

A metric for the Evaluation of Restricted Domain Ontologies

Mireya Tovar¹, David Pinto¹, Azucena Montes², Gabriel González-Serna³

¹ Benemérita Universidad Autónoma de Puebla,
Facultad de Ciencias de la Computación, Puebla,
Mexico

² Instituto Tecnológico de Tlalpan, TecNM, Mexico City,
Mexico

³ Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET), Cuernavaca,
Mexico

{mtovar, dpinto}@cs.buap.mx, {amr, gabriel}@cenidet.edu.mx

Abstract. In this article we propose a metric for the automatic evaluation of restricted domain ontologies. The metric is defined in terms of the evaluation of different lexico-syntactic, statistical and semantic approaches. A syntactic approach employed is the use of lexical syntactic patterns, other approaches as grouping by formal concept analysis, similarity, latent semantic analysis and dependence graphs are used as well. These approaches focus on reference corpora to find evidence of the validity of concepts and semantic relationships stored in the target ontology. The proposed evaluation approach is able to provide a score obtained through the metric, which is based on the accuracy measure used for each ontology evaluated. The score is associated in some way with the ontology quality. This score is given with a certain degree of reliability, and it is obtained by comparing the results given against the evaluation of human experts and a baseline.

Keywords. Ontology evaluation, formal concept analysis, syntactic patterns.

1 Introduction

In recent years, the amount of information produced in the Web and in local repositories has increased significantly, therefore analyzing, categorizing and retrieving information it has becoming a much more difficult task, in particular if we do not take into account the semantics of each document.

Since most of these information sources appear in an unstructured or semi-structured manner, it is necessary to process that from a semantic point of view. Ontologies play an important role in the semantic Web as they are resources that allow to capture the explicit knowledge in data through concepts and relationships; giving users and computers the opportunity to understand the data exchanged.

An ontology is defined as “an explicit and formal specification of a shared conceptualization” [12]. Normally, this type of semantic resource is made up of concepts or classes, relationships, instances, attributes, axioms, restrictions, rules and events. Domain ontologies are a type of knowledge representation that it is possible to organize in taxonomic and ontological structures of concepts for some area or domain of specific knowledge.

Automatic learning or generation of ontologies, is a process that can support the automatic or semi-automatic construction of ontologies for the knowledge engineer. Nowadays, there exist diverse computational systems for automatic generation of ontologies, nevertheless, in most of the cases they lack of an automatic components evaluation process, and in consequence, the quality of these semantic resources is normally unknown.

The evaluation of ontologies is a task that consists on measuring the quality of these semantic resources. The final aim of the ontology evaluation task is to facilitate the work of the knowledge engineer or the domain expert by checking the quality of the ontology.

This objective is helpful because when the ontology has a considerable size, this task becomes to be very complex in time (hours-person). The process of evaluation is far to be trivial, since it is necessary to choose the elements of the ontology that should be considered for measuring its quality, as well as the specific criteria to meet.

In this paper we assume that a reference corpus semantically associated with the domain ontology exists. The aim is “to evaluate” the quality of the relationships and concepts of the ontology using human experts and a baseline for such purpose. The obtained results are thereafter used by an integration metric that issues a quantitative result for the target ontology.

The main contributions of this paper are the following ones: 1) the evaluation of relation class-inclusion present in the ontology and in the corpus of domain, by means of lexical and syntactic patterns, latent semantic analysis, formal analysis of concepts, and similarity; 2) the evaluation of ontological or non-taxonomical relations by means of grammatical analysis, semantic latent analysis, formal concept analysis and similarity; and 3) a metric for the evaluation of the ontology quality.

The remaining of this paper is organized as follows. In Section 2, some works related to the evaluation of ontologies are shown. In Section 3 a methodology for the evaluation of restricted domain ontologies is presented. In the Section 4 the obtained results are shown. Finally, some conclusions are drawn in Section 5.

2 Related Work

Ontology evaluation approaches are normally classified in literature as presented in [3].

1. **Evaluation developed by human beings following criteria, standards and requirements:** there are certain characteristics or criteria defined that allow to evaluate the ontology providing a numerical score or ranking [27]. Some of the following characteristics are considered: completeness, correctness, legibility and flexibility [5, 10]. Other criteria for the evaluation of the content are developed manually by experts of domain, as the followings: consistency, completeness, concision, expansion capacity, sensibility [18].
2. **Evaluation based on an implementation or a task:** it consists of proving the performance of the ontology in an implementation, i.e. it tries to measure how much the ontology helps to improve the results of certain task, evaluating in some way how functional is that ontology in real applications. For example, to answer questions of the user using an ontology [25, 31], or using the ontology to improve the performance of a semantic search engine to recover relevant documents [14].
3. **Evaluation based on gold standard**, the quality of the ontology is expressed by the similarity that exists between the one constructed automatically a another ontology built manually (known as the gold standard ontology) [25, 29, 30, 39, 19, 8].
4. **Evaluation based on a reference corpus:** in this case, the quality of the ontology is represented by the opportunity it has to cover the topic of a corpus, as the criterion of completeness [11]. The evaluation approach focuses in the functional dimension of an ontology, which is compared with the content of a corpus of texts that are representative for the domain. The content of the corpus is analyzed by using natural language techniques in order to identify terms and semantic relations. In [4], a probabilistic approach is used to compare the concepts of an ontology with a set of important terms identified in the body of a reference text (extended by adding two levels of hypernyms from WordNet). The purpose is to detect in a

set of five ontologies, which adapts best to the artist domain.

In this article we are interested in the last type of evaluation for ontologies, that means, the evaluation based on a reference corpus considering, in this case, the criterion of accuracy.

3 Methodology for the Evaluation of Ontologies

In this section a methodology for the automatic evaluation of domain ontologies is presented. We use the evaluation approach based on reference corpus and the correctness criterion. The different phases that integrate the methodology are enumerated and thereafter each phase are explained into detail so that they can be further easily reproduced.

We assume as initial condition, that the restricted domain ontology has being structurally well designed, and that the reference corpus corresponding to the ontology domain exist and it is available for the evaluation process. The term “structurally well designed”, refers to the fact that the ontology must have not any syntactic errors, neither of design nor of inconsistency. Additionally, the ontology must avoid any redundancy in concepts, since the evaluation task is carried out when the ontology is already finished and not during the process of design or building.

