

Revista Electrónica de Investigación Educativa

Vol. 16, Núm. 1, 2014

Un modelo para la organización semiautomática de contenido educativo desde repositorios abiertos de objetos de aprendizaje

A Model for Semi-Automatic Composition of Educational Content from Open Repositories of Learning Objects

Paula Andrea Rodríguez Marín (1)
parodriguezma@unal.edu.co

Julián Moreno Cadavid (1)
jmoreno1@unal.edu.co

Néstor Darío Duque Méndez (2)
ndduqueme@unal.edu.co

Demetrio Arturo Ovalle Carranza (1)
dovalle@unal.edu.co
Universidad Nacional de Colombia Sede Medellín

Ricardo Silveira (3)
ricardo.silveira@ufsc.br

- (1) Universidad Nacional de Colombia, Sede Medellín
- (2) Universidad Nacional de Colombia, Sede Manizales
- (3) Universidade Federal de Santa Catarina

Universidad Nacional de Colombia, Campus La Nubia
Km. 9, vía al Aeropuerto La Nubia
Bloque Y, primer piso
Manizales, Caldas, Colombia

(Recibido: 11 de septiembre de 2012;
aceptado para su publicación: 20 de agosto de 2013)

Resumen

Los repositorios de Objetos de Aprendizaje (OA) son importantes para la construcción de contenidos educativos y deben facilitar los procesos de búsqueda, recuperación y organización de OA con el fin de atender objetivos educativos. Sin embargo, estos procesos demandan tiempo y no proveen los resultados esperados; por ello, el objetivo de este artículo es proponer un modelo para la organización semiautomática de contenidos recuperados automáticamente de repositorios abiertos. Para el desarrollo del modelo se analizaron medidas de similitud, mientras que para su calibración y validación se realizaron experimentos de comparación con los resultados obtenidos por docentes. Los resultados experimentales demuestran que al usar un valor de k (selección de OA) de 3, el porcentaje de coincidencias entre el modelo y los expertos supera el 75%. Se concluye entonces que el modelo propuesto permite ahorrar tiempo y esfuerzo al docente para seleccionar OA al realizar un proceso de prefiltrado.

Palabras clave: Repositorios de objetos de aprendizaje, Recursos educativos abiertos, Tecnología educativa, Teoría educativa.

Abstract

Learning objects (LOs) repositories are important in building educational content and should allow search, retrieval and composition processes to be successfully developed to reach educational goals. However, such processes require so much time-consuming and not always provide the desired results. Thus, the aim of this paper is to propose a model for the semiautomatic composition of LOs, which are automatically recovered from open repositories. For the development of model, various text similarity measures are discussed, while for calibration and validation some comparison experiments were performed using the results obtained by teachers. Experimental results show that when using a value of k (number of LOs selected) of at least 3, the percentage of similarities between the model and such made by experts exceeds 75%. To conclude, it can be established that the model proposed allows teachers to save time and effort for LOs selection by performing a pre-filter process.

Keywords: Learning Object Repositories, Open Educational Resources, Educational Technology, Educational Theory.

I. Introducción

El crecimiento de la información digital, la computación de alta velocidad y las redes ubicuas ha permitido el acceso a más información y, entre ella, a miles de recursos educativos. Esto ha potenciado el diseño de propuestas novedosas de enseñanza-aprendizaje, para compartir materiales y navegar a través de ellos (Peña *et al.*, 2002).

En particular, la compatibilidad de las tecnologías de la información con los procesos de enseñanza-aprendizaje ha propiciado la aparición de nuevas alternativas para la generación de cursos, una de estas alternativas son los Objetos de Aprendizaje (OA), los cuales permiten la construcción y distribución personalizada de contenidos, así como la reutilización de los mismos en nuevos contextos. Desde su surgimiento, hace más de 15 años, su objetivo ha sido la producción de contenido educativo mediante un proceso de organización basado en elementos pequeños, unidades elementales de aprendizaje a los que se les dio el nombre de OA. Una característica fundamental de estos elementos es que poseen una intención educativa que busca definir una interacción eficaz con el estudiante y apoyar su proceso de aprendizaje (Polsani, 2003). De una manera más formal, el Institute of Electrical and Electronics Engineers establece que un OA puede considerarse como una entidad digital con características de diseño instruccional, que puede ser usado, reutilizado o referenciado durante el aprendizaje asistido por computador con el objetivo de generar conocimientos, habilidades, actitudes y competencias en función de las necesidades del alumno (Learning Technology Standards Committee, 2002). Como complemento a la anterior definición, algunos autores agregan que los OA tienen como requisitos funcionales la accesibilidad, reutilización e interoperabilidad; adicionalmente, poseen metadatos que los describen e identifican, facilitando su búsqueda y recuperación. (Ouyang y Zhu, 2008), (Betancur, Moreno y Ovalle, 2009).

