

# Wikificación automática de objetos de aprendizaje

Claudia Pérez Martínez <sup>a,b,c</sup> and Alexander Gelbukh <sup>c</sup>

<sup>a</sup> *Universidad Autónoma del Estado de México, Centro Universitario Valle de Chalco, México*

<sup>b</sup> *Tecnológico de Estudios Superiores de Chalco, México*

<sup>c</sup> *Centro de Investigación en Computación, Instituto Politécnico Nacional, México*

**Abstract.** Wiley's definition of learning objects, "any digital resource that can be reused to support learning," is sufficiently general to permit us to include some tools in this resource, as is defined in the LOM standard where it refers to the possibility of adding software tools. We give a methodological proposal for enriching the contents of the learning objects, basing on the concept of wikification developed by Milne & Witten and other authors. Wikification is a process of linking phrases of the text with Wikipedia entries relevant for this phrase. This methodology has been defined as a series of steps, one of which is the calculation of the quality of the context of the phrase. We propose a new form of performing this calculation, taking into account the information in the object's metadata.

**Keywords:** Learning object, word sense disambiguation, *wikification*, natural language processing.

**PACS:** 00.01.20.+x, 00.01.40.Di, 00.01.40.gb, 00.01.50.ht.

**Resumen.** La definición de Wiley para objetos de aprendizaje "Cualquier recurso digital que pueda ser utilizado como soporte al aprendizaje", es lo suficientemente amplia para permitir la inclusión de utilidades a este recurso, tal como se ha establecido en el estándar LOM al referir la posibilidad de la agregación de herramientas de software. En este documento se hace una propuesta metodológica para el enriquecimiento del contenido de los objetos de aprendizaje, basada en el concepto de wikificación que Milne y Witten y otros autores han desarrollado. La wikificación es un proceso de vincular frases de un texto a los artículos de Wikipedia que hablan sobre tal frase. Tal metodología se ha definido como una serie de pasos, siendo uno de ellos el cálculo de la calidad del contexto de la frase. El objetivo de este trabajo es proponer una nueva forma de realizar el cálculo de la calidad del contexto de la frase, tomando en cuenta la información que proporcionan los metadatos del objeto de aprendizaje.

**Palabras clave:** Objeto de Aprendizaje, desambiguación del sentido de las palabras, wikificación, procesamiento de lenguaje natural.

**PACS:** RPACS 00.01.20.+x, 00.01.40.Di, 00.01.40.gb, 00.01.50.ht.

## INTRODUCCIÓN

Los Objetos de Aprendizaje (OA), se han constituido como recursos para la instrucción basada en el uso de la computadora. No existe una definición única para los OAs, pero una de las primeras definiciones la presentó formalmente IEEE como una "entidad, digital o no digital que puede ser utilizada, reutilizada o referenciada durante el aprendizaje apoyado en tecnología" [8]. A través del tiempo, la definición de OA ha tenido aportaciones de diferentes autores que especifican más su descripción. Wiley señala que un OA es "Cualquier recurso digital que pueda ser utilizado como soporte al aprendizaje" [18].

Aunque las definiciones y presentación de un OA son muy diversas, todos contienen un elemento común, los metadatos, se adjuntan al OA en un archivo formato XML, construido bajo estándares internacionales, que proporciona información acerca del OA en forma de descriptores o campos de información, por ejemplo: título, descripción, url, formato, etc., como se detallará posteriormente. En la práctica, los objetos de aprendizaje se presentan como un documento de Word, una página HTML, un archivo con extensión pdf, una animación en flash, o cualquier archivo en formato digital que pretende ayudar a aprender y/o enseñar algún conocimiento, como puede observarse en diferentes repositorios de objetos de aprendizaje de libre acceso [6], [3], [7]. Los OAs, desde la perspectiva de generación de conocimiento, no supone sólo el uso de la información contenida en el objeto, sino una "transformación de la misma para la generación de nuevos insumos informativos que alimentan la misma red de

objetos” [4]. Bajo este concepto se presenta la idea de enriquecer automáticamente un objeto de aprendizaje con información proveniente de la Enciclopedia digital más grande existente, Wikipedia [17].

En este documento se hace una propuesta metodológica para el enriquecimiento automático del contenido de un OA, denominado wikificación, con fundamento en la metodología de Milne y Witten [15], quienes han desarrollado previamente una metodología que identifica términos relevantes en un texto no estructurado y genera referencias a Wikipedia de manera automática; aquí la desambiguación de términos es quizá el reto más importante en este proceso, y consiste en seleccionar de un conjunto de artículos de Wikipedia, aquél que corresponda al contexto de un término específico; para lo cual, se considera contexto el conjunto de palabras circundantes al término que se pretende desambiguar. Por ejemplo, para el término “puma” se deberá seleccionar el artículo de Wikipedia que corresponde a un mamífero si el contexto indica que se está tratando acerca de animales; sin embargo, quizá deberá seleccionarse el artículo referente al equipo de fútbol denominado “Pumas” si el contexto indica que se está hablando de deportes.

La propuesta en este documento es utilizar la información contenida en los metadatos, específicamente, el título, la descripción y las palabras clave, con el objetivo de reforzar el contexto del término a desambiguar. Se ha supuesto que una frase dentro del objeto de aprendizaje tendrá una relación significativa, en la mayoría de los casos, con las palabras clave y con el título y la descripción de tal OA.

