

Precedencia Difusa y Generación de Itinerarios Docentes en Sistemas LCMS

Juan David Alarcón Serrano¹, José Raúl Fernández del Castillo Díez¹,
León A. González Sotos¹ y José Javier Martínez Herraiz¹

Departamento de Ciencias de la Computación, Universidad de Alcalá,
jd.alarcon@gmail.com, {joseraul.castillo, leon.gonzalez,
josej.martinez}@uah.es
<http://www.cc.uah.es>

Abstract. Los sistemas de búsqueda e indexación de contenidos han experimentado un gran avance en las últimas décadas. Los motores de búsqueda en internet han liderado la investigación en este campo consiguiendo resultados excelentes. En este trabajo se expone una propuesta de sistema que utilizando técnicas de análisis e indexación de contenidos sea capaz de generar Itinerarios Docentes de forma automática. El principal problema a resolver es cómo tomar la decisión de cuál es el mejor antecesor y sucesor para un Objeto Docente del grafo. La solución que se propone pasa por el uso de la teoría de la lógica difusa y la definición de un operador de agregación que para un determinado Objeto Docente nos calcule el grado de pertenencia como antecesor o sucesor para el resto de Objetos docentes.

1 Introducción

El rápido desarrollo del modelo conocido como e-formación o *e-learning*, cada vez más valorado, no como sustituto de la formación presencial tradicional, sino más como un complemento que se ha de adaptar según las necesidades y nivel de madurez del público receptor de esta formación [9], ha provocado que primero se desarrolle y luego se analicen las mejores soluciones. En el momento actual, todo está por hacer. Infinidad de esfuerzos se centran en conseguir hacer un verdadero sistema de *e-learning*. El *e-learning* no consiste en la distribución de temarios multimedia. Ese problema ya está resuelto. Ahora se busca conseguir sistemas que creen sensación de grupo, que sean capaces de entender las necesidades de los alumnos y de adaptarse a ellas [7], se buscan estándares adecuados como DC [5], SCORM [6], IMS [3], etc.

Los LCMS ¹ son sistemas basados en web que se utilizan para crear, aprobar, publicar, administrar y almacenar recursos educativos y cursos en línea. Los principales usuarios son los diseñadores instruccionales que utilizan los contenidos para estructurar los cursos, los profesores que utilizan los contenidos

¹ Learning Content Managment System

para complementar su material de clase e incluso los alumnos, que en algún momento pueden acceder a la herramienta para desarrollar sus tareas o completar sus conocimientos [10].

Los contenidos se almacenan como objetos descritos e identificables de forma única [8]. El objetivo de la creación de los objetos docentes, LOs², es que sean interoperables y reusables. La clasificación y la estructuración de los LO en los repositorios facilitan este objetivo. Si suponemos resuelto el problema de la generación de los LOs, otro problema al que hay que enfrentarse es organizar, clasificar y poner a disposición de los potenciales usuarios el LO producido.

La esencia de los LOs es ser utilizados en procesos de aprendizaje. En este artículo se propone un sistema LCMS capaz de recuperar información del contenido del LO, de sus metadatos y de su “historia” para poder construir de forma automática Itinearios Docentes en respuesta a las necesidades del usuario. Es decir, se pretende que el usuario del LCMS pueda preguntar al sistema cosas como:

1. quiero aprender a resolver sistemas de ecuaciones de primer grado
2. quiero conocer las guerras púnicas a nivel de 1º de bachillerato
3. quiero conocer las guerras púnicas a nivel de 3º de carrera
4. quiero saber programar en Java

Como se muestra en el artículo [7] la lógica difusa es una potente herramienta para el *e-learning* ya que permite personalizar los cursos. El LCMS que se propone utiliza un sistema de reglas difusas que evalúan los posibles antecesores y sucesores para un determinado LO.

En la sección 2 se describe la solución que se propone basada en el artículo “Organizing Encyclopedic Knowledge based on the Web and its Application to Question Answering” [4]. En la sección 3 se desarrolla el funcionamiento del sistema de recuperación de información y el sistema de reglas para la toma de decisiones. A continuación, en la sección 4 se describen los resultados que se prevén. Y por último en la sección 5 se abordan las conclusiones y trabajos futuros.