Como ocurre con cualquier recurso digital pensado para un uso masivo, existen bibliotecas especializadas para la administración de OA, conocidas como Repositorios de Objetos de Aprendizaje (ROA), los cuales pueden variar en varios aspectos, por ejemplo, en su arquitectura (centralizada o distribuida) y en su política de acceso (libre o restringida). Más aún, con el fin de maximizar el número de OA a los que un usuario puede tener acceso, los ROA pueden conectarse a través de federaciones con el fin de compartir los recursos educativos que poseen (Li, 2010). Una federación sirve para facilitar los procesos de acceso a los OA disponibles a partir de los ROA registrados en ella a través de un punto único de acceso, así como para la administración unificada de las aplicaciones de apoyo que puedan existir (Van de Sompel y Chute, 2008). Un ejemplo a nivel latinoamericano de federación es la FEB (Federação Educa Brasil), cuyo objetivo es la centralización de diversos repositorios institucionales a partir de un solo portal de búsqueda.

Queda claro que existen múltiples fuentes y puntos de entrada para acceder a los OA. La tabla I muestra un listado de algunos ROA y federaciones para el caso específico de Iberoamérica. Si bien esto puede entenderse como una oportunidad para los docentes en el sentido que pone a su disposición recursos que pueden ser utilizados y reutilizados para sus prácticas educativas, sean presenciales, virtuales o mixtas; lo cierto es que también se convierte en un desafío. Por una parte, dado un objetivo educativo específico (entiéndase, por ejemplo, “comprender el proceso celular de la transcitosis”), puede resultar tediosa la búsqueda y selección de OA relacionados. Por otra parte, una vez recuperados debe realizarse un proceso minucioso de filtrado para descartar aquellos que no sean relevantes, dejando únicamente aquellos OA que en realidad se mapeen con el objetivo educativo. Por último, debe tenerse en cuenta que este procedimiento completo no se realiza una única vez, sino que debe repetirse para

cada uno de los objetivos educativos que hagan parte de la práctica del docente, llámese sesión de clase, unidad/tema/módulo, o inclusive un curso completo.

Tabla I. Repositorios y federaciones de repositorios de Objetos de Aprendizaje Latinoamérica

Descripción	País	Ubicación
Aprendiendo con objetos de aprendizaje-APROA	Chile	http://www.aproa.cl
Banco nacional de objetos de aprendizaje e informativos-Colombia aprende	Colombia	http://64.76.190.172/drupalM
Colaboración para la construcción de objetos de aprendizaje-COBA	México	http://ccoba.cuaed.unam.mx
Comunidad latinoamericana de objetos de aprendizaje-LACLO	Latinoamérica	http://www.laclo.org
Federação educa Brasil-FEB	Brasil	http://feb.ufrgs.br/feb
Federación de repositorios de objetos de aprendizaje de Colombia-FROAC	Colombia	http://froac.manizales.unal.edu.co
Iniciativa colaborativa de objetos de aprendizaje utilizables y reutilizables-ICOA-URÚ	Venezuela	https://roa.mppeu.gob.ve
Objetos de aprendizaje-ODEA	Uruguay	http://www.ceibal.edu.uy
Repositorio de objetos de aprendizaje de la Universidad de Sevilla-RODAS	España	http://rodas.us.es

Dentro del contexto previamente introducido, el objetivo de este artículo es presentar una propuesta para la organización semiautomática de contenido educativo a partir de OA almacenados en repositorios. Dichos repositorios pueden ser locales o remotos, y encontrarse o no federados. De hecho, las únicas condiciones exigidas son: 1) ser accesibles vía web; 2) ser abiertos, es decir, que puedan ser consultados sin restricciones de acceso; y 3) contar con metadatos descriptivos para los OA.

II. Trabajos relacionados

A continuación se mencionan algunos trabajos relacionados con la organización automática o semiautomática de contenido educativo. Cabe señalar que no necesariamente, en todos los casos, dichos trabajos incorporan específicamente el uso de OA, aunque sí hacen alusión a recursos de baja granularidad.