The corpus can belong to any specific text domain, in other words, scientific publications, reports of projects, books, medical notes, etc. The most important thing of the corpus is that it has to be balanced for the task it will be required, that is to say, it is necessary to ensure that the texts of the input corpus should be diverse and that they should correspond to the domain; besides to have a reasonable amount of texts [9].

The methodology considers the following phases:

1. **Automatic preprocessing of information.** In this phase, the concepts and ontological relationships are extracted from the ontology. The documents or sentences of the domain corpus of reference are filtered according to the ontological concepts and relationships. For this purpose, we use an information retrieval system which allow us to improve the quality in the different approaches by having mainly relevant information of the domain corpus.
2. **Automatic discovery of candidate terms and/or ontological relationships.** In this phase, the approaches employed for the discovery of concepts and ontological relationships in the domain corpus are implemented. Some approaches are: lexical-syntactic patterns, formal concept analysis, similarity, dependency analysis and latent semantic analysis. The purpose of this phase is to find evidence of the ontological relationships and concepts in the reference corpus.
3. **Evaluation of the ontology.** In this phase, we propose metrics for evaluating the domain ontology, thus offering a way to measure the quality of the target semantic resource.

In the following sections we describe every phase that integrates the methodology for the automatic evaluation of domain ontologies, as well as the approaches of discovery designed in the second phase. The metrics proposed for the evaluation of the ontology are also described.

3.1 Automatic Preprocessing of Information

The reference domain corpus is made up by unstructured documents (raw texts) of scientific domain written in natural language. It is necessary, in this phase, to use the levels of the processing of the natural language and a system to retrieve the information, to improve the quality of the information of the reference corpus and thus to obtain a better performance in the results of the approaches of discovery proposed in the following phase.

In the phase of automatic preprocessing of information, the following actions are performed:

1. Ontology preprocessing:

- (a) Extracting concepts and relationships (triplets) of the ontology. In this case, Jena¹ is used for extracting the concepts and relationships of the domain ontology, which are expressed in the OWL format². The properties of the ontology used to extract the relationship are: subClassOf and objectProperty. The SRO triplets are composed by two concepts: Subject (S) and object (O), as well as a relationship (R).
- (b) Building of queries from the triplets. In this stage two types of queries are built: the first type contains the words that integrate the concept, the second type contains the words of both concepts that are part of the relationship without considering the terms of the same one.
- (c) The following operations are applied to the queries: Removing of stop words, truncation, tagging of parts of speech, and omission of morphologic errors.

2. Data preprocessing.

- (a) Removing of special symbols and/or not printable characters, stop words like: prepositions, articles, etc.
- (b) Splitting of the corpus in sentences, considering those sentences separated by a point mark.
- (c) Removing of punctuation symbols.
- (d) Application of the algorithm of Porter stemming with the purpose of grouping together sentences that contain the same concept, but written in some of its morphologic variants [28]. For example, the concept “subfields of artificial intelligence”, belonging to the domain of artificial intelligence, could appear in the corpus as: “subfield of artificial intelligence” or “subfields of Artificial Intelligence”. The query built for

this concept would be: “subfield artificial intelligence” that is the result of omitting stop words and applying stemming.

- (e) Applicating tagging of parts of speech using: FreeLing³ [26], TreeTagger⁴ [32], and analyzers of syntactic dependencies as: minipar⁵ [17] and the Stanford POS tagger⁶ [7].
- (f) Removing morphologic errors produced by the PoS tagger.

3. Construction of the filtered corpus. The Boolean Information Retrieval System (SRIB) uses queries obtained from the target ontology and the preprocessed reference corpus to construct a subcorpus with relevant information to the concepts of the ontology. The logical operator used in the information retrieval system is AND [20].

3.2 Automatic Discovery of Candidate Terms and/or Ontological Relationships

In this phase of the methodology, the information from the domain corpus is used to discover new ontological relationships (class-inclusion and non-taxonomic), extracted from the ontology. Thus, in this section there are several approaches that allow us to find evidence of the relationship and/or concept in the reference domain corpus.

Each approach assigns a weight w to the relationship. If the approach finds evidence of the relationship ($w = 1$), otherwise it assigns the weight of zero ($w = 0$).

The first approach uses lexical-syntactic patterns (LSP), that according to the state of the art research in the field, allow to identify taxonomic relationships (class-inclusion) in the corpus. The second approach uses the analysis of syntactic dependencies to discover non-taxonomic relationships. The third approach uses formal concept analysis (FCA), for class-inclusion relationships and it has been extended for non-taxonomic relationships. The fourth approach uses the latent

³<http://nlp.lsi.upc.edu/freeling/>

⁴<http://www.cis.uni-muenchen.de/schmid/tools/TreeTagger/>

⁵<http://webdocs.cs.ualberta.ca/lindek/minipar.htm>

⁶<http://nlp.stanford.edu/software/lex-parser.shtml>

¹<http://jena.apache.org/>
²<http://www.w3.org/TR/owl-features/>

semantic analysis (LSA), to identify class-inclusion and non-taxonomic relationships.

Finally, similarity measures are used to measure the correlation between the concepts that form the class-inclusion and non-taxonomic relation. These approaches are presented hereafter.

3.2.1 Discovery Approach based on the Use of Lexical-Syntactic Patterns

In order to indentify taxonomic relationships in the corpus, 107 lexical-syntactic patterns were obtained from the literature used to define taxonomical, functional, singular collective, plural collective and individual relations, which altogether allow to discover *class-inclusion* relationships [33, 2, 6, 13, 15, 22, 23, 24, 38, 16].

Therefore, in this research the discovery of class-inclusion relationships was carried out since they include the taxonomical relationships [1]. The approach based on lexical-syntactic patterns considers two types of behavior in the concepts for this type of relationship:

1. The words of one of the concepts are included or subsumed in the second concept, as in the case of example 1 and 2, of Table 1 [2].
2. The concepts are different (see examples 3 and 4, of Table 1).

The operations that are carried out in this approach are listed as follows:

1. Preprocessing of the subcorpus, the concepts and class-inclusion relations (Results obtained from phase 1).
2. Preprocessing applied to lexical-syntactic patterns.
3. Elimination of tagging morphological errors in the subcorpus.
4. Construction of regular expressions from the lexical-syntactic patterns.

