

**UNIVERSIDAD DE LOS ANDES
FACULTAD DE INGENIERÍA
DIVISIÓN DE ESTUDIOS DE POSTGRADO
POSTGRADO EN COMPUTACIÓN**



**“APRENDIZAJE AUTOMÁTICO DE ONTOLOGÍAS PARA UN MARCO
ONTOLÓGICO DINÁMICO EN LA WEB SEMÁNTICA”**

Autor: Eduard Gilberto Puerto Cuadros

Tutor: Dr. José Aguilar

**Trabajo de grado presentado ante la ilustre Universidad de Los Andes como
requisito parcial para optar al grado de *Magister Scientiae en Computación***

Mérida, Julio, 2010

RESUMEN

El propósito de la tesis es el diseño e implementación de un sistema de aprendizaje automático de conocimiento para un marco ontológico dinámico semántico, a partir de diversas fuentes de información. El marco ontológico dinámico semántico (MODS: Marco Ontológico Dinámico Semántico) permite interpretar consultas en lenguaje natural (español) a la web semántica. MODS está constituido por cinco estructuras de conocimiento que modelan el fenómeno del proceso de interpretación; *una ontología lingüística* para la gramática del español, *un lexicón* (extendido con onomasticon¹) para la información léxica, *una base de hechos* (repositorio de episodios o experiencia del sistema), *una ontología de tareas* que modela los procesos de análisis lingüístico tradicionales ajustados al procesamiento del lenguaje natural, y *una ontología interpretativa* para el análisis del contexto.

MODS requiere de un sistema de aprendizaje automático para la actualización de sus estructuras de conocimiento. El componente de aprendizaje usa información recuperada de la web e información suministrada implícitamente por el usuario en la consulta, como fuentes de información/conocimiento para sus métodos de aprendizaje. Para cada fuente en particular se definen algoritmos especializados de aprendizaje. El sistema de aprendizaje mapea la fuente de aprendizaje (información no estructurada o semi-estructurada), con las técnicas de descubrimiento de conocimiento (algoritmos de aprendizaje), y con las estructuras de conocimiento que son impactadas dentro del MODS.

Particularmente, se diseña y se elabora un prototipo del componente de aprendizaje, el cual considera dos elementos a aprender, por consiguiente, con dos métodos de aprendizaje diferentes: uno para categorías léxicas (verbo,

¹ abarca las composiciones compuestas por listas de palabras de ciertas categorías. En nuestro caso se usa para referirnos a listas de nombres propios.

sustantivo, adjetivo, adverbio) presentes en la consulta del usuario, y otro para relaciones de dominio o no taxonómicas (*causa, causado_por, operado_por, etc*) descubiertas de la información recuperada de la web. El primer método impacta al lexicón, y el segundo impacta a las ontologías interpretativa y lingüística.

Palabras claves: procesamiento de lenguaje natural, aprendizaje de ontologías, semántica ontológica, web semántica.

ÍNDICE GENERAL

	Página
INTRODUCCIÓN	1
1.1 GENERALIDADES	1
1.2 ANTECEDENTES	3
1.3 OBJETIVOS	7
1.3.1 Objetivo General	7
1.3.2 Objetivos Específicos	7
1.4 ORGANIZACIÓN DEL TRABAJO	8
2. MARCO TEORICO	9
2.1 APRENDIZAJE ONTOLÓGICO	9
2.1.1 Aprendizaje de ontologías en general	9
2.1.2 Aprendizaje de información morfosintáctica desde el lenguaje natural	12
2.1.3 Aprendizaje de relaciones no-taxonómicas	14
2.2 MARCO ONTOLÓGICO DINAMICO SEMÁNTICO: MODS	17
2.2.1 Pasos generales para la interpretación de la consulta en el MODS	20
2.2.2 Lexicón del MODS	21
2.2.3. Ontología lingüística del MODS	24
2.2.4 Ontología interpretativa del MODS	25
2.2.5 Ontología de tareas del MODS	27
3. ARQUITECTURA PARA EL APRENDIZAJE AUTOMÁTICO	35
3.1 PROCESO DE DESARROLLO DE ONTOLOGÍAS.	35
3.2 VISIÓN GENERAL DE LA ARQUITECTURA DE APRENDIZAJE	37
3.2.1 Aprendizaje semántico a partir de diferentes métodos y fuentes	41
3.2.2 Aprendizaje morfosintáctico del MODS	42
3.3 DETALLES DEL COMPONENTE DE APRENDIZAJE	43
3.3.1 Detalles del aprendizaje morfosintáctico	48
3.3.2 Aprendizaje semántico: caso de relaciones no taxonómica	55
4. CONSIDERACIONES PARA EL APRENDIZAJE MORFOLÓGICO	61

