

Evaluación de relaciones ontológicas en corpora de dominio restringido

Mireya Tovar^{1,2}, David Pinto², Azucena Montes^{1,3}, Gabriel González-Serna¹, and Darnes Vilariño²

¹ Centro Nacional de Investigación y Desarrollo Tecnológico, Cuernavaca, Morelos, México

² Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias de la Computación, Puebla, México

³ Universidad Nacional Autónoma de México, Grupo de Ingeniería Lingüística, México

{mtovar,gabriel}@cenidet.edu.mx, {dpinto,darnes}@cs.buap.mx, AMontesR@iingen.unam.mx

Resumen. En este artículo proponemos una evaluación automática de relaciones en ontologías de dominio restringido. En particular, usamos varios patrones léxico sintácticos con la finalidad de evaluar las relaciones class-inclusion y relaciones ontológicas que contiene la ontología. Nuestro enfoque se centra en un corpus de referencia para encontrar evidencia de la validez de la relación. El enfoque es capaz de proporcionar una medida de exactitud para cada ontología evaluada, un valor asociado de alguna manera con la calidad de las relaciones de la ontología. Esta puntuación se da con cierto grado de confiabilidad, obtenida mediante la comparación de los resultados dados por el enfoque contra de la evaluación de expertos humanos y un baseline.

Palabras clave. Evaluación de relaciones, patrones léxico sintácticos, ontologías de dominio restringido.

Evaluation of Ontological Relations in Corpora of Restricted Domain

Abstract. In this paper we propose a new approach for automatic evaluation of relations in ontologies of restricted domain. In particular, we use various lexico-syntactic patterns with the aim of evaluating the class-inclusion and ontological relations that the ontology holds. Our approach focuses on a reference corpus for finding evidence of the relation validity. The approach is capable to provide an accuracy measure for each ontology evaluated, a value associated in some way with the quality of the ontology relations. This score is given with a certain degree of reliability, and it is obtained by comparing the results given by our approach against the evaluation of human experts and a baseline.

Keywords. Evaluation of relations, lexico-syntactic patterns, ontologies of restricted domain.

1. Introducción

La Web en la actualidad es el espacio que almacena una gran cantidad de información del mundo. La mayor parte de esta información se encuentra expresada en lenguaje natural, es decir, carece de una estructura que permita asociar fácilmente el significado de sus componentes, por lo que cada vez más es necesario el uso de herramientas semánticas para el tratamiento de este tipo de información. Las ontologías juegan un papel importante en la Web Semántica, puesto que son recursos que permiten capturar el conocimiento explícito en los datos, por medio de conceptos y relaciones; dando la posibilidad a los usuarios y máquinas de entender los datos intercambiados.

Una ontología se define como "una especificación explícita y formal de una conceptualización compartida" [13]. En general, este tipo de recurso semántico está formado por conceptos o clases, relaciones, instancias, atributos, axiomas, restricciones, reglas y eventos. Las ontologías de dominio son un sistema de representación del conocimiento que se puede organizar en estructuras taxonómicas y ontológicas de conceptos de algún área o dominio de conocimiento específico.

El aprendizaje de ontologías o generación automática de ontologías, es un proceso que puede facilitar la construcción automática o semiautomática de ontologías para el ingeniero de conocimiento.

En la actualidad existen propuestas de sistemas computacionales para la generación automática de ontologías, sin embargo, en la mayoría de los casos carecen de una evaluación automática, por lo que regularmente se desconoce la calidad de los recursos semánticos que estos sistemas generan.

La evaluación de ontologías es una tarea que consiste en medir la calidad de dichos recursos. El objetivo final de la evaluación de la ontología es facilitar la labor del ingeniero del conocimiento o del experto del dominio a verificar la calidad de la misma, debido a que cuando la ontología es de un tamaño considerable, esta tarea consume mucho tiempo (horas-persona). El proceso de evaluación no suele ser trivial, pues es necesario elegir qué elementos de la ontología deberían ser considerados en el proceso de medición de la calidad de la misma, así como los criterios específicos a usar.

Las propuestas de evaluación según la literatura se categorizan en [3].

a) *Evaluación realizada por humanos siguiendo criterios, estándares y requerimientos*, se definen ciertas características o criterios que permiten evaluar la ontología proporcionando un puntaje o ranking numérico [29]. Algunas características que se consideran son: completitud, correctitud, legibilidad y flexibilidad [5, 11]. Otros criterios, para la evaluación del contenido, son realizados manualmente por expertos de dominio, como es: consistencia, completitud, concisión, la capacidad de expansión, sensibilidad [22]. b) *Evaluación basada en una aplicación o en una tarea*, consiste en probar el desempeño de la ontología en una aplicación, es decir, trata de medir qué tanto la ontología ayuda a mejorar los resultados de una cierta tarea. Normalmente, miden la dimensión funcional de una ontología. Por ejemplo, responder a preguntas del usuario al usar una ontología [32, 38] o el uso de la ontología para mejorar el desempeño de un motor de búsqueda semántico al recuperar documentos relevantes a la consulta [16]. c) *Evaluación basada en gold estándar*, la calidad de la ontología se expresa por la similitud que

existe con respecto a otra ontología construida manualmente, es decir, una ontología gold estándar [32, 33, 36, 46, 23, 10]. Otro de los enfoques considerados es el enfoque d) *Evaluación basada en un corpus de referencia*, en este caso, la calidad de la ontología se representa por la oportunidad que tiene para cubrir el tópico de un corpus, como el criterio de completitud [12]. Se enfocan en la dimensión funcional de una ontología, la cual es comparada con el contenido de un corpus de textos que son representativos para el dominio. El contenido del corpus es analizado con técnicas de lenguaje natural, para identificar términos y relaciones semánticas. En [4], se utiliza un enfoque probabilístico para comparar los conceptos de una ontología con un conjunto de términos importantes identificados en el corpus de referencia (extendidos al añadir dos niveles de hiperónimos desde WordNet). La finalidad es detectar de un conjunto de cinco ontologías, cuál se adapta mejor al dominio de artistas.

En este artículo estamos interesados en el último tipo de evaluación de ontologías, es decir, la evaluación basada en un corpus de referencia considerando, en este caso, el criterio de exactitud.

Por el momento, en este artículo únicamente nos enfocamos en evaluar las relaciones de la ontología basándonos en la evidencia que existe de éstas en el corpus asociado al mismo dominio de la ontología (evaluación basada en corpus). De los trabajos revisados en la literatura, la mayoría de los enfoques de evaluación de ontologías miden la calidad de las ontologías que han sido creadas automáticamente utilizando una ontología gold estándar. En nuestro caso, sin embargo, no se está creando una nueva ontología, y tampoco se tiene una ontología gold estándar, por lo tanto, es difícil tener un mecanismo de comparación con otros trabajos presentados en el estado del arte. Así, hemos optado por evaluar la calidad de los enfoques propuestos, usando dos criterios básicos: 1) un baseline, y 2) una validación sobre un subconjunto de muestras del corpus asociado utilizando expertos del dominio como evaluadores humanos.

