actividad.

Por tanto, el etiquetado automático puede ser más eficiente, menos costoso y más consistente que un etiquetado manual. Existen también trabajos (Greenberg, 2004) que explican que los mejores resultado se obtienen combinando el etiquetado automático con el etiquetado manual.

Existen algunas herramientas de extracción automática de metadatos (Pire et al., 2012), entre las que podemos destacar:

- SAXEF. Es un sistema creado en la Universidad de Palermo (Italia) que extrae indicadores didácticos de una página web.
- Tan Way Yuen System. Es un sistema desarrollado en una tesis de Master desarrollada por Tan Way Yuen. Este sistema está basado en el estándar LOM.
- Looking4LO. Desarrollado por la Universidad de la República (Uruguay), es capaz de extraer metadatos de diferentes tipos de archivos.
- MAGIC. Es un sistema desarrollado por el centro de investigación IBM. Trabaja con el estándar SCORM

Para el desarrollo de nuestro proyecto nos pusimos en contacto con algunos de los creadores de estas herramientas, pero o bien habían abandonado la vía de investigación o estaban actualizando la herramienta por lo que no se pudieron hacer pruebas en entornos reales. En cualquier caso, no encontramos la implementación práctica de ninguna herramienta de extracción automática de metadatos.

Este motivo nos llevó a desarrollar una herramienta que consta de tres partes principales:

- Primera parte: obtener los materiales a etiquetar.
- Segunda parte: añadir metadatos a estos materiales.
- Tercera parte: montar un buscador para organizar los elementos categorizados.

A continuación explicaremos cada una de las partes.

Primera parte: obtener los materiales a etiquetar

La primera parte es obtener los materiales a etiquetar. En un entorno real los materiales serán suministrados directamente a la herramienta, pero en nuestras pruebas queríamos testear la herramienta con diferentes elementos, no necesariamente objetos de aprendizaje, por lo que se analizaron los RSS de varias webs para obtener sus materiales y etiquetarlos.

La finalidad principal es que todo sea lo más automático posible y transparente al usuario. Por ello los materiales se obtienen usando una librería que facilita la obtención de los elementos compartidos por RSS de una página web (<http://www.simplepie.org/>).



Imagen 1. Obtención de los materiales vía RSS.

Segunda parte: añadir metadatos a estos materiales

Una vez tenemos los elementos almacenados pasaremos a la segunda fase del proyecto, el etiquetado de los materiales. Para ello se usó OpenCalais (<http://www.opencalais.com/>), una herramienta de etiquetado automático. Aunque esta herramienta no está directamente pensada para etiquetar objetos de aprendizaje ni soporta estándares de e-learning, es una de las herramientas más avanzadas que hemos encontrado a la hora de añadir semántica a recursos.

Usando la API de OpenCalais podemos añadir metadatos a nuestros recursos electrónicos en formato texto, HTML o XML. Actualmente soporta tres idiomas, castellano, inglés y francés, pero el motor más desarrollado es el de inglés, por lo que las pruebas se hicieron en este idioma.

OpenCalais añade metadatos a los elementos a diferentes niveles. Por un lado, identifica elementos (Persona, Ciudad, Relación, Cargo) añadiéndoles semántica, y por otro lado añade metadatos para categorizar el elemento globalmente. Entre estos metadatos podemos destacar dos:

- *Document Categorization*: estas etiquetas intentan darnos una idea del tema del elemento. Coloquialmente sería decir “de que va” el recurso.
- *Social Tags*: estas etiqueta, que se añadieron en la última versión de OpenCalais, intentan emular cómo una persona categorizaría un recurso.

En esta proyecto, lo que hicimos fue añadir metadatos a los elementos y empaquetarlos en un fichero JSON. JSON es un formato para almacenar e intercambiar datos. Una de sus ventajas es

su simplicidad, lo que facilita analizar datos almacenados en este formato. (<http://es.wikipedia.org/wiki/JSON>).



Imagen 2: Flujo de trabajo del etiquetado de los materiales RSS usando OpenCalais.