Duque (2005) propone un modelo de generación de cursos virtuales orientado por las metas u objetivos educativos que el estudiante espera lograr. El modelo emplea técnicas de planificación inteligente y considera los logros obtenidos, así como el estilo de aprendizaje del estudiante. Posteriormente, Baldiris y Duque Méndez (2012) proponen la generación de planes instruccionales a partir de micro-contextos de competencias educativas. En este trabajo se aplican técnicas de similitud para obtener los OA más adecuados aprovechando los microcontextos de sus metadatos. Mavrommatis (2008) propone un método para la creación de experiencias educativas por medio de OA, los cuales son seleccionados mediante unas medidas de similitud, más específicamente el coeficiente de superposición (*Overlap Coeficient*) para las propiedades funcionales y la distancia euclidiana normalizada para las propiedades no funcionales. Su principal aporte es un esbozo de técnicas de recuperación de

información con principios de diseño instruccional. Tseng *et al.* (2008) proponen la arquitectura de un sistema de aprendizaje adaptativo modular que tiene como objetivo dividir y transformar materiales didácticos en OA. Dicho sistema almacena para cada estudiante sus logros y el nivel de aprendizaje como insumo para construir un curso individual. Tales cursos dan la oportunidad a los estudiantes con bajo nivel de conocimientos, de aprender de los OA de un nivel de aprendizaje más alto, a través de varios índices como la eficiencia del aprendizaje y la tasa alcanzada en las pruebas, entre otros. Huang *et al.* (2008) proponen un marco de referencia para el proceso de generación de cursos estandarizados que incluye los siguientes elementos: un portal de e-learning LMS, una herramienta de autor para edición de recursos educativos, un sistema de formación de navegación que colecta información de la historia de aprendizaje del usuario y da sugerencias de cursos y, finalmente, un agente de organización de materiales que se encarga de ordenar la secuencia del curso. Verbert *et al.* (2012) proponen un estudio exploratorio que se basa en un enfoque semiautomático para apoyar a los docentes en la generación de recursos educativos a partir de la reutilización y secuenciación de actividades de aprendizaje y recursos relacionados. El principal aporte de este trabajo radica en un algoritmo recursivo que recibe como entrada la secuencia actual de actividades y la compara con patrones de diseño de aprendizaje existentes.

III. Modelo propuesto

Tal como se evidenció en la sección anterior, existen diferentes aproximaciones y propuestas para la organización automática o semiautomática de contenido educativo. Algunos investigadores consideran, para llevar a cabo dicha organización, la caracterización del dominio de conocimiento de interés, trátase de un curso completo o de una fracción del mismo (Mavrommatis, 2008; Verbert *et al.*, 2012). Otros toman en cuenta la caracterización del estudiante a partir de uno o más aspectos, tal como el nivel alcanzado o el estilo de aprendizaje (Tseng *et al.*, 2008). Algunos, incluso, consideran importante utilizar ambas caracterizaciones (Duque, 2005); (Huang, Chen, Huang, Jeng y Kuo, 2008; Baldiris y Duque, 2012).

Entre estas tres aproximaciones, la propuesta de este artículo se adhiere a la primera; es decir, parte de una caracterización del dominio de conocimiento, pero no caracteriza a los estudiantes a quienes dicho dominio se impartiría. La razón es de carácter práctico: por más recursos con que cuente un determinado repositorio, la cantidad de ellos que mapea un objetivo educativo (OE) específico suele ser muy baja, lo cual dificulta que los mismos mapeen, además, diversos “perfiles” de estudiantes. Esta suposición fue corroborada de manera informal en los experimentos que se presentan en el siguiente apartado; el esquema general del modelo se presenta en la Figura 1.

1. Se especifican los n OE específicos que se desean abordar y se estructuran como un árbol jerárquico.
2. Para cada OE $_i$ en las hojas del árbol:
 - 2.1 Se accede a uno o más repositorios y se recopilan los OA relacionados
 - 2.2 De los OA recolectados se selecciona aquél o aquellos más relevantes según criterio de similitud de los metadatos
3. Se devuelve el árbol instanciando en cada hoja con los OA seleccionados

Figura 1. Modelo propuesto

En el primer proceso se requiere que el docente o el diseñador instruccional encargados de la construcción del contenido educativo, especifiquen claramente cuáles son los OE que se desean abordar. Dicha especificación no sólo consiste en la definición de cada OE como tal, sino también de la definición de la relación de los mismos mediante una estructura de árbol jerárquico. Así, por ejemplo, dentro del contexto de un curso de biología, una especificación hipotética de OE esperados se presenta en la Figura 2, mostrando el detalle de uno de ellos. Este ejemplo es meramente ilustrativo, por lo que la especificación puede no corresponder con un curso real.