5. Evaluation of class-inclusion relationships (see Algorithm 1). The algorithm receives the subcorpus, the set of class-inclusion relations and lexical-syntactic patterns. The algorithm assigns to w the value of 1 ($w = 1$), if there is evidence of the relation in the subcorpus using this approach. The algorithm considers the two types of behavior mentioned before. In the first case, if the concept is included in the second one, then $w = 1$, otherwise it is verified if there is agreement between both concepts, if so, $w = 1$, (see examples 5 and 6, from Table 1). Otherwise it is determined if there is a lexical-syntactic pattern that joins both concepts, then $w = 1$ otherwise $w = 0$.
6. In the case of the concepts, only are considered as valid those ones that are associated with the triplet whose weight w is equal to 1.

Algorithm 1 Evaluation of class-inclusion relationships using LSP

Input: $Subcorpus = \{\{C_{1,1}|C_{1,2}|l_{1,1}|...|l_{1,k}|...\}, \dots\}$,
 $R = \{R_1, R_2, \dots, R_i, \dots\}$, $R_i = (C_{i,1}, C_{i,2})$,
Output: $CI = \{(w, R_1), (w, R_2), \dots\}$,

```

for all  $r_i$  in  $R$  do
  if  $substrLeft(C_{i,1}, C_{i,2}) == 1$  then
     $w = search(C_{i,1}, Corpus)$ ,
  else
    if  $substrRight(C_{i,1}, C_{i,2}) == 1$  then
       $w = search(C_{i,1}, Corpus)$ ,
      if  $w == 0$  then
         $w = concordance(C_{i,1}, C_{i,2}, Corpus)$ ,
      end if
    else
       $w = applyLSP(r_i, patterns, Corpus)$ ,
    end if
  end if
end for

```

3.2.2 Discovery Approach based on the Analysis of Syntactical Dependencies

In the approach based on the analysis of syntactical dependencies (ASD), the triple SRO (concept-relation -concept), is considered an argumental

Table 1. Examples of class-inclusion relationships in the artificial intelligence domain

No.	Concept ₁	Concept ₂	Sentence
1	human natural language	language	Natural language processing (NLP) is a field of computer science and linguistics concerned with the interactions between computers and human natural languages.
2	problems of ai	problem	The central problems of AI include such traits as reasoning, knowledge, planning, learning, communication, perception and the ability to move and manipulate objects.
3	knowledge representation	tree	Other knowledge representations are trees, graphs and hypergraphs, by means of which the connections among fundamental concepts and derivative concepts can be shown.
4	kr	data structure	Reminder a KR is not a data structure.
5	import package	package	This will enable SCORM conformant systems to <i>import and export packages</i> that can be used by other SCORM conformant systems.
6	redundant knowledge	knowledge	By being consistent, the KR can eliminate <i>redundant or conflicting knowledge</i> .

structure where the concepts (arguments), link a situation (relation), that requires both to be valid in the corpus at a semantic level, unlike the syntactic level where the arguments are not necessary. In order to prove what was mentioned above, we use a structure of dependencies to describe the syntactic structure in the sentence (by a dependency graph), associated to the triplet, identifying the dependency between the words of the concepts and the semantic relationship.

The approach uses the graph of dependencies that the parser (freeLing), generates with the associated sentences to the non-taxonomic relation. It is verified that the triplet (query), forms a substructure of dependencies between the concepts and the relationships contained in the sentence dependency tree; if so $w = 1$, otherwise $w = 0$, (for further information see [36]). In the case of the concepts, the approach considers as valid those whose answer of the approach is $w = 1$.

3.2.3 Approach based on Similarity

Another of the approaches used in the present methodology is the one based on similarity, considering several measures of similarity to determine the correlation exists between a pair of related concepts (triple). The similarity measure is used for generalization under the assumption that words semantically similar operate similarly

[21]. In case we need to calculate the similarity of two words, these are represented as vectors in a multi-dimensional space. The procedure followed in this approach is the following:

1. Preprocessing. The filtered corpus obtained by using an information retrieval system and queries associated to the ontological concepts is preprocessed by removing punctuation symbols, stop words and by applying the Porter stemmer, resulting in a set of types or vocabulary.
2. Vectorial representation. We use the frequency of types associated to each ontological concept as attribute of a vector for representing the information (concept).
3. Similarity measure. The similarity value is determinated for each pair of vectors (concepts). The similarity measures considered in this phase are shown in Table 2, which has been extended to n dimensions.
4. Degree of similarity. At this stage, several criteria are considered for assigning a weight w to each triple:
 - (a) If the cosine similarity value for the pair of concepts exceeds 0.40, the relationship takes the weight $w = 1$, otherwise $w = 0$ (Sim-cos).

- (b) If the cosine similarity value for the pair of concepts exceeds a threshold (see note), the relationship takes the weight $w = 1$, otherwise $w = 0$ (Sim-cos.u).
- (c) If the average similarity value of all the measures of similarity, for the pair of concepts exceeds 0.40, the relationship takes the weight $w = 1$, otherwise $w = 0$.
- (d) If the average similarity value of all the measures of similarity, for the pair of concepts exceeds a threshold (see note), the relationship takes the weight $w = 1$, otherwise $w = 0$.

5. In the case of concepts, Only those concepts associated with the triple whose weight w is equal to 1 are considered valid.

Note: The threshold is calculated as the average of the similarity results of all relationships processed divided by two.

Table 2, shows some measures employed to calculate the similarity of binary vectors, that is, vectors containing 0 or 1 values [21].

Table 2. Similarity measures for binary vectors [21]

Similarity measure	Definition
Matching coefficient	$X \cap Y$
Dice coefficient	$\frac{2 X \cap Y }{ X + Y }$
Jaccard (or Tanimoto) coefficient	$\frac{ X \cap Y }{ X \cup Y }$
Overlap coefficient	$\frac{ X \cap Y }{\min(X , Y)}$
Cosine	$\frac{ X \cap Y }{\sqrt{ X \times Y }}$

3.2.4 Approach based on Formal Concept Analysis

Another approach used in the methodology is the Formal Concept Analysis (FCA), method (for more information see [37], [35]). This method is used at the context level to extract the existing relationship between a set of objects (concepts) and attributes (properties). In this approach, objects represent the concepts that were extracted from the ontology in the previous phase, whereas the attributes or properties are the verbs that exist in the sentences associated to the concept (context), that are

obtained from the syntactic dependencies obtained with different morphological taggers.