Tercera parte: montar un buscador para organizar los elementos categorizados

Una vez tenemos los elementos etiquetados necesitamos alguna herramienta que nos permita acceder a ellos. En nuestro caso se usó una herramienta que permite crear buscadores por facetas a partir de ficheros JSON.

En un buscador por facetas el usuario va seleccionando categorías de manera que el aplicativo nos va mostrando los elementos que pertenecen a estas categorías. Estos buscadores son muy útiles cuando el usuario no conoce el vocabulario de búsqueda. Otra de las ventajas de estos buscadores es que no nos puede devolver búsquedas vacías (Juárez, Flores y Gil, 2009).

En nuestro caso las categorías (facetas) eran los metadatos devueltos por OpenCalais, y para ello se usó Exhibit (<http://www.simile-widgets.org/exhibit3/>), un framework que nos permite organizar grandes cantidades de datos de manera que se facilite el acceso a éstos. Exhibit ofrece diferentes formas de organizar la información, entre ellas crear un buscador por facetas a partir de unos datos de origen, que pueden venir de distintas fuentes. En nuestro caso, al ser una prueba de concepto, generamos un fichero JSON y a continuación lo copiamos en el entorno de trabajo de Exhibit, A continuación podemos ver una captura de Exhibit mostrando algunos de los artículos etiquetados. Podemos ver un listado con varios elementos y sus etiquetas. A la derecha podemos ver las diferentes facetas por las que podemos filtrar el contenido.