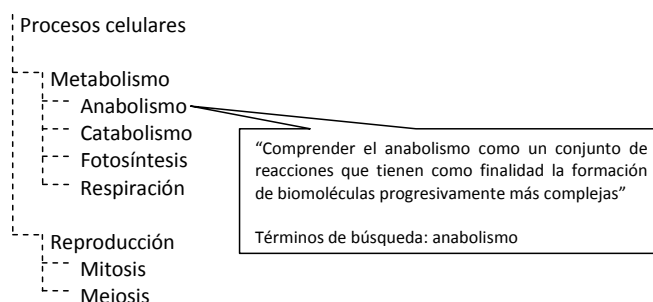


Figura 2. Ejemplo de especificación de OE

Una vez se tienen especificados los OE (como en la Figura 2), se procede a hacer una selección de OA relacionados con cada uno de ellos, accediendo a uno o más repositorios utilizando como criterios de búsqueda los términos definidos para cada OE. Este proceso puede llevarse a cabo de manera completamente automática siempre y cuando dichos repositorios cumplan las condiciones descritas previamente (sean accesibles vía web, sean abiertos, y cuenten con metadatos descriptivos para los OA) pero además cuenten con mecanismos de consulta a nivel de máquina: tipo web-service, SQL, OAI-CAT, entre otros.

Luego de obtener para cada OE $_i$ un conjunto C_i de OA, con cardinalidad igual o superior a 1, se procede a seleccionar aquél o aquellos más adecuados para satisfacer tal objetivo. Dicha selección es el corazón del modelo propuesto y está basada en la comparación de la descripción del OE con tres metadatos específicos de los OA: título, descripción y palabras clave. Al igual que en algunos de los trabajos descritos en la

sección anterior, dicha comparación se realiza mediante medidas de similitud de textos, las cuales son usadas generalmente en contextos como la recuperación de información, la minería de texto (*text-mining*), la minería web (*web-mining*), los sistemas de clasificación (*clustering*), o la detección de copias de documentos, entre otros (Kim y Choi, 1999).

En total, el modelo considera inicialmente ocho medidas de similitud (ver Tabla II). Si bien las interpretaciones matemáticas de las fórmulas empleadas son diferentes en cada caso, lo que buscan todas es una medida cuantitativa del grado de semejanza o diferencia de dos cadenas de texto Q y D. Para una explicación detallada de estas medidas se puede consultar a Amón y Jiménez (2010).

Tabla II. Medidas de similitud evaluadas

Medida	Fórmula	Descripción
Coseno (Salton <i>et al.</i> , 1975)	$\frac{QD}{ Q D }$	Coseno del ángulo que forman los vectores de las dos cadenas en un espacio N-dimensional.
Coeficiente de Dice (Dice, 1945)	$\frac{2 Q \cap D }{ Q + D }$	El doble de la cantidad de términos comunes dividido por la cantidad total de términos.
Coeficiente de Overlap (Haibo <i>et al.</i> , 2008)	$\frac{ Q \cap D }{\min(Q , D)}$	Cantidad de términos comunes dividido por la menor cantidad de términos.
Coeficiente de Jaccard (Real y Vargas, 1996).	$\frac{ Q \cap D }{ Q \cup D }$	Cantidad de términos comunes dividido por la cantidad de términos comunes y no comunes en los dos documentos.
Distancia booleana (Grossman y Frieder, 2004)	$\sqrt{Q \cap D}$	Define valores binarios en función de si los términos de una cadena están o no en la otra.
Distancia vectorial (Salton <i>et al.</i> , 1975)	$Q \cdot D$	Es similar a la distancia booleana pero define valores positivos y no binarios a los términos similares.
Distancia de Hamming (Hamming, 1950)	$\frac{ Q \sim D }{ Q }$	Cantidad de términos que se deben cambiar para que una cadena se convierta en la otra.
Distancia euclidiana (Bonny y Renou, 2002)	$ Q - D $	Distancia n-dimensional entre los vectores que representan ambas cadenas.

Dentro del modelo Q se refiere a la descripción del OE, mientras que D se refiere a los metadatos: título, descripción, y palabras clave, separadamente. Cabe señalar que para cada medida de similitud se debe realizar un proceso preliminar de eliminación de palabras irrelevantes (*stop-words*) para reducir la dimensión de las comparaciones. Ahora bien, como cada medida tiene una naturaleza diferente (la medida del coseno, por ejemplo, entrega un valor más fuerte cuando las cadenas comparadas tienen la misma palabra varias veces, puesto que tiene en cuenta la frecuencia de cada término para hacer el cálculo), resulta difícil determinar cuál es la mejor de todas para comparar los OE contra los metadatos de los OA. Por esta razón, más que elegir una de ellas, el modelo propuesto usa una combinación de las mismas para realizar finalmente la selección.