En este documento se presenta el trabajo relacionado con la wikificación, describiendo las dos metodologías que se han encontrado en la literatura: la de Mihálcea y Csomai [14] y la de Milne y Witten [15]. En ellas se identificará la importancia de la evaluación de la calidad del contexto y de la desambiguación en el proceso de Wikificación. Posteriormente se presenta teoría relacionada con la descripción de un Objeto de Aprendizaje y la estructura estándar de los metadatos. Con esta información se procede a describir la propuesta metodológica motivo de este documento, en la cual se mejora la calidad del contexto incluyendo para ello algunos metadatos del objeto de aprendizaje

## TRABAJO RELACIONADO

La *wikificación* es un proceso que realizan los editores de Wikipedia (*wikipedians*) sobre el texto de sus artículos, y consiste en seleccionar de ahí palabras o frases relevantes y vincularlas hacia algún artículo de Wikipedia que las describan. Esta vinculación implica, en algunos casos, seleccionar de entre varios artículos de Wikipedia con el mismo nombre pero con diferentes significados, el más adecuado de acuerdo al contexto de la frase. A esta frase que tiene más de un posible vínculo en Wikipedia se le denomina frase ambigua. El *wikipedian* entonces ha realizado una desambiguación, y lo ha hecho a través de un proceso de inferencia basado en un conocimiento amplio que él posee que va más allá del contenido del documento que ha escrito.

Llevar a cabo el proceso de wikificación de manera automática, se vuelve un proceso mucho más difícil, ya que la computadora no tiene ningún conocimiento, ya que para esta máquina, cada documento que posee, solo es un conjunto de palabras. Lo anterior sugiere que, de alguna manera, se tiene que hacer que la computadora pueda “entender” esa información que está almacenando, y que representa el conocimiento de un humano. Algunos autores han realizado este proceso a través de técnicas basadas en el uso de herramientas de inteligencia artificial obteniendo diferentes grados de eficiencia.

Hasta donde se ha investigado, no se tiene conocimiento de que se haya realizado un proceso de “*wikificación automática*” del contenido completo de un Objeto de Aprendizaje, visto como una funcionalidad más de estos recursos digitales. Sin embargo, se han encontrado dos propuestas de wikificación de textos no estructurados, en 2007, Mihálcea y Csomai [12], proponen una metodología de wikificación, cuyo proceso implica dos fases: selección de frases relevantes y desambiguación, para probar sus resultados, utilizaron diferentes tipos de documentos, incluyendo objetos de aprendizaje. Una propuesta más reciente y con mejores resultados, es la de Milne y Witten, quienes utilizan algoritmos de aprendizaje automático para desambiguar y posteriormente seleccionar las palabras a ser wikificadas, sin embargo no se tiene referencia de que hayan aplicado esta metodología para objetos de aprendizaje. A continuación se describen estas dos metodologías.

### Wikificación por generación y vinculación de palabras clave

Mihálcea y Csomai [14], crearon en 2007 un sistema denominado “Wikify”, y lo que hace es que, ingresando al sistema Wikify un texto no estructurado, este sistema identifica los conceptos importantes de dicho documento, devolviendo como salida las palabras clave del documento y los vínculos hacia artículos de Wikipedia de estas

palabras clave. Es decir, *Wikify* identifica las palabras clave de un documento y las *wikifica*. Para su funcionamiento, el sistema tiene dos funciones: extracción de palabras clave y desambiguación, las cuales hacen uso de Wikipedia.

#### *Extracción de Palabras Clave*

Las palabras clave son palabras o frases que son representativas del contenido del texto analizado. En el proceso de extracción de frases candidatas a ser palabras clave, se seleccionan todos los n-gramas (conjunto de una o más palabras) que son válidos dentro de algún vocabulario, y se calcula su ranking, para lo cual se utilizaron tres métodos: Tf-idf, X2-independence test y Key-phraseness.

#### *Desambiguación del Sentido de las Palabras*

Se proponen dos diferentes métodos de desambiguación: a) un método basado en conocimiento, este tipo de métodos se caracterizan por utilizar diccionarios, y b) un método basado en probabilidades, caracterizado por obtener información de cantidades grandes de datos. Aquí se utiliza la palabra ‘sentido’ para especificar cada uno de los vínculos (artículos de Wikipedia) hacia donde puede dirigirse una sola frase o palabra.

El proceso completo del sistema está dividido en cuatro fases:

- a) Si el documento es hipertexto, el sistema separa las etiquetas HTML del cuerpo del texto.
- b) El sistema pasa al módulo de extracción de palabras clave, las cuales identifican y marcan las palabras importantes y frases en el texto, determinadas por la probabilidad de que sean seleccionadas como vínculos.
- c) El módulo de extracción de palabras claves pasa el texto seleccionado hacia el módulo de desambiguación del sentido de las palabras, donde se resuelve la ambigüedad de los vínculos y completa las anotaciones con la referencia al artículo de Wikipedia correcto, utiliza un clasificador cuyas entradas son las características de la frase, sus palabras circundantes y son comparadas con ejemplos de entrenamiento obtenidos de toda Wikipedia.
- d) El sistema reconstruye la estructura del hipertexto original, incluyendo las nuevas referencias a los vínculos en el texto.

Este enfoque requiere un esfuerzo computacional considerable, la mayor precisión de esta metodología en la fase de extracción de palabras clave es de un 53.39 % en precisión, 55.9% en *recall*; en el proceso de desambiguación se obtiene una precisión del 94.33%, *recall* de 70.34%.

### **Wikificación a través de aprendizaje automático**

Milne y Witten [16] han desarrollado una metodología para enriquecer textos con vínculos hacia artículos de Wikipedia. Su metodología retoma algunas características del enfoque de Mihalcea y Csomai [14] sin embargo, la eficiencia y *recall* de este nuevo enfoque, ha resultado mucho mejor y con menos carga computacional. Las dos diferencias principales entre los dos enfoques son que en el segundo enfoque: a) los artículos de Wikipedia son utilizados para aprender cuáles términos deberían y cuáles no deberían ser vinculados (selección), y b) Se toman en cuenta los términos del contexto circundante para hacerlo, además de que consideran más características para seleccionar el sentido más adecuado de las palabras.