The approach considers two variants in the selection of properties to construct the incidence matrix, which requires the FCA, method to obtain the formal concepts. The incidence matrix is a binary relation made up of the concepts and verbs extracted from the existing syntactic dependencies in the sentences associated with the concept.

The difference between the two variants is the type of syntactic dependency analyzer used in the preprocessing stage of the previous phase to obtain the properties. The first variant uses the minipar tagger [17], while the second one uses the Stanford tagger [7]. For each variant, a set of dependency connection was selected manually, in order to extract the words associated with those relations.

The criterion of selection of attributes consists of the application of patterns for the extraction of verbs associated with the pair of concepts. For example, if we consider the pattern $C:i:V *$ (of the minipar tagger), all the verbs placed in “*” are selected, which are usually major verbs in the sentence.

In the case of Stanford patterns, for example, of the relation $dobj(*, -)$, we get the word placed in “*” and the word placed in “-” is omitted.

Table 3 shows the patterns used for the extraction of attributes in the documents POS tagged with minipar (FCA min), or Stanford (FCA sfd_0 , FCA sfd_2 , FCA sfd_3). In the case of the variant FCA min, the $C: i: V *$ pattern is used for both types of semantic relations (class-inclusion and non-taxonomic).

The FCA variant named sfd_0 only applies to class-inclusion relations and selects the main verbs in the sentence (root (*, *), cop (*, *)). In the variant FCA named sfd_2 , all verbs for the identification of non-taxonomic relations are selected. In the FCA sfd_3 variant, the verbs that define the non-taxonomic relations of the ontology in the extensional part of the formal concept are searched for, provided that the pair of concepts is in the intentional part (see Table 3).

The formal concepts obtained by the FCA system must associate the concepts of the ontology with some common verb, which allows the

Table 3. Patterns or names of relationships used by each variant

Variant	Pattern or name of the relation
FCA min	C:i:V *
FCA sfd_0	root(*,*), cop(*,*)
FCA sfd_2	nsubj(*,-), prep(*,-), root(*,*), dobj(*,-), acomp(*,-), advcl(*,*), agent(*,-), aux(*,*), auxpass(*,*), cop(*,*), csubj(-,*), csubjpass(*,-), dobj(*,-), expl(*,-), iobj(*,-), cop(*,*), nsubjpass(*,-), parataxis(-,*), pcomp(-,*), prep(*,-), prt(*,*), tmod(*,-), vmod(-,*)
FCA sfd_3	

approach to assign it to the relationship $w = 1$. In the case of concepts, the approach assigns 1 if $w = 1$, otherwise it assigns the value of 0.

3.2.5 Approach based on Latent Semantic Analysis

The last of the approaches used in second phase uses the method of Latent Semantic Analysis (LSA), in order to identify the semantic relations between concepts [34]. LSA assumes that words in the same semantic field tend to appear together or in similar contexts. In this case, it is considered that concepts that are semantically related can appear in the same sentence, or in different sentences but sharing information in common.

Starting from this assumption, we present an algorithm that considers for the weight assignment w , the result of the cosine similarity measure [21]. The approach executes the following steps:

1. Preprocessing both the reference domain corpus and the target ontologies. The domain corpus is split up into sentences and stopwords (such as prepositions, articles, etc.) are removed. The Porter stemmer algorithm [28], is also applied to the words contained in these sentences. The same process is applied to each of the concepts of the ontology in order to keep consistency in the terminology representation (stopword removing and the Porter stemming algorithm).

2. Application of the LSA algorithm to decrease the dimensionality of the context matrix. In this case, the S-Space⁷ package and the LSA⁸ algorithm are used. The algorithm receives as parameters the sentences of the domain corpus and the K dimensions (which in this case are 300 dimensions). The output of the LSA algorithm are semantic vectors of dimension K for each word identified by LSA in the corpus.
3. Extraction of concepts. The words obtained by the LSA method are clustered by cosine similarity to form the concepts of the ontology (LSA-cos).
4. Vocabulary reduction (vectors), of the LSA matrix. Only the concepts obtained in the previous step are placed in the new file which is the input to the next step. The rest of the words in the original matrix are removed.
5. Calculation of cosine similarity. The concepts obtained are used to determine the degree of similarity between each pair of concepts that will be part of the class-inclusion and non-taxonomic relations.
6. Calculation of threshold and weight w assigned to the relation. The threshold is calculated as the sum of the similarities between the total of relationships divided by 2. If the value of the similarity degree of the relation is greater than the threshold, the relation takes the weight of $w = 1$, otherwise it is equal to zero ($w = 0$).
7. In the case of concepts, if $w = 1$, the approach assigns the weight of 1 to the concept, otherwise the concept is assigned a zero value.

The third phase of the methodology for the evaluation of the ontology is described below.

3.3 Evaluation of the Ontology

The quality of the ontology is determined in the third phase of the proposed architecture of

⁷<https://github.com/fozziethebeat/S-Space>

⁸<http://code.google.com/p/airhead-research/wiki/LatentSemanticAnalysis>

solution. Since the ontology is made up of triplets with the *SRO* structure, where *S* and *O*, are concepts and *R*, is some type of semantic relation (class-inclusion or non-taxonomic), then we can use this information for determining whether the triple is correct or not, based on the evidence that it exists in the reference corpus. The metric considers the results of the approaches for each type of relationship and determines the degree to which the ontology is correct.

The metric (*M*), that determines the quality of the ontology is presented in Equation 1. As it can be observed, the metric is made up of the product of three matrices, $Matrix_C$, $Matrix_E$ and $Matrix_I$, which are defined as follows:

$$M(O) = Matrix_C \ Matrix_E \ Matrix_I. \quad (1)$$

The $Matrix_C$ contains the results of the accuracy measure (*A*), of each approach (E_i), for each type of semantic relationship (class-inclusion (*CI*), and non-taxonomic (*NT*); *n* is the total of approaches used, for this reason its dimensions are $2 \times n$. Equation 2, shows the structure of this matrix:

$$Matrix_C = \begin{bmatrix} A(E_1, CI) & \dots & A(E_n, CI) \\ A(E_1, NT) & \dots & A(E_n, NT) \end{bmatrix}. \quad (2)$$

The accuracy measure (*A*), applied to each approach (E_i), of the second phase and the type of semantic relationships (*R*), are presented in Equation 3. The semantic relationship, as mentioned above, may be the class-inclusion (*CI*), or the non-taxonomic one (*NT*):

$$A(E_i, R) = \frac{\sum_{i=1}^{|R|} Reliability(Triple_i)}{|R|}. \quad (3)$$

The measure of reliability is defined as shown in Equation 4, in which each triple is a linear combination: $\alpha * qual(C_{i,1}) + \beta * qual(C_{i,2}) + \gamma * qual(R_i)$, with the restriction $\alpha + \beta + \gamma = 1$. The measure has the following characteristics:

1. The quality of the first concept ($qual(C_{i,1})$).
2. The quality of the second concept ($(qual(C_{i,2}))$.
3. The quality of the relation between the two concepts ($qual(R_i)$).

