

# Ontological learning for a dynamic semantics ontological framework

Tania Rodríguez <sup>a</sup> & José Aguilar <sup>b</sup>

<sup>a</sup> Centro de Estudios en Microelectrónica y Sistemas Distribuidos (CEMISID), Universidad de Los Andes Mérida, Venezuela [taniana@ula.ve](mailto:taniana@ula.ve),

<sup>b</sup> Centro de Estudios en Microelectrónica y Sistemas Distribuidos (CEMISID), Universidad de Los Andes Mérida, Venezuela [aguilar@ula.ve](mailto:aguilar@ula.ve)

Received: October 1th, 2013. Received in revised form: June 16th, 2014. Accepted: July 7th, 2014

## Abstract:

In this paper we propose an Ontological Learning architecture, which is one of the key components of the Dynamic Semantic Ontological Framework (MODS) for the Semantic Web. This architecture supports the automatic acquisition of lexical and semantic knowledge. In particular, it allows the acquisition of knowledge of terms (words), concepts (taxonomic, not taxonomic) relations, production rules, or axiom. The architecture establishes where each acquired knowledge must be incorporated into the structures of MODS: its interpretive ontology, language ontology, or lexicon. Furthermore, the paper presents an example of use of architecture to the case of semantic learning

*Keywords:* Natural Language Processing, Ontological Learning, Semantic Web.

## Aprendizaje ontológico para el marco ontológico dinámico semántico

### Resumen:

En este trabajo se propone una arquitectura de Aprendizaje Ontológico, que es uno de los componentes claves del Marco Ontológico Dinámico Semántico (MODS) para la Web Semántica. Esta arquitectura general soporta la adquisición automática de conocimiento léxico y semántico. En particular, permite la adquisición de conocimiento sobre términos (palabras), conceptos (taxonómicos, no taxonómicos), relaciones entre ellos, reglas de producción o axiomas. La arquitectura establece donde cada conocimiento adquirido debe ser incorporado en las estructuras del MODS, ya sea en su ontología interpretativa, ontología lingüística, o lexicon. Además, el trabajo presenta un ejemplo de uso de la arquitectura para el caso del aprendizaje semántico.

*Palabras clave:* Procesamiento del Lenguaje Natural, Aprendizaje Ontológico, Web Semántica.

### 1. Introducción

En este trabajo se describe la arquitectura del componente de aprendizaje ontológico utilizado en mods. El mods es un sistema que permite el procesamiento ontológico de consultas, expresadas en lenguaje natural, para la web [1]. Mods está compuesto por una ontología lingüística del lenguaje español, una ontología para las tareas de procesamiento de la consulta, y una ontología interpretativa del perfil del usuario, además de un lexicon. El marco ontológico es dinámico, en el sentido que se actualiza a través de mecanismos de aprendizaje automático, para adaptarse a los cambios de la web y al perfil de los usuarios. Algunos de los objetivos que se persiguen con mods son: usar el lenguaje natural español para realizar

consultas sobre la web, explotar el contenido semántico sobre la web en procesos de razonamiento automático con el fin de optimizar los procesos de búsqueda, aprender sobre el uso y el contenido de la web, entre otras cosas.

Para lograr el proceso de adaptación del mods, se requiere de un componente de aprendizaje que permita el proceso de adquisición de conocimiento, el cual actualizará el contenido del resto de sus componentes. Sin este componente de aprendizaje, el mods se reduce a un simple sistema de interpretación de consultas, usando un conocimiento inicialmente definido en sus componentes. Así, la importancia de este componente tiene que ver con la capacidad que se le confiere al mods para adaptarse a su entorno (web y usuarios). En este trabajo se presenta la

arquitectura genérica del componente de aprendizaje. Además, se presenta los dos macro algoritmos, para el caso del aprendizaje morfosintáctico y del aprendizaje semántico.

A continuación se describen algunos trabajos previos, vinculados al problema de aprendizaje: *doddle-owl* es un sistema que soporta el aprendizaje de relaciones taxonómicas y no taxonómicas, usando métodos estadísticos (análisis de co-ocurrencia), *word net* [3] y textos específicos a un dominio dado [6]. *Text2onto* es el oficial sucesor de *text toonto*, el cual es una herramienta de aprendizaje de ontologías desde texto [10], capaz de identificar términos, sinónimos, conceptos, relaciones taxonómicas y axiomas. Además, tiene incorporado un modelo probabilístico ontológico, que permite registrar la evolución de una ontología (cambios en su corpus) [7]. *Cameleon* es un sistema que encuentra relaciones léxicas taxonómicas y no taxonómicas en textos planos, por medio de patrones léxico-sintácticos [8]. *Hasti* es un sistema para la construcción automática de ontologías, usa como entrada información no estructurada en la forma de texto en lenguaje natural persa. *Hasti* no usa conocimiento previo, es decir, construye las ontologías desde cero. Usa un lexicón que se encuentra inicialmente vacío, y va creciendo incrementalmente en la medida que va aprendiendo nuevas palabras. *Hasti* aprende conceptos, relaciones conceptuales taxonómicas, no taxonómicas, y axiomas. *Hasti* usa un enfoque simbólico híbrido (una combinación de lógica y métodos heurísticos, entre otros) [9]. *Ontolearn*, es un sistema que permite enriquecer una ontología de dominio con conceptos y relaciones. Usa aprendizaje de máquina, el cual ofrece algunas técnicas para el descubrimiento de conocimiento (patrones) basadas en cadenas de markov [11]

Este trabajo se diferencia de los trabajos previos, en primer lugar, porque forma parte de un componente de aprendizaje donde otros aspectos se aprenden (léxicos, semánticos, coloquiales, entidades nombradas); en segundo lugar, porque el resultado de la adquisición de conocimiento alimenta a específicos componentes del mods (en este caso, al lexicón y a la ontología interpretativa); y tercero, porque su fuente de aprendizaje es la consulta en lenguaje natural o el resultado arrojado por la web para dicha consulta. Este trabajo está organizado de la siguiente manera, en la sección 2 se presenta el marco ontológico dinámico semántico, en la sección 3 se presenta el aprendizaje ontológico, en la sección 4 se presentan algunos ejemplos de uso de esa arquitectura en mods, específicamente para el caso de aprendizaje semántico, y finalmente, en la sección 5, se presentan las conclusiones y trabajos futuros.

