

Sistemas de Información Inteligentes en el Dominio Educación: Extracción y Calidad de Metadatos de Objetos de Aprendizaje

Claudia Deco^{1,2}, Cristina Bender^{1,2}, Ana Casali^{1,3}, Jorge Saer²

¹ Facultad de Ciencias Exactas, Ingeniería y Agrimensura,
Universidad Nacional de Rosario, Rosario, Argentina.
{ deco, bender, acasali }@fceia.unr.edu.ar

² Facultad de Química e Ingeniería, Universidad Católica Argentina, Rosario, Argentina.
{ cdeco, cbender, jsaer }@uca.edu.ar

³ Centro Internacional Franco Argentino de Ciencias de la Información y de Sistemas CIFASIS

Resumen

Esta línea de investigación trata el desarrollo de Sistemas de Información Inteligentes que asistan a un docente o a un alumno en la búsqueda de objetos de aprendizaje que le ayuden en el proceso educativo. Los objetos de aprendizaje son materiales digitales los que como unidad o agrupación permiten o facilitan alcanzar un objetivo educacional. Estos objetos se describen mediante metadatos, y se almacenan en diferentes repositorios. Por medio de los metadatos se puede llevar a cabo un primer acercamiento con el objeto y conocer sus principales características. Esto significa que estamos tratando con un contenido normalizado y estandarizado. El objetivo de esta línea de I+D es el desarrollo de un conjunto de herramientas que faciliten la creación, descripción, búsqueda y re-uso de objetos de aprendizaje, haciendo hincapié en la extracción automática de metadatos y en la calidad de los metadatos, dado que esto es fundamental para la búsqueda y recomendación automática de este tipo de objeto.

Palabras Claves: Sistemas de Información Inteligentes, Sistemas Recomendadores, Objetos de Aprendizaje, Extracción automática de metadatos, Calidad de metadatos.

Contexto

Esta línea de I+D se está llevando a cabo a través de proyectos de la Universidad Nacional de Rosario (UNR), y de la Universidad Católica Argentina (UCA).

Los PIDs de la UNR involucrados son:

- [ING348] *Búsqueda personalizada de información en Bases de Datos*, dirigido por Claudia Deco. (2011-2014), en evaluación.
- [ING308] *Sistemas multiagentes y sus aplicaciones en la educación*, dirigido por Ana Casali. (2010-2013).

El Proyecto de Investigación Institucional de la UCA involucrado es:

- [PII UCA PERS] *Mejora de la enseñanza de las ciencias en la carrera de Ingeniería Ambiental utilizando un sistema de recomendación de búsqueda personalizada de recursos educativo*, dirigido por Claudia Deco. (2009-2012).

Se encuentra en evaluación el proyecto PICTO-2010-0143: *Hacia el desarrollo y utilización de Repositorios de Acceso Abierto para Objetos Digitales Educativos en el contexto de las universidades públicas de la región centro-este de Argentina*, dirigido por Patricia San Martín y Ana Casali, y presentado en el llamado a Proyectos de

Investigación Científica y Tecnológica Orientados del Fondo Nacional para la Investigación Científica y Tecnológica (FONCYT).

En estos proyectos se trabaja en conjunto con el grupo Concepción del Sistemas de Información, de la Universidad de la República, cuya directora es la Dra. Regina Motz. Además, se trabaja con investigadores de otras universidades latinoamericanas en el marco de la Comunidad Latinoamericana de Objetos de Aprendizaje (LACLO).

Introducción

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. En particular, con el desarrollo de la Web y su utilización masiva, se tiene una amplia gama de posibilidades de acceso a material útil e interesante para ser empleado tanto por un alumno que desea aprender un tema, como por un docente que desea preparar material didáctico. Sin embargo, se advierte una sobrecarga de información convirtiendo la selección de la información que les interesa en una tarea tediosa, que insume mucho tiempo y que es difícil de realizar sin la asistencia de herramientas de búsqueda intuitiva y eficiente. Este problema también sucede en el escenario de la búsqueda en repositorios de recursos educativos.