#### *Enfoque de Desambiguación de Términos.*

El enfoque de vinculación de Milne y Witten, es primero desambiguar para después seleccionar los términos que deberían ser vinculados. Se utiliza un clasificador para desambiguar los usos de los links encontrados en los artículos de wikipedia, y funciona de la siguiente manera:

Para cada término, palabra o frase vinculada, en un texto de Wikipedia, un “*wikipedian*” ha hecho personalmente una selección de un artículo de wikipedia, hacia el cual se vinculará tal término. Lo cual constituye un ejemplo positivo para una muestra de entrenamiento para un clasificador; con los artículos no elegidos para el mismo término, palabra o frase vinculada, se constituyen ejemplos negativos para la muestra de entrenamiento. El conjunto de ejemplos positivos y negativos constituirá la muestra de entrenamiento para el algoritmo de aprendizaje automático de desambiguación.

Como se muestra en la FIGURA 1, el resultado del proceso de desambiguación será un conjunto de asociaciones entre los términos en el documento y los artículos de Wikipedia que lo describen.

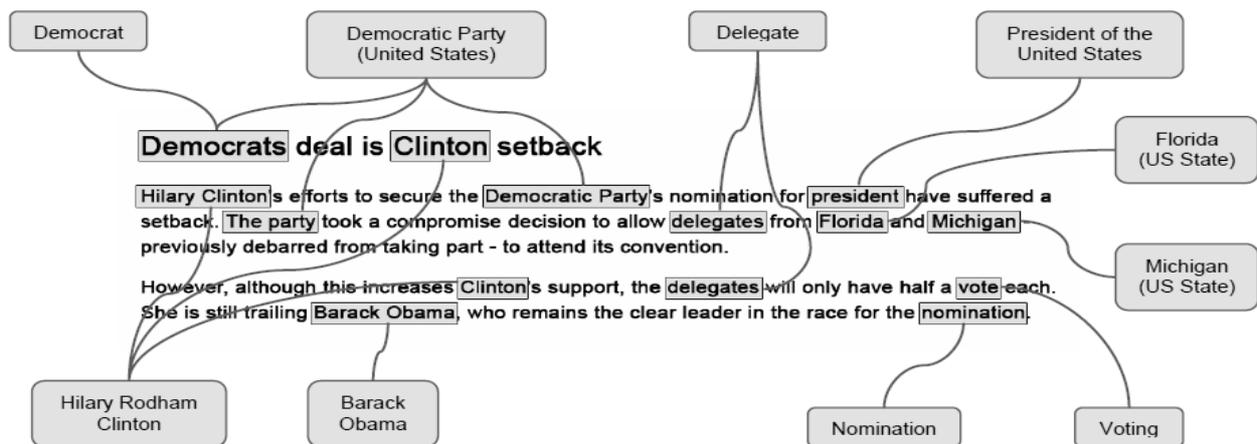


FIGURA 1. Asociación de Frases del Documento con Artículos Wikipedia Apropriados [11].

El desempeño de este desambiguador fue 96.6% de *recall*, de 97% de precisión. Lo que representa una mejora substancial con respecto a la propuesta de Csomani y Mihalcea, presentada en la sección anterior.

## OBJETO DE APRENDIZAJE

Como consecuencia a cambios producidos por la innovación tecnológica, por ejemplo Internet, se ha dado una transición dentro de la forma en la que se diseñan, desarrollan y entregan los materiales educativos, y es la llegada de una tecnología instruccional denominada Objetos de Aprendizaje, la cual dirige sus esfuerzos a llevar a cabo el diseño instruccional de una forma diferente, aprovechando las características de reusabilidad, generatividad, adaptatividad y escalabilidad, que éstos objetos de Aprendizaje poseen [18].

Los objetos de aprendizaje son elementos de instrucción basada en computadora que se construyen bajo el paradigma orientado a objetos. Estos objetos son entendidos como entidades digitales susceptibles de transportarse a través de Internet, significando que muchas personas pueden acceder y usarlo de manera simultánea (de manera opuesta a los medios de instrucción tradicionales) [18].

Los Objetos de Aprendizaje se han definido de formas diversas. La IEEE definió un Objeto de Aprendizaje como una “entidad, digital o no digital que puede ser utilizada, reutilizada o referenciada durante el aprendizaje apoyado en tecnología” [8]. Posteriormente Wiley ha delimitado esta definición señalando que un OA es “Cualquier recurso digital que pueda ser utilizado como soporte al aprendizaje” [18]. Y así, más autores han especializado las características de un OA. Un OA “debe tener un objetivo de aprendizaje, una unidad de instrucción que enseñe el objetivo y una unidad de evaluación que mida el objetivo” [2].

En la vida práctica, los objetos de aprendizaje se presentan como un documento de Word, un archivo con extensión pdf, una animación en flash, entre los más comunes, los cuales pretenden ayudar a aprender y/o a enseñar a alguien algún conocimiento. Una característica común a todos estos objetos es que cuentan con un archivo en formato HTML denominado metadatos, el cual será descrito en los párrafos siguientes. Actualmente, se encuentran objetos de aprendizaje disponibles de manera gratuita a través de repositorios, por ejemplo, FLOR, ARIADNE, GLOBE, MERLOT, entre muchos otros. La FIGURA 2 muestra un objeto de aprendizaje y sus metadatos.

Una de las finalidades del uso de metadatos es la estandarización, la cual es necesaria por causas distintas: a) cuando los recursos son creados deben considerarse tecnologías, políticas y formatos compatibles; b) cuando esos recursos son incluidos en un repositorio, y deben ser descritos, se utilizarán esquemas que aseguren su fácil localización y compatibilidad con otros sistemas de metadatos; c) cuando esos recursos sean utilizados y tengan que incorporarse a diferentes servicios, repositorios, plataformas y aplicaciones en un contexto dado; y d) cuando los sistemas involucrados en un entorno tengan que interoperar con otros para cumplir sus funciones o ampliar sus capacidades [5].

