Use of Lexico-Syntactic Patterns for the Evaluation of Taxonomic Relations

Mireya Tovar[1,2], David Pinto[2], Azucena Montes[1,3], Gabriel González[1], Darnes Vilariño[2], and Beatriz Beltrán[2]

[1] Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET), Mexico
[2] Faculty of Computer Science, Benemérita Universidad Autónoma de Puebla, Mexico
[3] Engineering Institute, Universidad Nacional Autónoma de Mexico
Puebla, México
{mtovar,amontes,gabriel}@cenidet.edu.mx,
{dpinto,darnes,bbeltran}@cs.buap.mx

Abstract. In this paper we present an approach for the evaluation of taxonomic relations of restricted domain ontologies. We use the evidence found in corpora associated to the ontology domain for determining the validity of the taxonomic relations. Our approach employs lexico-syntactic patterns for evaluating taxonomic relations in which the concepts are totally different, and it uses a particular technique based on subsumption for those relations in which one concept is completely included in the other one. The integration of these two techniques has allowed to automatically evaluate taxonomic relations for two ontologies of restricted domain. The performance obtained was about 70% for one ontology of the e-learning domain, whereas we obtained around 88% for the ontology associated to the artificial intelligence domain.

Keywords: Lexico-syntactic patterns, Ontology evaluation, Taxonomic relations.

1 Introduction

There is a huge amount of information that is uploaded every day to the World Wide Web, thus arising the need for automatic tools able to understand the meaning of such information. However, one of the central problems of constructing such tools is that this information remains unstructured nowadays, despite the effort of different communities for give a semantic sense to the World Wide Web. In fact, the Semantic Web research direction attempts to tackle this problem by incorporating semantic to the web data, so that it can be processed directly or indirectly by machines in order to transform it into a data network [18]. For this purpose, it has been proposed to use some knowledge structures such as ontologies for giving semantic and some structure to unstructured data. An ontology, from the computer science perspective, is "an explicit specification of a conceptualiation" [8]. An ontology typically includes classes, instances,

J.F. Martínez-Trinidad et al. (Eds.): MCPR 2014, LNCS 8495, pp. 331–340, 2014.

attributes, relations, constraints, rules, events and axioms. Even thought the ontologies may be structured with taxonomic and/or non-taxonomic relations, in this paper we focus the evaluation over the taxonomic relations, which normally are referred as relations of type "is-a" (hypernym/hyponymy or subsumption).

There are plenty of research works in literature that address the problem of automatic construction of ontologies. The major of those works evaluate the ontology created by using a gold standard, which in fact, it is supposed to be created by one expert. Using this approach, it is assumed that the expert has created the ontology in a correct way, but there is not a guaranty of such thing. Thus, we consider very important to investigate the manner of evaluate automatically the quality of these kind of resources, which are continuously been used in the framework of the semantic web.

Our approach attemps to find evidence of the relations to be evaluated in a reference corpus (associated to the same domain of the ontology), and therefore, we needed to analyze the different approaches reported in literature for automatic identification of ontology relations. A number of classification methods have been addressed for identifying relations between concepts or instances [4,5,16]. For instance, for identifying whether or not a given instance (a pair of words *flower*:*tulip*) belongs to a specific relation (class-inclusion) [21]. Other approaches identify the degree of semantic similarity between a set of word pairs in which it is known that belong to a certain semantic class (semantic relation) [10,20,19]. For the purpose of this paper, we focus our analysis on those techniques that identify taxonomic relations.

The remaining of this paper is structured as follows. Section 2 describes more into detail the lexico-syntactic patterns found in literature. In Section 3 we present the model proposed for addressing the problem aforementioned. Section 4 shows and discusses the results obtained by the presented approach. Finally, in Section 5 the findings and the future work are given.

2 Lexico-Syntactic Patterns

A seminal work in the task of automatic identification of hypernyms from raw texts is the one presented by Hearst [9]. She proposed six lexico-syntactic patterns, which actually are known as Hearst's patterns, that have been widely used in other works. In [1], for example, the authors obtaind co-hyponyms by using the Hearst's patterns, but there are other approaches such as the following ones: [17,7,3,13]. Even though these patterns behave well on the above mentioned task, it is important to notice that they may be adjusted to work better in particular domains, which is our case.