## 2. Marco Ontológico Dinámico Semántico (MODS)

El MODS es una propuesta novedosa que permite el análisis y la realización de consultas en lenguaje natural en la Web Semántica. La consulta para el MODS, más que una petición de información, es un elemento cargado de información útil para formarse una idea del tipo de usuario, y aproximarse, de manera sucesiva, a una respuesta que satisfaga cada vez más las necesidades del mismo. MODS tiene el desafío de interpretar y formalizar la consulta realizada por el usuario en lenguaje natural, como refinar

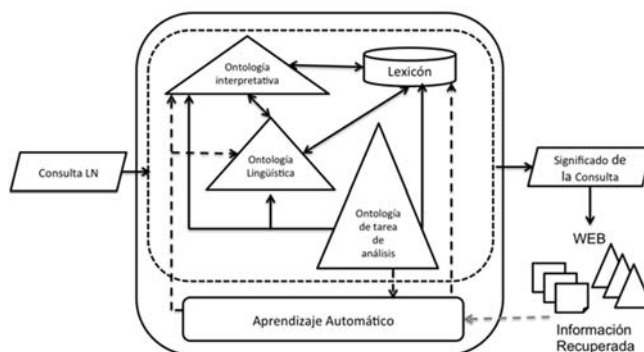


Figura 1. Marco Ontológico Semántico  
Fuente: Rodríguez, T. et al., 2010

sus esquemas internos frente a la dinámica de la Web para tratar futuras consultas (para lo cual requiere de mecanismos de adaptación). Específicamente, en [1] se propone la arquitectura de MODS. A continuación presentamos dicha arquitectura (ver Fig. 1).

De manera general, MODS transforma la consulta en un lenguaje ontológico, utilizando sus diferentes componentes: el lexicón, la ontología lingüística, la ontología de tareas y la ontología de dominio. De esta manera, MODS utiliza mecanismos de la semántica ontológica y herramientas del procesamiento del lenguaje natural para el procesamiento de las consultas de los usuarios. Una descripción más detallada de la arquitectura (Fig. 1) es la siguiente: la ontología de tareas modela las tareas de procesamiento de la consulta en lenguaje natural (análisis léxico-morfológico, análisis sintáctico-semántico, análisis pragmático); la ontología lingüística especifica la gramática del lenguaje español, y cuenta con una extensión de derivaciones coloquiales y con un lexicón para caracterizar al lenguaje español, que a su vez contiene un *onomasticon* (se define como *onomasticon* como una colección de nombres propios y/o términos especializados y/o coloquiales); la ontología interpretativa modela el conocimiento sobre el contexto específico del usuario (es una ontología de alto nivel, con especializaciones/extensiones basadas en ontologías de dominio externas al MODS), donde esta ontología interpretativa, está compuesta con una ontología del usuario que describe el uso del sistema que va realizando cada usuario, lo que permite incorporar a la consulta formal las características propias del usuario (contextualización) para intentar delimitar la respuesta de la web. Finalmente, otro componente clave para la adaptabilidad del MODS a la dinámica de la web y del usuario, es el componente de aprendizaje de ontologías que se propone en este trabajo, cuyo fin es permitir que las ontologías evolucionen a la par con la usabilidad del sistema.

## 3. Arquitectura de aprendizaje ontológico

El aprendizaje ontológico es usado para la adaptabilidad del MODS. Este componente recibe como entrada (ver Fig. 2):

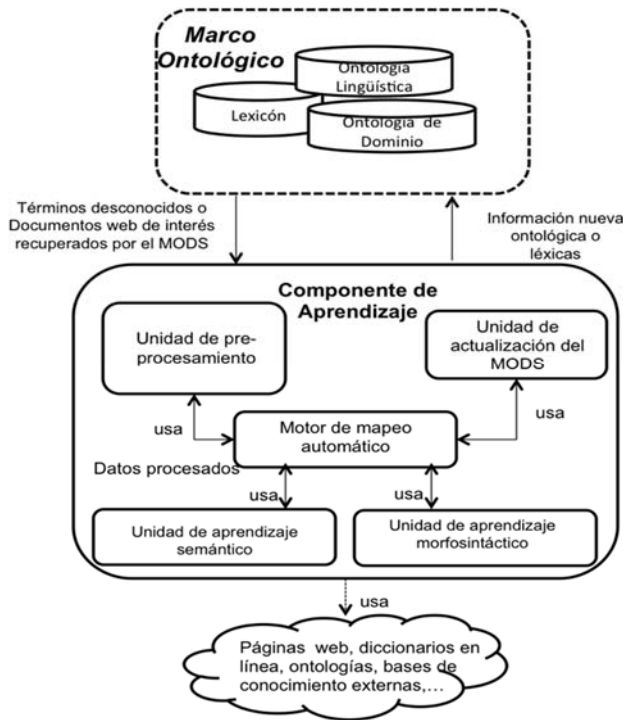


Figura 2. Componente de Aprendizaje  
Fuente: adaptado por La cruz, Rodríguez, Aguilar, 2012

Términos desconocidos o documentos Web. El primer caso ocurre con el análisis de la consulta, cuando se encuentra frente a un término desconocido por el lexicón del MODS; en este caso, el MODS invoca el aprendizaje ontológico morfosintáctico [4]. El segundo caso ocurre cuando el proceso de interpretación de la consulta es exitoso, lo que resulta en un grupo de información recuperada de la Web por la consulta generada por MODS. En este caso, el MOD invoca el aprendizaje ontológico semántico [5], el cual aprende nuevos conceptos, relaciones taxonómicas y no taxonómicas.