Se conoce como Objeto de Aprendizaje (OA) a todo recurso digital que apoya a la educación y que puede ser reutilizado, es decir, es un conjunto de materiales los que como unidad o agrupación permiten o facilitan alcanzar un objetivo educacional; por ejemplo: textos, imágenes, cortos de video o audio, pequeñas aplicaciones Web, y páginas Web. El objetivo principal de los repositorios de objetos de aprendizaje es organizar los OAs y mantenerlos disponibles para diferentes usos. Ejemplos de estos repositorios son: FLOR¹, Ariadne² y OER Commons³. Los objetos almacenados en estos repositorios se describen mediante metadatos, que en general siguen el estándar LOM⁴.

Los resultados de la búsqueda no siempre es el

esperado por el usuario si la consulta se realiza sólo considerando palabras claves, porque un material dado no es el adecuado para todos los usuarios. Esto se debe a que los usuarios poseen distintas características y preferencias personales, que deberían también ser consideradas. Así, el uso de los metadatos, además de las palabras claves usadas para expresar la consulta, permite recuperar aquellos objetos que mejor satisfagan las necesidades de información del usuario, teniendo en cuenta sus características y preferencias individuales, mejorando así la precisión de la búsqueda.

En los últimos años, los sistemas recomendadores surgen para ayudar a resolver problemas de este tipo, puesto que son capaces de seleccionar, de forma automática y personalizada, el material que mejor se adapte a las preferencias o necesidades de un usuario [1]. El uso de los sistemas multiagentes se ha incrementado para abordar el desarrollo de sistemas distribuidos complejos y en particular, se ha utilizado para el desarrollo de sistemas recomendadores. Esta tecnología de agentes es importante a la hora de modelar diferentes características que se espera de estos sistemas. La base para el razonamiento del sistema recomendador es el uso de los metadatos de los objetos de aprendizaje y de los metadatos que describen el perfil del usuario.

Por esto, para mejorar los resultados de la recomendación es fundamental contar con los metadatos de los OAs y que éstos sean de calidad. Por este motivo, en esta línea de investigación, se propone por un lado la extracción automática de metadatos y por otro lado medir la calidad de los mismos.

Respecto a la extracción automática de metadatos se ha analizado el estado del arte de las técnicas de extracción automática de metadatos y diferentes sistemas de extracción. Algunos ejemplos de sistemas para la extracción de metadatos son: SAXEF (System for Automatic eXtraction of E-learning object Features) [2] y Looking4LO [3]. También existen propuestas para la extracción automática de metadatos de páginas HTML tales como las presentadas en [4] y [5]. Algunos aspectos relevantes para el análisis y diseño de sistemas de extracción automática son el tipo de archivo (por ejemplo html, txt, ppt, pdf, doc, etc.), los metadatos que extraen (son de particular interés los metadatos educacionales: tipo y nivel de interactividad, densidad semántica, dificultad, etc.) y las técnicas y recursos utilizados

¹ <http://www.laclo.org>

² <http://www.ariadne-eu.org>

³ <http://www.oercommons.org>

⁴ Estándar LOM 1484.12.1-2002. <http://ltsc.ieee.org>

(por ejemplo herramientas para el procesamiento de lenguaje natural (NLP), ontologías, etc.).

La calidad en objetos de aprendizaje puede abordarse desde distintas perspectivas y en diversos momentos de su propio ciclo de vida. Con relación a la calidad desde la perspectiva educativa del OA, ésta se cumplirá en mayor grado cuanto más se logre el objetivo de obtener un aprendizaje significativo por parte del usuario, es decir el alumno. Un aspecto medular para lograr lo anterior es asegurando la calidad de contenido del objeto. Esta perspectiva no es abordada en este proyecto. La calidad de diseño se refiere a todo lo relacionado con el diseño del objeto de aprendizaje como es la selección de metadatos adecuados y de los elementos que lo deben integrar (textos explicativos, gráficas, fotografías, música, texto hablado, videos, animaciones, evaluaciones, módulos de experimentación, hipervínculos) para que sea eficaz en el proceso de aprendizaje dependiendo del contexto en que se aplique. La calidad de concordancia se refiere al grado de cumplimiento de los estándares y recomendaciones definidas para el OA.

Se propone analizar la calidad de los metadatos de los objetos de aprendizaje con el fin de que los recursos de mayor calidad escalen a las primeras posiciones en los resultados de búsqueda y que además se personalicen al perfil del usuario que realiza la búsqueda.