A lo largo del tiempo se han establecido estándares por parte de diferentes instituciones para los Objetos de Aprendizaje. Así fue creado el estándar IEEE Std 1484.12.1-2002, IEEE Standard for Learning Object Metadata [9]. El estándar agrupa a los metadatos en categorías, agrupando en cada una de ellas un conjunto de descriptores del OA [3], algunos descriptores que se muestran en la TABLA 1. Su utilidad en este documento se describe en secciones posteriores.

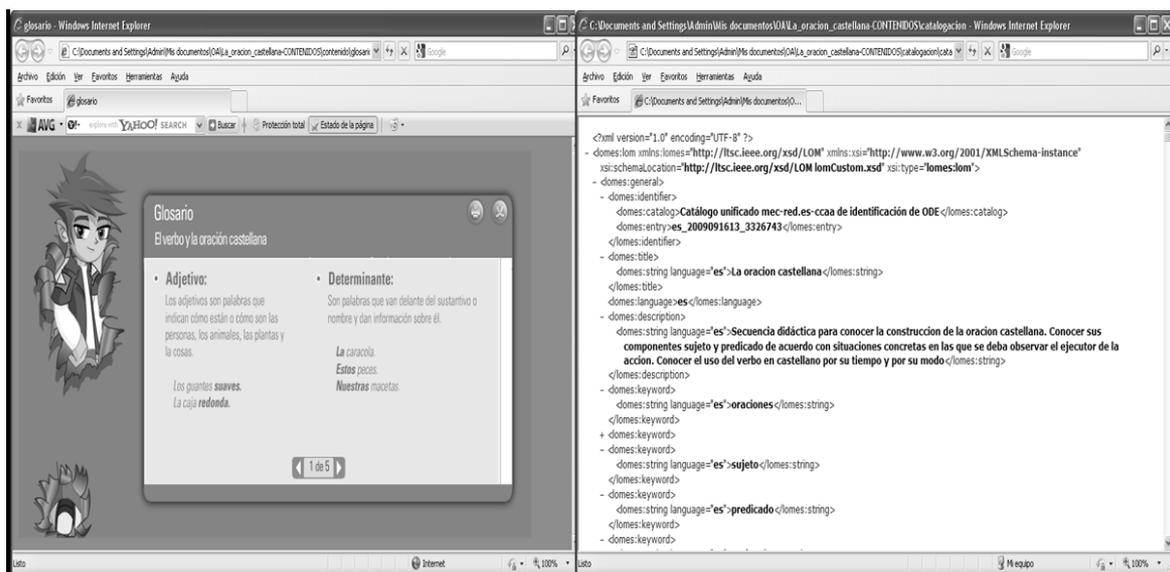


FIGURA 2. Presentación y metadatos del Objeto de aprendizaje “La oración castellana” [1]

TABLA (1). Algunos descriptores útiles para la desambiguación del contenido de un OA [3]

Categoría	Metadato	Descripción
General	Título	Nombre asignado al OA
General	Descripción	Descripción textual del contenido del OA
General	Palabras clave	Palabras clave o frases que describen los tópicos de este objeto de aprendizaje, estos datos no deben ser usados para características que pueden ser descritas por otros elementos.
Uso educativo	Contexto	Entorno dentro del cual el estudiante usa este OA (Escuela, educación superior, entrenamiento, otros)
Uso educativo	Rango de edad típica	Edad típica del usuario esperado
Uso educativo	Lenguaje	Lenguaje del usuario esperado
Clasificación	Trayectoria taxonómica	Trayectoria taxonómica en sistema específico de clasificación.
Clasificación	Fuente	Nombre del sistema de clasificación
Clasificación	Taxon	Término particular dentro de una taxonomía.

## PROPUESTA DE ALGORITMO PARA LA WIKIFICACIÓN DE OBJETOS DE APRENDIZAJE

Como se ha descrito previamente el proceso de *wikificar* implica seleccionar las palabras o frases relevantes para que éstas sean vinculadas a artículos de Wikipedia que las describan. Cuando un *wikipedian* realiza este trabajo puede encontrar que la frase relevante tiene asociados más de un vínculo en Wikipedia, visualizando entonces una ventana de desambiguación. Por ejemplo si se requiere vincular la palabra “puma”, Wikipedia tiene una página de desambiguación para esta palabra, como se muestra en la FIGURA 3.

El *wikipedian* seleccionará el vínculo de la definición a la frase que se relacione lógicamente con el contexto de su artículo. El *wikipedian* humano ha establecido en el artículo del “Puma concolor” los vínculos hacia frases relevantes, como se muestra en la FIGURA 4.

La definición más concreta de Wiley [18] para objetos de aprendizaje “Cualquier recurso digital que pueda ser utilizado como soporte al aprendizaje”, permite inclusión de muchos elementos a este recurso. Uno de los claros objetivos de un Objeto de aprendizaje es satisfacer necesidades de aprendizaje, no importando la modalidad de éste. Atendiendo a la naturaleza iterativa e incremental de construcción bajo el paradigma orientado a objetos, se debe reconsiderar cada vez la definición de los requerimientos, en este caso particular, satisfacer necesidades de

