

Deco, Claudia ; Bender, Cristina ; Saer, Jorge

Ponderación de metadatos de recursos educativos como forma de mejorar los resultados de una búsqueda

Energeia, Año 9, N° 9, 2011

Este documento está disponible en la Biblioteca Digital de la Universidad Católica Argentina, repositorio institucional desarrollado por la Biblioteca Central "San Benito Abad". Su objetivo es difundir y preservar la producción intelectual de la institución.

La Biblioteca posee la autorización del autor para su divulgación en línea.

Cómo citar el documento:

Deco, C., Bender, C., Saer, J.. Ponderación de metadatos de recursos educativos como forma de mejorar los resultados de una búsqueda [en línea]. *Energeia*, 9(9), 2011. Disponible en:
<http://bibliotecadigital.uca.edu.ar/repositorio/revistas/ponderacion-metadatos-recursos-educativos-busqueda.pdf>

(Se recomienda indicar fecha de consulta al final de la cita. Ej: [Fecha de consulta: 19 de agosto de 2010]).

Proyecto “Mejora de la enseñanza de las ciencias en la carrera de Ingeniería Ambiental utilizando un sistema de recomendación de búsqueda personalizada de recursos educativos.”

Departamento de Investigación Institucional – Facultad de Química e Ingeniería “Fray Rogelio Bacon”
Pontificia Universidad Católica Argentina – Campus Rosario

Ponderación de metadatos de recursos educativos como forma de mejorar los resultados de una búsqueda

Claudia Deco¹, Cristina Bender¹, Jorge Saer¹

¹ Facultad de Química e Ingeniería “Fray Rogelio Bacon”, Pontificia Universidad Católica Argentina, Campus Rosario

Resumen. En el dominio de la educación existe gran cantidad y diversidad de material que puede ser utilizado en la enseñanza y que constituye una importante contribución al proceso enseñanza-aprendizaje. Mucho de este material es accesible a través de diferentes repositorios de objetos de aprendizaje, donde cada objeto tiene metadatos descriptivos. Estos metadatos permiten recuperar aquellos objetos que satisfacen no sólo el tema de la consulta, sino también el perfil de usuario, teniendo en cuenta sus características y preferencias. En este trabajo se analizan repositorios de objetos de aprendizaje y herramientas para la extracción automática de metadatos. Además, se presenta una propuesta para ordenar los resultados, aplicando ponderaciones a cada metadato en función del cumplimiento del estándar de metadatos para objetos de aprendizaje (LOM), y ponderar categorías de metadatos para ajustar los resultados al tipo de usuario.

Palabras clave: objetos de aprendizaje, metadatos, extracción de información.

Abstract. In educational domain there is great quantity and diversity of material that can be used in teaching and constitutes an important contribution to teaching-learning process. Much of this material is accessible through different learning object repositories, where each object has descriptive metadata. This metadata let the user retrieve objects that satisfy not only the query topic, but also his/her characteristics and preferences. In this paper we analyze learning object repositories and tools for automatic metadata extraction. We propose to sort the query results applying weights to each metadata. For this, we propose to analyze if metadata follow the metadata standard for learning objects (LOM), and also to weight each LOM metadata category. So, we can adjust results to the user profile.

Keywords: learning objects, metadata, information extraction.

Introducción

Existe actualmente gran cantidad y diversidad de material educativo disponible en la Web que puede ser utilizado y reutilizado en el proceso enseñanza-aprendizaje. Sin embargo, un mismo material puede no ser el adecuado para todos los usuarios, porque los usuarios poseen distintos estilos de aprendizaje así como características y preferencias personales. El proceso enseñanza-aprendizaje puede ser, entonces, mejorado si se utiliza material educativo acorde al perfil de cada usuario. La propuesta es mejorar la enseñanza de las ciencias utilizando un sistema recomendador de objetos de aprendizaje que personalice los resultados de una búsqueda según un perfil de usuario, considerando sus características y preferencias en el momento de la búsqueda.