Para que dicha combinación sea posible, lo primero que debe hacerse es llevar todas las medidas a una misma escala, específicamente a [0.1] indicando el límite superior una semejanza perfecta entre las cadenas de texto según la medida. Una vez realizado este procedimiento una combinación de medidas puede definirse como:

$$\sum_{i=1}^8 P_i M_i$$

Donde M_i es el valor arrojado por la medida i , mientras que P_i es el peso que se le da a dicha medida en la combinación (0 ó 1). Según esta fórmula, existen $2^8 - 1 = 255$ posibles combinaciones de las medidas consideradas.

Para determinar el valor del vector P se realizó una calibración experimental que consistió en tomar 10 OE de dominios de conocimiento diversos y 20 OA relacionados con cada uno. Uno de tales objetivos fue por ejemplo: “Comprender el cálculo del perímetro de figuras geométricas básicas”. Para cada objetivo un grupo de expertos valoró los OA relacionados dándoles un orden de relevancia entre 1 y 20 (siendo 1 el más relevante y 20 el más irrelevante) valiéndose únicamente de la información contenida en los metadatos: título, descripción y palabras clave. Posteriormente se llevó a cabo un proceso iterativo para cada una de las 255 combinaciones calculando el promedio entre las tres medidas (una para cada metadato) para cada OA. Cabe señalar que, dado que el estándar de metadatos permite campos multivariados tanto para descripción como para palabras clave, el proceso anterior consideró el mayor valor de similitud en caso de existir varias descripciones o varias palabras clave para un mismo OA. Finalmente, se contrastaron para cada OE los resultados arrojados tanto por el experto como por cada una de las combinaciones. De esta manera se encontró que el valor de P que minimizó la diferencia fue $\{1, 0, 0, 0, 0, 1, 0, 1\}$, que corresponde a la combinación de coseno, distancia vectorial, y distancia euclidiana. La Figura 3 muestra dicha comparación.

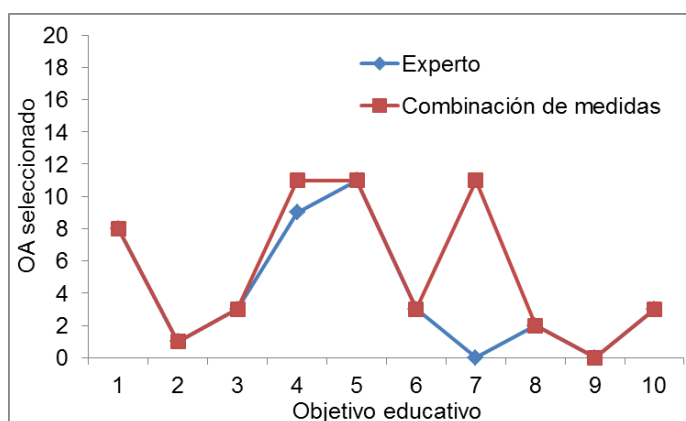


Figura 3. Comparación de la selección automática de OA vs. Selección de un experto

Como puede observarse, la selección automática arrojada por la combinación elegida coincidió perfectamente en 8 de los 10 casos de prueba.

Luego de terminado el procedimiento de selección, es decir, que para cada OE se haya encontrado un conjunto de OA relacionados desde uno o más repositorios y que para cada OA en dichos conjuntos, se calcule la medida de similitud agregada, la organización final se limita a escoger los k OA con los mayores valores de dichas medidas. La elección del valor de k no es una tarea fácil. En efecto, un valor demasiado pequeño facilita la labor del docente en el sentido que realiza un filtro muy exhaustivo, pero tiene como riesgo dejar por fuera OA potencialmente útiles. En el otro lado del espectro, un valor grande de este parámetro evitaría descartar OA potencialmente útiles pero aumentaría el esfuerzo posterior del docente de revisarlos. Considerando este dilema se recomienda un valor intermedio para k de 5 ± 2 . Aunque este valor podía parecer alto, es importante considerar que dependiendo del dominio de conocimiento y de las fuentes consultadas, una búsqueda de OA para un OE puede llegar a retornar decenas, sino centenas de resultados.

Volviendo al ejemplo presentado en la Figura 2, y considerando un valor de $k=3$, el resultado final al aplicar el modelo sería algo la estructura jerárquica que se presenta en la Figura 4.