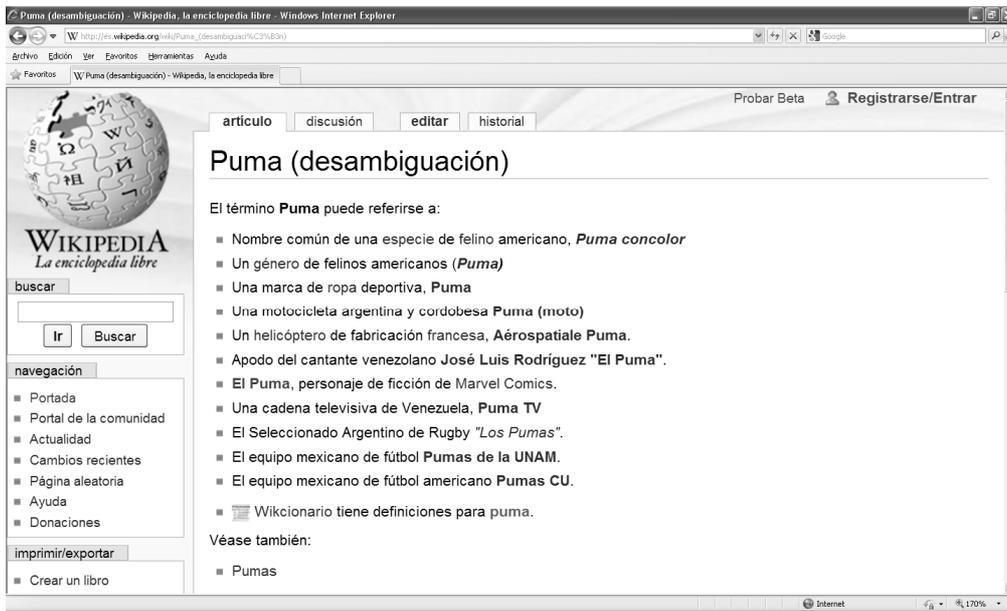


FIGURA 3. Página de desambiguación para la palabra “Puma” [17].

aprendizaje, como se ha establecido en el estándar LOM [9], en el cual especifica la inclusión de herramientas de software en el desarrollo de Objetos de aprendizaje. Por tal motivo, se propone el enriquecimiento automático de contenido de un objeto de aprendizaje, como una más de las funcionalidades de un objeto de aprendizaje. Diversos autores aceptan tres funcionalidades básicas del Objeto de aprendizaje, las cuales son: tutoría, ejercitación y evaluación [2], la wikificación del objeto de aprendizaje podría mirarse como una especificación del caso de uso Tutoría, y su utilización pretende contribuir a lograr mejores resultados en esta función. La FIGURA 5 esquematiza la propuesta.

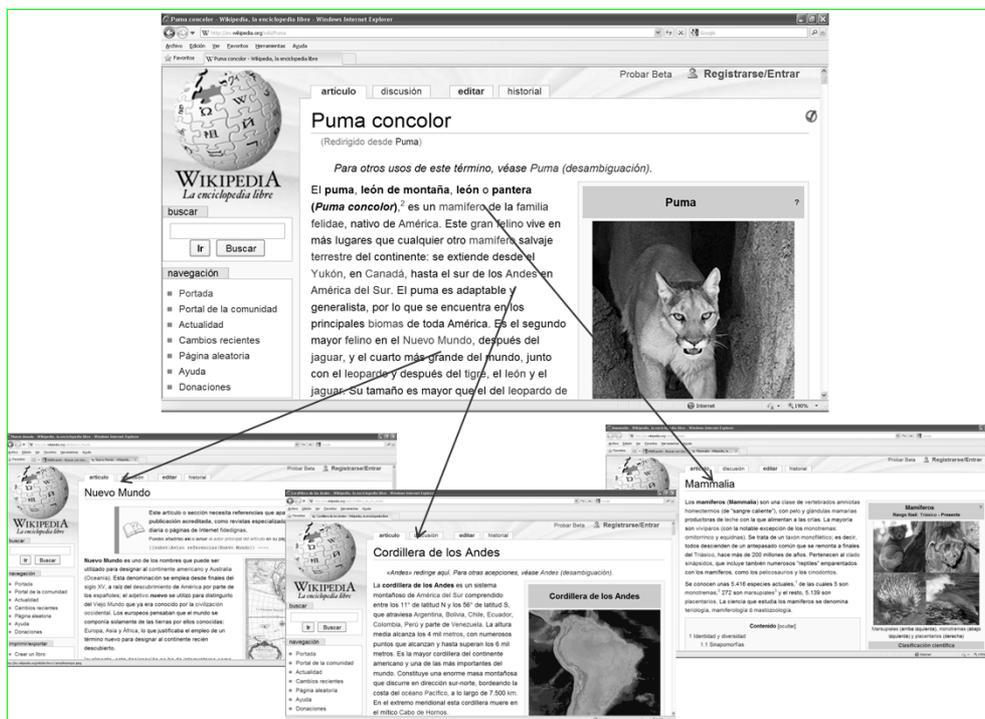


FIGURA 4. Proceso de wikificar un artículo de Wikipedia [17]