Un objeto de aprendizaje es todo recurso digital que apoya a la educación y que puede ser reutilizado [1]. Estos recursos pueden ser empleados tanto por un alumno que desea aprender un tema como por los docentes para la preparación de material didáctico acorde a la temática a abordar, el tipo de material buscado, el nivel académico del curso y las características de cada grupo de alumnos. Los objetos de aprendizaje están almacenados en repositorios estructurados como una base de datos con metadatos asociados y disponibles vía Web. Algunos ejemplos de repositorios son: MERLOT (Multimedia Educational Resource for Learning and Online Teaching)¹, CAREO (Campus Alberta Repository of Educational Objects)², FLOR (Federación Latinoamericana de Repositorios)³, OER Commons (Open Educational Resources)⁴, entre otros.

Los metadatos son un conjunto de atributos necesarios para describir las características de un recurso. Para lograr una buena interconexión entre repositorios y facilitar el desarrollo de sistemas de búsqueda, como es el caso de un sistema recomendador, es importante que los metadatos estén escritos y organizados según un estándar. El estándar de metadatos para los objetos de aprendizaje es LOM (Learning Object Metadata) acreditado por el IEEE 1484.12.1-2002 [2].

En este trabajo se presenta una propuesta que aporte a evaluar la calidad de los metadatos de los objetos de aprendizaje. Para esto, se propone analizar si los objetos poseen metadatos no vacíos y asignados correctamente, con lo cual se puede mejorar el resultado de la recomendación de los recursos educativos. En este sentido, se propone

¹ <http://www.merlot.org/>

² <http://www.careo.org/>

³ <http://ariadne.cti.espol.edu.ec/FederatedClient>

⁴ <http://www.oercommons.org/>

ordenar los resultados de una búsqueda mediante la aplicación de pesos a cada metadato en función del cumplimiento del estándar LOM, y la aplicación de una ponderación de las categorías en las que LOM clasifica los metadatos de forma que ajuste el ordenado al tipo de usuario (profesor o estudiante).

Repositorios y Metadatos

Realizado un relevamiento de repositorios existentes de objetos de aprendizaje y los estándares que utilizan para describir sus objetos, en la actualidad la mayoría de los repositorios trabajan con el esquema de metadatos LOM [2]. Sin embargo, al analizar los repositorios, se observa la falta de información en algunos metadatos. Por ejemplo, OER Commons en la mayoría de los objetos de aprendizaje cuenta únicamente con metadatos que especifican la edad y el idioma del destinatario típico, sin contar por ejemplo con el tiempo aproximado que puede tomar trabajar con el objeto, o el grado de interacción requerido por el mismo; pero los registros cuentan con etiquetas (las cuales no son parte del estándar LOM) que permiten especificar datos educativos como el tipo de material buscado y el contexto académico. El repositorio FLOR cuenta con algunos metadatos educativos incluidos en LOM, como Tipo de Recurso Educativo (Learning Resource Type), Rol del Usuario Final (Intended End User Role) y Contexto (Context). Pero, falta la descripción de los metadatos Tipo de Interactividad (Interactivity Type) y Dificultad (Difficulty), los cuales son necesarios para efectuar una adecuada recomendación. ARIADNE en cambio, cuenta con los valores correspondientes en la mayoría de los metadatos educativos de sus objetos.

La falta de información en los metadatos, especialmente en los educativos, de los objetos de aprendizaje es un problema. Además, preparar objetos de aprendizaje con metadatos adecuados es laborioso, y a veces lleva a una falta de calidad en los metadatos. Así mismo, existen distintos problemas que surgen relacionados con los metadatos cargados. Por ejemplo, en objetos revisados se ha encontrado que: el valor del metadato idioma de un objeto fue establecido usando únicamente el título del objeto pero el cuerpo del objeto estaba en otro idioma; un documento fue clasificado como Texto en el metadato Tipo de Recurso Educativo, pero éste era un código de un programa de computación y por lo tanto debería ser clasificado mejor como Ejercicio o Ejemplo; entre otras discrepancias.

Frente a este tipo de problemas: falta de información o discrepancias respecto al documento que describen los metadatos, y debido a la importancia de los metadatos para la recuperación personalizada se analizaron distintas herramientas para la extracción automática de metadatos.

Extracción automática de metadatos

Cada herramienta para la extracción de metadatos tiene sus propios objetivos, arquitectura y técnicas. En esta sección se analizan cuatro sistemas: SAXEF, TWYS, Looking4LO, y MAGIC, considerando los siguientes puntos: tipos de archivo que tratan, metadatos que extraen y técnicas utilizadas para la extracción automática.