La arquitectura de aprendizaje ontológico consta de cinco componentes: el primer componente es la *unidad de pre-procesamiento*, que caracteriza la información de entrada, sea un término desconocido o información recuperada de la Web. El segundo componente es la *unidad de aprendizaje semántico*, el cual contiene un repositorio de técnicas de aprendizaje de conceptos, relaciones taxonómicas y no taxonómicas, a partir de documentos. El tercer componente es la *unidad de aprendizaje morfosintáctico*, este componente utiliza diccionarios en línea para determinar las estructuras léxicas o sintácticas de términos desconocidos (este componente se encuentra implementado en su totalidad en [4]). El cuarto componente es la *unidad de actualización* que determina el elemento del MODS que se va actualizar, según el aprendizaje que se esté realizando: al lexicón, a la ontología interpretativa, o a la ontología lingüística. Finalmente, el *motor de mapeo automático*, es el que se encarga de llamar a las unidades de aprendizaje según la entrada al sistema (determina cuál de los dos aprendizajes se va a utilizar).

Tabla 1.  
Requerimientos de aprendizaje de información morfosintáctica.

Categoría	G	N	T	M	Ti	A	V	P
Sustantivo	x	x	x					
Adjetivo	x	x	x					
Adverbio			x					
Verbo				x	x	x	x	x

Fuente: Los autores

A continuación describimos las dos unidades de aprendizaje, que definen los dos tipos de aprendizaje del MODS.

### 3.1. Unidad de Aprendizaje Morfosintáctica

Consiste en una serie de métodos de aprendizaje para la adquisición, ampliación y refinamiento de la información morfológica y sintáctica almacenada en el lexicón, sin los cuales MODS no podría robustecerse en su capacidad de interpretación del lenguaje natural. Así, esta unidad tiene como objetivo enriquecer el lexicón. Las estructuras del lenguaje natural principales que tienen que ver con el aprendizaje morfosintáctico son las unidades léxicas de alto nivel (verbos, sustantivos, adjetivos, adverbios).

Según la categoría lingüística a aprender, esta unidad de aprendizaje requiere de un conjunto de conocimientos a adquirir, la Tabla 1 los define.

Donde G es el Género, N: Número, T: Tipo, M: Modo, Ti: Tiempo, A: Aspecto, V: Voz, P: Persona.

Dicha información es caracterizada en la siguiente función, que constituye la interfaz del MODS con el componente de aprendizaje:

*lex\_mor (componente léxico, categorías, tipo, género, número, modo, tiempo, aspecto, voz, persona, instancia\_ontologia\_linguistica).*

Esta unidad está formada por cuatro fases (ver Fig. 3).

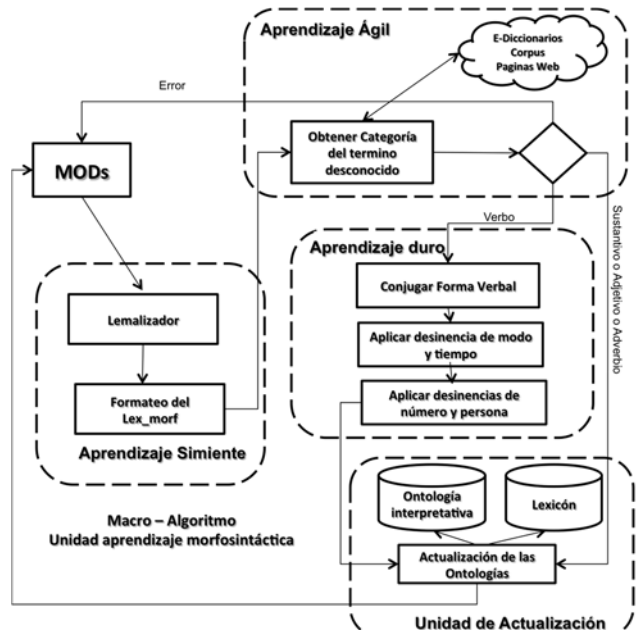


Figura 3. Macro Algoritmo de la Unidad de Aprendizaje Morfosintáctica  
Fuente: Los autores

La fase de *aprendizaje simiente* que es de pre-procesamiento y caracterización, recibe como entrada un término desconocido, y halla la información morfosintáctica básica del término, tal como su forma canónica, afijos (prefijos y sufijos), etc. Con esta información se caracteriza el término. La forma canónica de un verbo en infinitivo es el mismo, y para un verbo conjugado su forma infinitiva: para un sustantivo, adjetivo o adverbio en plural, su forma canónica es su singular. En esta fase se hace uso de una librería de algoritmos diseñados para tales fines: un lematizador y un analizador de constituyentes, respectivamente. En el caso de los adverbios, muchos son derivados de los adjetivos (por ejemplo, felizmente es derivado del adjetivo feliz). Es así que para el aprendizaje simiente de un adverbio se separa el sufijo o el prefijo del núcleo. En cuanto a los adjetivos, son palabras que nombran o indican cualidades, rasgos y propiedades de los nombres o sustantivos a los que acompañan. El tratamiento de aprendizaje simiente para el adjetivo consiste en aprender su forma canónica. Las demás unidades léxicas de orden inferior, tales como pronombres, determinantes etc., se encuentran predefinidos en MODS, y no son objeto de aprendizaje.

