

Uso de los Lenguajes de Marca para el Registro del Conocimiento Contenido en un Plan de Estudios

Yuri J. Landa
Universidad de Lima
Surco, Lima 33, Perú

RESUMEN

El presente estudio muestra una estructura de clases y propiedades adecuadas para el registro del conocimiento contenido en un plan de estudios, desarrollada utilizando el Web Ontology Language (OWL). Para tal fin, se analizaron las características de los datos que contiene un plan de estudios, más desde una perspectiva académica que administrativa, para precisar cómo pueden ser almacenados en una estructura que permita organizar el conocimiento.

Esta estructura de clases y propiedades ha sido concebida para construir una base de conocimientos, que luego servirá para atender a aplicaciones que cubrirán los requerimientos de alumnos, profesores y administradores de un plan de estudios. A modo de prueba se registró una parte del plan de estudios de la Facultad de Economía de la Universidad de Lima.

Palabras claves: Lenguajes de marca, Ontologías, Plan de estudios, Bases de conocimientos, Relaciones semánticas.

1. INTRODUCCIÓN

En la actualidad, la universidad tiene a su alcance una variedad de sistemas de cómputo que le facilitan la gestión de los datos que genera su actividad administrativa. Sin embargo, no dispone de software que le permita administrar su actividad académica, no en su aspecto operativo, sino en lo que se refiere a la gestión del conocimiento.

Normalmente, cuando una facultad crea un plan de estudios el administrador desarrolla su contenido hasta cierto nivel de detalle, llegando a lo más hasta la descripción general de cada asignatura conocida como sumilla. Sin embargo, preguntas tales como, ¿cómo se despliegan los contenidos de las asignaturas?, ¿cuál es el nivel de profundidad que abarcan?, ¿cuánto tiempo se le asigna a los temas y conceptos?, son aspectos que quedan en manos de los docentes, de lo cual los administradores del plan muchas veces no reciben mayor información; más aún si la recibieran, con las herramientas tecnológicas actuales no les sería fácil procesarlas.

En ese contexto, nuestro trabajo se centró en la creación de una estructura de datos que permitiera registrar en una ontología el conocimiento que contiene un plan de estudios, desde la descripción general del plan hasta los contenidos desarrollados dentro de las asignaturas. Tal estructura debe permitir formar una base de conocimientos que podrá ser utilizada en el futuro para

darle al administrador del plan una mayor capacidad de gestión del conocimiento. Asimismo, los profesores podrán tener a disposición una herramienta para la consulta de conceptos, mientras que los alumnos un nuevo material educativo.

Para tal fin mostramos en las secciones siguientes las ventajas que traen consigo el uso de los lenguajes de marca desarrollados por la World Wide Web Consortium (W3C) para la creación de ontologías. Luego evaluamos las características del conocimiento que contiene un plan de estudios desde una perspectiva informática. Seguidamente describiremos las clases y propiedades que se adecuan a nuestros requerimientos, y finalmente mostraremos algunas conclusiones.

2. LOS LENGUAJES DE MARCA

La W3C viene desarrollando un conjunto de lenguajes con la finalidad de incrementar al máximo la interoperabilidad de la Web, es decir, permitir el intercambio de datos entre organizaciones de diferentes campos de la actividad económica y social, con el menor grado de intervención humana posible.

La base de estos lenguajes es el uso de marcas o etiquetas con las cuales se crean las estructuras de datos que luego se procesarán. Estos lenguajes almacenan sus datos y programas en formato de texto, lo que trae consigo ventajas tales como, una mayor portabilidad de los programas para ser implementados en cualquier arquitectura de hardware o software; el uso y reuso amplio de los programas ya que son de código abierto; facilidad de mantenimiento debido a que los archivos son más entendibles por humanos, y además porque los datos y programas pueden ser creados y modificados por cualquier editor de textos.

Una de las áreas de trabajo de la W3C es el desarrollo de la Web semántica, que busca transformar la Web actual haciendo posible que las máquinas puedan intercambiar información de manera casi automática, con la menor intervención del hombre. Para ello, la intención es trasladarle a una computadora la capacidad de “comprender” la información que tienen registrada, y de incorporar en ellas ciertas reglas de “raciocinio”, para con ello permitirles tomar decisiones [1].