El sistema SAXEF (System for Automatic eXtraction of E-learning object Features) [3] extrae automáticamente atributos texto/multimedia de cualquier página Web, produciendo una tarjeta de identificación de E-learning (EIC) que permite a los profesores evaluar fácilmente cuándo una página es de su interés. Las EICs se organizan en una base de datos y se muestran a través de una interfaz gráfica indicando el tema principal y sus conexiones. Este sistema elimina las palabras no significativas de un texto e identifica las palabras relevantes que se encuentran dentro de etiquetas como <title> (título) y <meta> (palabras clave) en un archivo HTML. A cada palabra relevante le asigna un peso, que es utilizado para determinar el tema principal y los secundarios. Además, procesa las áreas textuales y multimedias. El área textual se determina por el producto entre el número de caracteres de la página y el área ocupada por cada carácter. El área multimedia se determina por la suma de las áreas (en pixeles) de los objetos multimedia presentes. Asumiendo que el área de la página Web es la suma del área textual y el área multimedia, el cociente entre el área textual y el área total resultará en un índice analítico y el cociente entre el área multimedia y el área total es el índice multimedia. De esta forma permite obtener indicadores de cuán analítico o multimedial es un recurso.

El sistema TWYS [4] extrae metadatos de objetos de aprendizaje desde páginas HTML, según el estándar LOM. Algunos metadatos LOM pueden ser extraídos directamente desde la información encontrada en las páginas Web mientras que otros metadatos LOM requieren reglas o métodos no triviales para determinar su valor. Este sistema provee un método de Mapeo directo, para el primer caso, utilizando técnicas tales como determinación de Stopwords, Term Frequency Weighting (TF-IDF) y Ontologías. Para el segundo caso, propone el uso de Reglas basadas en heurísticas, utilizando dos técnicas: Verificación de la Existencia de Etiquetas HTML para saber cuándo cierta etiqueta HTML existe en la página; y Conteo Estático de HTML y Contenido para conocer la cantidad de veces que aparecen ciertas etiquetas HTML y ciertas palabras en una página Web. TWYS recolecta páginas y las almacena en una base de datos de archivos HTML. Luego, separa el contenido HTML de las cabeceras y de las etiquetas HTML. Finalmente, extrae y genera registros LOM por medio de los métodos y heurísticas nombrados previamente.

Looking4LO [5] es un sistema genérico y flexible capaz de extraer objetos de aprendizaje con sus respectivos metadatos de archivos XML y HTML, documentos Word, presentaciones PowerPoint, archivos PDF, y paquetes SCORM⁵. SCORM es un estándar que permite a los sistemas de aprendizaje en línea importar y reutilizar contenidos de aprendizaje, teniendo en cuenta la descripción del objeto según LOM y el uso de reglas de secuenciamiento y de navegación. Este sistema es un prototipo que puede extraer objetos que cubren una cierta temática (matemática, historia, etc.) y proveerlos en forma automática con metadatos. Para la extracción, utiliza la plataforma GATE⁶ (General Architecture for Text Engineering), produciendo un documento enriquecido con anotaciones. Luego analiza el

⁵ <http://www.adlnet.org/>

⁶ <http://gate.ac.uk/>

contenido del documento, utilizando técnicas de procesamiento de lenguaje natural (NLP) para identificar los conceptos en el documento que pertenecen al dominio de una ontología que modela el área temática. Utiliza luego un conjunto de reglas contextuales para identificar las ocurrencias de los objetos. Finalmente, empaqueta los objetos resultantes en el formato deseado. Durante este proceso de recuperación, los metadatos extraídos para cada objeto son: Autor (Author), Tiempo de Lectura (Read time), Nivel de Interactividad (Interactivity Level) y detecta si el objeto tiene imágenes o no.

MAGIC (Metadata Automated Generation for Instructional Content) [6] identifica, segmenta y genera metadatos de acuerdo al estándar SCORM. La extracción se hace sobre diferentes tipos de archivos (video, audio, texto, etc.) aplicando un conjunto de métodos de extracción de información. La extracción de metadatos de textos la realiza con herramientas lingüísticas (tokenización, búsqueda léxica, análisis morfológico, extracción de las frases). El resultado es un archivo de metadatos LOM en formato XML.