En [43] se presenta una metodología para la evaluación automática de ontologías de dominio. Se muestra la evaluación de conceptos, siguiendo

una metodología híbrida, que hace uso de patrones léxico morfológicos detectados en la escritura de los conceptos y en el uso de medidas de similitud para la identificación de relaciones jerárquicas. Es importante mencionar que en este artículo sólo se presenta la segunda fase de la metodología, la cual consiste en la evaluación automática de las relaciones existentes en las ontologías.

La aproximación propuesta asume que el corpus de referencia está asociado semánticamente a la ontología de dominio. En una primera etapa, el propósito es "evaluar" la calidad del sistema de validación de las relaciones de la ontología, tanto por expertos humanos como a través de su comparación con un baseline.

La evaluación de las relaciones se lleva a cabo de la siguiente forma: a) Identificación de relaciones class-inclusion de la ontología en el corpus de dominio a partir de patrones léxico sintácticos existentes en la literatura, y b) Identificación de relaciones ontológicas en el corpus de referencia por medio del análisis de dependencias sintácticas para la misma ontología de dominio.

Las principales contribuciones de este artículo son: (1) la evaluación de relaciones class-inclusion existentes en la ontología y en el corpus de dominio, por medio de patrones léxico sintácticos y (2) la evaluación de relaciones ontológicas por medio del análisis gramatical.

El resto de este artículo se organiza como sigue: en la sección 2 se presentan los tipos de relaciones existentes en ontologías y se discuten algunos trabajos relacionados con el descubrimiento de relaciones en las ontologías. En la sección 3 se presenta nuestra propuesta de evaluación de relaciones. En la sección 4 se muestran los resultados de la evaluación. Finalmente, las conclusiones se presentan en la sección 5.

2. Relaciones

Las relaciones, de acuerdo a [1], se clasifican en 10 clases o familias: class-inclusion, part-whole, similar, contrast, attribute, nonattribute, case relations, cause-purpose, space-time y reference. Cada una con un conjunto específico de relaciones como miembros. En particular, la clase class-inclusion contiene cinco subclases: taxonómica,

funcional, singular colectiva, plural colectiva y clase individual [1], [17]. Para este trabajo de investigación, estamos interesados en las relaciones class-inclusion. Las relaciones taxonómicas son consideradas relaciones jerárquicas que asocian una entidad de un tipo específico (hipónimo) a otra entidad de un tipo más general (llamado hiperónimo) [37]. La clase part-whole también es considerada como relación jerárquica [37]. Las restantes 8 clases se consideran relaciones ontológicas. Las relaciones de hiperonimia también son conocidas como relaciones is-a, class-inclusion o subsunción [35]. Por ejemplo, en las oraciones "*Las emociones son estados de ánimo que expresa el ser humano cuando ocurre algo en su vida que tiene una significación especial para él. Alegría, tristeza, ira, miedo, asombro, desagrado o interés son emociones que se pueden observar desde los primeros meses de vida.*", se identifica el concepto "emoción" que contiene un hiperónimo: estado de ánimo y los hipónimos: ira, tristeza y alegría (ver figura 1).

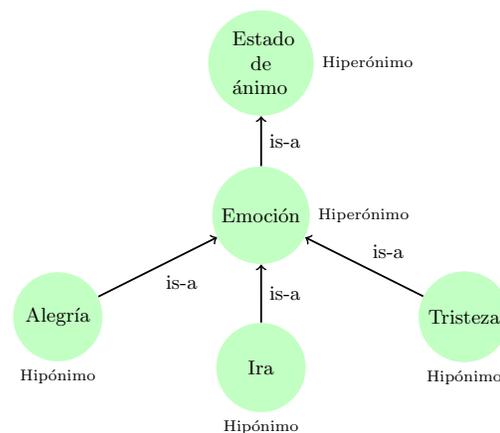


Fig. 1. Ejemplo de relación de hiperonimia e hiponimia

Las relaciones de hiponimia e hiperonimia han sido ampliamente estudiadas en lingüística computacional. Uno de los primeros enfoques que hace uso de patrones léxico sintácticos para descubrir automáticamente hiperónimos, desde textos, fue propuesto por [14]. Estos patrones han sido utilizados en otros trabajos, como por ejemplo [33, 9, 25] y se conocen ampliamente como patrones de Hearst. En [26] se presentan patrones léxico sintácticos para la identificación de relacio-

nes de hiponimia en la lengua rumana. En [2] se obtienen co-hipónimos a partir del uso de los patrones de Hearts.

Para facilitar el proceso de la creación de ontologías, la lingüística computacional ha desarrollado métodos que permiten extraer elementos de la ontología desde el lenguaje natural [34]. Se ha centrado principalmente en la extracción de relaciones; los elementos lingüísticos utilizados en algunos casos para este propósito son sustantivos, adjetivos, preposiciones y verbos [15]. En el ejemplo de la figura 1, los hipónimos pueden ser sustantivos (alegría, ira, tristeza, emoción) formados por una sola palabra o por varias palabras como el hiperónimo “estado de ánimo”.

Con la finalidad de identificar relaciones class-inclusión en textos se obtuvieron 106 patrones léxico sintácticos de la literatura utilizados para definir relaciones taxonómicas, funcionales, singular colectivas, plural colectivas y clase individual [2, 9, 14, 17, 20, 25, 26, 27, 44]. De los 106 patrones sólo 15 fueron encontrados en los corporas de dominio considerados en la etapa de experimentación de esta investigación (ver tabla 1). En la misma tabla, la columna 2 muestra el patrón léxico sintáctico, la columna 3 es una representación extendida del mismo que posteriormente se utiliza para producir una expresión regular que será usada en la extracción de relaciones tipo class-inclusion. Las variables X , Y y Z son sustituidas por los conceptos que se están revisando.

Las relaciones ontológicas o relaciones no taxonómicas, pueden ser categorizadas por verbos que aparecen con frecuencia en una vecindad de n palabras entre los conceptos [18]. Otros consideran que los conceptos que están semánticamente relacionados, tienden a estar cercanos en el texto, usan métodos estadísticos para descubrir relaciones en textos anotados [42]. Una medida de asociación entre un verbo y un par de conceptos es la frecuencia condicional. Otra medida es *above expectation* (AE) que compara la frecuencia condicional con el producto de dos frecuencias condicionales más simples [18]. Otra métrica es la correlación-semántica definida como el producto de la probabilidad condicional $P(C_j|C_i)$ con la suma de la medida de información mutua $I(C_i, C_j)$ utilizada para descubrir instancias y relaciones [42].