Después sigue la fase de *aprendizaje ágil*, cuyo objetivo es validar y aprender información gramatical en función del término. Para ello recibe la salida generada en el aprendizaje simiente, y devuelve alguna de las siguientes salidas:

**Caso 1:** Ocurre cuando la categoría descubierta de la palabra desconocida es un sustantivo, adjetivo o adverbio.

**Caso 2:** Ocurre cuando la categoría de la palabra descubierta es un verbo.

**Caso 3:** Ocurre cuando no se ha encontrado ninguna información relacionada con la palabra desconocida.

Según sea el caso realiza diferentes tareas, en el caso 1 va a la *unidad de actualización* para actualizar el lexicón, en el caso dos, entra en la fase de aprendizaje duro, el cual soporta el aprendizaje morfosintáctico de verbos, y consiste en la conjugación de los verbos, tanto regulares como irregulares, del español. En el caso 3 envía un mensaje de error al MODS.

En general, una de las tareas más complejas dentro del aprendizaje morfosintáctico es el aprendizaje de verbos. En este trabajo se ha desarrollado para el *aprendizaje duro* un conjunto de algoritmos que pueden diferenciar entre verbos regulares e irregulares, y desplegar la conjugación de los verbos regulares y parte de los irregulares del español, a partir de su forma en infinitivo. En general, el verbo es una categoría léxica que, al conjugarse puede variar en persona, número, tiempo, modo, aspecto y voz. Desde un punto de vista morfológico, el verbo consta de dos partes:

- la raíz, que es la parte que nos aporta el significado del verbo, tal y como lo podemos encontrar en el diccionario (información léxica).
- las desinencias, que son los morfemas flexivos que nos dan la información referente a la persona, número, tiempo, modo, aspecto y voz (información gramatical).

De este modo, en una forma verbal como “amo”, se pueden distinguir las siguientes partes: am-, raíz cuyo

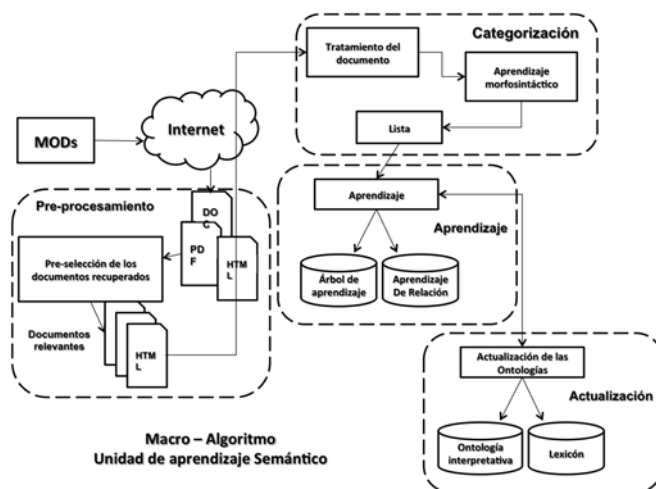


Figura 4. Esquema conceptual de la Unidad de Aprendizaje Semántico  
Fuente: Los autores

significado es “tener amor a alguien o algo”, y la desinencia que nos revela que la forma verbal amo se corresponde con la primera persona (persona) del singular (número), del presente (tiempo) del indicativo (modo) de la voz activa.

Finalmente, la fase de la Unidad de Actualización del lexicón, recibe como entrada la función `lex_mor` y actualiza el lexicón.

### 3.2. Unidad de Aprendizaje Semántico

Este aprendizaje cubre el aprendizaje de conceptos, relaciones taxonómicas, relaciones de dominio, relaciones no taxonómicas, propiedades y axiomas. En el caso del MODS el aprendizaje de relaciones no taxonómicas es fundamental, y consiste en el descubrimiento de verbos relacionados con el dominio, en la extracción de conceptos que están relacionados no taxonómicamente, en la extracción de relaciones de marcado (por ejemplo, las etiquetas puestas a un producto que hablan sobre él), de reglas de comportamiento, entre otras cosas. El aprendizaje de información semántica es fundamental para la ontología interpretativa (dominio) y lingüística.

En la Fig. 4 se muestra el esquema conceptual del aprendizaje semántico, el cual se describe brevemente a continuación.

La *Unidad de Aprendizaje Semántico* ocurre cuando el proceso de interpretación de la consulta es exitoso, generando como insumo, un conjunto de enlaces web, a partir de la consulta hecha por el MODS. Ese conjunto de enlaces es la entrada a esta unidad. La información de esos enlaces puede traer consigo información heterogénea: textos, diccionarios, bases de conocimiento, información semi-estructurada (anotaciones RDF de recursos WEB, esquema XML), información relacional, ontologías, entre otros.

Esta información pasa tres fases. Se comienza por la fase de *pre-procesamiento* para caracterizar la entrada. Para ello se realiza una clasificación débil, la cual consiste en una preselección de los documentos recuperados, tomando como base los términos de la consulta. El criterio utilizado para la

selección de los documentos relevantes es que el 100 % de los términos de la consulta estén en el documento. Otro factor que se toman en el pre-procesamiento, es que solo se van a tomar los documentos no estructurados que están en el formato HTML. En decir, la clasificación débil es un filtro sobre los documentos recuperados, basados en los criterios antes descritos.

Luego se pasa a la fase de *categorización*, comenzando con el tratamiento de cada documento, que consiste en separar todos los términos de los documentos relevantes en *tokens*, y por cada *tokens* se busca si existe en el lexicón. En caso que no exista, se genera una tabla de frecuencia de los términos a aprender, con su respectiva frecuencia de aparición en los documentos relevantes. A cada término almacenado en la tabla de frecuencia, se le realiza el análisis morfosintáctico llamando a la unidad de aprendizaje morfosintáctica, para obtener la categoría de cada término (sustantivo, verbo), y generar la lista de los términos a aprender.