Respecto al tipo de archivos, todos los sistemas presentados aquí procesan páginas HTML. Looking4LO también procesa archivos SCORM, los cuales son estructurados y tienen algunos campos de metadatos ya clasificados. Algunos tipos de archivos no estructurados, es decir que no presentan etiquetas o metadatos, tales como TXT, PDF y PPT, son tratados por Looking4LO y MAGIC. Entre los archivos de tipo video y audio (AVI, MPG, MP3, MP4, WMA), MAGIC es el único de los sistemas presentados aquí que puede manipular esta clase de formatos.

Con referencia al poder de extracción de metadatos, SAXEF no sigue al estándar LOM, ya que produce una tarjeta propia de identificación con información definida por los desarrolladores del sistema. TWYS extrae muchos de los metadatos del estándar LOM mientras que MAGIC extrae un subconjunto menor de metadatos LOM. También Looking4LO extrae algunos campos LOM. En particular, SAXEF, TWYS y Looking4LO extraen algunos metadatos educacionales (tales como contenido teórico o práctico, nivel multimedia, tipo de interactividad, dificultad).

Respecto a las técnicas y recursos utilizados para la extracción automática, SAXEF utiliza stopwords, reglas de mapeo directo, reglas heurísticas de mapeo y medidas estadísticas. TWYS usa ontologías, stopwords, TF/IDF, reglas de mapeo directo y reglas heurísticas de mapeo. Looking4LO emplea ontologías y herramientas de procesamiento de lenguaje natural al igual que MAGIC.

De este relevamiento de herramientas para la extracción de metadatos podemos concluir que existen herramientas pero que en general no son capaces aún de obtener los valores referidos a los metadatos educativos, que justamente son los de interés para poder producir una buena recomendación en forma automática.

Propuesta de Ponderación de Metadatos

Para lograr un ordenado adecuado a las necesidades de información del usuario, se propone utilizar los metadatos de los objetos a los fines de contrastarlos con el perfil del usuario que realiza la búsqueda. Un aspecto de interés para este ordenado es evaluar si los objetos de aprendizaje tienen valores cargados en los metadatos y en caso de tenerlos, si éstos cumplen el estándar de metadatos, en nuestro caso el estándar LOM [2]. En este sentido, la propuesta es calificar la instancia de metadatos de un objeto. Esto es, tener en cuenta la posibilidad de que existan metadatos faltantes o incompletos, y en el caso que el metadato esté completo considerar si cumple o no las especificaciones definidas en LOM.

El estándar LOM especifica la sintaxis y la semántica de un conjunto mínimo de metadatos necesario para identificar, administrar, localizar y evaluar un objeto de aprendizaje en forma completa y adecuada. Agrupa los metadatos en nueve categorías: General; Life Cycle; Meta-metadata; Technical; Educational; Rights; Relation; Annotation; y Classification. Las categorías Educational y Life Cycle contienen metadatos descriptivos del objeto. Technical y Rights son metadatos administrativos y las categorías Classification y Relation tienen metadatos sobre la estructura del objeto. Para cada metadato, LOM provee: el nombre; una definición; el tamaño (cantidad de valores permitido); indica si es relevante el orden de los valores; un ejemplo ilustrativo; define un espacio de valores con el conjunto de valores permitidos para el metadato (normalmente en forma de un vocabulario o una referencia a otro estándar); y el tipo de datos (tales como LangString, Date, Duration, Vocabulary, CharacterString o Undefined). Un Vocabulary (vocabulario) es una lista recomendada de valores apropiados, de manera de tener el máximo grado de interoperabilidad semántica y maximizar la probabilidad de que otros usuarios o sistemas comprendan dichos metadatos. El vocabulario de un metadato puede ser ordenado o no ordenado, según la relevancia o no del orden de los valores en la lista. Los metadatos con espacio de valores enumerados se componen de fuente (source) y valor (value); por ejemplo para source = LOMv1.0, el espacio de valores es: draft, final, revised, unavailable.