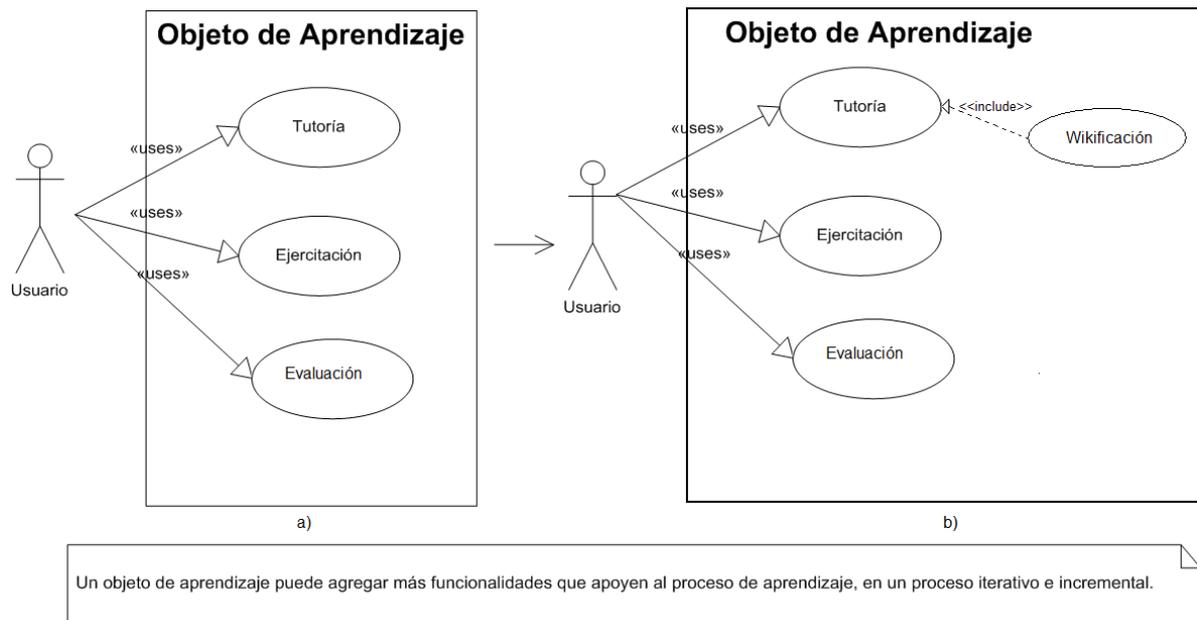


FIGURA 5. Diagrama de Casos de Uso de un OA: a) Actual, b) Propuesta de una Nueva Funcionalidad para los OAs.

## METODOLOGÍA

Aquí se especificarán los pasos de la metodología propuesta de wikificación de un objeto de aprendizaje. La nueva metodología modifica la forma de evaluar el contexto, involucrando para ello la relación semántica de los n-gramas obtenidos de los metadatos.

### Descripción

1. *Definición del contexto* El contexto inicial se constituirá con el conjunto de palabras circundantes al n-grama ambiguo, que vinculen al menos a un artículo de Wikipedia cuya probabilidad de ser vinculado (*commonness* o *link probability*) sobrepase un umbral muy pequeño, con la finalidad de eliminar palabras sin sentido, ver ecuación (1). En el ejemplo de la FIGURA 6 se muestra un segmento de un Objeto de aprendizaje en donde la palabra relaciones es ambigua en Wikipedia, es decir, tiene más de una definición.

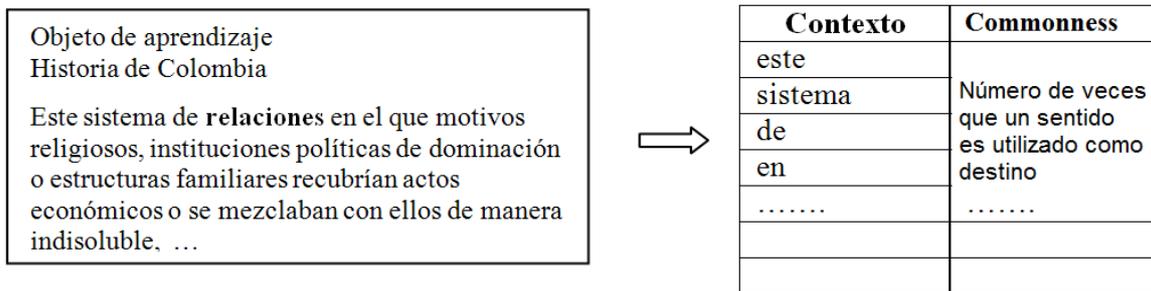
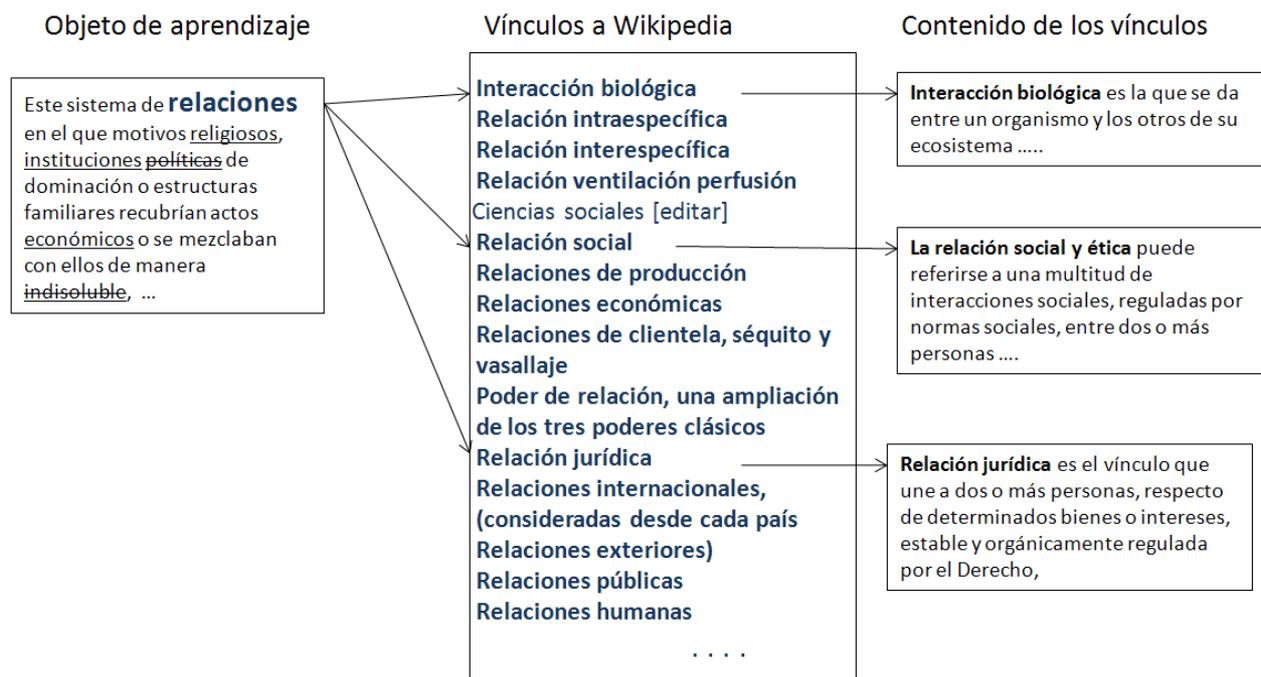


FIGURA 6. Generación del Contexto

$$link\ probability = \frac{\# \text{ de veces que la frase se utiliza como ancla}}{\# \text{ total de artículos que mencionan la frase}} \quad (1)$$

2. *Comparación de cada sentido posible con el contexto* Aquí el concepto 'sentido' se refiere a cada uno de los artículos de wikipedia posibles para un mismo término. La figura muestra en el cuadro izquierdo el contenido del



**FIGURA 7.** Identificación de los *Sentidos Candidatos* de un Término Ambiguo

OA y a la derecha algunos de los vínculos sugeridos por Wikipedia para la palabra ‘relaciones’, éstos constituyen los sentidos candidatos, ver FIGURA 7.