Al margen de las posibilidades de éxito de los planes de la W3C en lo que se refiere al intercambio automático de datos, los lenguajes de marca que nos ofrecen permiten el registro de la información de una organización con un sentido más semántico que informático, lo que hace posible su uso en áreas como el registro del conocimiento. Los lenguajes orientados al registro

semántico de la información son el Lenguaje Extendido de Marcas (XML), el Marco de Descripción de Recursos (RDF) y el Lenguaje de Ontologías Web (OWL) [2].

El Lenguaje Extendido de Marcas (XML) proporciona la sintaxis básica para el resto de lenguajes. Su potencial se apoya en el hecho de que, a través de la creación de etiquetas, cualquier organización puede crear las estructuras de datos necesarias para registrar su información, para luego procesarla o distribuirla, en redes locales o extensas, utilizando el protocolo www. Esta característica convierte al XML en un lenguaje universal de intercambio de información.

El lenguaje RDF ofrece un modelo de datos en forma de tripleta compuesto de un sujeto, un predicado y un objeto, con el cual se pueden crear estructuras de datos definidas en base a las relaciones que se pueden formar entre sus elementos. En este modelo, el sujeto es un dato (recurso) sobre el cual se desea hacer la declaración y estará representado por una clase; el predicado es una propiedad o característica relativa al sujeto, un atributo en la tecnología de objetos; y el objeto es un valor concreto de la propiedad. El objeto puede ser otra clase o un valor literal. En este modelo la propiedad o relación que se forma entre dos datos constituye el aspecto central de la estructura que se quiere definir.

Estas relaciones entre datos se pueden refinar con la ayuda de un conjunto de clases y atributos que son parte del lenguaje. Entonces, en la medida que se está definiendo un dato a través de las relaciones que establece con otros datos, se está almacenando su significado o su semántica. Es así como se puede utilizar el modelo que ofrece el RDF para digitalizar la información contenida en cualquier campo específico de la actividad humana, como las comunicaciones, el comercio, la industria, la salud, etc., más allá de los procesos habituales que a la fecha ya están digitalizados en estas mismas áreas, como planillas, ventas, inventarios, y similares.

El Lenguaje de Ontologías Web (OWL) utiliza la sintaxis del RDF pero amplía las características y restricciones de las propiedades que se usarán para describir las relaciones entre clases. Este lenguaje ha sido desarrollado para crear ontologías, es decir, para poder describir todos los conceptos que abarcan una esfera específica del conocimiento, incluyendo las relaciones que se forman entre tales conceptos. En este contexto entonces, las clases definidas contienen en su mayoría conceptos que forman parte de algún área de la ciencia o técnica, mientras que las propiedades son las diferentes relaciones que se pueden formar entre ellas.

En OWL las clases se definen con la ayuda de subclases, restricciones y propiedades. Una propiedad expresa una relación binaria que puede ser de dos tipos: una datatype property, que permite expresar relaciones entre instancias de una clase y literales RDF o tipos de datos XML; o una object property, que permiten relacionar instancias de dos clases.

Para precisar el uso de una propiedad se debe determinar

su dominio y rango, es decir, qué clases o partes de una clase pueden utilizar la propiedad, y, de qué clase se obtiene el conjunto de valores que resulta de la relación. Asimismo, una propiedad puede ser declarada como inversa de otra, en la medida que se quiera que el lenguaje deduzca que el dominio de una propiedad es el rango de la otra y viceversa. Del mismo modo, una propiedad también puede ser declarada transitiva cuando se requiera que el lenguaje vincule clases a través de la propiedad, sin que esta relación se haya ingresado explícitamente.

El presente trabajo ha hecho uso de estas características del lenguaje OWL para registrar en una ontología el conocimiento contenido plan de estudios.

3. EL PLAN DE ESTUDIOS

Un plan de estudios no es otra cosa que un conjunto de términos relacionados entre sí que representan un ámbito del conocimiento. Sin embargo, desde una perspectiva informática podemos distinguir tres elementos que conforman el conocimiento: en primer lugar los signos, que son las palabras, fórmulas o gráficos que describen el conocimiento; en segundo lugar las relaciones que se forman entre signos, que determinan su posición dentro de la estructura del conocimiento, y en tercer lugar el significado de estos signos, es decir, la imagen que se forma en la mente de un individuo asociada a un concepto.

Mientras que la digitalización de los signos, gracias a los sistemas de codificación como ASCII o UNICODE ya es un problema resuelto, los lenguajes de marca ofrecen una herramienta que permite digitalizar las relaciones que se forman entre los signos que representan conceptos, lo que a su vez permitirá mejorar el registro de su significado. Nuestro trabajo se concentra entonces en la definición de las relaciones que se forman entre los signos que componen un plan de estudios.