Para puntuar el cumplimiento del estándar para cada metadato se propone el siguiente criterio. Si el metadato tiene valor y su tamaño corresponde al definido en el estándar, entonces se considera el tipo de dato del metadato, y se asignan los siguientes puntajes: $p=1$ si el tipo de dato detectado es *LangString* o *CharacterString* o *Undefined*; $0 < p \leq 1$ si el tipo de dato corresponde a un *Vocabulary* y pertenece al correspondiente espacio de valores; $p=1$ si el tipo de dato es *Date* o *Duration* y el formato del valor almacenado sigue el estándar; $p=0$ si el metadato no contiene valor, o si su tamaño no corresponde al definido. Asignado este puntaje, se propone calcular un valor que mida el cumplimiento de cada categoría de LOM por parte de cada objeto, por medio de un promedio evaluado como la suma de los valores obtenidos en cada uno de los metadatos incluidos en una categoría c dividida la cantidad de metadatos de dicha categoría. Esto arroja un valor de cumplimiento entre 0 y 1 para cada categoría de metadatos de un objeto de aprendizaje.

Hay dos metadatos que deben ser analizados con mayor profundidad: el correspondiente a las palabras claves

(*Keywords*) que describen al objeto, y los correspondientes al idioma en que está escrito el objeto (*Language* de la categoría *General*) y al idioma en que deberían estar escritos los metadatos del tipo *LangString* (*Language* de la categoría *Meta-Metadata*). En este sentido, se propone evaluar la coincidencia o no de las palabras claves provistas en los metadatos correspondientes con respecto a palabras claves obtenidas a través de una extracción automática de términos de los documentos, así como la coincidencia del idioma asignado en los metadatos y el idioma del objeto. Los valores obtenidos a partir de lo que se describe a continuación, afectarán directamente el valor de cumplimiento del metadato.

Las palabras claves contenidas en el metadato *Keywords* se expanden utilizando Wordnet [7], obteniendo para cada objeto recuperado un conjunto con palabras, frases, y una familia de sinónimos, hiperónimos e hipónimos asociados a cada una. El análisis de la morfología de cada palabra clave extraída utilizando un algoritmo de tagueo morfológico [8], permite determinar si está en singular o plural y obtener un conjunto de palabras asociadas. Además, se filtran las palabras no significativas (*Stopwords*). El resultado es entonces un conjunto de palabras asociado a cada una de las palabras claves de este metadato. Para establecer si existe coincidencia entre palabras del documento y este conjunto de palabras claves del objeto, en primer lugar se extraen las palabras contenidas en el texto del archivo HTML. Para definir si el término extraído es una palabra clave o no, se utiliza un algoritmo que cuenta las ocurrencias de las palabras en el texto. Si w_i es el número total de veces que aparece la palabra i en el texto, y w es el número total de palabras en el texto; entonces el valor de la densidad de la palabra i es: $D_i = w_i/w$. Luego, se consideran sólo aquellas palabras claves que tengan una densidad mayor a una dada. Siguiendo a [9], se propone que los valores de densidad para determinar las palabras claves del documento sean: mayores al 20% para los términos del título y mayores al 2% para los términos del cuerpo. Así se representa que las palabras claves obtenidas de los títulos son más significativas que las contenidas en el texto del documento, y se obtiene un conjunto de palabras claves descriptivas a partir del contenido del documento. Si este conjunto coincide en su totalidad con las palabras claves contenidas en el metadato *Keywords* se tiene el valor óptimo, esto es el valor 1, para la coincidencia. En el caso de que los conjuntos no coincidan, se utiliza el conjunto de palabras asociado a cada una de las palabras claves de este metadato y se lo contrasta con el conjunto de palabras claves obtenidas del documento. Esto permite asignar un valor de coincidencia entre 0 y 1 en la asignación de las palabras claves, y éste será el puntaje de este metadato.