Ciaramita introduce en [7] un enfoque no supervisado, para extraer relaciones candidatas entre entidades nombradas por medio del uso de patrones generados a partir del análisis de dependencias. Aplican la prueba χ^2 para seleccionar las relaciones más fuertemente relacionadas con el par de entidades nombradas y realizan una evaluación manual de las mismas. Schutz identifica tripletas altamente relevantes por medio del procesamiento lingüístico (dependencias y reconocimiento de entidades nombradas) y estadístico (prueba χ^2) para el dominio de football [39].

Desde el campo de extracción de información, [40] usa un modelo de predicado-argumento (*predicate-argument*), que frecuentemente es utilizado como SVO (patrón Sujeto-Verbo-Objeto) y medidas de similitud semántica de WordNet para determinar la similitud entre pares de patrones [41]. Por otro lado, [12] utiliza un sistema de análisis morfológico, sintáctico y basado en reglas, para extraer tripletas SVO desde textos. Selecciona el verbo que formará a la relación en base a dos criterios: si sólo hay un verbo se selecciona la forma básica del mismo, si hay una expresión verbal, se descarta el verbo auxiliar y se toma el verbo principal.

En nuestro caso, estamos interesados en la evaluación de relaciones ontológicas por medio del análisis de dependencias y verificamos la existencia de la relación en los corpus de dominio.

En el trabajo desarrollado por Zarrouk, se utiliza un mecanismo de inferencia para proponer nuevas relaciones a partir de las existentes en una red léxica. Las relaciones inferidas son evaluadas por un sistema de votación (juegos con un propósito) y expertos que realizan la validación/invalidación directa de la relación [45].

3. Enfoque propuesto para la evaluación de relaciones en ontologías de dominio

El enfoque que se propone en este artículo para evaluar relaciones de las ontologías de dominio restringido consiste en la validación de dichas relaciones usando un corpus de referencia asociado

Tabla 1. Patrones léxico sintácticos

No.	Patrón literatura	Patrón extendido
12	NP such as (NP,)* (or and the) NP	X such as (.+) (or and the) Y
13	NP 's NP	X 's Y
15	such NP as (NP,)*	such (.+) (X) (.+) as (Y (,) (.+))+
42	NP (is are) NP	X (is are) Y
43	NP (is are) (a an) NP	X (is are) (a an) Y
46	NP such as (NP,)* (or and) NP	Y such as (.+) (X)? (.+) (and or) (X)?
50	NP (classify (in into) comprise contain compose (of)? group (in into) divide (in into) fall (in into) belong (to)) NP	Y (classify (in into) comprise contain compose group (in into) divide (in into) fall (in into) belong (to)) X
86	NP (and or) (another other) NP	X (and or) (another other) Y
92	NP , such as (NP,)* (or and the)? NP	X (,)? such as (Z,)* (or and the)? Y
94	NP NP , is (a an the) NP	X Z (,)? is (a an the) Y
96	NP , (is are) (NP,)* (or and the) NP	X (,)? (is are) (Z,)* (or and the) Y
97	(NP,)* (or and the) (NP,)* is (a an the) NP	(Z,X)* (or and the) (Z, X)* is (a an the) Y
98	NP , including NP	X (,)? including Y
104	NP as (NP,*) (or and the) NP	X as (Z,*) (or and the) Y
106	NP, for example, is (a an the) NP	X, for example, is (a an the) Y

al mismo dominio de la ontología. En este sentido, empleamos varios métodos para la extracción automática de relaciones ontológicas (class-inclusion y otras) y posteriormente comparamos si dichas relaciones han sido consideradas en la ontología a evaluar. Si bien, estamos haciendo un proceso de extracción automática de relaciones ontológicas, nuestro objetivo no es crear nuevas ontologías, sino, tal y como hemos comentado, buscar evidencia de las relaciones presentes en la ontología a evaluar en el corpus de referencia asociado a dicha ontología. Así, estamos usando lo que se denomina el criterio de "correctitud", el cual especifica si la información almacenada en la ontología es verdadera, independientemente del dominio de interés [30], [5].

El enfoque propuesto de evaluación se divide en cuatro etapas:

- Tratamiento de la información.
- Verificación de relaciones class-inclusion.
- Verificación de relaciones ontológicas.
- Evaluación.

La figura 2 muestra de manera gráfica estas etapas, las cuales se describen a continuación de manera detallada.

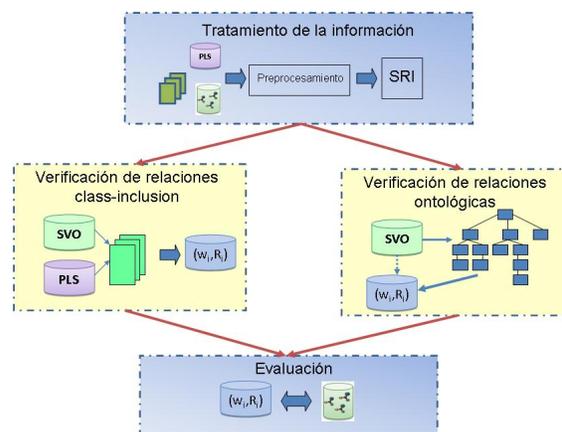


Fig. 2. Enfoque propuesto para la evaluación de ontologías

3.1. Tratamiento de la información

En la etapa de tratamiento de la información (ver figura 3), se realizan las siguientes acciones:

1. Extracción de conceptos y relaciones. En este caso se utiliza Jena¹ para extraer los conceptos y relaciones de la ontología de dominio²

¹ <http://jena.apache.org/>

² Las propiedades que se usan de la ontología para extraer las relaciones son `subClassOf` y `objectProperty`.

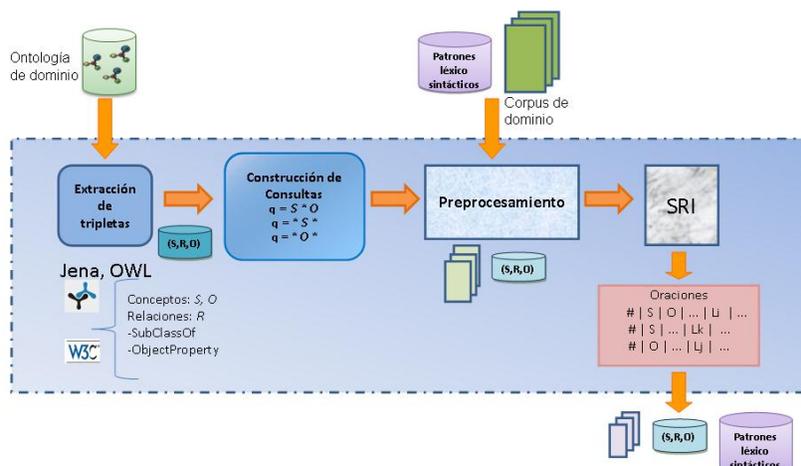


Fig. 3. Fase de tratamiento de la información

que están expresados en el formato OWL³. Las tripletas SRO están constituidas por dos conceptos: sujeto (S) y objeto (O), así como de una relación (R).