La fase de aprendizaje involucra un conjunto de tareas:

- Aprendizaje de relaciones taxonómicas, que en nuestro caso es la taxonomía de conceptos definidos en la ontología interpretativa del MODS, el cual usa un árbol de aprendizaje como representación (ver Fig. 5). Este árbol está compuesto por Entidades, que representan los objetos físicos y abstractos (normalmente son los sustantivos), y por Eventos, que representan una acción (normalmente son los verbos). Por cada concepto a aprender se establece una función numérica de pesos, utilizada para reforzar el concepto que se está aprendiendo La función está definida con la siguiente fórmula:

$$Peso = peso + incremento (incr),$$

Donde  $incr = \alpha * incr$ , siendo  $\alpha$  el factor de aprendizaje.

La misma fórmula usa para decrementar el peso. Ese peso es usado en el arco que enlaza el tipo de nodo evento o entidad, con el concepto a aprender. Si ese peso tiene un valor superior a un umbral dado, después de culminar el proceso de aprendizaje, es incorporado al MODS, de lo contrario se descarta.

- Aprendizaje de relaciones no taxonómicas, el cual consiste en aprender las relaciones entre los conceptos. Los tipos de relación a aprender pueden ser las relaciones binarias, donde solo se relaciona dos conceptos (por ejemplo, José dicta Inteligencia Artificial), relaciones entre sinónimos, de polisemia.

En particular, el procedimiento general que se sigue en esta fase es el siguiente

- Para cada elemento de la lista de aprendizaje, se busca el documento donde se encuentra el elemento, y se extrae del documento la frase donde es usado.
- Luego se realiza el análisis morfosintáctico y la etiquetación de la frase.
- Con el resultado, se busca el patrón en la base de datos de plantillas de frases (patrones léxicos-sintáctico).
- En caso de que se encuentre ese patrón, se actualiza el árbol de aprendizaje (el cual es una jerarquía de concepto). Esa actualización consiste en modificar los

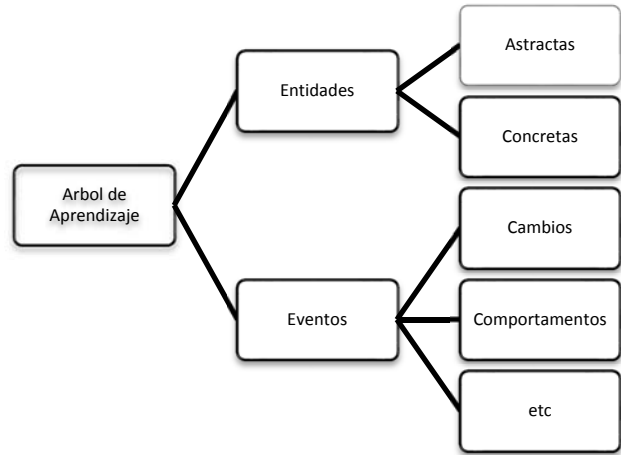


Figura 5. Árbol de Aprendizaje  
Fuente: Los autores

pesos del árbol, reforzando los pesos de los arcos que representan que los conceptos son usados según lo establecido en la plantilla, debilitando el resto de los casos que representan usos distintos del concepto a lo establecido en la plantilla.

- Si no se encuentra se actualiza el árbol de relación (ya que es una relación entre dos conceptos).

En la sección 4 se presenta un ejemplo de cómo se comporta esta unidad.

Por último, en la fase de actualización se actualiza la ontología interpretativa.

#### 4. Ejemplo de la unidad de aprendizaje semántico

En esta sección se desarrolla un ejemplo de la unidad de aprendizaje semántico. Para el caso del aprendizaje morfosintáctico, ya existen ejemplos en [4]. El aprendizaje semántico ocurre cuando el proceso de interpretación de la consulta es exitoso.

Supongamos una información recuperada de la Web, por la siguiente consulta “Profesores and pertenecen and ‘Universidad de Los Andes’ and Mérida and Venezuela filetype: html”, realizada por el MODS. En la Fig. 6 se muestra una parte de los documentos recuperados.

El siguiente paso es la preselección de los documentos recuperados (fase de pre-procesamiento): los documentos 1, 2, 4 y 6 son seleccionados como documentos relevantes, por tener todos los términos de la consulta (criterios de la clasificación débil), mientras que los documentos 3 y 5 son rechazados.

Siguiendo con la fase de categorización, se genera una tabla de frecuencia con los términos a aprender y su categoría (sustantivo o verbo), ver Tabla 2.

Luego se pasa a la fase de aprendizaje, que se detalla a continuación para un caso en particular; por ejemplo, uno de los términos que se encuentra en la tabla de términos a aprender es Universidad, con una frecuencia de 35 y con categoría “sustantivo”.

El procedimiento que se sigue es el siguiente:

- Aproximadamente 34.600 resultados (0,28 segundos)
- Reseña (Biografía) - Web del Profesor - ULA**  
webdelprofesor.ula.ve/ingenieria/oscaror/Personal/Resena\_Biografia.html ...  
El 1º de Abril de ese mismo año ingreso como **Profesor Contratado** adscrito al ...  
Ingeniero Electricista, **Universidad de Los Andes (ULA), Mérida, Venezuela, 1975. ...**  
**Pertenece** al Programa de Promoción al Investigador (PPI) nivel Candidato: ...
  - vitae\_herbert\_hoeger - CeCaICULA - Universidad de Los Andes**  
www.cecalc.ula.ve/.../miembros/.../vitae\_herbert\_hoeger\_08\_99\_esp.htm...  
**Universidad de Los Andes, Mérida, Venezuela. 1984. ... Profesor Asociado** en el  
Departamento de Investigación de Operaciones. Universidad de ... **SOCIEDADES**  
**CIENTÍFICAS Y PROFESIONALES A LAS QUE PERTENECE.** Colegio de ...
  - metadata\_san\_juan**  
www.cecalc.ula.ve/proyectos/redbo/metadata/metadata\_san\_juan.html ...  
Instituto de Investigaciones Agropecuarias de la **Universidad de Los Andes**, Sector  
Santa Rosa., **Mérida, Mérida, 5101 Venezuela ...**
  - Fortalezas Caracter Resiliencia Estudiantes Medicina Universidad ...**  
biblioteca.universia.net/html...merida-venezuela/.../56220252.html ...  
**Pertenece** a: Biblos-e Archivo (Universidad Autónoma de Madrid) ... **Universidad de los**  
**Andes, Mérida, Venezuela** Tesis doctoral inédita leída en la ... inédita leída en la  
Universidad Autónoma de Madrid, Facultad de Formación de **Profesor.**
  - Comando Simón Bolívar integró a estudiantes y profesores**  
elmeridenazo.com.ve/meridaaldia/noticia11050.htm?idCont=1,1,1...  
20/03/2013 - El **profesor** José María Andrés, coordinador del voluntariado universitario  
con **Henrique Capriles** en **Mérida**, ofreció una charla, destacando ... 500 personas que  
**pertenece**n a la universidad, pero necesitan más personas que ... porque con el actual  
régimen la **Universidad venezolana** está amenazada.
  - DINASTÍA BIENVENIDA: MÉRIDA: HA MUERTO EL ESCULTOR ...**  
dinastiabienvenida.blogspot.com/.../merida-ha-muerto-el-escultor-manue...  
05/03/2010 - Doctor "Honoris Causa" de la **Universidad de los Andes de Mérida ...**  
afincado en la ciudad **venezolana**, a cuya universidad **pertenece**, desde hace más ...  
Ha sido **profesor** de dibujo y escultura en la Escuela Antonio Esteban ...

Figura 6. Extracto de los documentos recuperados de la consulta realizada por el MODS.  
Fuente Resultado obtenido del motor de búsqueda google. Fecha: 12/08/2013.

Tabla 2.  
Ejemplo de la tabla de frecuencias

Término	Frecuencia	Categoría
Universidad	35	Sustantivo
Organización	13	Sustantivo
Dicta	7	Verbo
Mérida	4	Sustantivo

Fuente: Los autores

- Se extrae de uno de los documentos relevantes la frase donde se encuentra la palabra Universidad, obteniéndose

la siguiente frase: “Universidad de Los Andes es una Universidad”;

Se realiza el análisis morfosintáctico de la frase, que consiste en obtener la información morfológica y la estructura sintáctica, utilizando en este caso la gramática en español. Luego de este análisis, se obtiene lo siguiente para nuestro ejemplo:

Universidad [NP] de [prep] “Los Andes” [NP] “es una” Universidad [NP]

Donde:

NP: Nombre Propio

Prep: preposición

La regla (patrón) para Nombre Propio es la siguiente:

NP=> NP | NP prep NP

Por lo tanto, la frase anterior queda de la siguiente manera

“Universidad de Los Andes” [NP] “es una” Universidad [NP]

Obteniendo el siguiente patrón:

NP “es una” NP

- El patrón obtenido se busca en la base de datos de las plantillas de patrones léxicos-sintáctico. Un ejemplo de las plantillas que podrían estar en esa base de datos se muestra en la Tabla 3.
- Si encuentra el patrón determina que tipo de relación es. En nuestro ejemplo es una relación es\_ una (Universidad, Universidad de Los Andes),
- Luego busca en el árbol de aprendizaje de la Fig. 5 el concepto Universidad, si no encuentra el concepto ve que tipo de concepto es. En este caso se dedujo que es un nombre propio. Por la regla 11 se sabe que puede ser una cosa, persona, animal o un lugar; y por la regla 13 que debe ir en la rama de Entidades y Concreta. Por lo tanto, se inserta el concepto en la rama de entidades, con su respectivo peso inicial de 0.25, tal como se muestra en la Fig. 7.

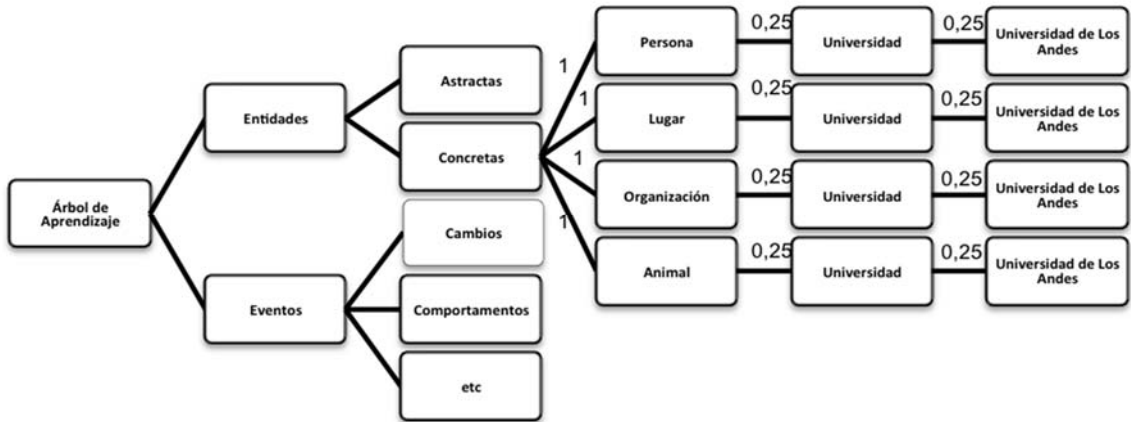


Figura 7. Árbol de Aprendizaje después de considerar los términos Universidad y Universidad de Los Andes en la frase.  
Fuente: Los autores