2 Solución que se plantea

Tomando como base el sistema que plantea [4], el sistema dispondrá de tres módulos: recuperación, extracción y organización. No obstante, la funcionalidad difiere sensiblemente de la que ellos plantean. En esta sección se describen los módulos que implementará el sistema. La ejecución de estos módulos deberá ser planificada para que se ejecute en los momentos de menor carga del sistema.

2.1 Recuperación

La fuente de LOs para el sistema no será la Web. El sistema que se plantea es *per se* un almacén de LOs y serán estos los que habrá que analizar. Por lo

² Learning Objects

tanto, nuestro universo de estudio de documentos se acota significativamente. Este módulo tendrá la misión de planificar cual es el siguiente LO que hay que analizar, es decir, establecerá un orden para los LOs en función de prioridades:

1. los más utilizados
2. los mejor valorados
3. los de un determinado formato
4. fecha más reciente
5. etc.

2.2 Extracción

Los LOs almacenados en el sistema, tendrán asociados unos metadatos, en concreto Dublin Core; también tendrán un formato adecuado y fundamentalmente no hay que perder de vista que se está hablando de un LCMS. Todos los contenidos tendrán una buena estructura ya que existen procesos de revisión y validación previos a la aceptación y publicación de contenidos [11].

Vuelven a aparecer unos límites claros en nuestro universo de estudio que favorecen una buena calidad en los resultados de este paso. Este módulo se encargará de recuperar información del documento estudiando su: estructura y contenido. Tendrá en cuenta palabras clave, definiciones, referencias bibliográficas y referencias a otros LOs (HDL [2]), tendrá que generar una estructura de índice inverso (se pueden utilizar sistemas como Lucene [1]) sobre el texto. Con este análisis el sistema será capaz de completar algunos metadatos del conjunto de DC como por ejemplo: palabras clave, formato, tipo de recurso, lengua o nivel.

2.3 Organización

El estándar SCORM prevé una organización más compleja entre LOs de lo que sería un típico índice secuencial. Lo que se pretende en este sistema es conseguir definir grafos dirigidos donde los nodos son LOs y los arcos son relaciones entre los nodos. SCORM permite trabajar con esta información. Es necesario analizar la información obtenida en el paso anterior para establecer las relaciones entre los LOs.

Por el contexto en el que se plantea el sistema se han podido acotar ciertas características, como por ejemplo, temas de formato o que son contenidos docentes. Sin embargo, es imposible establecer unos límites claros a la hora de decidir si A es antecesor de B para cualquier A y B perteneciente al conjunto de LOs almacenados en el LCMS, es decir:

$$A, B \in LOs, A \rightarrow B \quad (1)$$

La lógica difusa soluciona este problema. En la sección 3 se van a especificar un conjunto de reglas en lenguaje natural que permiten tomar esta decisión. Posteriormente se traducirán estas reglas en una función de pertenencia tal que:

$$F_B(A) = t_a \quad (2)$$

donde t_a es el grado de pertenencia de A al conjunto de predecesores de B

2.4 Formalización y gestión del conocimiento y lógica difusa

Muchos de los marcos utilizados en la representación y gestión del conocimiento están basados en la lógica formal y en el cálculo de predicados de primer orden, que proporcionan un buen modelo para razonar con información cierta, pero que en cuanto se intenta aplicarlos al mundo real pueden resultar ineficaces o inviables. Ello se debe a frecuentes contingencias derivadas de la “calidad” de la información que se maneje. Así, puede ocurrir que falte información relevante, debida a la insuficiencia de las observaciones, a que la información disponible sea errónea debido a la fiabilidad limitada de los instrumentos de medida, situación frecuente en los procesos psicológicos inherentes al proceso de enseñanza-aprendizaje, como lo es el que tenga que involucrar apreciaciones subjetivas, ser resultado de una conjunción de informaciones conflictivas o redundantes etc.