There are other works proposing lexico-syntactic patterns, such as the one presented in [14], in which the authors focused on the romanian language. In [7], they propose a methodology that combine two techniques for the extraction of hyponymy, meronymy, co-hyponymy and near-synonymy in texts the Italian Wikipedia, i. e., lexico-syntactic patterns and statistical distributional systems. They use only five lexico-syntactic patterns, achieving good experimental results in this type of semantic relations. In [17], an automatic classifier for the hypernym and hyponym relation identification is built. It based in the use of dependency paths for some lexico-syntactic patterns.

In our case, we have collected 106 lexico-syntactic patterns, associated with the identification of taxonomic relations, from nine different sources [10,9,1,3,13,14,12,15,22]. Although we have found useful only 16 of these when we evaluated the two target ontologies, the rest of them may be useful for future investigations on ontology evaluation or automatic ontology learning tasks. Therefore, in Table 1, it can be seen all these lexico-syntactic patterns compiled.

Table 1. Database of lexico-syntactic patterns useful for detecting taxonomic relations

<table>
<tr><th>No.</th><th>Lexico-syntactic patterns</th></tr>
<tr><td>12</td><td>NP such as (NP,)* (or|and|the) NP</td></tr>
<tr><td>13</td><td>NP ’s NP</td></tr>
<tr><td>15</td><td>such NP as (NP,)*</td></tr>
<tr><td>42</td><td>NP (is | are) NP</td></tr>
<tr><td>43</td><td>NP (is | are) (a | an) NP</td></tr>
<tr><td>46</td><td>NP such as (NP,)* (or|and) NP</td></tr>
<tr><td>50</td><td>NP (classify (in | into) | comprise | contain | compose (of)? | group (in | into) | divide (in | into) | fall (in | into) | belong (to)) NP</td></tr>
<tr><td>86</td><td>NP (and | or) (another | other) NP</td></tr>
<tr><td>92</td><td>NP (,)? such as (NP,)* (or|and|the)? NP</td></tr>
<tr><td>94</td><td>NP NP , is (a|an|the) NP</td></tr>
<tr><td>96</td><td>NP , (is | are) (NP,)* (or | and|the) NP</td></tr>
<tr><td>97</td><td>(NP,)* (or|and|the) (NP,)* (is | are) (a|an|the) NP</td></tr>
<tr><td>98</td><td>NP , including NP</td></tr>
<tr><td>104</td><td>NP as (NP,*) (or| and| the) NP</td></tr>
<tr><td>106</td><td>NP, for example, is (a|an|the) NP</td></tr>
</table>

3 Evaluation of Taxonomic Relations

The evaluation process proposed is based on finding evidence in a reference corpus using the “correctness” criterion [2]. We assume that there exist such a collection of documents associated to the ontology domain (reference corpus) from which it is possible to find evidence of the correctness of the taxonomic relations held by the ontology. This evidence is found throught the use of lexico-syntactic patterns.

The approach proposed the following three steps:

– Pre-processing stage: In this step, all data (ontology, reference corpus and lexico-syntactic patterns) receive a special treatment in order to have normalized information, representing them by their lemmas. For this purpose, we use the FreeLing PoS tagger[1]. An information retrieval system is used for

[1] http://nlp.lsi.upc.edu/freeling/

filtering those documents which contain information refering the two concepts of any of the relations extracted from the ontology to be evaluated[2].
- Discovering of taxonomic relations: For practical purposes, the lexico-syntactic patterns are transformed into regular expressions, which are used for discovering evidence of the ontology taxonomic relations in its reference corpus.
- Evaluation: Our system provide a score for evaluating the ontology by using the accuracy formulae: Accuracy(ontology) $= \frac{|S(R)|}{|R|}$, where $|S(R)|$ is the total number of relations from which our system considers that exist evidence in the reference corpus, and $|R|$ is the number of taxonomic relations in the ontology to be evaluated. This score, need to be evaluated in order to determine the quality of the approach presented. For this purpose, we compare the results obtained by our system with respect to those results obtained by human experts.

In order to evaluate the taxonomic relations in the ontology, we consider two different situations:

1. The two concepts of a given taxonomic relation are completely different. In this case, we propose to use our bank of lexico-syntactic patterns for finding evidence of relation validity in the reference corpus.
2. One of the two concepts (X) of a given taxonomic relation is part of the other concept (Y). In this case, we propose a subsumption technique [1], which basically searches evidence of the hyponym Y in the reference corpus.

Examples of these types of situations are given in the first three rows of Table 2. The last row of this Table shows an example of a semantic relation that exist in the ontology but the evidence in the reference corpus indicates that the relation is not taxonomic.