4.1 ANÁLISIS DEL VERBO	61
4.1.1 Conjugación del verbo	63
4.2 PASOS PARA AVERIGUAR SI UN VERBO CUALQUIERA ES REGULAR O IRREGULAR	64
4.2.1 Conjugación irregular	66
4.2.2 Condiciones que debe cumplir un verbo para ser irregular	69
5. DISEÑO DEL PROTOTIPO DEL APRENDIZAJE AUTOMÁTICO DE ONTOLOGÍAS	71
5.1 DIAGRAMA DE CASOS DE USO DEL COMPONENTE DE APRENDIZAJE DE ONTOLOGÍAS	71
5.2 DIAGRAMA DE ACTIVIDADES DEL SISTEMA DE APRENDIZAJE DE ONTOLOGÍAS	74
5.2.1 Diagrama de actividades: pre-procesar información	75
5.2.2 Diagrama de actividades: gestionar aprendizaje	76
5.2.3 Diagrama de actividades: actualizar aprendizaje	77
5.3. DISEÑO DE CLASES DEL COMPONENTE DE APRENDIZAJE DE ONTOLOGÍAS	78
5.4 DIAGRAMAS DE COMPONENTES DEL COMPONENTE DE APRENDIZAJE DE ONTOLOGÍAS	79
5.4.1 Sistema de Gestión de aprendizaje de ontologías	80
5.4.2 Sistema de prueba de información léxica	80
5.4.3 Sistema de procesamiento de la información	81
5.4.4 Sistema de actualización del MODS	82
5.5 DISEÑO ESPECÍFICO DEL PROTOTIPO PARA EL APRENDIZAJE MORFOSINTÁCTICO: CASO CONJUGACIÓN DEL VERBO	83
5.5.1 Diagrama de caso de uso: Conjugación verbo	84
5.5.2 Diagrama de actividades: conjugación verbo	85
5.5.3 Diagrama de clase para el módulo de conjugación de verbo	86
5.5.4 Conjugación de irregular con patrón <i>Dao</i> y <i>Dei</i>	87
6. CASOS DE ETUDIO DEL APRENDIZAJE MORFOSINTÁCTICO	89
6.1 APRENDIZAJE A PARTIR DE UN SUSTANTIVO	89

6.2 APRENDIZAJE A PARTIR DE UN VERBO	91
6.3 APRENDIZAJE A PARTIR DE UN ADJETIVO	94
6.4 APRENDIZAJE A PARTIR DE UN ADVERBIO	96
6.5 TRAZA DE SALIDA PARA UNA PALABRA NO VALIDA	97
6.6 TRAZA DE SALIDA PARA UNA PALABRA DESCONOCIDA	98
7. CONCLUSIONES Y RECOMENDACIONES	100
7.1 CONCLUSIONES	100
7.2 RECOMENDACIONES	103

ÍNDICE DE TABLAS

	Página
Tabla 1.	Sistemas de aprendizaje de ontologías 11
Tabla 2.	Sistemas de aprendizaje de términos 12
Tabla 3.	Sistemas que soportan el aprendizaje de relaciones no- taxonómicas 16
Tabla 4.	Tipos especificados para el lexicón 22
Tabla 5.	Llenado de tipos para verbo golpear 22
Tabla 6.	Extensión del tipo <i>morf</i> en el tiempo futuro para el verbo golpear 23
Tabla 7.	Tipos definidos para la ontología lingüística 24
Tabla 8.	Tipos definidos para la ontología lingüística 24
Tabla 9.	Componente de una ontología computacional 25
Tabla 10.	Tipos definidos para la ontología interpretativa 26
Tabla 11.	Tipos definidos para la ontología de tareas 27
Tabla 12.	Valores de argumentos de componente <i>lex_mor()</i> para <i>relación</i> 30
Tabla 13.	Clasificación de XML para el aprendizaje semántico a partir de deferentes métodos y fuentes. 41
Tabla 14.	Requerimientos de aprendizaje de información morfosintáctica 42
Tabla 15.	Métodos y técnicas para el componente de aprendizaje del MODS 48
Tabla 16.	Formación de adverbios 50
Tabla 17.	Formas no personales de un verbo 62
Tabla 18.	Flexiones de voz 62
Tabla 19.	Pasos generales para la formación de cualquier verbo regular 64
Tabla 20.	Condiciones de irregularidad de un verbo 69
Tabla 21.	Descripción de caso de uso realizar aprendizaje ontológico 72
Tabla 22.	Descripción de caso de uso realizar aprendizaje léxico 72
Tabla 23.	Descripción de caso de uso realizar aprendizaje semántico 72
Tabla 24.	Descripción de caso de uso realizar gestión de errores 73

ÍNDICE DE FIGURAS

	Página
Figura 1. Marco ontológico dinámico semántico	19
Figura 2. Taxonomía de la ontología de tareas	27
Figura 3. Árbol de traducción generado después del análisis sintáctico a la sentencia " <i>Juan es bueno</i> "	32
Figura 4. Árbol semántico generado después del análisis semántico	33
Figura 5. Pasos para el desarrollo de una ontología lingüística por niveles	35
Figura 6. Arquitectura general del componente de aprendizaje de ontologías	38
Figura 7. Arquitectura aprendizaje morfosintáctico del componente léxico	49
Figura 8. Módulo de aprendizaje ágil	51
Figura 9. Proceso de aprendizaje de relaciones no taxonómicas	58
Figura 10. Diagrama de casos de uso del componente de aprendizaje	71
Figura 11. Diagrama de actividades: aprendizaje de ontología	74
Figura 12. Diagrama de sub-actividades: pre-procesar información	75
Figura 13. Diagrama de sub-actividades: gestionar aprendizaje	76
Figura 14. Diagrama de sub-actividades: actualizar aprendizaje	77
Figura 15. Diagrama de clases: componente de aprendizaje de ontologías	78
Figura 16. Diagrama de componentes del sistema de aprendizaje	79
Figura 17. Diagrama de paquetes del sistema de gestión de aprendizaje	80
Figura 18. Diagrama de paquetes del sistema prueba de información	81
Figura 19. Diagrama de paquetes para el sistema de actualización	82
Figura 20. Diagrama de casos de uso para el módulo conjugar verbo	83
Figura 21. Diagrama de casos de uso para el módulo conjugar verbo	84
Figura 22. Diagrama de paquetes para el sistema de actualización	85
Figura 23. Diagrama de clase para el módulo de conjugación de verbo	86
Figura 24. Respuesta de un diccionario	90