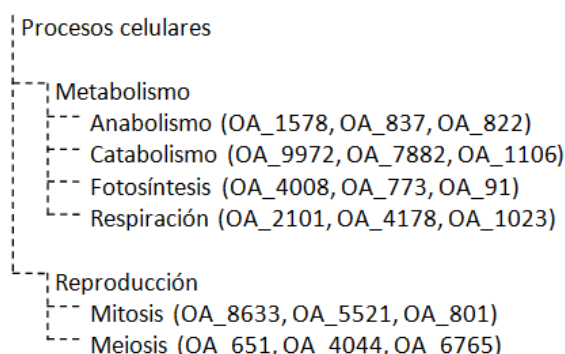


Figura 4. Ejemplo de organización de contenido educativo a partir de OA seleccionados

IV. Validación

A pesar de que el modelo propuesto es genérico, es decir, puede ser utilizado en diversos dominios del conocimiento y a nivel de detalle que se requiera, con el fin de hacer la validación se aplicó para un caso en el área de Ciencias de la Computación y se seleccionó como fuente de OA la federación de repositorios Educa Brasil-FEB (<http://feb.ufrgs.br/>). Se definió una estructura hipotética de OE como la que se presenta en la Figura 5.

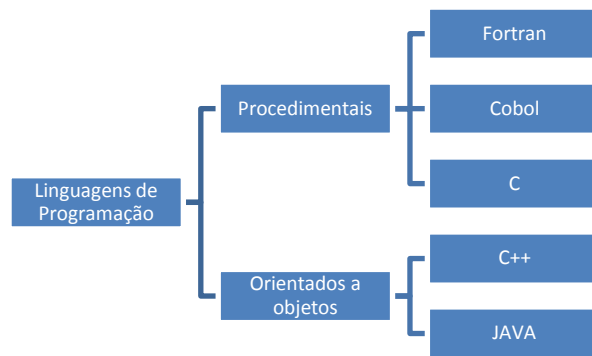


Figura 5. Estructura de un curso básico de lenguajes de programación

Por cuestiones de disponibilidad de los expertos que participaron en la validación, ésta se llevó a cabo para un solo OE específico: “Java”. Dicho objetivo fue descrito en más detalle como “*Aprendizagem de JAVA como um linguagem de programação orientado a objetos*”, y los términos de búsqueda empleados fueron precisamente “Java” y “*programação orientada a objetos*”. Del proceso de recuperación de OA se tomaron los primeros 10 resultados (ver Tabla III). El orden relativo en que se presentaron a los expertos fue alterado aleatoriamente respecto al orden arrojado en la búsqueda para evitar algún tipo de sesgo. Nótese que la descripción del objetivo está acorde con el repositorio FEB, y por tanto la mayoría los OA recuperados se encuentran en portugués –observe que el OA2 está en inglés.

Tabla III. Primeros 10 OA recuperados para el objetivo educativo “Java”

OA	Título
0	Ambiente de execução para aplicações escritas em Java no sistema EPOS
1	Curso de Java
2	Making java work for microcontroller applications
3	Sistema Aldeia: programação paralela e distribuída em Java sobre Infiniband e DECK
4	Javarray: uma arquitetura reconfigurável para o aumento de performance e economia de energia de aplicações embarcadas baseadas em Java
5	Padrões de projeto para desenvolver aplicações distribuídas com CORBA e Java
6	Execução distribuída de programas funcionais usando a máquina virtual Java
7	Programação funcional usando Java
8	Ambiente visual para programação distribuída em Java
9	JEduc: reflexão sobre a linguagem java na educação

A partir de estos OA, se pidió a 15 expertos –profesores del área de Ciencias de la computación– que seleccionaran los 3 objetos que, según su criterio, fueran los más apropiados para satisfacer el OE correspondiente. Específicamente se les pidió que para el OA más adecuado dieran una calificación máxima de 3, al segundo de 2 y al tercero de 1. Aunque los expertos provenían de ciudades y países diferentes (Porto Alegre y Florianópolis (Brasil), Lima (Perú), Manizales y Medellín (Colombia), todos estaban familiarizados con el idioma portugués. Los resultados de estas calificaciones se muestran en la Figura 6.

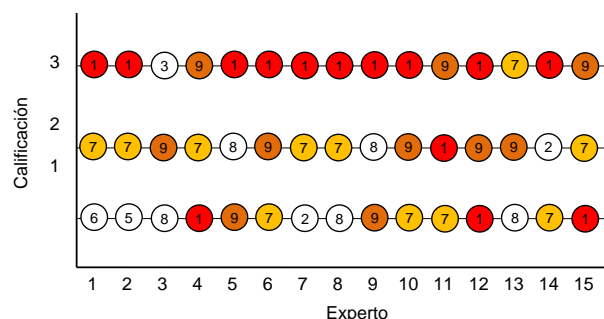


Figura 6. Selección de OA por parte de los expertos

Con el fin de realizar un comparativo, en la Tabla IV se muestra para cada OA tanto la valoración dada por el modelo propuesto como la valoración dada por los expertos. En el caso de que un profesor no haya seleccionado un OA como uno de los tres más relevantes, se asume una calificación de 0.