4 Experimental Results

In this section we present the datasets, so as the results obtained in the experiments. In order to have a better understanding of the particular lexico-syntactic patterns applied in the evaluation, in the first part of the results subsection, we show the frequency of their occurrence in the reference corpus.

4.1 Dataset

In Table 3 we present the number of concepts (C) and taxonomic relations (R) of the two ontologies evaluated in this paper. The following characteristics of their references corpus are also given: number of documents (D), number of tokens (T), vocabulary dimensionality (V), and the number of sentences filtered (O)

[2] We used Jena for extracting the taxonomic relations from the ontology (`http://jena.apache.org/`)

Table 2. Examples of taxonomic relations in the artificial intelligence domain

Num.	$Concept_1$	$Concept_2$	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.

Table 3. Datasets

Domain	Ontology		Reference corpus			
	C	R	D	T	V	O
AI	276	205	8	10,805	2,180	464
SCORM	1,461	1,038	36	32,644	2,154	1,632

by the information retrieval system (S). As can be seen, the two domains used in the experiments are: Artificial Intelligence (AI), and the standard e-Learning SCORM ($SCORM$)[3] [23].

As we mentioned before, we requested human experts to evaluate the validity of the ontology taxonomic relations, according to different sentences obtained from the reference corpus. This manual evaluation was used to determine the performance of our approach. The results obtained are shown in the following subsection.

4.2 Results

In Table 4 we show the frequency of occurrence of the lexico-syntactic patterns that found evidence of taxonomic relations in the reference corpora. They are sorted according to their frequence in descending order.

In order to evaluate the stability of occurrence frequency of the bank of patterns given in Table 4, we used the Kendall tau correlation coefficient [11] that determines the degree in which the two lists matches, according to the descending order established.

The Kendall tau coefficient (τ) is calculated as $\tau = \frac{2P}{(k(k-1))/2} - 1$, where k is the number of items, and P is the number of concordant pairs obtained as

[3] The two ontologies together with their reference corpus can be downloaded from `http://azouaq.athabascau.ca/goldstandards.htm`

Table 4. Results of lexico-syntactic patterns

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

the sum, over all the items, of those items ranked after the given item by both rankings.

The Kendall tau coefficient value lies between -1 and 1, and high values imply a high agreement between the two rankings. Therefore, if the agreement (disagreement) between the two rankings is perfect, then the coefficient will have the value of 1 (-1). In case of obtaining the value 0, then it is said that the rankings are completely independent.

By ordering the lexico-syntactic patterns in descending order, we obtain a Kendall tau equal to 0.733, which means that exist a high agreement in the order obtained in the two reference corpus. This means that there exist a consistency in the application of these patterns, independently of being applied in different domains. This fact is true, at least for the two ontologies used in the experiments.

Table 5 show the result obtained by the approach when the AI ontology is evaluated using the accuracy criterion. The three last columns indicate the quality of the system prediction according to three human experts (E_1, E_2 and E_3). We consider that the quality obtained (91%, 82% and 86%) is a good result, however, we need to investigate the reason because we were not able to detect the remaining percentage of taxonomic relations. In general, our approach assigns an accuracy of 0.87% to the quality of the AI ontology.

Table 5. Accuracy of the AI ontology, and quality of the system prediction

Accuracy	*Quality(E_1)*	*Quality(E_2)*	*Quality(E_3)*
0.87	0.91	0.82	0.86

Table 6 shows the result obtained by the approach when the SCORM ontology is evaluated using the accuracy criterion. Again, the last columns indicate the quality of the system prediction according to three human experts (E_1, E_2 and E_3). According to the human experts, the accuracy result we obtained for the SCORM ontology (0.59) is less reliable than the one we obtained for the AI ontology. The experts assigned a quality value less than 80%. Despite this result, we assume that our system is capable of give a valuable accuracy that provides a clue for the quality of the target ontology.

Table 6. Accuracy of the SCORM ontology, and quality of the system prediction

Accuracy	*Quality(E_1)*	*Quality(E_2)*	*Quality(E_3)*
0.59	0.78	0.69	0.72

In order to validate the results obtained by the approach presented here, we have evaluated the agreement between each human expert evaluation (also named raters) and the system result, by using the Cohen's Kappa coefficient [6]. This measure is calculated as shown in Eq.(1); $Pr(a)$ is the relative observed agreement among one rater and the system result, and $Pr(e)$ is the hypothetical probability of chance agreement, using the observed data to calculate the probabilities of each result randomly saying each category. If the rater and the system are in complete agreement then $\kappa = 1$. If there is no agreement between them other than what would be expected by chance (as defined by Pr(e)), $\kappa = 0$.

$$\kappa = \frac{Pr(a) - Pr(e)}{1 - Pr(e)} \tag{1}$$

The results obtained by the Cohen's kappa statistical measure are shown in Table 7.