Por su propia índole, estas “deficiencias”, que hay que tener en cuenta si se quiere ser realista, no siempre pueden resolverse buscando la información necesaria ni depurando la disponible, tareas con frecuencia costosas si acaso posibles. Además hay otros aspectos que no pueden ser tratados por medio de la lógica formal ordinaria, entre los que están algunos que no se pueden soslayar en el ámbito que aquí nos interesa, como lo son el hecho de que el proceso de enseñanza-aprendizaje es esencialmente dinámico, lo que conlleva la necesidad de tener muy en cuenta el orden de ejecución y la temporalidad de las tareas, la causalidad que pueda relacionar unas con otras, la relevancia de las mismas dependiendo del contexto y la fase en que se esté y otra serie de factores que afectan a la organización del conocimiento y que tienen que ser tenidos en cuenta en ocasiones en los sistemas LCMS. Es claro que en afirmaciones del tipo “si se hace ... entonces ocurrirá que...” frecuentes en el modo de expresión ordinario están involucradas la temporalidad y la causalidad, que trascienden el alcance expresivo de la lógica clásica.

Aparte de estas incidencias en la organización del conocimiento, en el contexto de nuestro problema de gestión de los recursos docentes, puede haber otras relativas al conocimiento mismo con el que que rara vez es en este campo se puede considerar totalmente apodíctico ni categórico, dado que con frecuencia hay que tener en cuenta que el conocimiento puede ser incierto, las reglas pueden tener excepciones, puede faltar información o ser ésta confusa, inconsistente, vaga, ambigua o imprecisa debido entre otras causas a la índole del tema y a que estas son características habituales del lenguaje ordinario en que se expone.

En estas circunstancias la formalización de los procesos de razonamiento no la podemos realizar solamente por medio de la lógica formal ordinaria, que no dispone de modos de representación ni de inferencia para predicados vagos. Una breve lectura del sistema de reglas que se plantea en 3.3 muestra que es conveniente usar técnicas de lógica difusa. La veracidad, entendida como correspondencia de los predicados con los hechos no debe ser entendida en el ámbito de los objetos docentes de un modo dicotómico, sino gradual, circunstancia que abunda en la conveniencia del empleo de la lógica alternativa basada en los conjuntos difusos y la modelización en ellos basada, que ofrece la posibilidad de ponderar

niveles de certeza mediante distribuciones de posibilidad, con términos y niveles de veracidad graduales pero representables en esquemas numéricos computables.

3 Propuesta de sistema

En esta sección se desgana el diseño del sistema propuesto. Inicialmente se describirá el modelo de datos capaz de almacenar la información necesaria sobre cada LO; a continuación se explicará el sistema de reglas (lógica difusa) que se aplicará para decidir si un LO A es antecesor de otro LO B ; por último, se describe el sistema de representación que se aplicará para las relaciones que se descubran entre LOs.

3.1 Modelo de datos

Para un LO A , el módulo de extracción de información tendrá que llevar a cabo las siguientes tareas:

- extraerá el conjunto de términos que se definen en el mismo. Para decidir qué terminos se definen en un LO se aplicará la propuesta de [4].
- A continuación asociará una lista de términos que se utilizan en su representación. De la misma manera se extraerán las referencias tanto hipervínculos como referencias bibliográficas. De esta forma se construira una tabla de terminos definidos, terminos utilizados en las definiciones y referencias.

Cruzando las tablas de teminos definidos con la tabla de terminos utilizados en definiciones de otro LO se podrá decidir si es o no su antecesor. Tomando como modelo base el que se utiliza en ETSIISpace [11], para cada LO se propone almacenar además, la siguiente información:

- HDL identifier[2] al modo en que se usan en ETSIISpace[11]
- Lista de terminos definidos
- Lista de terminos utilizados en las definiciones
- Lista de referencias bibliográficas
- Lista de referencias a otros HDLs
- Historico de uso en otros Itinerarios Docentes. En concreto la lista de HDLs de otros LOs para los que ha sido antecesor.

No hay que perder de vista que además de esta información y dado que estamos trabajando con un LCMS también disponemos de los metadatos. En este caso, del conjunto de metadatos DC.