Tabla 3.  
Base de Datos de patrones léxicos-sintáctico

Patrón	Ejemplo de instancias de reglas
1. FN <sub>i</sub> {,FN <sub>i</sub> }*{,} u otros FN <sub>n</sub>	1. Por el patrón 1. "Jose
2. FN <sub>i</sub> {,FN <sub>i</sub> }*{,} u otras FN <sub>n</sub>	Aguilar, Gerard Paez, u otros
3. FN <sub>i</sub> es un FN <sub>2</sub>	profesores"
4. FN <sub>i</sub> es una FN <sub>2</sub>	2. Por el patrón 5.
5. Es_una(FN <sub>2</sub> ,FN <sub>1</sub> )	"Universidad de Los Andes" es una
6. Es_un(FN <sub>2</sub> ,FN <sub>1</sub> )	Universidad
7. U_otros(FN <sub>n</sub> ,FN <sub>i</sub> ) para i=1 hasta n-1	
8. U_otras(FN <sub>n</sub> ,FN <sub>i</sub> ) para i=1 hasta n-1	
9. FN->NP	
10. NP-> NP de NP	
11. NP->Persona   Organización   Animal   Lugar	
12. FN->N	
13.NPsem-> Entidades -Concretas	

Fuente: Los autores

Donde FN-> Frase Nominal. NP-> Nombre propio. NPsem-. Semántica del Nombre Propio.

En una segunda iteración volvemos a buscar el término en los documentos relevantes y recomenzamos de nuevo:

1. Por ejemplo, se encuentra en otro documento relevante la siguiente frase: "Universidad es una Organización".
2. Realizando el proceso anterior se tiene: NP es una NP, en este caso es una relación es\_una (Organización, Universidad)
3. Además, Universidad y Organización están en el árbol de aprendizaje, por lo que lo que se hace es incrementar los pesos de los conceptos donde están universidad y organización, y disminuir los pesos en los otros lugares donde se encuentra universidad pero no organización. En concreto, donde está Organización y Universidad se incrementa a 0,26, y los otros, tales como persona y Universidad, Lugar y Universidad, etc. se disminuye a 0,24 (ver Fig. 8)

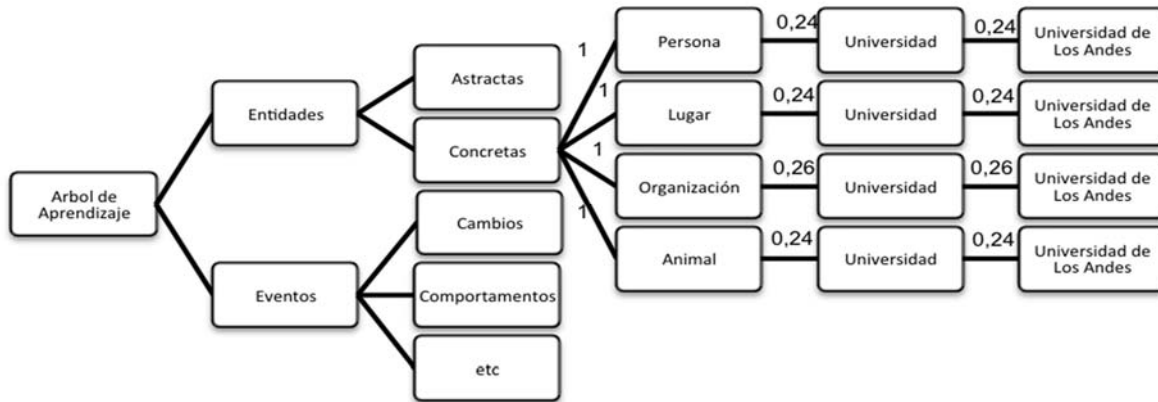


Figura 8. Árbol de Aprendizaje después de considerar los términos Organización y Universidad que aparecen en la frase  
Fuente: Los autores

Supongamos que en otro documento relevante se encuentra nuevamente la frase: "Universidad es una Organización". Esto hará nuevamente reforzar el concepto de que Universidad es una Organización, lo que implica la actualización de los pesos otra vez; es decir, donde esta organización y universidad se actualiza a 0,27, y los demás se disminuye. Este proceso se sigue haciendo cada vez que se encuentre en algún documento relevante frases donde aparece el término Universidad, reforzando y debilitando a las ramas respectivas.

Para la fase de actualización, se seleccionan del árbol de aprendizaje los conceptos cuyos pesos de sus ramas sean mayor a 0,90. Los conceptos que están en ese rango, son candidatos para la actualización de la ontología Interpretativa del MODS. Entonces, esta fase se encarga de actualizar, tanto al lexicón como a la ontología Interpretativa, con estos términos.

## 5. Conclusiones

En este trabajo se presentó la arquitectura de aprendizaje ontológico para el MODS, capaz de adquirir nuevo conocimiento en función de la información suministrada por el proceso de análisis de una consulta en lenguaje natural, y/o de la información recuperada desde la web por dicha consulta. Esta arquitectura, a partir del nuevo conocimiento adquirido, permite potenciar las estructuras del MODS. La arquitectura de aprendizaje ontológico se caracteriza por integrar diferentes métodos y técnicas de descubrimiento de conocimiento léxico, morfológico, sintáctico, semántico y pragmático. Además, a través de la arquitectura es posible explotar cualquier recurso computacional, sea una página web, una base de datos, una ontología.

Así, se ha diseñado una arquitectura novedosa con capacidad para soportar una variedad de elementos de aprendizaje.

Lograr su implementación completa ha mostrado ser un reto de gran complejidad. Una de las razones es que estamos frente a uno de los fenómenos más interesantes en la inteligencia humana, el aprendizaje lingüístico. En términos computacionales, esto implica la construcción de varios algoritmos que emulen los procesos que para los humanos son fáciles, como por ejemplo, saber si una palabra es un verbo o un sustantivo (por ejemplo: comer y pan). Esto tiene que ver con lo que se ha denominado aprendizaje morfosintáctico. Otro ejemplo sería saber si es un concepto o una relación (por ejemplo, Juan dicta Matemáticas), este caso ocurre dentro de lo que se ha denominado aprendizaje semántico.