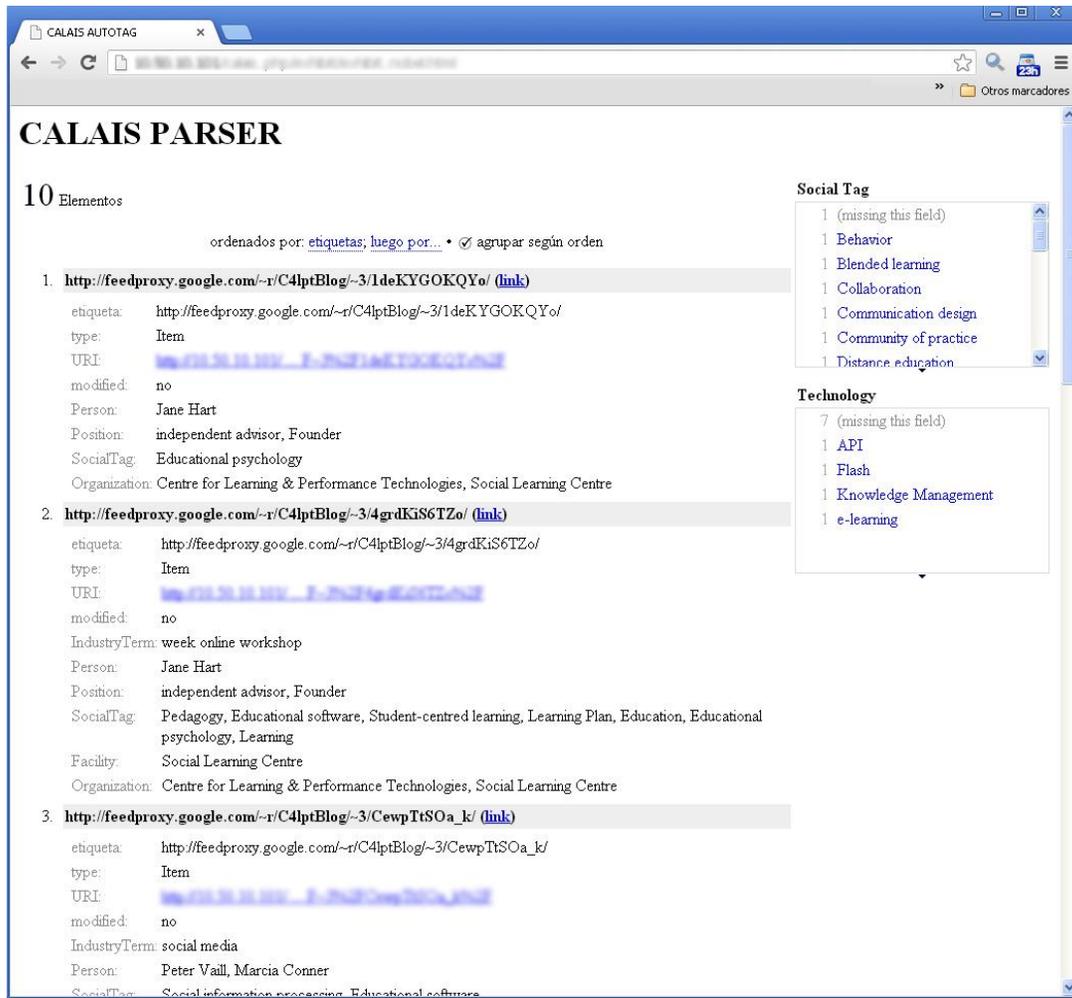


Imagen 3: Prototipo del buscador por facetas (Exhibit).

3. CONCLUSIONES

Añadir semántica a la web es un reto que muchos autores están intentando desarrollar pero que, como hemos visto en este trabajo, todavía se encuentra en una fase de desarrollo sin acabar de concretar. Hemos encontrado una importante falta de herramientas sólidas para realizar etiquetados automáticos, seguramente por la dificultad que entraña desarrollar este tipo de herramientas, y que corrobora la idea que todavía queda mucho camino por recorrer.

Desde nuestro punto de vista, creemos que el análisis desarrollado en nuestro proyecto demostramos la importancia del etiquetado de los objetos de aprendizaje para favorecer su reusabilidad y su interoperabilidad, un hecho que sin lugar a dudas favorece el desarrollo de los procesos formativos online.

El objetivo del proyecto era ofrecer unas líneas de lo que podría ser un sistema automático de etiquetado de objetos de aprendizaje, avanzando un paso más y diseñando un sistema de búsqueda basado en facetas, donde estas facetas serían los metadatos.

Parece claro que el problema principal en el etiquetado de materiales repercute en su correcta ejecución, sobretodo en categorías que podríamos definir como más “subjetivas”, como podría ser el nivel educativo recomendado del recurso, su nivel de dificultad, etc. En este caso, vemos que es muy difícil que el etiquetado no deba hacer manualmente.

En conclusión se ha observado que aún se tienen que mejorar mucho las herramientas de etiquetado automático y que seguramente la solución final consista en usar los dos sistemas, un etiquetado automático complementado con una manual.

4. REFERENCIAS

GAGNÉ, R.M. (1987). Las condiciones del aprendizaje. México DF: Nueva Editorial Interamericana.

GREENBERG, J. (2004, in press). Metadata Extraction and Harvesting: A Comparison of Two Automatic Metadata Generation Applications. *Journal of Internet Cataloging*, 6(4): 59-82. Disponible en <http://www.ils.unc.edu/mrc/pdf/greenberg04metadata.pdf> [Fecha de consulta: 20/12/12]

IEEE LEARNING TECHNOLOGY STANDARDS COMMITTEE (LTSC) (2001). IEEE P1484.12 Learning Object Metadata Working Group. WG12 Home page. Scope & Purpose. Disponible en <http://ltsc.ieee.org/wg12/>. [Fecha de consulta: 27/11/12].

JUÁREZ, J.; FLORES, Ò. (2008). La mejora del uso de las TIC en la docencia: etiquetar contenidos educativos para facilitar su reutilización. Actas del V Congreso Internacional “Docencia Universitaria e Innovación” ,Lleida, 2, 3 y 4 de julio de 2008. ISBN: 978-84-8458-279-3. Disponible en http://www.ice.udl.cat/aside/observatori/documents/10_cidui08_1749_119.pdf [Fecha de consulta: 27/11/12].