Table 7. Agreement between experts and system results

ontology	*Cohen's kappa*		
	E_1	E_2	E_3
AI	0.51	0.18	0.09
SCORM	0.54	0.23	0.36

The interpretation of the results obtained by the Cohen's kappa coefficient follows. Two human experts show light agreement ($0.01 \leq \kappa \leq 0.20$), whereas one expert show moderate agreement ($0.41 \leq \kappa \leq 0.60$) for the AI ontology. In the case of the SCORM ontology, again the expert E_1 obtained a moderate agreement, whereas the other two human experts showed fair agreement with $0.21 \leq \kappa \leq 0.40$.

The results presented above were obtained with samples of the taxonomic relations because of the great effort needed for manually evaluate their validity.

For the AI ontology we used 205 relations, whereas the SCORM ontology was evaluated only with 169 relations. Actually, we only provide samples of the reference corpora to the human experts for validating the taxonomic relation, which may bias the overall result. Therefore, in order to have a complete evaluation of the two ontologies, we have calculated the accuracy for both ontologies, but in this case considering all the sentences associated to the relations to be evaluated. Table 8 shows the number of taxonomic relation evaluated ($TaxRel$), the number of taxonomic relations found by the system ($TaxRelFound$), and the accuracy assigned to each ontology (*Accuracy*).

Table 8. Accuracy given to the two ontologies

Ontology	$TaxRel$	$TaxRelFound$	*Accuracy*
AI	205	181	88.29%
SCORM	1038	731	70.42%

As can be seen, the system obtain a slightly better accuracy for the AI ontology. This result is obtained because, in this case, the system have a greater number of sentences associated to each relation, therefore, having more oportunity to find evidence of the validity of the taxonomic relation. The SCORM ontology accuracy obtained was significantly better because, in this case we have evaluated a greater number of relations (1038) compared with those used in the sample evaluation (169). Besides this fact, in this last experiment, we have used a greater number of sentences which improves the oportunity of finding evidence in the reference corpus.

5 Conclusions

Evaluating the quality of ontologies is a very challenging topic that need to be addressed by the computational linguistic community. In this paper we have presented an approach based on lexico-syntactic patterns and reference corpora that allows to determine the accuracy of taxonomic relation of a given ontology of restricted domain. The experiments show that there exist a high agreement in the frequency of occurrence for the patterns used in the evaluation process, independently of the ontology domain. The approach assigned an accuracy of 88.29% for the AI ontology and 70.42% for the SCORM ontology, which reflects in some way the quality of the ontology. These results should be read in terms of the quality of our system, that was evaluated by two experts obtaining an average of 80% of reliability. As future work, we plan to evaluate the reliability of the system considering a greater number of relations. Besides that, we would like to use other ontologies for the evaluation process.