En el ámbito de un plan de estudios la relación entre dos términos más fácil de identificar es la de todo-parte, u holonimia-meronimia, es decir, cuando un término está compuesto de otros, o se divide en otros, dando origen a una jerarquía de términos. En este sentido, un plan de estudios está dividido en áreas académicas, cada área en subáreas académicas, cada subárea en asignaturas, cada asignatura en temas, cada tema en conceptos y cada concepto en subconceptos, y así sucesivamente hasta lograr el grado de detalle que uno pretenda registrar. Entonces, podemos ver al plan de estudios como una gran estructura jerárquica de términos que representan el conocimiento, organizado en niveles, desde lo más general hasta el concepto más específico, y vinculados entre sí, a través de relaciones de holonimia.

Otro tipo de relación que se forma entre los términos de un plan de estudios es la relación de requisito, es decir, cuando la asimilación de una idea sólo será posible si se ha asimilado una idea previa. Esta relación es declarada explícitamente en un plan de estudios a nivel de asignaturas, sin embargo, esta misma relación se puede plantear a nivel de conceptos.

Según las teorías del aprendizaje los conceptos que se asimilan en el proceso del aprendizaje están formados por proposiciones. Una proposición es la unión de dos o más palabras para describir un concepto. Entonces, el aprendizaje de conceptos se realiza a través del aprendizaje de proposiciones [3]. En la medida que asimilamos más proposiciones de un tema, más sabemos sobre él. Y, en la medida que relacionamos proposiciones nuevas con las que tenemos registradas en nuestra memoria, mayor significado tiene el concepto para el estudiante, por lo tanto más fácil es su aprendizaje.

En nuestro caso, podemos digitalizar las proposiciones almacenando las relaciones entre conceptos. Para lograrlo, debemos reducir las relaciones entre conceptos a oraciones formadas por pocas palabras y que tengan la forma sujeto, predicado y objeto. En esta relación, tanto el sujeto como el objeto serían conceptos, mientras que el predicado representaría un tipo de relación entre ambos.

De esta manera podemos, por ejemplo, definir la relación “se divide en”, y formar con ella una estructura jerárquica que atraviese todo el plan de estudios, tal como se muestra en la figura 1.

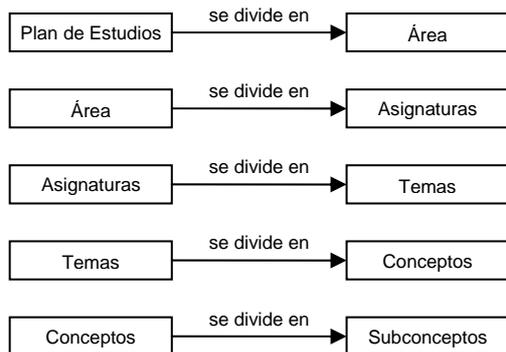


Figura 1

En la medida que podemos registrar todas las relaciones de holonimia que contiene un plan, habremos creado una taxonomía de términos que formarán el cuerpo principal de nuestra ontología. A esta relación se le puede añadir la relación de requisitos entre términos, luego relaciones de sinonimia, y así, ir paso a paso consolidando nuestra ontología.

4. ESTRUCTURA DE LAS CLASES

La estructura de clases que representará el conocimiento en una ontología ha sido definida como parte de un sistema mayor que integrará a la ontología como una base de conocimientos, la que será actualizada y consultada por diferentes aplicaciones. Estas aplicaciones deben responder a las necesidades de tres actores, a saber, el administrador del plan, el profesor y el alumno, tal como se muestra en la figura 2.

El administrador del plan debe disponer de una aplicación que le permita crear el plan de estudios, las áreas académicas, y las asignaturas hasta la sumilla de la

misma. Al mismo tiempo estará interesado en tener una aplicación que le permita administrar cambios curriculares, en donde debería poder eliminar y crear asignaturas, unir o dividir asignaturas, declarar asignaturas equivalentes, etc.