3. The quality of the relation between the two concepts ($qual(R_i)$).

The quality of the relation between two concepts is determined as the weight *w* that assigns the approach (Phase two) to the relation R_i . The quality of each concept is determined as the weight that assigns the approach to the concepts that made up the relation.

The proposed equation for the evaluation of triple is presented in Equation 4:

$$Reliability(T_i) = \begin{cases} 1, & \text{If } \alpha \times qual(C_{i,1}) + \\ & \beta \times qual(C_{i,2}) + \\ & \gamma \times qual(R_i) > 0.75, \\ 0, & \text{otherwise,} \end{cases} \quad (4)$$

where
 $T_i = (C_{i,1}, R_i, C_{i,2})$ is a triple of the ontology,
 $C_{i,1}$ y $C_{i,2}$ are concepts,
 R_i ontological relationship.

In the case of the quality of the semantic relationship ($qual(R_i)$), we consider the measure of accuracy that, considering the total of relations proposed by the approach ($E_i(R)$), with the format (w_k, R_k) , where $1 \leq k \leq n$, and the total ontology relation ($n = |R|$), where R are the class-inclusion or non-taxonomic relations, w_i is the weight that the approach assigns to the relationship analyzed:

$$qual(R) = \frac{|E_i(R)|}{|R|}. \quad (5)$$

$Matrix_E$, is the external coefficient matrix of dimensions $n \times 2$ which assigns a weight for each approach normalized between 0 and 1, i. e., $a_1 + a_2 + \dots + a_n = 1$, and $b_1 + b_2 + \dots + b_n = 1$. Finally the $Matrix_I$, are internal coefficients, with dimensions 1×2 , which allows normalizing the results of class-inclusion and non-taxonomic relationships between 0 and 1, that is, $d_1 + d_2 = 1$:

$$Matrix_E = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \\ \dots & \dots \\ a_n & b_n \end{bmatrix}, \quad (6)$$

$$Matrix_I = [d_1 \ d_2]. \quad (7)$$

The experimental results are presented in the following section.

4 Experimental Results

In this section we present the dataset used together with the results obtained in the experiments.

4.1 Dataset

The knowledge domains considered in the experiments are: artificial intelligence (*AI*), standard e-Learning SCORM (SCORM)⁹ [40] and the OIL taxonomy of the oil domain with their corresponding reference domain corpus. Each ontology contains a number of concepts (*C*), class-inclusion (*S*), relationships, and non-taxonomic (*R*), relationships (see Table 4).

Table 4. Dataset, ontologies

Domain	Ontology		
	<i>C</i>	<i>S</i>	<i>R</i>
AI	276	205	61
SCORM	1,461	1,038	759
OIL	48	37	-

The documents (*D*), of the domain corpora were used to determine the amount of sentences (*O*), total tokens or words (*T*), of these sentences, the vocabulary or types (*V*), of the sentences, and the number of filtered sentences (*Of*), by the information retrieval system (see Table 5).

Table 5. Dataset, corpora

Domain	Corpora				
	<i>D</i>	<i>O</i>	<i>T</i>	<i>V</i>	<i>Of</i>
AI	8	475	11,370	1,510	415
SCORM	36	1,621	34,497	1,325	1,606
OIL	577	546,118	10,290,107	168,554	157,276

As mentioned above, a validation by human experts is also carried out. The sentences manually checked by the experts are approximately one or two per relation (OSE y ORE); these sentences in some cases were manually selected and in other cases randomly selected. In the case of class-inclusion (*S*) relations, in the ontologies AI and OIL were checked 100%, but in the case of the SCORM ontology only 10% (SE), as well as

⁹Ontologies and their corpus are available on the page: <http://www.site.uottawa.ca/azouaq/goldstandards.htm>

for non-taxonomic (*R*) relation reviewed by experts (RE).

Table 6 presents the amount of information evaluated by domain experts.

Table 6. Dataset (experts)

Domain	Class-inclusion			Non-taxonomic		
	<i>S</i>	<i>SE</i>	<i>OSE</i>	<i>R</i>	<i>RE</i>	<i>ORE</i>
AI	205	205	312	61	61	110
SCORM	1,038	100	159	759	189	309
OIL	37	37	75	-	-	-

Only a subset of data is validated due to the large human-hour effort required to manually evaluate the validity of each relation.

4.2 Baseline

In order to have a reference value on the evaluation of the ontology, we have built a baseline value for the validation of semantic relationships. The proposed process is to validate all the semantic relationships whose concepts are closely related in the reference corpus. In this sense, if two concepts associated with a semantic relationship appear together in the same context, we assume that both concepts are related. It is clear that this relationship might be different to the one established in the ontology, however, we only use this measure as a baseline.

In order to measure the degree of co-relation, we use the concept of mutual information, which is outlined below. Given a triple (*S, R, O*) with *R* the semantic relation between the concepts *S* and *O*, the mutual information between the two concepts is measured as:

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

The complete corpus and the subcorpus used by the experts were preprocessed, removing punctuation symbols and considering the stemmed version (using Porter stemmer), and the ontologic concepts. The obtained results are shown in Table 7.

Table 7. Baseline results obtained by using the mutual information correlation coefficient

Ontology	Type of relation	Subcorpus	Corpus
AI	Class-inclusion	55.61	23.90
	Non-taxonomic	47.54	16.39
SCORM	Class-inclusion	14.00	25.53
	Non-taxonomic	41.27	38.34
OIL	Class-inclusion	21.62	56.76
	Non-taxonomic	-	-

4.3 Experimental Results for Class-Inclusion Relationships

In this section we present the results obtained from the approach, using the accuracy criteria (A), when we evaluate the ontologies and their quality (C), in the prediction of the approach and according to three human experts (H_1 , H_2 y H_3), and the baseline.