References

1. Bhatt, B., Bhattacharyya, P.: Domain specific ontology extractor for indian languages. In: Proceedings of the 10th Workshop on Asian Language Resources, pp. 75–84. The COLING 2012 Organizing Committee, Mumbai (December 2012), `http://www.aclweb.org/anthology/W12-5209`
2. Cantador, I., Fernández, M., Castells, P.: A collaborative recommendation framework for ontology evaluation and reuse. In: Actas de International Workshop on Recommender Systems, en la 17th European Conference on Artificial Intelligence (ECAI 2006), Riva del Garda, Italia, pp. 67–71 (2006)
3. de Cea, G.A., de Mon, I.A., Montiel-Ponsoda, E.: From linguistic patterns to ontology structures. In: 8th International Conference on Terminology and Artificial Intelligence (2009)
4. Celli, F.: Unitn: Part-of-speech counting in relation extraction. In: Proceedings of the 5th International Workshop on Semantic Evaluation, SemEval 2010, pp. 198–201. Association for Computational Linguistics, Stroudsburg (2010), `http://dl.acm.org/citation.cfm?id=1859664.1859707`
5. Chen, Y., Lan, M., Su, J., Zhou, Z.M., Xu, Y.: Ecnu: Effective semantic relations classification without complicated features or multiple external corpora. In: Proceedings of the 5th International Workshop on Semantic Evaluation, SemEval 2010, pp. 226–229. Association for Computational Linguistics, Stroudsburg (2010), `http://dl.acm.org/citation.cfm?id=1859664.1859714`
6. Cohen, J.: A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20(1), 37–46 (1960), `http://epm.sagepub.com/cgi/content/refs/20/1/37`
7. Giovannetti, E., Marchi, S., Montemagni, S.: Combining statistical techniques and lexico-syntactic patterns for semantic relations extraction from text. In: Gangemi, A., Keizer, J., Presutti, V., Stoermer, H. (eds.) SWAP. CEUR Workshop Proceedings, vol. 426, CEUR-WS.org (2008)
8. Gruber, T.R.: Towards Principles for the Design of Ontologies Used for Knowledge Sharing. In: Guarino, N., Poli, R. (eds.) Formal Ontology in Conceptual Analysis and Knowledge Representation. Kluwer Academic Publishers, Deventer (1993)
9. Hearst, M.A.: Automatic acquisition of hyponyms from large text corpora. In: Proceedings of the 14th International Conference on Computational Linguistics, pp. 539–545 (1992)
10. Jurgens, D., Mohammad, S., Turney, P., Holyoak, K.: Semeval-2012 task 2: Measuring degrees of relational similarity. In: *SEM 2012: The First Joint Conference on Lexical and Computational Semantics – vol. 1: Proceedings of the Main Conference and the Shared Task, and vol. 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012), June 7-8, pp. 356–364. Association for Computational Linguistics, Montréal (2012), `http://www.aclweb.org/anthology/S12-1047`
11. Kendall, M.G.: A new measure of rank correlation. Biometrika 30(1/2), 81–93 (1938)
12. Klaussner, C., Zhekova, D.: Lexico-syntactic patterns for automatic ontology building. In: Temnikova, I., Nikolova, I., Konstantinova, N. (eds.) RANLP Student Research Workshop, pp. 109–114. RANLP 2011 Organising Committee (2011)
13. Maynard, D., Funk, A., Peters, W.: Sprat: a tool for automatic semantic pattern-based ontology population. In: International Conference for Digital Libraries and the Semantic Web (2009)

14. Mititelu, V.B.: Hyponymy patterns in romanian. Memoirs of the Scientific Sections of the Romanian Academy, vol. XXXIV, pp. 31–40 (2011)
15. Montiel-Ponsoda, E., Aguado de Cea, G.: Using natural language patterns for the development of ontologies. In: Researching Specialized Languages, pp. 332–345 (2008)
16. Negri, M., Kouylekov, M.: Fbk_nk: A wordnetbased system for multiway classification of semantic relations. In: Proceedings of the 5th International Workshop on Semantic Evaluation, SemEval 2010, pp. 202–205. Association for Computational Linguistics, Stroudsburg (2010), http://dl.acm.org/citation.cfm?id=1859664.1859708
17. Snow, R., Jurafsky, D., Ng, A.Y.: Learning syntactic patterns for automatic hypernym discovery. In: NIPS (2004)
18. Solís, S.: La Web Semántica. Lulu Enterprises Incorporated (2007)
19. Tovar, M., Pinto, D., Montes, A., Vilariño, D.: Determining the degree of semantic similarity using prototype vectors. In: Carrasco-Ochoa, J.A., Martínez-Trinidad, J.F., Rodríguez, J.S., di Baja, G.S. (eds.) MCPR 2012. LNCS, vol. 7914, pp. 364–373. Springer, Heidelberg (2013)
20. Tovar, M., Reyes, J.A., Montes, A., Vilariño, D., Pinto, D., León, S.: Buap: A first approximation to relational similarity measuring. In: Proceedings of the 6th International Workshop on Semantic Evaluation (SemEval 2012), June 7-8, pp. 502–505. Association for Computational Linguistics, Montréal (2012), http://www.aclweb.org/anthology/S12-1071
21. Turney, P.D.: Similarity of semantic relations. Computational Linguistics 32, 379–416 (2006)
22. Volkova, S., Caragea, D., Hsu, W., Drouhard, J., Fowles, L.: Boosting biomedical entity extraction by using syntactic patterns for semantic relation discovery. In: 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), vol. 1, pp. 272–278 (2010)
23. Zouaq, A., Gasevic, D., Hatala, M.: Linguistic patterns for information extraction in ontocmaps. In: Blomqvist, E., Gangemi, A., Hammar, K., del Carmen Suárez-Figueroa, M. (eds.) WOP. CEUR Workshop Proceedings, vol. 929, CEUR-WS.org (2012)