La calidad de los metadatos se analiza en una primera etapa verificando y/o validando si los metadatos siguen el estándar LOM en cuanto a su presencia, tamaño, formato, pertinencia de los valores de cada uno de los metadatos que describen al objeto. Por ejemplo, en el caso del metadato Language que contiene el idioma o idiomas en que está escrito el objeto, se valida en forma automática la concordancia de este o estos valores del metadato con el contenido del objeto. Asimismo, se valida la concordancia de las palabras claves asignadas, con respecto al metadato Keywords. El objetivo es priorizar el análisis automático sobre un análisis manual de expertos o usuarios finales, con el fin de que el usuario no tenga participación al momento de calificar la calidad desde el punto de vista estructural y de contenido sino que la tenga solo desde el punto de vista pedagógico.

Se está desarrollando un prototipo de aplicación web que permite realizar búsquedas en un repositorio mediante palabras claves y que evalúa

la calidad de los metadatos de cada objeto de aprendizaje de forma que se adapten de una manera más precisa al perfil del usuario que realiza la búsqueda.

Así, teniendo objetos de aprendizaje con metadatos no vacíos y asignados correctamente, se puede mejorar los resultados de la recomendación de recursos educativos.

Líneas de investigación y desarrollo

Para alcanzar los objetivos que proporcionen una ayuda a los problemas mencionados en la introducción, se han planteado distintas líneas de investigación las cuales se interrelacionan.

En el proyecto [ING308] se plantea utilizar sistemas multiagentes diseñando a sus componentes con arquitecturas de agentes que los capaciten para actuar de forma flexible y eficiente, en entornos multiagentes. En este sentido se ha propuesto el diseño de agentes BDI graduados (g-BDI) en un entorno social integrado por otros agentes. Resultados preliminares de la arquitectura g-BDI pueden verse en [6]. Se ha utilizado este modelo de agente para diseñar e implementar agentes recomendadores en el dominio de la educación [1].

Paralelamente, se trabaja en la expansión semántica de la búsqueda en el contexto del proyecto [ING348]. El objetivo de esta línea es producir la estrategia de búsqueda temática. Cuando el usuario hace una consulta, ingresa un conjunto de términos que describen el tema de su interés. Luego es necesario un proceso que desambigüe estos términos y los expanda semánticamente incorporando sinónimos y conceptos relacionados. Además, se trabaja en cómo utilizar las características y preferencias del usuario (docentes o alumnos) para que un sistema recomendador no sólo recupere los recursos que respondan temáticamente a la consulta del usuario, sino que presente aquellos que sean los más adecuados de acuerdo a su perfil. Resultados pueden verse en [7].

El objetivo del proyecto [PII UCA PERS] es mejorar el proceso de enseñanza-aprendizaje de las ciencias mediante la recomendación de objetos de aprendizaje. Así, se ayuda al docente en la selección y 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. Resultados parciales se muestran en [8].

Uno de los problemas encontrados en estos proyectos, en la búsqueda de objetos de aprendizaje en los distintos repositorios, es la ausencia de valores en los metadatos, así como inconsistencia del valor del metadato asignado con el contenido real del objeto. En particular se observa además la falta de valores en los metadatos correspondientes a la categoría Educacional, que agrupa las características pedagógicas y educacionales del objeto, por ejemplo aquellos que indican el tipo de interactividad requerido por el objeto y el nivel de dificultad del mismo. Por esto, actualmente se está trabajando en la extracción y validación automática de estos metadatos.

Resultados Obtenidos / Esperados

Entre los resultados obtenidos en esta línea de investigación se encuentran:

- Diseño de la arquitectura del sistema recomendador como un sistema multiagente.
- Diseño de algunos de los agentes que componen el sistema recomendador.
- Implementación de un prototipo de estos agentes utilizando gBDI.
- Evaluación de los agentes implementados mediante el análisis de casos de uso: se obtuvieron resultados promisorios.
- Análisis de técnicas para la extracción automática de metadatos.
- Diseño de una base de datos local de metadatos de recursos obtenidos mediante consultas realizadas en repositorios de objetos de aprendizaje.
- Establecimiento de un conjunto inicial de metadatos para establecer una evaluación de la calidad de los mismos.