In the case of the class-inclusion relationships, the approach lexical-syntactic patterns (LSP), obtains an accuracy of 88.78% for the quality of the ontology AI (see Table 8), but the average amount (Avg), that the experts assigned was 88.48% with an error range of 1.48%. This result indicates that the approach is more accurate since it is very close to the average result given by the experts.

In the case of the approach LSA-cos, the system error value is 3.57%, which indicates that this approach is also close to the experts answers. The approach that has a higher error range is the one named sim-cos_u (similarity coseno-umbral), with an error range of 7.87% with respect to the average value of the experts. These results show that the approaches behaved correctly with respect to the experts answers. Furthermore they overpassed the results of the baseline.

Table 8. Accuracy of the AI ontology and the quality of predictions of approaches for class-inclusion relationships

Approach	A	$C(H_1)$	$C(H_2)$	$C(H_3)$	Avg
LSP	88.78	89.76	84.39	88.29	87.48
Sim-cos	90.24	83.41	80.98	87.80	84.07
Sim-cos_u	98.05	90.24	86.83	95.61	90.89
FCA min	95.61	89.76	85.37	94.15	89.76
FCA std ₀	100.00	92.20	88.78	97.56	92.85
LSA-cos	94.15	90.24	89.76	92.68	90.89
Baseline	56.00	57.00	51.00	55.00	54.00

In the case of the SCORM ontology, for class-inclusion relations, the LSP approach obtained 54% accuracy if 100 ontology relationships are considered to be correct. However, the experts assigned an average value of 74.33% of quality to the approach (see Table 9).

The results provided by the experts indicate that the approach agrees with at least 70 expert responses. Also, It is observed that the LSA-cos approach continues providing good results for this ontology. All average results provided by the experts for these approaches exceed the baseline. Considering these results, it is observed that the approach identifies other ontology relationships which are not class-inclusion, according to the answers given by the experts.

Table 9. Accuracy of SCORM ontology and quality predictions of approaches for class-inclusion relationships

Approach	A	$C(H_1)$	$C(H_2)$	$C(H_3)$	Avg
LSP	54.00	76.00	70.00	77.00	74.33
Sim-cos	89.00	65.00	77.00	66.00	69.33
Sim-cos_u	93.00	67.00	81.00	68.00	72.00
FCA min	89.00	65.00	75.00	64.00	68.00
FCA std ₀	98.00	70.00	84.00	69.00	74.33
LSA-cos	92.00	70.00	84.00	69.00	74.33
Baseline	14.00	42.00	30.00	45.00	39.00

The experimental results of each approach for the OIL ontology are shown in Table 10. In this case, the LSP approach obtained 14 correct class-inclusion relations of the 37 that exist in the ontology considered as valid, producing a 37.84% of accuracy.

However, the average manual evaluation of experts indicates that the approach obtains a quality of 61.26% for this type of relationship. All results of the average value obtained exceeded the baseline.

4.4 Experimental Results for Non-Taxonomic Relationships

In the case of non-taxonomic relationships, the Syntactic Dependency Analysis (SDA) approach, for the AI ontology, obtained 88.52% of accuracy, while the experts assigned 82%, 83% and 86% of quality to the ontology (the average of the three

Table 10. Accuracy of OIL ontology and quality predictions of approaches for class-inclusion relationships

Approach	<i>A</i>	<i>C(H</i> ₁ <i>)</i>	<i>C(H</i> ₂ <i>)</i>	<i>C(H</i> ₃ <i>)</i>	<i>Avg</i>
LSP	37.84	72.97	64.86	45.95	61.26
Sim-cos	91.89	67.57	81.08	100.00	82.88
Sim-cos.u	91.89	67.57	81.08	100.00	82.88
FCA min	86.49	62.16	75.68	94.59	77.48
FCA sfd ₀	91.89	67.57	81.08	100.00	82.88
LSA-cos	89.19	70.27	83.78	91.89	81.98
Baseline	21.62	62.16	48.65	29.73	46.85

results is 84.15%) (see Table 11), which is not very far from the result *A*.

It is observed that the FCA sfd₂ approach is closer to the average answers of the experts, but obtains an error of 8% with respect to result *A*. Again, it can be seen that the average results exceeded the baseline.

Table 11. Accuracy of the AI ontology and the quality of the predictions of approaches for non-taxonomic relationships

Approach	<i>A</i>	<i>C(H</i> ₁ <i>)</i>	<i>C(H</i> ₂ <i>)</i>	<i>C(H</i> ₃ <i>)</i>	<i>Avg</i>
SDA	88.52	81.97	86.89	83.61	84.15
Sim-cos	93.44	86.89	88.52	85.25	86.89
Sim-cos.u	98.36	88.52	93.44	90.16	90.71
FCA min	93.44	80.33	88.52	85.25	84.70
FCA sfd ₂	100.00	86.89	95.08	91.80	91.26
FCA sfd ₃	95.08	81.97	90.16	90.16	87.43
LSA-cos	90.16	83.61	85.25	85.25	84.70
Baseline	48.00	51.00	46.00	52.00	50.00

In the case of the SCORM ontology, the SDA approach obtained 86% accuracy and 83%, 87%, and 80%, of the quality was predicted by the experts (see Table 12). Again, the FCA sfd₂ approach achieves the average of 91%, but it has an 8% error when compared to the result *A*. All average results exceeded the baseline.

4.5 Experimental Results of the Evaluation Metric

The results of the approaches presented in Tables 8, 9 and 10, for class-inclusion relationships and the results of Tables 11 and 12, for non-taxonomic relationships are used in the metric of evaluation.

The experimental results of the metric (*M*), are shown in Table 13 for each domain ontology (*O*), where *M(S)*, is the result of the metric for the

Table 12. Accuracy of the SCORM ontology and the quality of the predictions of the approaches for non-taxonomic relationships

Approach	<i>A</i>	<i>C(H</i> ₁ <i>)</i>	<i>C(H</i> ₂ <i>)</i>	<i>C(H</i> ₃ <i>)</i>	<i>Avg</i>
SDA	86.24	82.54	84.66	84.13	83.77
Sim-cos	96.30	84.13	92.59	85.71	87.48
Sim-cos.u	98.41	85.19	94.71	86.77	88.89
FCA min	96.30	85.19	93.65	88.89	89.24
FCA sfd ₂	99.47	87.30	96.83	88.89	91.01
FCA sfd ₃	90.48	82.54	89.95	83.07	85.19
LSA-cos	86.77	77.78	85.19	78.31	80.42
Baseline	41.00	47.00	43.00	43.00	44.33

automated evaluation system considering only the data validated by the experts, *M(H_i)*, with *i* = 1, 2, 3 and *P*, is the average of the results obtained by the experts.