3.2 Modelo de representación

La representación para las relaciones entre los LOs se hará mediante un grafo dirigido. Lo arcos estarán caracterizados por un peso. Este peso representa la calidad de la relación, es decir, si por ejemplo en nuestro sistema disponemos

de un LO A que explica como sumar y otro C que explica como resolver una potencia del tipo x^y , se puede establecer una relación entre $A \rightarrow C$ con un peso/distancia grande; son necesarios unos conocimientos intermedios antes de poder pasar de la suma a la resolución de potencias. Por el contrario, tenemos el anterior LO A y ahora disponemos de otro B que nos enseña a multiplicar. La distancia entre $A \rightarrow B$ será menor. También aparecerá una relación $B \rightarrow C$ con una distancia relativamente pequeña.

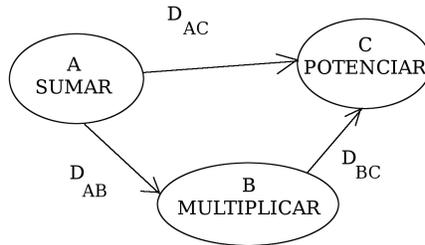


Fig. 1. Ejemplo de grafo de LOs

Esto permite jugar con algoritmos para el recorrido del grafo en función del tipo de Itinerarios Docentes que se desee obtener. Por ejemplo, utilizando el algoritmo de *búsqueda avariciosa* se conseguirían como resultados Itinerarios Docentes muy en detalle, recorriendo la mayor cantidad de nodos; también se podría jugar con distintas heurísticas o favoreciendo que se usen el mínimo número de nodos. Las posibilidades son grandes y habría que estudiar en detalle este aspecto, por esto se deja propuesto como trabajo futuro (consultar sección 5).

3.3 Sistema de reglas

$A, B, C \in LOs$ / A es antecesor de B si:

1. B hace algunas referencias a A en la bibliografía o mediante HDL³

$$\alpha(x) = \begin{cases} \frac{1}{4}x & (0 < x < 4) \\ 1 & (x > 4) \end{cases} \quad (3)$$

x es el número de veces que se referencia en el texto al LO A .

2. B utiliza para sus definiciones bastantes términos definidos en A

$$\beta(n_{BA}) = \begin{cases} \frac{10n_{BA}-2}{4} & (0.2 < n_{BA} < 0.6) \\ 1 & (n_{BA} > 0.6) \\ 0 & (n_{BA} < 0.2) \end{cases} \quad (4)$$

³ Mediante la referencia bibliográfica se tendrá que resolver cuál es el HDL del objeto al que se referencia

x es el porcentaje de palabras clave utilizadas por B que se definen en A

$$\begin{aligned} W_B &\equiv \text{Conjunto de palabras clave de } B \\ D_A &\equiv \text{Conjunto de palabras definidas en } A \\ \tau_{BA} &= |W_B \cap D_A| \equiv \text{Cardinal de la intersección entre ambos conjuntos} \\ n_{BA} &= \frac{\tau_{BA}}{|W_B|} \end{aligned} \quad (5)$$

3. A a sido utilizado más veces como antecesor de B

$$\gamma(x) = \begin{cases} \frac{1}{10}x - \frac{1}{10} & (1 < x < 10) \\ 1 & (x > 10) \\ 0 & (x < 1) \end{cases} \quad (6)$$

el sistema almacenará un historico con las veces que A ha sido utilizado como antecesor de B

4. A y B son de niveles parecidos

$$\delta(x) = \begin{cases} -\frac{1}{2}x + \frac{3}{2} & (0 < x < 2) \\ 0 & (x > 2) \end{cases} \quad (7)$$

x es la diferencia de nivel entre A y B en valor absoluto. El nivel se almacena como metadato DC.

5. A y B están en el mismo idioma, función *crisp*⁴

$$\mu(A, B) = \begin{cases} 1 & \text{si} \\ 0 & \text{no} \end{cases} \quad (8)$$

el idioma se almacena como metadato DC

6. A y B pertenecen a la misma categoría, función *crisp*

$$\nu(A, B) = \begin{cases} 1 & \text{si} \\ 0 & \text{no} \end{cases} \quad (9)$$

la categoría se almacena como metadato DC

El peso en los arcos representa la calidad de la relación entre LOs. El cálculo final para el valor de la distancia sería el que se puede ver en la ecuación 10

$$\begin{aligned} F_B(A) &= \frac{\alpha c_1 + \beta c_2 + \gamma c_3 + \delta c_4 + \mu c_5 + \nu c_6}{6} \\ c_1 + c_2 + c_3 + c_4 + c_5 + c_6 &= 1 \end{aligned} \quad (10)$$