2. Construcción de consultas. Las palabras que forman a los conceptos de la ontología son utilizadas para construir las dos consultas que se usarán en el sistema de recuperación de información.
3. Preprocesamiento. A los documentos se les aplica las siguientes acciones:
 - a) Eliminación de símbolos especiales y/o caracteres no imprimibles, palabras cerradas como: preposiciones, artículos, etc.
 - b) División del corpus en oraciones, considerando aquellas oraciones separadas por punto, cuya longitud sea mayor a 30 caracteres.
 - c) Aplicación del algoritmo de stemming de Porter, con la finalidad de agrupar oraciones que contienen el mismo concepto, pero escrito en alguno de sus variantes morfológicas [31]. Por ejemplo, el concepto “subfields of artificial intelligence”, del dominio de inteligencia artificial, puede aparecer en el corpus como: “subfield

of artificial intelligence” o “subfields of Artificial Intelligence”. La consulta construida para este concepto sería: “subfield artificial intellig” que es el resultado de eliminar palabras cerradas y del truncamiento.

- d) Aplicación de un etiquetador de partes de la oración (FreeLing⁴ [28]).
- e) Eliminación de errores morfológicos producidos por FreeLing.
- f) Eliminación de símbolos de puntuación.

En el caso de las consultas se realizan las siguientes acciones: eliminación de palabras cerradas, truncamiento, etiquetado de partes de la oración y eliminación de errores morfológicos. A los patrones léxico sintácticos sólo se les aplica el etiquetado de partes de la oración.

4. Sistema de Recuperación de Información (SRI). Una vez que los documentos son trasladados a oraciones, los conceptos y las relaciones son preprocesados, utilizamos un SRI booleano AND [24] para asociar las oraciones a los conceptos y posteriormente estas oraciones asociadas se utilizan en la intersección de las oraciones que contienen todas las palabras de los dos conceptos, sin considerar en la intersección a la relación.

³<http://www.w3.org/TR/owl-features/>

⁴<http://nlp.lsi.upc.edu/freeling/>

3.2. Verificación de relaciones tipo class-inclusion

La técnica empleada para la identificación de relaciones tipo class-inclusion en el corpus de dominio, se propone en el algoritmo 1. Este algoritmo considera dos tipos de comportamiento en este tipo de relaciones:

1. Las palabras de uno de los conceptos están incluidos o subsumidos en el segundo concepto, como es el caso del ejemplo 1 y 2 de la tabla 2 [2].
2. Los conceptos son diferentes (ver ejemplos 3 y 4 de la tabla 2).

Algoritmo 1 Verificación de relaciones class-inclusion

Input: $Corpus = \{\{C_{1,1}|C_{1,2}|l_{1,1}|\dots|l_{1,k}|\dots\}, \dots\}$,

$R = \{R_1, R_2, \dots, R_i, \dots\}, R_i = (C_{i,1}, C_{i,2})$

Output: $CI = \{(w, R_1), (w, R_2), \dots\}$

for all r_i **in** R **do**

if $substrIzq(C_{i,1}, C_{i,2}) == 1$ **then**

$w = buscar(C_{i,1}, Corpus)$

else

if $substrDer(C_{i,1}, C_{i,2}) == 1$ **then**

$w = buscar(C_{i,1}, Corpus)$

else

$w = aplicaPLS(r_i, patrones, Corpus)$

end if

end if

end for

Los datos de entrada del algoritmo son los conceptos de cada relación R , el corpus de dominio $Corpus$, y los patrones léxico sintácticos trasladados a expresiones regulares, en el formato de FreeLing [28]. La salida es la lista de relaciones CI que fueron encontradas en el corpus, con un peso asociado w que indica si existe evidencia en el corpus para la relación ($w = 1$) o si se carece de evidencia ($w = 0$). En el algoritmo, la función $substrIzq(C_{i,1}, C_{i,2})$ determina si el concepto $C_{i,2}$ está incluido a la izquierda del concepto $C_{i,1}$, la función $substrDer(C_{i,1}, C_{i,2})$ determina si el concepto $C_{i,2}$ está incluido a la derecha del concepto $C_{i,1}$, la función $buscar$ transforma el concepto C_i

en una expresión regular y lo busca en las oraciones asociadas al mismo ($Corpus$) y la función $aplicaPLS$, transforma los patrones léxico sintácticos en expresiones regulares que se aplican a las oraciones de la relación R .

Considerando el ejemplo 3 de la tabla 2, el algoritmo aplicaría el patrón 42, sustituyendo $C_{3,1} = knowledge\ representation$ en X y el $C_{3,2} = tree$ en Y , la respuesta del algoritmo en este ejemplo es $\{(1, R_3)\}$. Si consideramos todos los ejemplos de la tabla 2, el algoritmo regresaría como respuesta $CI = \{(1, R_1), (1, R_2), (1, R_3), (0, R_4)\}$.

3.3. Verificación de relaciones ontológicas

Para la verificación de relaciones ontológicas se propone el algoritmo 2. Los datos de entrada son las líneas del corpus asociadas a los conceptos de dominio ($Corpus$), las relaciones ontológicas (R) y los conceptos (T_i) de las tripletas. El algoritmo realiza un análisis de dependencias [28] por cada oración asociada a los dos conceptos que forman la relación ($dependenciasFreeLing(c_i)$). Además, se construye una consulta ($consulta(t_i)$) con el formato " $Concepto_1\ Relacion\ Concepto_2$ " para corroborar si hay o no evidencia de la relación en la oración. Esta consulta se utiliza para recorrer el árbol de dependencias, haciendo una expansión de nodos ($analizaNT(q_i, d_i)$). La ruta creada debe llegar a nodos hoja considerados como nodos terminales que indicarían que hay evidencia de la relación ($w = 1$), de lo contrario, no existe evidencia ($w = 0$).

Por ejemplo, para verificar la veracidad de la relación 1 de la tabla 3, el algoritmo produce un árbol de dependencias de la oración⁵ (ver figura 4), construye una consulta "*intelligent agent may learn to achieve goal*" que utiliza para recorrer el árbol (ver figura 5). La función $analizaNT(q_1, d_1)$ identifica los niveles del árbol, el número de nodo (#), los nodos hijos, los nodos terminales ($\{0\}$) y los nodos padre para cada nodo del árbol (los cuales llamamos atributos), además revisa que la relación incluya por lo menos un verbo. La función asocia a cada palabra de la consulta q_1 sus atributos correspondientes y realiza una expansión de nodos (exp). Cada nodo expandido, debe producir