En este trabajo se describe en forma general los dos macro algoritmos de la unidad de aprendizaje, y en particular, se muestra el uso del de la unidad de aprendizaje semántico, ya que el de la unidad morfosintáctica ya fue presentado en [4].

Según los resultados obtenidos, vemos que la unidad de aprendizaje semántico realiza el aprendizaje de conceptos, relaciones taxonómicas y no taxonómicas, en documentos no estructurados que son recuperados de la Web tradicional.

Entre los trabajos futuros se tiene la implementación de la unidad de aprendizaje semántico, y la integración de ese componente de aprendizaje con el MODS.

## Agradecimientos

Al proyecto CDCHT I-1237-10-02-AA, y su proyecto satélite I-1239-10-02-ED, de la Universidad de Los Andes, y al Programa de Cooperación de Postgrados (PCP) entre Venezuela y Francia, titulado “Tareas de Supervisión y Mantenimiento en Entornos Distribuidos Organizacionales”, contrato Nro. 2010000307.

## Referencias

- [1] Rodríguez, T. Puerto, E. y Aguilar, J., Dynamic semantic ontological framework for web semantics. Proceeding of the 9<sup>th</sup> WSEAS(CIMMACS’10, Mérida-Venezuela), pp. 91-98, 2010.
- [2] Randal, D. Howard, S. and Szolovits, P., What is a knowledge representation? AI Magazine [En Línea], 14 (1), pp. 17-33, 1993. [Fecha de consulta marzo 09 de 2010] Disponible en: <http://groups.csail.mit.edu/medg/ftp/psz/k-rep.html>
- [3] Gómez, A., Fernández, M. y Corcho, O., Ontological engineering: with examples from the areas of knowledge management, e-commerce and the Semantic Web, 1<sup>ra</sup> ed., London, Springer, 2004. ISBN 1852335513.
- [4] Puerto, E., Aguilar, J. and Rodríguez, T., Automatic learning of ontologies for the Semantic Web: Experiment lexical learning. Revista Respuestas [en línea]. 17 (2), pp. 5-12, 2012, ISSN 0122-820X. Disponible en: <http://www.ufps.edu.co/ufpsnuevo/revista-respuesta/index2.php>
- [5] Lacruz, S., Rodríguez, T. y Aguilar, J., Informe Técnico: Aprendizaje semántico para el marco ontológico dinámico semántico. Centro de Estudios en Microelectrónica y Sistemas Distribuidos (CEMISID), Universidad de Los Andes, Mérida Venezuela, 2013.
- [6] Takeschi, M., Yoshihiro, S., Naoki, S. and Naoky, F., DOODLE-OWL: OWL-based Semi-Automatic ontology development environment, Prepared for 3<sup>rd</sup> International Semantic Web Conference, pp. 33-36, 2004.
- [7] Cimiano, P. and Volker, J., Text2Onto - A Framework for ontology learning and data-driven change discovery, Proceedings of the 10<sup>th</sup> International Conference on Applications of Natural Language to

Information Systems (NLDB), vol. 3513 of Lecture Notes in Computer Science, Alicante, Spain, pp. 227-238. 2005. Available at: <https://code.google.com/p/text2onto/>

- [8] Aussenac, N. y Seguela, P., Les relations sémantiques: Du linguistique au formel. Cahiers de grammaire. N° spécial sur la linguistique de corpus. A. Condamines (Ed.) Vol. 25. Toulouse: Presse de l’UTM, 2000.
- [9] Shamsfard, M. y Barforoush, A., An introduction to Hasti: An ontology learning system, Proceedings of the iasted international conference artificial intelligence and soft computing, Acta Press, Calgary, Canada, pp 242-247, 2002.
- [10] Valencia, R., Un entorno para la extracción incremental de conocimiento desde texto en lenguaje natural. Tesis Doctoral, Departamento de Ingeniería de la Información y las Comunicaciones, Universidad de Murcia, España, 2005.
- [11] Missikoff, M. Navigli R. y Velardi, P., The usable ontology: An environment for building and assessing a domain ontology, Research paper at International Semantic Web Conference (ISWC), Sardinia, Italia, 2002.

**T. Rodríguez**, es graduada como Ing. de Sistemas en 1993, MSc. en Computación en 2000, ambos en La Universidad de Los Andes, Venezuela. Actualmente candidata a Dr en Ciencias Aplicadas de la Facultad de Ingeniería de la Universidad de los Andes, Mérida Venezuela; en los últimos 4 años se desempeña como investigadora en el Centro Estudios en Microelectrónica y Sistemas Distribuidos (CEMISID) en el área de Ontologías, Semántica Web. Aprendizaje Ontológico. Se desempeña como administradora general tecnología de información en el postgrado en computación de la Universidad de Los Andes, Venezuela. Página Web: <http://taniana.novacorp.co/>

**J. Aguilar**, es graduado como Ing. de Sistemas en 1987 en la Universidad de Los Andes, Mérida, Venezuela. MSc. en Informática en 1991 en Paul Sabatier, Toulouse, Francia. Dr. en Ciencias Computacionales en 1995 en la Universidad Rene Descartes, Paris, Francia. Es Profesor Titular del Departamento de Computación de la Universidad de los Andes, Venezuela e investigador del Centro Estudios en Microelectrónica y Sistemas Distribuidos (CEMISID) de la misma Universidad. Ha publicado más de 350 artículos y 8 libros, en el campo de sistemas distribuidos y paralelos, computación inteligente, automatización industrial, gestión en ciencia y tecnología. Actualmente se desempeña como Jefe del Departamento de Computación de la Escuela de Ingeniería de Sistemas de la Universidad de Los Andes, Venezuela y Coordinador del Programa Doctoral en Ciencias Aplicadas de la Facultad de Ingeniería de la misma universidad. Página Web: <http://www.ing.ula.ve/~aguilar/>