JUÁREZ, J.; FLORES, Ò.; GIL, R. (2009). Faceted browsing in repositories of learning objects. Actas de la V Conferencia Internacional sobre Multimedia y Tecnologías de la Información y la Comunicación en Educación (m-ICTE'2009). Lisboa, 22 - 24 d'abril de 2009. Disponible en http://www.ice.udl.cat/aside/observatori/documents/11_micte09_facetes.pdf [Fecha de consulta: 05/12/12]

L'ALLIER, J. J. (1997). Frame of Reference. NETg's Map to Its Products. Their Structures and Core Beliefs. Disponible en <http://web.archive.org/web/20020615192443/www.netg.com/research/whitepapers/frameref.asp>. [Fecha de consulta: 16/11/12].

MORALES, R.; AGÜERA, A. S. (2002). Capacitación basada en objetos reusables de aprendizaje. Boletín IIE, enero-febrero de 2002, 23-28. Disponible en <http://www.iie.org.mx/2002a/tendencias.pdf> [Fecha de consulta: 28/11/12].

NAHARRO, S. (et al.) (2007). "Los objetos de aprendizaje como recurso de calidad para la docencia: criterios de validación de objetos en la Universidad Politécnica de Valencia". En Actas del IV Simposio Pluridisciplinar sobre Diseño, Evaluación y Desarrollo de Contenidos Educativos Reutilizables. Bilbao: Universidad del País Vasco. 19, 20 y 21 de Septiembre 2007. ISBN: 978-84-8373-998-1. Disponible en <http://spdece07.ehu.es/actas/Naharro.pdf> [Fecha de consulta 24/04/08].

PIRE, T.; DECO, C.; CASALI, A.; ESPINASE, B. (2012) Extracción Automática de Metadatos de Objetos de Aprendizaje: un estudio comparativo. VI Congreso de Tecnología en Educación y Educación en Tecnología. Disponible en <http://sedici.unlp.edu.ar/handle/10915/18418> [Fecha de consulta: 27/11/12].

ROY, D.; SARKAR, S.; GHOSE, S. (2010). A Comparative Study of Learning Object Metadata, Learning Material Repositories, Metadata Annotation & an Automatic Metadata Annotation Tool. En *Joshi, M.; Boley, H.; Akerkar, R. (Eds.) (2010) Advances in Semantic Computing*. TMFR e-Book, v. 2, p. 103-126. Disponible en <http://www.tmrfindia.org/eseries/ebookV2-C6.pdf> [Fecha de consulta: 04/12/12].

SONNTAG, M. (2004). METADATA in E-Learning Applications: Automatic Extraction and Reuse. In: Christian Hofer, Gerhard Chroust (Eds.): IDIMT-2004. 12th Interdisciplinary Information Management Talks. Linz: Universitätsverlag Rudolf Trauner, p. 219-231 (ISBN 3-85487-665-3). Disponible en http://www.fim.uni-linz.ac.at/Publications/Sonntag/Metadata_in_E-Learning.pdf [Fecha de consulta: 04/12/12].

WILEY, D. (2000). Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy, a WILEY, D. (editor) (2000). *The Instructional Use of Learning Objects*. Disponible en <http://www.reusability.org/read>. [Fecha de consulta: 20/11/12].

WISCONSIN ONLINE RESOURCE CENTER [en línea]. What are learning objects? Disponible en <http://www.wisc-online.com/> [Fecha de consulta: 20/11/12].

Para citar este artículo:

JUÁREZ, J. & FLORES, O. (2013). Sistema de etiquetado automático de contenidos educativos para favorecer su reusabilidad e interoperabilidad. *EDUTEC, Revista Electrónica de Tecnología Educativa*, 46. Recuperado el dd/mm/aa de http://edutec.rediris.es/Revelec2/Revelec46/sistema_etiquetado_contenidos_educativos_reusabilidad_interoperabilidad.html

Fecha de recepción: 2013-02-06

Fecha de aceptación: 2013-11-22

Fecha de publicación: 2013-12-20