⁵<http://nlp.lsi.upc.edu/freeling/demo/demo.php>

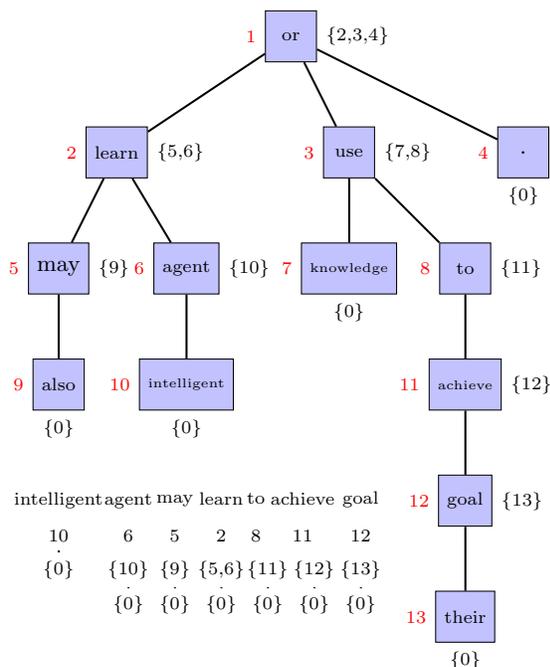


Fig. 5. Ejemplo de procesamiento del algoritmo 2

3.4. Evaluación

La medida de evaluación que se considera es la de exactitud (E), que considera el total de relaciones propuestas por el sistema ($S(R)$) con el formato (w_i, R_i) donde $1 \leq i \leq n$, y el total de relaciones de la ontología ($n = |R|$), donde R son las relaciones class-inclusion y/o ontológicas.

$$E = \frac{|S(R)|}{|R|} \tag{1}$$

Para validar los resultados obtenidos por el enfoque propuesto realizamos una evaluación por expertos de dominio.

Por medio del coeficiente kappa κ medimos la concordancia que existe entre los resultados del enfoque para cada tipo de relación, con respecto a las respuestas de los expertos [8]. El coeficiente kappa corresponde a la proporción de concordancia observada sobre el total de las observaciones, eliminando la concordancia al azar (ver ecuación 2). Los valores de kappa se encuentran en el rango de -1 a 1, si el valor de kappa es cercano a 1 indica un mayor grado de concordancia interobservada,

si kappa es cero indica una concordancia observada al azar [6].

$$\kappa = \frac{\Sigma \text{concordancia_observada} - \Sigma \text{concordancia_azar}}{\text{total_observaciones} - \Sigma \text{concordancia_azar}} \tag{2}$$

4. Resultados experimentales

En esta sección se presentan los datos utilizados y los resultados obtenidos en los experimentos.

4.1. Datos de prueba

Los dominios de las ontologías consideradas en los experimentos son: inteligencia artificial (AI) y estándar e-Learning SCORM (SCORM)⁶ [47]. Cada ontología contiene un número determinado de conceptos (C), relaciones tipo class-inclusion (S) y relaciones ontológicas (R). Los documentos (D) de los corpora de dominio fueron utilizados para determinar la cantidad de tokens (T), el vocabulario (V) y el número de oraciones filtradas (O) por el sistema de recuperación de información (ver tabla 4).

Tabla 4. Conjunto de datos

Dominio	Ontología			Corpora			
	C	S	R	D	T	V	O
AI	276	205	61	8	10,805	2,180	464
SCORM	1,461	1,038	759	36	32,644	2,154	1,632

4.2. Baseline

Con la finalidad de tener un valor de referencia sobre la evaluación de la ontología, hemos construido un valor base (baseline) sobre la validación de las relaciones ontológicas. El proceso propuesto consiste de validar todas aquellas relaciones ontológicas, cuyos conceptos se encuentren estrechamente co-relacionados en el corpus de referencia. En este sentido, si dos conceptos asociados a una relación ontológica aparecen ambos en el mismo contexto, asumimos que dichos conceptos

⁶Las ontologías y sus corpus correspondientes están disponibles en la página <http://azouaq.athabascau.ca/golds-standards.htm>

Tabla 5. Resultados de los patrones léxico sintácticos

No.	Patrón léxico sintáctico p	$fr(p, AI)$	$fr(p, SCORM)$
96	NP , is (NP,)* (or and the) NP	7	55
43	NP (is are) (a an) NP	5	24
92	NP (,)? such as (NP,)* (or and the)? NP	7	13
97	(NP,)* (or and the) (NP,)* is (a an the) NP	4	12
46	NP such as (NP,)* (or and) NP	4	7
42	NP (is are) NP	2	6
12	NP such as (NP,)* (or and the) NP	1	4
94	NP NP , is (a an the) NP	3	2
15	such NP as (NP,)*	0	1
50	NP (classify (in into) comprise contain compose (of)? group (in into) divide (in into) fall (in into) belong (to)) NP	0	1
86	NP (and or) (another other) NP	0	1
98	NP , including NP	1	1
104	NP as (NP,*) (or and the) NP	0	1
13	NP 's NP	1	0
106	NP, for example, is (a an the) NP	1	0

están relacionados. Está claro que dicha relación puede ser distinta a la que se establece en la ontología, sin embargo, esta medida la usamos únicamente como baseline.

Para medir el grado de co-relación, empleamos el concepto de información mutua, la cual se describe a continuación.

Dada una tripleta (S, R, O) , con R la relación ontológica entre los conceptos S y O , la información mutua entre los dos conceptos se mide como:

$$PMI(S, O) = \frac{Prob(S, O)}{Prob(S) * Prob(O)}$$

El corpus fue pre-procesado, eliminando signos de puntuación y considerando la versión truncada (usando el truncador de Porter) del corpus y de los conceptos ontológicos. Los resultados obtenidos se muestran en la Tabla 6.

4.3. Resultados obtenidos por el sistema

Como se mencionó anteriormente se hizo una evaluación manual por expertos, lo que nos permite determinar el desempeño de nuestro enfoque. Para determinar la estabilidad de los patrones léxico sintácticos, que encontraron evidencia de la

Tabla 6. Resultados baseline obtenidos mediante el coeficiente de co-relación de información mutua

Ontología	Tipo de relación	Subconjunto de Expertos	Corpus completo
AI	Class-Inclusion	55.61 %	23.90 %
	Ontológicas	47.54 %	16.39 %
SCORM	Class-Inclusion	14.00 %	25.53 %
	Ontológicas	41.27 %	38.34 %

relación class-inclusion en los corpora, utilizamos la medida del coeficiente de correlación tau (τ) de Kendall [19]. La cual determina el grado en el cual las dos listas de frecuencia $ft(p, corpus)$ (categorías) se igualan de acuerdo al orden descendente (ver tabla 5).

Los valores del coeficiente tau de Kendall se encuentran en el rango de -1 y 1, un alto valor indica un alto grado de acuerdo entre las dos categorías. Por lo tanto, si el acuerdo (desacuerdo) entre las dos categorías es perfecta, entonces el coeficiente tendrá el valor de 1 (-1). En caso de que el valor sea 0, indica que las categorías son completamente independientes.

En este caso, el resultado del coeficiente tau de Kendall es igual a 0.733, lo que significa que

existe un alto acuerdo en el orden obtenido en los dos corpus de referencia. Esto quiere decir que existe una consistencia en la aplicación de estos patrones, independientemente de si son aplicados en diferentes dominios. Este hecho es cierto, al menos para las dos ontologías utilizadas en los experimentos.