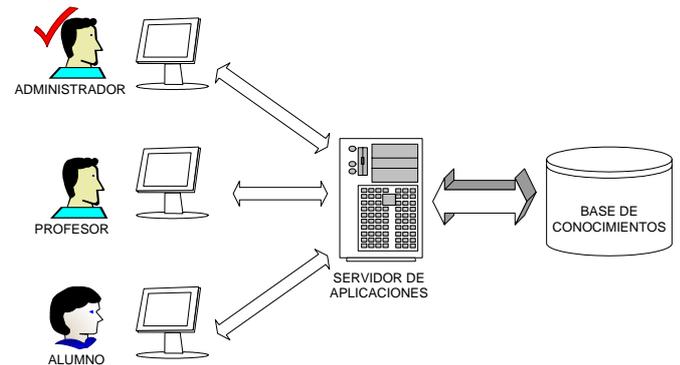


Figura 2

El profesor debe disponer de una aplicación con la que pueda añadir el contenido de las asignaturas, es decir, añadir los capítulos y subcapítulos del sílabo, los conceptos que se dictan en cada capítulo, la bibliografía y los ejercicios asociados a los conceptos. Asimismo, puede disponer de una aplicación que, a través de consultas, genere reportes en formato de texto o en formato gráfico con el contenido de las asignaturas y las relaciones que se forman entre los temas y conceptos de las mismas.

El alumno puede disponer de la misma aplicación anterior, pero adicionalmente puede implementarse motores de búsqueda de términos dentro de la ontología. Al mismo tiempo, se puede poner a disposición del alumno una aplicación de autoaprendizaje, en donde se usa el contenido de la ontología como material de enseñanza y se le ofrecen ejercicios, para que a través de sus aciertos o errores, la aplicación le indique qué conceptos ya aprendió satisfactoriamente o qué conceptos debe volver a revisar.

Para satisfacer tales requerimientos se crearon 8 clases, 18 Datatype Properties y 22 Object Properties. Detallaremos a continuación el papel de cada clase y la función de las propiedades más importantes.

La clase Ontología ha sido creada para corresponder a la forma cómo el lenguaje OWL organiza las etiquetas que representarán los datos a registrarse, a saber, todas deben estar dentro de una etiqueta raíz. Es así que se ha elegido a la clase Ontología como una superclase que contiene a todas las demás. Esta clase usa la propiedad “nombre largo” para registrar el nombre completo de cada instancia, y en la medida que el resto de clases son sus hijas, heredarán automáticamente la misma propiedad.

La clase Plan de Estudios ha sido creada para registrar los datos del plan, principalmente con fines administrativos. Contiene 6 Datatype Properties que registran su fecha de creación, la escuela universitaria a la que pertenece, el perfil u objetivos de formación que ofrece a los alumnos, la cantidad de créditos obligatorios y electivos que deben cumplir los alumnos y el número de niveles en los que están divididas las asignaturas. Además es dominio de la

Object Property “se divide en Área” que servirá para relacionar a esta clase con instancias de la clase Área.

La clase Área ha sido creada para registrar las características de un área académica. En este contexto, un área académica juega un rol tanto administrativo como académico-pedagógico. Según sus funciones administrativas, el área académica se encarga de organizar las asignaturas del plan, de coordinar las actividades de los profesores que dictan esas asignaturas y de plantear políticas generales de evaluación o de desarrollo de materiales. Por el lado académico, el área es también una sección del campo de dominio del plan, es decir, es una sección del conocimiento. Es así que se ha definido a la clase Área como dominio de la propiedad “se divide en Asignatura” para que se pueda formar el vínculo con las instancias de la clase Asignatura. Asimismo, es rango de la misma propiedad, lo que permite dividir un área académica en subáreas, según las particularidades del plan.

La clase Asignatura permitirá almacenar instancias que correspondan a las asignaturas que se dictan en un plan de estudios. Tiene asociadas propiedades que representan datos administrativos y propiedades que facilitan el registro del conocimiento. Así contiene 7 Datatype Properties que registran el carácter de la asignatura, su valor en créditos, las horas de clase semanales que usa, la sumilla o resumen de contenido, el nivel al que corresponde en el plan de estudios, y los materiales que utiliza. Es dominio de la Object Property “se divide en Tema”, con lo que registramos los temas contenidos en los sílabos de cada asignatura. También es dominio de la propiedad “es dictada por”, para hacer el vínculo con el o los profesores que la dictan. Finalmente es también dominio de la propiedad “es requisito de”, la que servirá para hacer el vínculo con las asignaturas de las cuales es

un requisito.

La clase tema ha sido creada para representar a través de sus instancias a las unidades de enseñanza en las que se divide el conocimiento contenido en una asignatura. Corresponden a los capítulos y subcapítulos que contiene el sílabo de una asignatura. Es dominio y rango de la Object Property “se divide en Concepto”, lo que servirá para dividir las instancias de la clase Tema en subtemas, o para relacionar cada tema con los conceptos en los que se desprende.