Tabla IV. Comparación modelo propuesto vs panel de expertos

Valoración		OA									
		0	1	2	3	4	5	6	7	8	9
Expertos	\bar{x}	0	2.267	0.2	0.2	0	0.133	0.067	1.267	0.467	1.4
	σ	0	1.163	0.561	0.775	0	0.352	0.258	0.961	0.743	1.183
Modelo		1.639	2.16	1.853	2.15	2.181	1.627	1.786	2.535	2.148	2.301

Como se puede inferir en la Tabla IV, con un valor de k igual a 3, el modelo propuesto coincide en un 67% con la opinión del panel de expertos, mientras que con un valor de 4 la coincidencia aumenta al 75%. Aunque estos resultados son muy alentadores, y en términos generales el orden relativo de selección por parte del modelo concuerda con la opinión de los expertos, existen algunos casos que escapan a esta tendencia. Un ejemplo de ellos es el OA4, el cual fue seleccionado como tercero por el modelo, mientras que ninguno de los expertos lo eligió. Este hecho no desvaloriza las bondades del modelo, simplemente ratifica la recomendación de un valor de k alrededor de 5 ± 2 : ni demasiado excluyente, ni demasiado amplio.

V. Conclusiones

Considerando los resultados experimentales reportados en la sección anterior y teniendo en cuenta que se realizó con un número limitado de docentes, se puede concluir que el modelo propuesto para la organización semiautomática de contenido educativo evidencia un acercamiento adecuado al seleccionar OA desde un repositorio a partir de la descripción de un OE específico o una colección de ellos. Lo anterior al compararlo contra los resultados que obtendría un sujeto real. De hecho, tal experimentación demostró que al usar un valor de k (cantidad de OA a seleccionar) de al menos 3, el porcentaje de coincidencias entre el modelo y un panel de expertos supera o iguala el 75%. Si bien este valor aún dista considerablemente de una coincidencia perfecta, lo cierto es que hay una ganancia considerable en términos del

esfuerzo requerido: mientras que para un sujeto real (considérese un docente por ejemplo), la selección de OA para un OE puede tomarle algunas decenas de minutos o incluso horas dependiendo de la cantidad de registros devueltos al hacer una búsqueda en uno o varios repositorios, esta misma tarea en el caso de una implementación computacional del modelo no tardaría más que unos pocos segundos. Esta comparación de esfuerzos se hace incluso más crítica cuando no se trata de un único OE, sino una colección de ellos como suele ser el caso cuando se prepara el material para un curso real. En otras palabras, pese a que la calidad en la selección de OA no es perfecta, y por tanto aún es requerida una intervención de una persona para seleccionar finalmente el o los OA más adecuados, la disminución del trabajo de revisión de esa persona se ve considerablemente disminuido por el uso del modelo.

Como complemento a la conclusión anterior es preciso mencionar que a pesar de haber realizado nuestro mejor trabajo para calibrar el modelo, lo cierto es que la moraleja, por así llamarla, encontrada durante esta investigación, es que en el contexto de la organización de contenido educativo jamás será posible suplantar completamente a un experto. Lo que sí se puede hacer es facilitar y agilizar su labor realizando un proceso de pre-filtrado que se traduce, de esta forma, en una disminución considerable de tiempo y esfuerzo.

Sin embargo, es importante señalar que antes que todo lo anterior sea cierto existe una clara limitación que debe superarse: la falta de calidad en los metadatos. Por más apropiado que sea un OA para una determinada necesidad, si éste no está adecuadamente especificado en sus metadatos, difícilmente podrá ser seleccionado con éxito del repositorio en el cual esté alojado o referenciado. Con esto en mente, la comunidad científica en el área ha venido trabajando de manera paralela en la creación de modelos para determinar la calidad de los metadatos (Tabares *et al.*, 2013), los cuales deberían ser un insumo previo al proceso de recuperación.

Como trabajo futuro existen varios frentes que se desean abordar. Por una parte, se pretende tener en cuenta más medidas de similitud de textos; esto con el fin de darle aún más robustez al modelo. Por otra parte, se quiere evaluar el uso del vector P utilizado en la combinación de medidas que sea continuo y no binario, esto para determinar si existen combinaciones parciales que mejoren los resultados. Así mismo, se desea realizar una calibración más exhaustiva de dicho vector realizando experimentos con una mayor cantidad tanto de OA como de expertos. Finalmente, se evaluará la posibilidad de emplear una combinación de medidas no de manera agregada para los metadatos considerados sino independientemente. Esto para determinar si existen diferencias significativas en la selección final al hacerlo de ese modo.