Ahora, presentamos los resultados obtenidos por el enfoque utilizando el criterio de exactitud (E) al evaluar las ontologías y la calidad (C) de la predicción del enfoque de acuerdo a tres expertos humanos (E_1 , E_2 y E_3) y el baseline.

El enfoque de evaluación de relaciones class-inclusion asigna una exactitud del 87% a la calidad de la ontología AI. Consideramos que la calidad obtenida (91%, 82% y 86%) para la relación class-inclusion es un buen resultado al compararlo también con el baseline, sin embargo, consideramos que es necesario investigar las razones por las cuales no encontramos el porcentaje restante. En el caso de las relaciones ontológicas, se presenta un comportamiento similar, logrando un 82%, 84% y 80% de calidad para la ontología AI y un 82% de exactitud para el enfoque propuesto (ver tabla 7). Logrando aproximadamente un 30% más que el baseline.

Tabla 7. Exactitud de la ontología AI y la calidad de las predicciones de los enfoques

Tipo de relación	E	$C(E_1)$	$C(E_2)$	$C(E_3)$
Class-inclusion	0.87	0.91	0.82	0.86
-- baseline	0.56	0.57	0.51	0.55
Ontológicas	0.82	0.82	0.84	0.80
-- baseline	0.48	0.51	0.46	0.52

En el caso de la ontología SCORM, el enfoque obtuvo el 59% de exactitud para las relaciones de tipo class-inclusion y los expertos asignaron un valor menor al 80% de calidad, superando al baseline en aproximadamente un 35% (ver tabla 8). De acuerdo a los expertos, el resultado que se obtiene para la ontología SCORM en este tipo de relación es menos confiable que el que se obtuvo para la ontología AI. A pesar de esto, consideramos que el enfoque es capaz de dar un valor de exactitud que puede dar un indicio de la calidad de la ontología. En el caso de las relaciones ontológicas el enfoque obtuvo un 86% de exactitud

y los expertos asignaron el 83%, 87% y el 80% de calidad a la predicción del enfoque resultados mayores al baseline.

Tabla 8. Exactitud para la ontología SCORM y la calidad de las predicciones de los enfoques

Tipo de relación	E	$C(E_1)$	$C(E_2)$	$C(E_3)$
Class-inclusion	0.59	0.78	0.69	0.72
-- baseline	0.14	0.42	0.30	0.45
Ontológicas	0.86	0.83	0.87	0.80
-- baseline	0.41	0.47	0.43	0.43

Los resultados del grado de concordancia que existe entre el sistema y los expertos de dominio para la ontología AI, en unos casos es leve (rango de 0.01 a 0.20) y en otros son aceptables (rango de 0.21 a 0.40), de acuerdo a los rangos de valoración del coeficiente kappa presentados en [21], [6] (ver tabla 9).

Tabla 9. Resultados coeficiente kappa para la ontología AI

Ontología	Tipo de relación	Coeficiente kappa		
		E_1	E_2	E_3
AI	Class-inclusion	0.51	0.18	0.09
	Ontológicas	0.32	0.30	0.33

Los resultados del grado de concordancia que existe entre el sistema y los expertos de dominio para la ontología SCORM es leve (rango de 0.01 a 0.20) para un caso de las relaciones ontológicas, pero para las relaciones class-inclusion son aceptables (rango de 0.21 a 0.40) considerando los rangos de valoración del coeficiente kappa presentados en [21], [6] (ver tabla 10).

Tabla 10. Resultados coeficiente kappa para la ontología SCORM

Ontología	Tipo de relación	Coeficiente kappa		
		E_1	E_2	E_3
SCORM	Class-inclusion	0.54	0.23	0.36
	Ontológicas	0.24	0.06	0.26

Los resultados presentados anteriormente fueron obtenidos con ejemplos de las relaciones, debido al gran esfuerzo que se necesita para evaluar manualmente su validez. Para la ontología AI

usamos 266 relaciones, mientras que la ontología SCORM fue evaluada con 285 relaciones. Por lo tanto, para tener una evaluación completa de las dos ontologías, calculamos su exactitud.

El enfoque al verificar las 266 relaciones semánticas de la ontología AI (*TotalRel*), encontró 231 relaciones válidas (*RelEnc*) que corresponde al 86.84 % de exactitud (ver tabla 11). Resultados superiores al baseline (ver columna 4 de la tabla 6).

Tabla 11. Resultados experimentales para la ontología AI

Tipo de relación	<i>TotalRel</i>	<i>RelEnc</i>	<i>Exactitud</i>
Class-inclusion	205	181	88.29 %
Ontológicas	61	50	81.97 %
<i>Total</i>	266	231	86.84 %

El resultado de la medida de exactitud para ontología e-Learning SCORM es del 76.52 % (ver tabla 12). Se observa que la exactitud disminuye con respecto a los resultados obtenidos en la ontología AI, pero la cantidad de relaciones evaluadas por el enfoque (1797 relaciones) es mayor que la cantidad de relaciones evaluadas para la ontología AI (266 relaciones). Resultados mayores al baseline (ver columna 4 de la tabla 6).

Tabla 12. Resultados experimentales para la ontología SCORM

Tipo de relación	<i>TotalRel</i>	<i>RelEnc</i>	<i>Exactitud</i>
Class-inclusion	1038	731	70.42 %
Ontológicas	759	644	84.84 %
<i>Total</i>	1797	1375	76.52 %

Como se puede apreciar, el enfoque obtiene una mejor exactitud para la ontología AI. Este resultado se debe a que el sistema tiene un número mayor de oraciones asociadas a cada relación, por lo que, tiene más oportunidad de encontrar evidencia de la validez de la misma en el corpus. La exactitud de la ontología SCORM fue significativamente mejor, porque, en este caso se evaluó una gran cantidad de relaciones 1797 comparada con las 285 usadas en la evaluación por expertos. Por otro lado, en este último experimento, usamos un número mayor de oraciones que proporcionan la oportunidad de

encontrar evidencia de la relación en el corpus de referencia.

5. Conclusiones

En este artículo se presenta un enfoque para la evaluación de relaciones class-inclusion y ontológicas en ontologías de dominio, inteligencia artificial y estándar e-Learning SCORM, considerando el criterio de exactitud. El enfoque propuesto evalúa dichas relaciones, basándose en la evidencia de la relación en el corpus de referencia por medio de patrones léxico sintácticos y el análisis de dependencias sintácticas. Los experimentos muestran que existe un alto acuerdo en la frecuencia de ocurrencia de los patrones utilizados en la evaluación, esto independientemente del corpus de dominio.

Para evaluar la calidad del enfoque propuesto se hizo una validación por expertos humanos y por el uso de un baseline.