La clase Concepto ha sido creada para registrar los términos que componen el conocimiento en su nivel de abstracción más bajo. En la medida que pretende representar el detalle del conocimiento, es la clase que más instancias tiene. Debido a que en una asignatura existen conceptos que se desarrollan utilizando otros conceptos, dentro de esta clase existen instancias de diferente grado de abstracción. Para eso se ha definido a la clase Concepto como dominio y rango de la Object Property “se divide en Subconcepto”, lo que nos permite almacenar conceptos con sus subconceptos, con los subconceptos de los últimos y así sucesivamente. Asimismo, es dominio de la propiedad “vinculado a Ejercicio”, para asociar, si es necesario, cada concepto a un grupo de ejercicios, que será utilizado por una aplicación de autoaprendizaje a desarrollarse en el futuro.

Por último tenemos a las clases Profesor y Ejercicios cuyas funciones respectivas se desprenden de las explicaciones anteriores.

La figura 3 muestra la estructura de las clases creadas con sus atributos o Datatype Properties utilizando el Lenguaje Unificado de Modelado UML.

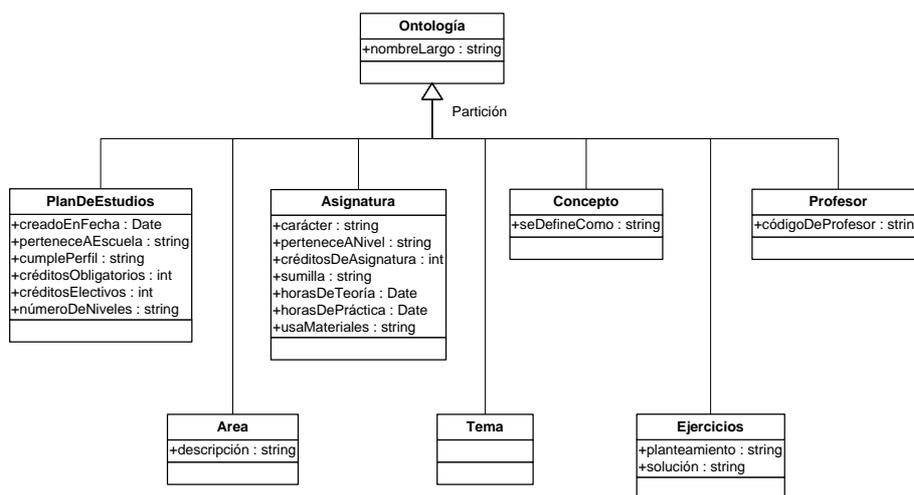


Figura 3

5. DESARROLLO DE LA ONTOLOGÍA

Luego de haber creado la estructura de clases y las propiedades, se crearon las instancias. Para ello se utilizó el editor de ontologías Protégé de la Universidad de Stanford.

A modo prueba se utilizó el plan de estudios de la

Facultad de Economía de la Universidad de Lima, vigente desde marzo de 2006 [4]. Se implementaron las características del plan, sus áreas académicas, y la descripción de la mayoría de las asignaturas, hasta el nivel de la sumilla. Para el registro del conocimiento contenido en las asignaturas se eligió, a modo de prueba, la asignatura Economía General del segundo nivel de estudios. Para cada tema de dicha asignatura se registró

en qué subtemas se divide, o qué conceptos usa para desarrollar el tema, o a qué casos se aplica el tema en cuestión. No todos los temas tienen subdivisiones, o casos de aplicación, pues no lo necesitan, pero nuestra estructura de datos está preparada para implementar estas características en los casos necesarios.

Seguidamente, se implementaron las instancias de la clase Concepto para la asignatura elegida. Se crearon conceptos derivados de los temas o capítulos del sílabo, y para cada concepto se registró su definición, si es requisito de otro concepto o tema, si tiene subdivisiones, qué conceptos usa para su desarrollo, qué ejercicios tiene asociados, entre otras propiedades. La estructura de datos definida permite crear conceptos y subconceptos sin límite, tanto como lo requiera el tema o lo decida el administrador de la ontología.