Referencias

Amón, I. y Jiménez, C. (2010). Funciones de similitud sobre cadenas de texto: una comparación basada en la naturaleza de los datos. *CONF-IRM Proceedings*, 58. Recuperado de <http://aisel.aisnet.org/confirm2010/58>

Baldiris, S. y Duque Méndez, N. (2012). Searching and positioning of contextualized *Learning Objects*, 5(3), 1-11.

Betancur, D., Moreno, J. y Ovalle, D. (2009). Modelo para la recomendación y recuperación de objetos de aprendizaje en entornos virtuales de enseñanza/aprendizaje. *Revista Avances en Sistemas e Informática*, 6(1), 45-56.

Bonny, J. y Renou, J. (2002). Euclidian distance-weighted smoothing for quantitative MRI: application to intervoxel anisotropy index mapping with DTI. *Journal of magnetic resonance*, 159(2), 183-9.

Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26(3), 297-302.

Duque, N. D. (2005). *Modelo de cursos virtuales adaptativos en un ambiente de planificacion inteligente*. Universidad Nacional de Colombia.

Grossman, D. y Frieder, O. (2004). *Information retrieval: algorithms and heuristics*. Berlín: Springer.

Haibo, G., Wenxue, H., Jianxin, C., Yong, Z. y Hui, M. (2008). *Pattern recognition of multivariate information based on non-statistical techniques*. International Conference on Information and Automation, 697-702. doi:10.1109/ICINFA.2008.4608088

Hamming, W. (1950). Error detecting and error correcting codes. *Bell System Technical Journal*, 29(2), 147-160.

Huang, Y.-M., Chen, J.-N., Huang, T.-C., Jeng, Y.-L. y Kuo, Y.-H. (2008). Standardized course generation process using Dynamic Fuzzy Petri Nets. *Expert Systems with Applications*, 34(1), 72-86. doi:10.1016/j.eswa.2006.08.030

Jaccard, P. 1908. Nouvelles recherches sur la distribution florale. *Bulletin de la Société Vaudoise des Sciences Naturelles*, 44, 223-270.

Kim, M. y Choi, K. (1999). A comparison of collocation-based similarity measures in query expansion. *Information processing & management*, 35(1), 19-30.

Learning Technology Standards Committee. (2002). *IEEE Standard for learning object metadata*. Nueva York: Institute of Electrical and Electronics Engineers.

Li, J. Z. (2010). *Quality, evaluation and recommendation for learning object*. International Conference on Educational and Information Technology, 533-537.

Mavrommatis, G. (2008). Learning objects and objectives towards automatic learning construction. *European Journal of Operational Research*, 187(3), 1449-1458. doi:10.1016/j.ejor.2006.09.024

Ouyang, Y. y Zhu, M. (2008). eLORM: learning object relationship mining-based

repository. *Online Information Review*, 32(2), 254-265. doi:10.1108/14684520810879863

Peña, C. I., Marzo, J., De la Rosa, J. L. y Fabregat, R. (2002). *Un sistema de tutoría inteligente adaptativo considerando estilos de aprendizaje*. Universidad de Girona, España.

Pithamber, R. P. (2003). Use and abuse of reusable learning objects. *Journal of Digital Information*, 3(4).

Real, R. y Vargas, J. M. (1996). The probabilistic basis of Jaccard's Index of similarity. *Systematic Biology*, 45(3), 380. doi:10.2307/2413572

Salton, G., Wong, A. y Yang, C. S. (1975). A vector space model for automatic indexing. *Communications of the ACM*, 18(11), 613-620.

Tabares, V., Duque, N., Moreno, J., Ovalle, D. y Vicari R. (2013) Evaluación de la calidad de metadatos en repositorios digitales de objetos de aprendizaje. *Revista interamericana de bibliotecología* (En dictamen).

Tseng, S. S., Su, J. M., Hwang, G. J., Hwang, G. H., Tsai, C. y Tsai, C. J. (2008). An object-oriented course framework for developing adaptive learning systems. *Educational Technology & Society*, 11(2), 171-191.

Van de Sompel, H. y Chute, R. (2008). The aDORe federation architecture: digital repositories at scale. *International Journal*, 9, 83-100. doi:10.1007/s00799-008-0048-7

Verbert, K., Ochoa, X., Derntl, M., Wolpers, M., Pardo, A. y Duval, E. (2012). Semi-automatic assembly of learning resources. *Computers & Education*, 59(4), 1257-1272. doi:10.1016/j.compedu.2012.06.005