Al comparar los resultados obtenidos por el enfoque con respecto al baseline se observa que los resultados de nuestra propuesta superan en aproximadamente un 30 % al baseline en el caso de la ontología AI y en un 35 % en el caso de la ontología SCORM lo que nos indica que el enfoque tiene un comportamiento favorable para estas dos ontologías. En el caso de la evaluación por expertos se midió el grado de acuerdo entre ellos y el enfoque fue aceptable de acuerdo a los rangos de valoración del coeficiente kappa. El enfoque asigna un 76.52 % de exactitud para la ontología SCORM y un 86.84 % de exactitud para la ontología AI, esto de alguna forma refleja la calidad de cada ontología.

Si bien, el enfoque considera para la evaluación relaciones tipo class-inclusion pero estas dos ontologías sólo consideran relaciones taxonómicas. Como trabajo a futuro, se planea evaluar la confiabilidad del enfoque propuesto considerando una cantidad mayor de relaciones. Además de considerar otras ontologías para el proceso de evaluación de la calidad de las mismas en sus correspondientes corpus de dominio.

Agradecimientos

Los autores agradecen el apoyo otorgado por la Benemérita Universidad Autónoma de Puebla, México y al Centro Nacional de Investigación y Desarrollo Tecnológico, Campus Cuernavaca, México para la realización de este trabajo de investigación, el cual ha sido parcialmente financiado por el Consejo Nacional de Ciencia y Tecnología (CONACYT) con el número de becario 54371, por el Programa para el Mejoramiento del Profesorado (PROMEP) con folio BUAP-792 y número de convenio PROMEP/103.5/12/4962, y a través del proyecto CONACYT 106625.

Referencias

1. **Bejar, I., Chaffin, R., & Embretson, S. (1991).** *Cognitive and Psychometric Analysis of Analogical Problem Solving*. Recent Research in Psychology Series. Springer London, Limited.
2. **Bhatt, B. & Bhattacharyya, P. (2012).** Domain specific ontology extractor for indian languages. *Proceedings of the 10th Workshop on Asian Language Resources*, The COLING 2012 Organizing Committee, Mumbai, India, pp. 75–84.
3. **Brank, J., Grobelnik, M., & Mladenić, D. (2005).** A survey of ontology evaluation techniques. *Proc. of 8th Int. multi-conf. Information Society*, pp. 166–169.
4. **Brewster, C., Alani, H., Dasmahapatra, S., & Wilks, Y. (2004).** Data driven ontology evaluation. *Proceedings of International Conference on Language Resources and Evaluation*, pp. .
5. **Cantador, I., Fernández, M., & Castells, P. (2006).** A collaborative recommendation framework for ontology evaluation and reuse. *Actas de International Workshop on Recommender Systems, en la 17th European Conference on Artificial Intelligence (ECAI 2006), Riva del Garda, Italia*, pp. 67–71.
6. **Cerda L, J. & Villarroel Del P, L. (2008).** Evaluación de la concordancia inter-observador en investigación pediátrica: Coeficiente de Kappa. *Revista chilena de pediatría*, Vol. 79, pp. 54 – 58.
7. **Ciaramita, M., Gangemi, A., Ratsch, E., Saric, J., & Rojas, I. (2005).** Unsupervised learning of semantic relations between concepts of a molecular biology ontology. *IJCAI*, Professional Book Center, pp. 659–664.
8. **Cohen, J. (1960).** A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, Vol. 20, No. 1, pp. 37–46.
9. **de Cea, G. A., de Mon, I. A., & Montiel-Ponsoda, E. (2009).** From linguistic patterns to ontology structures. *8th International Conference on Terminology and Artificial Intelligence*,.
10. **Dellschaft, K. & Staab, S. (2008).** Strategies for the evaluation of ontology learning. **Buitelaar, P. & Cimiano, P.**, editors, *Bridging the Gap between Text and Knowledge Selected Contributions to Ontology Learning and Population from Text*, IOS Press, Amsterdam.
11. **Gómez-Pérez, A. (2004).** *Ontology Evaluation*. International Handbooks on Information Systems. Springer.
12. **Grigonyté, G. (2010).** *Building and Evaluating Domain Ontologies: NLP Contributions*. Logos-Verlag.
13. **Gruber, T. R. (1993).** Towards Principles for the Design of Ontologies Used for Knowledge Sharing. **Guarino, N. & Poli, R.**, editors, *Formal Ontology in Conceptual Analysis and Knowledge Representation*, Kluwer Academic Publishers, Deventer, The Netherlands.
14. **Hearst, M. A. (1992).** Automatic acquisition of hyponyms from large text corpora. *Proceedings of the 14th International Conference on Computational Linguistics*, pp. 539–545.
15. **Jacobs, V. (2006).** *Using the semantics of prepositions for ontology learning*. Master's thesis, Utrecht University, the Netherlands.
16. **Jimenez Muñoz, R. J. (2013).** *Un sistema de búsqueda semántica de información para su uso en el dominio de recuperación mejorada en yacimientos petroleros*. Master's thesis, Fac. Ciencias de la Computación, BUAP, Puebla, Mex.
17. **Jurgens, D., Mohammad, S., Turney, P., & Holyoak, K. (2012).** Semeval-2012 task 2: Measuring degrees of relational similarity. **SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*, Association for Computational Linguistics, Montréal, Canada, pp. 356–364.
18. **Kavalec, M., Maedche, A., & Svátek, V. (2004).** Discovery of lexical entries for non-taxonomic relations in ontology learning. **van Emde Boas, P.**,