6. CONCLUSIONES

El lenguaje OWL permite crear clases y asignarle a las relaciones entre clases propiedades transitivas e inversas, con las cuales se pueden conectar todos los términos que componen un plan de estudios, formando así una red de conceptos. Esta red, utilizando el editor de ontologías Protégé, se puede convertir en una red de tablas navegables, y hacer más visibles las ventajas de esta nueva forma de registrar el conocimiento, en la medida que se puede consultar el plan desde sus aspectos más generales hasta el contenido detallado de las asignaturas. Más aún, para el estudiante, poder ver un concepto asociado con el que le antecede o con el que le precede le permite enfocar el conocimiento de manera similar a cómo el cerebro humano realiza el aprendizaje de conceptos, según los psicólogos del aprendizaje, convirtiendo a nuestra ontología en un nuevo y poderoso material educativo.

La estructura de clases y las relaciones creadas nos han permitido registrar el plan como una taxonomía de términos. Sin embargo el lenguaje OWL ofrece la posibilidad de incorporar a las relaciones definidas proposiciones lógicas, las que cuando son utilizadas por un motor de razonamiento permite evaluar inconsistencias en las instancias que se introducen a la ontología. Esta propiedad se puede utilizar para crear módulos de autoaprendizaje que evalúe el avance de un estudiante a través de su interacción con la ontología [5].

En el trabajo presentado se completó la ontología para una sola asignatura a modo de prueba, dentro del plan de estudios de una facultad. El paso que sigue es continuar registrando los contenidos de las demás asignaturas, para lo cual se necesita de la participación de los docentes relacionados con cada una de ellas. Esto será un proceso lento en la medida que el esfuerzo recaiga mayormente en los expertos de cada tema. Sin embargo, existen estudios en donde se evalúan las posibilidades de crear ontologías en forma automática o semiautomática, en base a contenidos de textos [6], [7]. En la medida que se desarrollen y difundan tales herramientas disminuirían las

horas hombre requeridas para el llenado de la ontología.

Uno de los problemas esenciales asociado con la construcción de ontologías se refiere a las posibilidades de trasladarle a una máquina la capacidad de representar relaciones semánticas. El presente trabajo se concentró en la creación de propiedades que representaran las relaciones de holonimia (todo-parte) y de requisito, sin embargo, es obvio que en el campo del conocimiento las relaciones que se pueden dar entre conceptos son múltiples. Bases de datos semánticas existentes como Wordnet y Eurowornet construyen sus ontologías respectivas en base a más de 50 relaciones semánticas. Siendo estas bases de datos de propósito general queda por evaluar cuáles de estas relaciones son adecuadas para el registro del conocimiento contenido en un plan de estudios.

7. REFERENCIAS

- [1] Berners-Lee, Tim; Hendler, James; Lasilla, Ora. La Red Semántica. En: Investigación y Ciencia. Barcelona, Julio 2001. Nro. 298.
- [2] Todas las especificaciones de los lenguajes mencionados se encuentran en: W3C. World Wide Web Consortium. <<http://www.w3.org/>>. Visitado en abril de 2006.
- [3] Novak, Joseph D. Conocimiento y Aprendizaje. Alianza Editorial. Madrid, 1998. Pag. 61.
- [4] La terminología curricular usada en esta parte del trabajo corresponde a las definiciones contenidas en el Reglamento General de Estudios de la Universidad de Lima. Sitio Web de la Universidad de Lima. Normas y Reglamentos. <<http://www.ulima.edu.pe/webulima.nsf/default/reg?OpenDocument&dn=12>>, visitado en abril 2006.
- [5] Díez, Fernando; Gil, Rafael. Nuevas Perspectivas en la Resolución de Matemáticas mediante el uso de Modelos Semánticos de Representación del Conocimiento. En: Revista Iberoamericana de Informática Educativa. Número 2, Julio-Diciembre 2005, pp. 1-12.
- [6] Sánchez-Cuadrado, Sonia; Lloréns, Juan; Morato, Jorge; Hurtado, José A. Extracción Automática de Relaciones Semánticas. Revista Iberoamericana de Sistemas, Cibernética e Informática Volumen 1 - Número 1. < <http://www.iiisci.org/journal/risci/pdfs/P338146.pdf> >, visitado en enero de 2007.
- [7] Serrano-Guerrero, Jesús; Olivás, José A.; De la Mata, Javier. Relaciones Borrosas Conceptuales para la Construcción de Ontologías con GUMSe. Revista Iberoamericana de Sistemas, Cibernética e Informática Volumen 2 - Número 2. < <http://www.iiisci.org/journal/risci/pdfs/P985449.pdf>>, visitado en enero de 2007.