3. Calcular la relación semántica. Para cada *sentido candidato* se evalúa su *relación semántica* con relación a todos los términos del contexto, el promedio de ellos constituirá su peso promedio (*weighted average*) de cada *sentido candidato*, cómo se aprecia en la FIGURA 8, véase la siguiente fórmula [16]:

$$relatedness(a,b) = \frac{\log(\max(|A|, |B|)) - \log(|A \cap B|)}{\log(|W|) - \log(\min(|A|, |B|))} \quad (2)$$

En la fórmula 1, las variables a y b son dos artículos de interés; A y B son conjuntos de todos los artículos que vinculan a y b, respectivamente; W, es el conjunto de todos los artículos de Wikipedia.

Como puede apreciarse en la FIGURA 8, se incluyen frases significativas (metadatos), que deberían estar fuertemente vinculadas con los artículos a los cuales se desea vincular un término. Esto constituye una diferencia con el algoritmo de Milne y Witten porque además se asignará un peso mayor a la relación semántica obtenida con los términos de los metadatos que a la relación con los términos del contexto.

5. Evaluación del contexto. Se obtiene al evaluar la suma de los valores asignados a cada término del contexto (Medida en la que se relaciona con el contexto, *relatedness*; y frecuencia de uso del término como vínculo de Wikipedia, *link probability*).

- Link probability de cada una de las frases del contexto.
- Relatedness de cada una de las frases del contexto con el resto del contexto, para lo cual se utiliza la ecuación (1) nuevamente.
- Milne & Witten evalúan el valor de *weighted average*, ecuación (3), realizando el promedio aritmético de *link probability* y *relatedness*, como se muestra en la TABLA 2.

$$\text{weighted average} = \text{prom}(\text{link probability}, \text{relatedness}) \quad (3)$$

Se generan instancias de entrenamiento para un clasificador, en el cual, se tienen ejemplos positivos, artículo seleccionado, y ejemplos negativos, que son los artículos no seleccionados para este término. Para cada ejemplo que se introduzca, el algoritmo entregará una respuesta acerca de que la probabilidad de un sentido sea válido o no. Si más de un sentido es válido para un término, se elegirá el que tenga la mayor probabilidad.

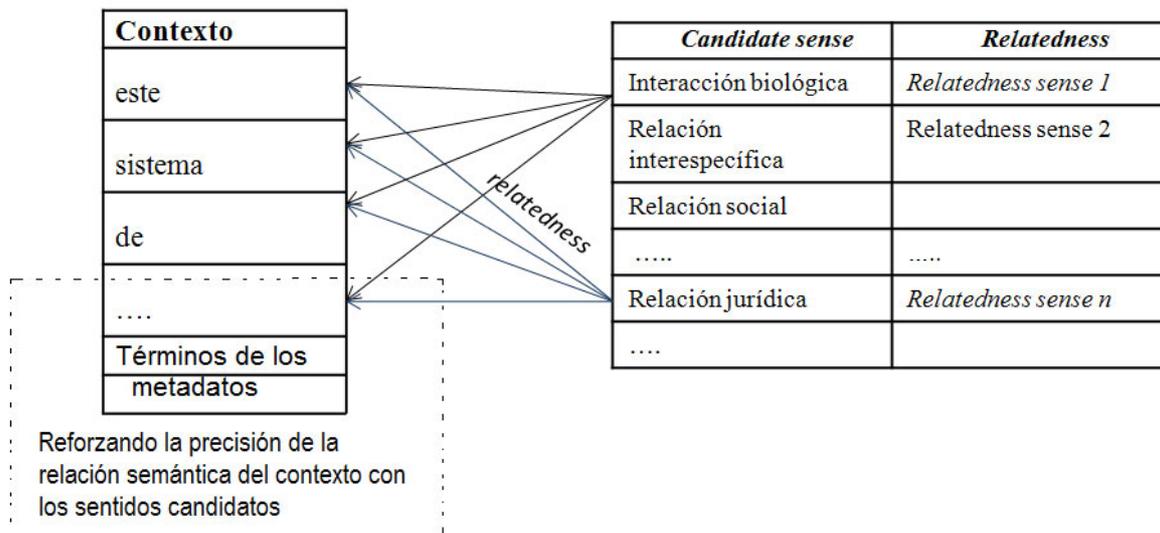


FIGURA 8. Cálculo de la Relación Semántica del Contexto y los Sentidos Candidatos.

TABLA 2. Evaluación del Contexto

Contexto	Link probability	Average semantic relatedness	Weight average de cada término del context
n-grama 1	Valor numérico		$prom(link\ probability, relatedness)$
n-grama 2		Valor numérico	$prom(link\ probability, relatedness)$
....			.....
n-grama $m$			$prom(link\ probability, relatedness)$

### Evaluación de la metodología

Para evaluar la eficiencia de la metodología propuesta, se realizará la wikificación manual de un conjunto de objetos de aprendizaje de diferentes repositorios. Estos objetos de aprendizaje serán sometidos a la wikificación automática con el programa creado a partir del algoritmo descrito.