According to the observed results, the AI ontology shows an acceptable result when obtaining an average quality of 88.22%, whose error is 6.7% with respect to the system (*M(S)*), indicating that the proposed result 94.31%, is found within the limits of acceptable error for the measurement of the quality of the ontology. In the case of the SCORM ontology, an error greater than 13% is obtained, so that the result of 89%, is within the average quality range of the ontology. Finally, for the OIL ontology, the error is 4.2% and the assigned evaluation value of 81.53% is within the acceptance range of the ontology quality.

Table 13. Results of metric evaluation applied to domain ontologies with data and results of experts

<i>O</i>	<i>M(S)</i>	<i>M(H</i> ₁ <i>)</i>	<i>M(H</i> ₂ <i>)</i>	<i>M(H</i> ₃ <i>)</i>	<i>M(P)</i>
AI	94.31	86.79	87.86	92.52	88.22
SCORM	89.63	76.18	84.79	76.97	79.31
OIL	81.53	68.02	77.93	88.74	78.23

4.6 Experimental Results of the Automatic Evaluation System

Considering the complete information of ontologies and corpora, the results of the approaches for the evaluation of class-inclusion relationships are presented in Table 14, and non-taxonomic relationships are presented in Table 15.

As can be seen, in most cases the baseline results are outperformed. The quality of the

Table 14. Experimental results for class-inclusion relationships

Ontology	Approach	A
AI	LSP	88.78
	Sim-cos	76.10
	Sim cos.u	97.07
	FCA min	96.59
	FCA sfd_0	100.00
	LSA-cos	87.32
	Baseline	23.90
SCORM	LSP	72.45
	Sim-cos	83.72
	Sim-cos.u	92.68
	FCA min	92.49
	FCA sfd_0	97.01
	LSA-cos	87.48
	Baseline	25.53
OIL	LSP	45.95
	Sim-cos	72.97
	Sim-cos.u	89.19
	FCA min	100.00
	FCA sfd_0	89.19
	LSA-cos	70.27
	Baseline	56.76

domain ontologies, when applying the metric of evaluation are shown in Table 16. Based on the obtained results, it can be observed that the AI ontology is the most stable one, obtaining more than 90% of quality.

On the other hand, the OIL ontology does not obtain good results with the LSP approach, but there is an improvement with the rest of the approaches in the percentage of accuracy, which allows the metric to assign the quality value of 77%.

The SCORM ontology, which is the one that has the highest number of relationships, achieves a quality of 88.44%, which based on the conditions of the same, is considered an acceptable result.

As can be seen, the results obtained for the AI ontology are 90.80%, of quality, a result that is considered to be in the error range of 6% that was obtained in the evaluation with domain experts, that is, the result of 90.80%, is in the margin of error of 6% with respect to the values of 88.22% and 94%. We consider this value acceptable for the AI ontology.

Table 15. Experimental results for non-taxonomic relationships of each ontology

Ontology	Approach	A
AI	SDA	88.52
	Sim-cos	72.13
	Sim-cos.u	98.36
	FCA-min	95.08
	FCA- sfd_2	100.00
	FCA- sfd_3	96.72
	LSA-cos	83.61
SCORM	Baseline	16.39
	SDA	86.82
	Sim-cos	78.13
	Sim-cos.u	94.99
	FCA-min	96.18
	FCA- sfd_2	98.95
	FCA- sfd_3	91.44
OIL	LSA-cos	78.26
	Baseline	38.34

Table 16. Results of the automatic evaluation for each ontology

Ontology	Evaluation
AI	90.80
SCORM	88.44
OIL	77.93

5 Conclusions

This paper presents a methodology for the automatic evaluation of ontologies of restricted domain, by means of natural language processing, extraction of information and linguistic tools.

The evaluation methodology is made up of 3 phases. In the first phase, the relationships and concepts of the domain ontology to be evaluated are extracted. Queries are built to retrieve relevant information from the reference corpus. On the recovered information, a preprocessing step is carried out executing operations such as removing of punctuation symbols and stop words.

In the second phase, five different approaches for the discovery of class-inclusion and non-taxonomic relationships were developed. The approaches allowed to assign to the relationship a score that indicates whether or not exist evidence of the presence of the same relationship in the reference corpus. In the same phase of the methodology, the concepts that are part of the relations discovered are extracted.

In the third phase, a metric was developed to measure the quality of the ontology that has to be evaluated. The metric is built based on the accuracy of the approaches developed in the second phase and provides a measure for the ontology quality.

In order to validate this metric, the results obtained by the methodology against standard gold values (ideals) obtained by the evaluation of three human experts are compared.

Additionally, the results are compared against a baseline value calculated using criteria of contextual similarity by means of a measure of correlation based on mutual information.

The validation of the developed methodology was carried out on three ontologies of restricted domain.

The methodology ensures that the artificial intelligence domain ontology achieves 90% quality, the SCORM ontology 88%, and the OIL ontology a 77%. The results offered by the methodology depend completely on the applied approaches. In fact, the metric proposed in this research work considers each of the approaches. In this case, if there is any approach that can not be fully applied to the evaluation process, then that metric will be affected, and therefore, the final result of the evaluation of the ontology quality will be affected as well.

However, in order to guarantee the quality of the developed methodology, it is important that a well-constructed reference corpus of the domain to be represented by the ontology is offered. As alternatives of improvement, is intended to carry out a study that allows to determine the ideal parameters that indicate the importance of each approach in the evaluation. As well as the use of other approaches that measure the quality of the ontology at the structure level.

Acknowledgment

This work is partially supported by the Sectoral Research Fund for Education with the CONACyT project 257357, by PRODEP-SEP ID 00570 (EXB-792) DSA/103.5/15/10854, and VIEP-BUAP project 00478.