Para evaluar el cumplimiento respecto al idioma, la propuesta es validar si los idiomas asignados en los metadatos respectivos son correctos, utilizando para ello un algoritmo de detección de idiomas, que permita dar un valor numérico que represente la coincidencia de los idiomas entre los asignados en los metadatos y el o los idiomas del objeto. Para esto, se calcula la distancia a la que se encuentran los idiomas asignados en los metadatos de los idiomas detectados mediante el algoritmo a partir del contenido del objeto. En este sentido, se tiene en cuenta que existen algunos idiomas que tienen la característica de poseer alguna palabra en su lenguaje de forma que dicha palabra ocurre con más frecuencia que otras palabras. Una de las maneras más comunes para expresar esta idea es la ley de Zipf [10]. Particularmente, esta ley es verdad para la ocurrencia de los N-gramas más comunes para un determinado idioma. Para la detección del idioma del contenido del objeto de aprendizaje se propone entonces utilizar N-gramas [11]. Un N-grama es una subsecuencia de n elementos de una secuencia dada. Los espacios en blanco son considerados como carácter. Se añaden espacios en blanco al principio y al final de la secuencia con el fin de ayudar a la igualdad entre el comienzo y el final de la palabra. De este modo los N-gramas de inicio y final son considerados de igual forma que si estuviesen por el medio. Por ejemplo, la palabra TEXT estará compuesta por los Bi-gramas (N=2): $_T$, TE , EX , XT , $T_$; y para N = 3 por los Tri-gramas: $_T$, $_TE$, TEX , EXT , $XT_$, $T_$. Entonces, dado un texto de prueba se halla el conjunto de N-gramas junto con sus frecuencias y se calculan las distancias entre este conjunto hallado contra todos los conjuntos de N-Gramas de muestra correspondientes a cada uno de los idiomas a evaluar, y se considera el idioma donde la distancia sea mínima. Es decir, el idioma más probable será aquel dónde su conjunto de N-Gramas relacionado esté más cerca al conjunto de N-Gramas del texto de prueba. Los conjuntos de N-Gramas de muestra son conjuntos hallados desde textos de muestra específicos durante el entrenamiento del algoritmo. En [12] se muestra que N=3 es el valor óptimo a considerar para los N-Gramas. Entonces, N=3 es el valor que utilizaremos en este trabajo, obteniendo no sólo el idioma más probable sino una lista de idiomas, ordenados del más probable al menos probable. Por cada objeto recuperado se extraen los valores asignados en los metadatos *General.language* y *MetaMetadata.language*. Luego se detectan los idiomas más probables del contenido del objeto en formato HTML. El resultado es un valor numérico que representa la coincidencia de idiomas.

Obtenido un valor de cumplimiento del estándar LOM por cada una de las categorías, y antes de obtener un valor del cumplimiento del estándar para todos los metadatos del objeto se puede hacer la siguiente consideración. Según cada tipo de usuario, la categoría en la cual está incluido el metadato adquiere distinta importancia. [13] presenta los resultados de dos encuestas. En una analiza cuáles son los metadatos más relevantes para los estudiantes al momento de realizar una búsqueda, y en la otra cuál es la información más solicitada por los profesores al momento de analizar un objeto. Así, obtienen una lista ordenada de metadatos para cada tipo de usuario, en orden descendente en base a la relevancia de cada metadato. A partir de estos resultados, se propone ponderar la importancia de cada categoría siguiendo el orden siguiente para Estudiantes: 1) *General*, 2) *Classification*, 3) *Technical*, 4) *Educational*, y 5) *Life Cycle*, y el siguiente para Profesores: 1) *Technical*, 2) *Classification*, 3) *Life Cycle*, y 4) *Annotation*. Así, la calificación total de los metadatos del objeto será el promedio ponderado del puntaje en cada una de las categorías, donde la ponderación viene dada por el tipo de usuario que realiza la consulta. Por ejemplo, para los usuarios de tipo estudiante, se propone valores de ponderación iguales a 1 en todas las categorías que no se encuentran en el orden descrito antes, y valores superiores a 1 para las que si se encuentran, con $1 < p_5 < p_4 < p_3 < p_2 < p_1$ donde el subíndice es el número de la categoría de LOM (1=*General*, 2=*Classification*, etc.). De esta manera los objetos de aprendizaje más relevantes para el perfil de cada usuario y con metadatos que cumplen con el estándar escalan a las primeras

posiciones en la lista ordenada con la que se muestran los resultados de la búsqueda.

Conclusiones

En este trabajo se presentó la importancia de los metadatos de los objetos de aprendizaje para dar una recomendación personalizada. El análisis de distintos repositorios ha mostrado que hay una falta de información en muchos de los metadatos que describen a sus objetos. Otro problema encontrado en los repositorios analizados es que, aunque se prevé la inclusión de metadatos educacionales en su diseño, éstos generalmente no están cargados. Estos metadatos son necesarios para hacer una selección de los objetos de aprendizaje no sólo por su temática sino por las características y preferencias del usuario.