4 Resultados previstos

Vamos a ilustrar las consideraciones anteriores con un pequeño prototipo, con LOs muy sencillos, relativos a la aritmética de los números naturales, intentando que sirva de concreción y muestre el tipo de resultados que obtendría el modelo en que venimos tratando. Supongamos los siguientes LOs A , B y C (son del mismo idioma, de la misma categoría, no hay información histórica sobre relaciones anteriores, $c_1 = c_2 = c_3 = c_4 = 1/4$ y $c_5 = c_6 = 0$).

⁴ sólo puede tomar valor de verdadero o falso

- A, Sumar y Restar
 - Sumar** *Dados dos conjuntos disjuntos X e Y , cuyos cardinales son x e y respectivamente, se define la suma de x e y como el cardinal del conjunto Z unión de X e Y*
 - Restar** *dos números x e y es calcular otro número z cuya suma con y dé x*
 - HDL, bibliografía** \emptyset
- B, Multiplicar y dividir
 - Multiplicar** *consiste, dados dos números, en sumar el primero de ellos (multiplicando) consigo mismo tantas veces como indique el segundo (multiplicador)*
 - Dividir** *consiste, dados dos números, en calcular cuantas veces (cociente) se puede restar el segundo de ellos (divisor) del primero de ellos (dividendo), mientras lo restante siga siendo mayor que el divisor, hasta que no lo sea y quede el resto*
 - HDL, bibliografía** A(3)
- C, Potencias
 - potencia** *un número (base) a otro (exponente) consiste en multiplicar consigo mismo al primero tantas veces como indique el segundo*
 - HDL, bibliografía** A(1), B(3)

$$\begin{aligned}
 D_A &= \{\text{sumar, restar}\} \\
 D_B &= \{\text{multiplicar, dividir}\} \\
 D_C &= \{\text{potencias}\} \\
 W_A &= \{\text{conjunto, cardinal, unión}\} \\
 W_B &= \{\text{sumar, restar, conjunto}\} \\
 W_C &= \{\text{multiplicar, base, exponente}\} \\
 B_A &= \{\emptyset\} \\
 B_B &= \{A(3)\} \\
 B_C &= \{A(1), B(3)\}
 \end{aligned}$$

$$\begin{aligned}
 F_A(B) &= 0 \\
 F_A(C) &= 0 \\
 F_B(A) &= \frac{1.166}{6} = 0.194 \\
 F_B(C) &= 0 \\
 F_C(A) &= \frac{0.50}{6} = 0.083 \\
 F_C(B) &= \frac{1.083}{6} = 0.180
 \end{aligned}$$

4.1 Problemas previstos

La calidad del sistema que se propone depende en gran medida de la calidad de la extracción de palabras clave de los LOs, la resolución de referencia bibliográficas a HDL y en general, que se cumplan las condiciones que se esperan de un LCMS. Un gran problema sería que los LOs fueran libros completos. En esta situación se complican los cálculos de precedencias y bajará la calidad de los resultados

en las búsquedas. Se entiende que un LO debería ser algo así como un artículo de la Wikipedia⁵.

Otro problema claro sería: ¿qué sucede si el LO A define un concepto que B utiliza y a su vez B define otros que usa A ?, es decir, ¿qué se hacen con los bucles? Este es un problema interesante que se podría resolver con un buen algoritmo de recorrido, utilizando una búsqueda heurística, en lugar de búsquedas avariciosas. Este sería un tema a tratar en trabajos posteriores (consultar 5).

5 Conclusiones y trabajos futuros

El sistema que se propone en este artículo es fruto de la aplicación de los sistemas de recuperación de información a los LCMS. Con la experiencia adquirida en el desarrollo de ETSIISpace [11] se ve como en los LCMS actuales hay una gran carencia en la solución de las búsquedas y recuperación de LOs. Las búsquedas son meras extensiones de un índice inverso, los metadatos, ya sean DC (Dublin Core) o LOM, rara vez se rellenan con seriedad, fundamentalmente por el tiempo necesario para ello; la información que se almacena en estos sistemas suelen carecer de granularidad, se suelen publicar capítulos o artículos completos para no perder tiempo en su posterior estructuración.