References

1. **Bejar, I. I., Chaffin, R., & Embretson, S. E. (1991).** *Cognitive and psychometric analysis of analogical problem solving*. Recent research in psychology. Springer-Verlag.
2. **Bhatt, B. & Bhattacharyya, P. (2012).** Domain specific ontology extractor for indian languages. *Proceedings of the 10th Workshop on Asian Language Resources*, The COLING 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*.
5. **Cantador, I., Ferández, 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. **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*.
7. **de Marneffe, M.-C., MacCartney, B., & Manning, C. D. (2006).** Generating typed dependency parses from phrase structure trees. *LREC*.
8. **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, Amstedam.
9. **Gangemi, A., Catenacci, C., Ciaramita, M., & Lehmann, J. (2006).** Modelling ontology evaluation and validation. *Proceedings of the 3rd European Semantic Web Conference (ESWC2006)*, vol. 4011 *LNCS*, Springer.
10. **Gómez-Pérez, A. (2004).** *Ontology Evaluation*. International Handbooks on Information Systems. Springer.
11. **Grigonyté, G. (2010).** *Building and Evaluating Domain Ontologies: NLP Contributions*. Logos-Verlag.

12. **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.
13. **Hearst, M. A. (1992).** Automatic acquisition of hyponyms from large text corpora. *Proceedings of the 14th International Conference on Computational Linguistics*, pp. 539–545.
14. **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.
15. **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.
16. **Klaussner, C. & Zheкова, D. (2011).** Lexico-syntactic patterns for automatic ontology building. *Proceedings of the Second Student Research Workshop associated with RANLP*, RANLP 2011 Organising Committee, Hissar, Bulgaria, pp. 109–114.
17. **Lin, D. (1998).** Dependency-based evaluation of minipar. *Proc. Workshop on the Evaluation of Parsing Systems*, Granada.
18. **Lovrencic, S. & Mirko, C. (2008).** Ontology evaluation - comprising verification and validation. *Proceedings of Central European Conference on Information and Intelligent Systems, CECIS - 2008*.
19. **Maedche, A. & Staab, S. (2002).** Measuring similarity between ontologies. *Proceedings of European Knowledge Acquisition Workshop (EKAW)*.
20. **Manning, C., Raghavan, P., & Schütze, H. (2008).** *Introduction to Information Retrieval*. Cambridge University Press.
21. **Manning, C. D. & Schütze, H. (1999).** *Foundations of statistical natural language processing*. MIT Press, Cambridge, MA, USA.
22. **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*.
23. **Mititelu, V. B. (2011).** Hyponymy patterns in romanian. *Memoirs of the Scientific Sections of the Romanian Academy*, Vol. XXXIV, pp. 31–40.
24. **Montiel-Ponsoda, E. & Aguado de Cea, G. (2008).** Using natural language patterns for the development of ontologies. In *Researching specialized languages*, pp. 332–345.
25. **Ortiz, R. & Alejandro, J. (2013).** *Creación automática de Ontologías a partir de Textos con un Enfoque Lingüístico*. Ph.D. thesis, Dept Ciencias Computacionales, Cenidet, Cuernavaca, Mor., Mex.
26. **Padró, L. & Stanilovsky, E. (2012).** Freeling 3.0: Towards wider multilinguality. *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)*, ELRA, Istanbul, Turkey.
27. **Pak, J. & Zhou, L. (2009).** A framework for ontology evaluation. *WEB*, pp. 10–18.
28. **Porter, M. F. (1997).** An algorithm for suffix stripping. In **Sparck Jones, K. & Willett, P.**, editors, *Readings in Information Retrieval*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 313–316.
29. **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.
30. **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*.
31. **Salem, S. & AbdelRahman, S. (2010).** A multiple-domain ontology builder. *Proceedings of the 23rd International Conference on Computational Linguistics, COLING '10*, Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 967–975.
32. **Schmid, H. (1994).** Probabilistic Part-of-Speech Tagging Using Decision Trees. *Proceedings of International Conference on New Methods in Language Processing*.
33. **Tovar, M., Pinto, D., Montes, A., González, G., Vilariño Ayala, D., & Beltrán, B. (2014).** Use of lexico-syntactic patterns for the evaluation of taxonomic relations. **Trinidad, J. F. M., Carrasco-Ochoa, J. A., Olvera-López, J. A., Rodríguez, J. S., & Suen, C. Y.**, editors, *Pattern Recognition*,

volume 8495 of *Lecture Notes in Computer Science*, Springer International Publishing, pp. 331–340.

34. **Tovar, M., Pinto, D., Montes, A., & González Serna, J. G. (2017).** An approach based in LSA for evaluation of ontological relations on domain corpora. **Carrasco-Ochoa, J. A., Trinidad, J. F. M., & Olvera-López, J. A.**, editors, *Pattern Recognition - 9th Mexican Conference, MCPR*, volume 10267 of *Lecture Notes in Computer Science*, Springer, pp. 225–233.

35. **Tovar, M., Pinto, D., Montes, A., González Serna, J. G., & Vilariño Ayala, D. (2015).** Patterns used to identify relations in corpus using formal concept analysis. **Carrasco-Ochoa, J. A., Trinidad, J. F. M., Azuela, J. H. S., Olvera-López, J. A., & Famili, F.**, editors, *Pattern Recognition, 7th Mexican Conference (MCPR)*, volume 9116 of *Lecture Notes in Computer Science*, Springer, pp. 236–245.

36. **Tovar Vidal, M., Pinto Avendaño, D., Montes Rendón, A., González Serna, J. G., & Vilariño Ayala, D. (2015).** Evaluation of ontological relations in corpora of restricted domain. *Computación y Sistemas*, Vol. 19, No. 1.

37. **Vidal, M. T., Avendaño, D. P., Rendón, A. M., Serna, J. G. G., & Ayala, D. V. (2015).** Identification of ontological relations in domain corpus using formal concept analysis. *Engineering Letters*, Vol. 23, No. 2.

38. **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.

39. **Zavitsanos, E., Palioras, 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.

40. **Zouaq, A., Gasevic, D., & Hatala, M. (2012).** Linguistic patterns for information extraction in ontocmaps. **Blomqvist, E., Gangemi, A., Hammar, K., & del Carmen Suárez-Figueroa, M.**, editors, *WOP*, volume 929 of *CEUR Workshop Proceedings*, CEUR-WS.org.

Article received on 01/08/2016; accepted on 12/10/2016.

Corresponding author is Mireya Tovar.