Se han seleccionado, para formar parte del contexto en el contenido de un OA, los siguientes descriptores de la TABLA 1: título, descripción, palabras clave. Los cuales tienen una relación estrecha con el contenido del objeto de aprendizaje. Bajo el supuesto de que las frases del contexto que tienen una mayor relación semántica con los descriptores, mejorarán de manera importante la calidad del contexto.

Se validará la calidad del contexto de la siguiente manera: se agregarán tres factores más para calcular el peso promedio, los cuales representan el promedio de la relación semántica del n-grama del contexto con los elementos de cada uno de los descriptores (título, descripción, palabras clave). Cada uno de los elementos que determinan la calidad del contexto se denominan factores.

Para evaluar la calidad del contexto, se le ha asignado un valor de ponderación  $\alpha_i$ . El cual podrá modificarse de acuerdo a los resultados por diferentes pruebas. De tal forma que el *weight average* o peso promedio será calculado como se muestra en la fórmula (5).

$$weighted\ average = \frac{factor_1\alpha_1 + factor_2\alpha_2 + factor_3\alpha_3 + factor_4\alpha_4 + factor_5\alpha_5 + factor_6\alpha_6}{\#elem} \quad (5)$$

Lo que se pretende es que:

- a) El contexto se determina también por la relación que guarda cada una de sus frases con los descriptores del OA.
- b) Los determinantes  $\alpha_i$  son números que le asignan menor o mayor peso a cada uno de los factores, deberá ser un valor previamente probado con alguna muestra lo suficientemente grande para que nos pueda dar algún indicador acerca de los valores óptimos.

Con esto se espera que la precisión del proceso de desambiguación se vea beneficiado. Finalmente, todos estos elementos serán introducidos en un clasificador cuya salida sea el sentido candidato más apropiado para el contexto presentado. Se realizarán diferentes pruebas asignando valores distintos a  $\alpha_i$ , hasta encontrar algún valor que proporcione los mejores resultados.

Se pretende mejorar la precisión y *recall* al incluir elementos más determinantes como lo son los descriptores del Objeto de Aprendizaje. En este momento se utiliza Wikipedia como recurso para enriquecer el contenido del Objeto de Aprendizaje, sin embargo, si el número y diversidad de Objetos de Aprendizaje es lo suficientemente grande, estos mismos recursos podrían ser la fuente de información para el enriquecimiento.

## AGRADECIMIENTOS

El trabajo fue parcialmente financiado por el Gobierno de México (SNI, SIP-IPN, COFAA-IPN, CONACYT 50206-H, and SIP-IPN 20100773). El trabajo fue hecho cuando la primera autora estaba en una estancia de investigación en el CIC-IPN gracias a la cooperación académica entre el CIC-IPN y UAEM.

## REFERENCIAS

1. AGREGA. *Repositorio de Objetos de Aprendizaje*. <http://www.proyectoagrega.es/default/Inicio>
2. Allier, J. *NETg's precision skilling: The linking of occupational skills descriptors to training interventions*, recuperado el 01 junio de 2010 de <http://www.netg.com/research/pskillpaper.htm>. 1998.
3. ARIADNE. *Foundation for European Knowledge Pool*. <http://www.ariadne-eu.org/#>
4. Chan, M. *Objetos de aprendizaje: una herramienta para la innovación educativa*. 2002. Documento mecanografiado, México: Universidad de Guadalajara.
5. DCMI. *Dublin Core Metadata Initiative*. Recuperado el 01 de abril de 2010 de <http://dublincore.org/>. 2010.
6. FLOR. *Repositorio Latinoamericano de Objetos de aprendizaje*, <http://ariadne.cti.espol.edu.ec/lacloFinder/>
7. GLOBE. *Global Learning Objects Brokered Exchange*, <http://globe.edna.edu.au/globe/go>
8. IEEE Learning Technology Standards Committee (LTSC), <http://www.ieeeltsc.org:8080/Plone>.
9. IEEE Learning Technology Standards Comitee. *Learning Object Metadata*. IEEE 1484.12.1. 2002.
10. López, C. *Los Repositorios de Objetos de Aprendizaje como soporte a un entorno e-learning*, Tesina doctoral, Universidad de Salamanca. (Director Francisco José García Peñalvo). 2005.
11. Medelyan, O., Witten, I.H. and Milne, D. Topic Indexing with Wikipedia. In Proceedings of the AAAI 2008 Workshop on Wikipedia and Artificial Intelligence (WIKIA,2008), Chicago, IL. 2008.
12. MERLOT. <http://www.merlot.org/merlot/index.htm>
13. Mihalcea R. *Using Wikipedia for Automatic Word Sense Disambiguation*. Proceedings of NAACL HLT, 2007, pp. 196–203.
14. Mihalcea, R. and Csomai, A. *Wikify!: linking documents to encyclopedic knowledge*. In Proceedings of the 16th ACM Conference on Information and Knowledge management (CIKM'07), Lisbon, Portugal, 2007, pp. 233-242.
15. Milne, D. and Witten, I.H. *Learning to link with Wikipedia*. In Proc. of the ACM Conference on Information and Knowledge Management (CIKM'08), Napa Valley, California, 2008.
16. Milne, D., and Witten, I.H. *An effective, low-cost measure of semantic relatedness obtained from Wikipedia links*. In Proceedings of the AAAI 2008 Workshop on Wikipedia and Artificial Intelligence (WIKIAI 2008), Chicago, IL.
17. Wikipedia, *la enciclopedia libre*. <http://es.wikipedia.org/wiki/Wikipedia:Portada>.
18. Wiley, D. A. *Connecting learning objects to instructional design theory: a definition, a metaphor, and a taxonomy*. En D.A. Wiley (Ed.) *The Instructional Use of Learning Objects*. Recuperado el 01 de abril de 2010, de <http://reusability.org/read/chapters/wiley.doc>.