Un gran número de metadatos podrían ser completados por el propio sistema aplicando técnicas de recuperación de información, lógica difusa y sistemas expertos. También el sistema nos podría apoyar en la labor de construir Itinerarios Docentes mediante asociaciones de LOs. Un LCMS no es una base de datos acotada a documentos científicos o docentes. Cuando se hace una búsqueda sobre un LCMS no se espera obtener un artículo o un recurso web, se espera obtener una explicación de un concepto que puede implicar un conjunto ordenado de varios LOs, un Itineario Docente.

Las líneas de trabajo que se abren en este campo serían las siguientes:

1. Estudio del estandar IMS/LD que incluye SCORM y las posibilidades de organización de LOs que ofrece. Es fundamental resolver el problema técnico del almacenamiento de los metadatos y las metaestructuras asociadas a cada elemento de una forma óptima.
2. Estudio del sistema de recuperación de información más adecuado para el contexto del sistema. Teniendo en cuenta que no sólo se dispone de documentos, también hay unos metadatos asociados y también se puede aprovechar la experiencia del sistema.
3. Sería interesante plantearse técnicas de inteligencia artificial, como por ejemplo redes neuronales, para ajustar los valores umbrales en la construcción de Itinerarios Docentes tomando como información de entrada otros cursos ya estructurados.
4. Estudiar en detalle las distintas posibilidades en cuanto algoritmos de recorrido del grafo de LOs del LCMS. Grado de ramificación, heurísticas, etc.

⁵ <http://www.wikipedia.org>

5. Refinar las funciones de pertenencia, por ejemplo, el número de términos sería más adecuado hacerlo en función del número de terminos utilizados en le LO, es decir, si el LO tiene 100 definiciones y utiliza dos terminos de otro LO la precedencia es débil, sin embargo, en un documento con 3 definiciones que utilice 2 terminos de otro tienen una dependencia mayor.
6. Sería interesante combinar el sistema que aquí se plantea con Mediawiki⁶. Los contenidos tienen una estructura y tamaño ideal para su análisis. Esta simbiosis permitiría mejorar la organización de los contenidos de los sistemas wiki y permitiría hacer búsquedas que devuelvan varios artículos con algún hilo argumental.

⁶ <http://www.mediawiki.org>

References

1. Apache lucene, 2007. <http://lucene.apache.org/>.
2. The handle system, 2007. <http://www.handle.net/>.
3. Ims global learning consortium, 2007. <http://www.imsglobal.org/>.
4. Atsushi Fujii and Tetsuya Ishikawa. Organizing encyclopedic knowledge based on the web and its application to question answering. *University of Library and Information Science*, 2001.
5. Dublin Core Metadata Initiative. Dublin core, 2007. <http://www.dublincore.org/>.
6. Advanced Distributed Learning. Scorm, 2007. <http://www.adlnet.gov/scorm/index.aspx>.
7. Jose-L Marzo, Clara Inés Peña, Josep Lluís de la Rosa, and Ramón Fabregat. *Un sistema de tutoría inteligente adaptativo considerando estilos de aprendizaje*. Universitat de Girona, 2002.
8. Francisco José García Peñalvo. Estado actual de los sistemas e-learning. *Teoría de la Educación, Educación y Cultura en la Sociedad de la Información*(6), 2002.
9. Francisco José García Peñalvo and Joaquín García Carrasco. Los espacios virtuales educativos en el ámbito de internet: un refuerzo a la formación tradicional. *Teoría de la Educación, Educación y Cultura en la Sociedad de la Información*(3), 2001.
10. Raghavan Rengarajan. Lcms and lms: Taking advantage of tight integration. *click2learn*, 2001.
11. Juan David Alarcón Serrano. *Repositorio de Objetos Docentes: Análisis y evaluación de la situación actual. Desarrollo de ETSIISpace*. Escuela Técnica Superior de Ingeniería Informática, Universidad de Alcalá, Noviembre 2006.