Entre los resultados esperados en esta línea de investigación se encuentran:

- Extraer y validar automáticamente ciertos metadatos educacionales de los objetos de aprendizaje.
- Implementar, experimentar y evaluar los componentes faltantes de la arquitectura diseñada.

- Integrar los distintos componentes.
- Transferir los resultados a la comunidad científico-educativa.

Formación de Recursos Humanos

El equipo de trabajo está integrado por la Dra. Ana Casali (investigadora de la Universidad Nacional de Rosario, y del CIFASIS), la Dra. Claudia Deco (investigadora de la Universidad Nacional de Rosario y de la Universidad Católica Argentina), la M. Sc. Cristina Bender (investigadora de la Universidad Nacional de Rosario y de la Universidad Católica Argentina), y el Ing. Jorge Saer (investigador de la Universidad Católica Argentina).

Dentro del marco de esta línea de investigación, se desarrolla actualmente la tesina de Maestría en Tecnología Informática aplicada en Educación, UNLP “Mejoramiento del diagnóstico del estudiante con modelos difusos” de Constanza Huapaya, dirigida por Ana Casali y codirigida por L. Lanzarini. Además, se desarrollan las siguientes tesinas de grado de la Licenciatura en Ciencias de la Computación, de la Universidad Nacional de Rosario:

- Búsqueda Semántica en Repositorios de Objetos de Aprendizaje. Alumno Marcos Belén, Directora: Claudia Deco y Cristina Bender. 2010.
- Análisis de calidad y confianza en los metadatos de objetos de aprendizaje recuperados. Damián Álvarez Machicote. Directora: Cristina Bender y Claudia Deco, en curso.
- Automatización del proceso de carga de metadatos educacionales en DSpace. Santiago Fontanarrosa. Directora: Ana Casali, en curso.

Asimismo, se han desarrollado las siguientes pasantías:

- Interfase Web para el sistema recomendador que interactúe con un federado de objetos de aprendizaje. Valeria Gerling, en el marco del proyecto JARDIN, 2010.
- Extracción de información para la generación de Objetos de Aprendizaje, Taihú Piré. Directora: Ana Casali (UNR) y Bernard Espinasse (LSIS, Universidad de Marsella,

Francia), en el marco de la cooperación entre UNR-CIFASIS y el LSIS, 2010.

Referencias

- [1]. Casali A., Gerling V., Deco C. y Bender C. Chapter 8: Recommender System for Personalized Retrieval of Learning Objects. In: Book of Educational Recommender Systems and Technologies: Practices and Challenges. ERSAT. Ed. Olga C. Santos and Jesus G. Boticario. aDeNu Research Group. UNED, Spain. ERSAT. 2011.
- [2]. Alfano, M., Lenzitti, B. and Visalli, N., SAXEF: A System for Automatic eXtraction of learning object Features. Journal of e-Learning and Knowledge Society, vol. 3, (2), 83-92, 2007.
- [3]. 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. pp 223-229. 2009.
- [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]. Bender C., Burgués M., Deco C. Generación automática de metadatos a partir de la extracción de información del contenido de documentos HTML. Conferencia conjunta Iberoamericana sobre Tecnologías para el Aprendizaje (CcITA 2009). Mérida, México. Julio 2009.
- [6]. Casali, A., Godo, Ll. y Sierra, C. g-BDI: A graded intentional agent model for practical reasoning, MDAI 2009. LNAI, Vicenç Torra, Yasuo Narukawa, Masahiro Inuiguchi (eds.), vol. 5861, Awaji Island, Japan, Springer, pp. 5-20, 2009.
- [7]. Casali A., Gerling V., Deco C. and Bender C. A multiagent system for personalized recommendation of learning objects. Congreso Iberoamericano de Informática Educativa - IE 2010. ISBN 978-956-19-0722-5. pp 503-510. Santiago, Chile. 2010.
- [8]. Casali A., Gerling V., Deco C. y Bender C. Sistema inteligente para la recomendación de objetos de aprendizaje **LACLO 2010 Best Papers**. Revista Generación Digital Vol 9, No 1. pp. 88-95. Colombia. 2011.