- Pokorný, J., Bieliková, M., & Stuller, J.**, editors, *SOFSEM*, volume 2932 of *Lecture Notes in Computer Science*, Springer, pp. 249–256.
19. **Kendall, M. G. (1938)**. A new measure of rank correlation. *Biometrika*, Vol. 30, No. 1/2, pp. 81–93.
 20. **Klaussner, C. & Zhekova, D. (2011)**. Lexico-syntactic patterns for automatic ontology building. *Proceedings of the Second Student Research Workshop associated with RANLP 2011*, RANLP 2011 Organising Committee, Hissar, Bulgaria, pp. 109–114.
 21. **Landis, J. R. & Koch, G. G. (1977)**. The measurement of observer agreement for categorical data. *Biometrics*, Vol. 33, No. 1, pp. 159–174.
 22. **Lovrencic, S. & Mirko, C. (2008)**. Ontology evaluation - comprising verification and validation. *Proceedings of Central European Conference on Information and Intelligent Systems, CECIIS - 2008*.
 23. **Maedche, A. & Staab, S. (2002)**. Measuring similarity between ontologies. *Proceedings of European Knowledge Acquisition Workshop (EKAW)*, pp. .
 24. **Manning, C., Raghavan, P., & Schütze, H. (2008)**. *Introduction to Information Retrieval*. An Introduction to Information Retrieval. Cambridge University Press.
 25. **Maynard, D., Funk, A., & Peters, W. (2009)**. Sprat: a tool for automatic semantic pattern-based ontology population. *International Conference for Digital Libraries and the Semantic Web*.
 26. **Mititelu, V. B. (2011)**. Hyponymy patterns in romanian. *Memoirs of the Scientific Sections of the Romanian Academy*, Vol. XXXIV, pp. 31–40.
 27. **Montiel-Ponsoda, E. & Aguado de Cea, G. (2008)**. Using natural language patterns for the development of ontologies. *Researching specialized languages*, pp. 332–345.
 28. **Padró, L. & Stanilovsky, E. (2012)**. Freeling 3.0: Towards wider multilinguality. *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)*, ELRA, Istanbul, Turkey.
 29. **Pak, J. & Zhou, L. (2009)**. A framework for ontology evaluation. **Sharman, R., Rao, H. R., & Raghu, T. S.**, editors, *WEB*, volume 52 of *Lecture Notes in Business Information Processing*, Springer, pp. 10–18.
 30. **Paslaru, E. (2005)**. Using context information to improve ontology reuse. *Doctoral Workshop at the 17th Conference on Advanced Information Systems Engineering CAiSE05*.
 31. **Porter, M. F. (1997)**. Readings in information retrieval. chapter An Algorithm for Suffix Stripping. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 313–316.
 32. **Reyes Ortiz, J. A. (2013)**. *Creación automática de Ontologías a partir de Textos con un Enfoque Lingüístico*. Ph.D. thesis, Dept Ciencias Computacionales, Cenidet, Cuernavaca, Mor., Mex.
 33. **Rios-Alvarado, A. B., López-Arévalo, I., & Sosa, V. J. S. (2013)**. Learning concept hierarchies from textual resources for ontologies construction. *Expert Syst. Appl.*, Vol. 40, No. 15, pp. 5907–5915.
 34. **Ruiz, J. L. J. (2001)**. *Iniciación a la Lingüística*. Editorial Club Universitario.
 35. **Ruiz-Casado, M., Alfonseca, E., & Castells, P. (2005)**. Automatic extraction of semantic relationships for wordnet by means of pattern learning from wikipedia. *NLDB*, Springer Verlag, pp. 67–79.
 36. **Sabou, M., Lopez, V., Motta, E., & Uren, V. (2006)**. Ontology selection: Ontology evaluation on the real semantic web. *Proceedings The 4th International EON Workshop, Evaluation of Ontologies for the Web*.
 37. **Saint-Dizier, P. & Viegas, E. (1995)**. *Computational Lexical Semantics*. Studies in Natural Language Processing. Cambridge University Press.
 38. **Salem, S. & AbdelRahman, S. (2010)**. A multiple-domain ontology builder. *Proceedings of the 23rd International Conference on Computational Linguistics*, Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 967–975.
 39. **Schutz, A. & Buitelaar, P. (2005)**. Relext: A tool for relation extraction from text in ontology extension. *Proceedings of the 4th International Semantic Web Conference (ISWC)*, pp. 1–5.
 40. **Stevenson, M. & Greenwood, M. A. (2005)**. A semantic approach to ie pattern induction. *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, ACL '05, Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 379–386.
 41. **Stevenson, M. & Greenwood, M. A. (2006)**. Learning Information Extraction Patterns Using WordNet. *Proceedings of the 5th Intl. Conf. on Language Resources and Evaluations (LREC)*, pp. 95–102.
 42. **Tegos, A., Karkaletsis, V., & Potamianos, A. (2008)**. Learning of semantic relations between ontology concepts using statistical techniques. *High-level Information Extraction Workshop 2008 (HLIE08)*, ECML-PKDD.

43. **Tovar, M., Montes, A., & Pinto, D. (2013).** Methodology for automatic evaluation of restricted domain ontologies. *Research in Computing Science, Special Issue: Advances in Pattern Recognition*, Vol. 61, pp. 63–72.
44. **Volkova, S., Caragea, D., Hsu, W., Drouhard, J., & Fowles, L. (2010).** Boosting biomedical entity extraction by using syntactic patterns for semantic relation discovery. *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*, volume 1, pp. 272–278.
45. **Zarrouk, M., Lafourcade, M., & Joubert, A. (2013).** Inference and reconciliation in a crowdsourced lexical-semantic network. *Computación y Sistemas*, Vol. 17, No. 2, pp. 147–159.
46. **Zavitsanos, E., Paliouras, G., & Vouros, G. A. (2011).** Gold standard evaluation of ontology learning methods through ontology transformation and alignment. *IEEE Trans. Knowl. Data Eng.*, Vol. 23, No. 11, pp. 1635–1648.
47. **Zouaq, A., Gasevic, D., & Hatala, M. (2012).** Linguistic patterns for information extraction in ontomaps. **Blomqvist, E., Gangemi, A., Hammar, K., & del Carmen Suárez-Figueroa, M., editors, WOP**, volume 929 of *CEUR Workshop Proceedings*, CEUR-WS.org.

Mireya Tovar Vidal es Maestra en Ciencias por el Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (2002). Actualmente estudiante de doctorado en el Centro Nacional de Investigación y Desarrollo Tecnológico. También es profesora investigadora de tiempo completo en la Facultad de Ciencias de la Computación de la Benemérita Universidad Autónoma de Puebla. Sus áreas de interés son: ontologías, Procesamiento de Lenguaje Natural, paralelismo y lógica.

Azucena Montes Rendón es Doctora en Ciencias por la Université Paris-Sorbonne, Francia (2002).

Profesora Investigadora de tiempo completo en el Centro Nacional de Investigación y Desarrollo Tecnológico de 2002 a 2012. Actualmente, se encuentra en la Universidad Nacional Autónoma de México en el grupo de Ingeniería Lingüística. Sus áreas de interés son: Semántica cognitiva, representación del conocimiento y lingüística computacional.

David E. Pinto es Doctor en Informática por la Universidad Politécnica de Valencia, España (2008). Actualmente es profesor investigador de tiempo completo en la Facultad de Ciencias de la Computación en la Benemérita Universidad Autónoma de Puebla, sus áreas de interés son: Procesamiento del Lenguaje Natural, recuperación de información y ontologías.

Gabriel González Serna es Doctor en Ciencias Computacionales por el Centro de Investigación en Computación en el Instituto Politécnico Nacional, Mexico, D.F. (2006). Actualmente, es profesor investigador de tiempo completo del Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET). Sus áreas de interés son: Modelado semántico, computación consciente del contexto, servicios de recomendación sensibles al contexto, cómputo ubicuo y HCI.

Darnes Vilariño Ayala es Doctora en Matemáticas en el área de optimización en la Universidad de la Habana Cuba (1997). Actualmente es profesora investigadora de tiempo completo en la Facultad de Ciencias de la Computación de la Benemérita Universidad Autónoma de Puebla. Sus áreas de interés son: Inteligencia Artificial, optimización y Procesamiento del Lenguaje Natural.

*Article received on 20/03/2014, accepted on 18/09/2014.
Corresponding author is Mireya Tovar.*