Se analizaron distintas herramientas para la extracción automática de metadatos, y se observó que cada una de ellas tiene sus propios objetivos, arquitectura y técnicas de extracción. Se presentaron brevemente algunos sistemas y se los comparó de acuerdo a tres características: tipos de archivos, metadatos extraídos y recursos de procesamiento utilizados. Consideramos que se debe incrementar el esfuerzo en la extracción automática de metadatos si se desea tener una correcta y completa información en los metadatos de los objetos. En particular, la extracción de metadatos educacionales es uno de los principales desafíos.

Se presentó una propuesta para ordenar los resultados de una búsqueda de objetos de aprendizaje, utilizando sus metadatos mediante la aplicación de pesos a los metadatos en función del cumplimiento del estándar LOM, y una ponderación a las categorías que lo ajuste al tipo de usuario (profesor / estudiante). En particular, se analizó el caso de metadatos conteniendo palabras claves descriptivas del contenido y de metadatos con información del idioma en que está escrito cada documento. Para evaluar si se mejora la precisión de la lista ordenada de objetos resultante de una búsqueda, se plantea verificar si los objetos de aprendizaje con mejor cumplimiento del estándar LOM en sus metadatos asignados escalan las primeras posiciones en esta lista ordenada. Se han realizado algunos casos de uso que han arrojado resultados promisorios.

Referencias

- [1] Wiley, D., "Connecting Learning Objects to Instructional Design Theory: A definition, a metaphor, and a taxonomy", in D. A. Wiley (ed.) *Instructional Use of Learning Objects*. Editorial Association for Instructional Technology, 2002.
- [2] IEEE-LOM: IEEE 1484.12.1, Standard for Learning Object Metadata. New York: Institute of Electrical and Electronics Engineers, Learning Technology Standards Committee. Disponible en <http://ltsc.ieee.org/wg12>. 2002.
- [3] Alfano, M., Lenzitti, B., Visalli, N., "SAXEF: A System for Automatic eX-traction of learning object Features". *Journal of e-Learning and Knowledge Society*, vol. 3, (2), 83-92, 2007.
- [4] Wai Yuen, T., "Automatic Extraction of Learning Object Metadata (LOM) from HTML Web Pages," Master of philosophy, City University of Hong Kong, May 2007.
- [5] Motz, R., Badell, C., Barrosa, M., Sum, R., Díaz, G., Castro, M.: *LooKIng4LO: Sistema Informático para la Extracción Automática de Objetos de Aprendizaje: Caso de Estudio*. IEEE-RITA(2009) 223-229.
- [6] Li, Y., Dorai, C., Farrell, R., "Creating MAGIC: system for generating learning object metadata for instructional content," in *MULTIMEDIA '05: Proceedings of the 13th annual ACM international conference on Multimedia*, pp. 367-370, New York, NY, USA, 2005.
- [7] Miller, G. "WordNet: A Lexical Database for English". *Communications of the ACM* Vol. 38, No. 11. 2010. pp 39-41
- [8] Brill, E. "A simple rule-based part of speech tagger". Department of Computer Science University of Pennsylvania Philadelphia, Pennsylvania 19104. 1993.
- [9] Rodriguez, I. Keywords - El Elemento Más Importante en SEO. Disponible en <http://www.miempresaenlinea.com/blog/post/keywords-seo.aspx>. 2009.
- [10] Zipf, G. K. *Human Behavior and the Principle of Least-Effort*. Editorial Addison-Wesley. 1949.
- [11] W.B. Cavnar and J.M. Trenkle. "N-gram based text categorization", in *Proceedings of the Third Annual Symposium on Document Analysis and Information Retrieval*, 1994, pp 161-169.
- [12] Ahmed, B. Cha, S.H. and Tappert, C. "Language Identification from Text Using N-gram Based Cumulative Frequency Addition", in *Proceedings of Student/Faculty Research Day, CSIS, Pace University, May 7th, 2004*
- [13] Woo, K., Maree, G., Gibbs, D., Hand, T., Kerr S. and Rich D. "User Perspectives on Learning Object Systems", disponible en <http://ausweb.scu.edu.au/aw04/papers/refereed/woo/paper.html>. 2004.