Proceedings of the World Congress on Engineering and Computer Science 2017 Vol I
WCECS 2017, October 25-27, 2017, San Francisco, USA

Improving Knowledge Extraction from Texts by Generating Possible Relations

N.F.Nabila, Nurlida Basir, A.Mamat, and Mustafa Mat Deris

***Abstract*—Existing research focus on extracting the concepts and relations within a single sentence or in subject-object-object pattern. However, a problem arises when either the object or subject of a sentence is "missing" or "uncertain", which will cause the domain texts to be improperly presented as the relationship between concepts is no extracted. This paper proposes a solution for the enrichment of the knowledge of domain text by finding all possible relations. The proposed method suggests the appropriate or the most likely term for an uncertain subject or object of a sentence using the probability theory. In addition, the method can extract the relations between concepts (i.e. subject and object) that appear not only in a single sentence, but also in different sentences by using a synonym of the predicates. The proposed method has been tested and evaluated with a collection of domain texts that describe tourism. Precision, recall, and f-score metrics have been used to evaluate the results of the experiments.**

***Index Terms*—Relation Extraction, Non-taxonomic, Ontology**

I. Introduction

THIS taxonomic relation represents all relations other than an "is-a" relation that exists in texts. In extracting a non-taxonomic relation, three approaches have been used. The first approach is to extract relationships between predefined entities such as person, location, organization, and the like. All instances of entities are identified using name entities tagging and these are used to determine the contextual patterns to label the relationships between the entities. There are several works that use this approach to identify non-taxonomic relation, such as [17], [20], [8] and [7]. [17] presented a technique called AutoSlog that automatically generates conceptual dictionaries. This system used CIRCUS and several heuristic tagged noun phrases. [20] proposed a system called EXDISCO, which extracts a group of verbs that appear with identified classes. [8] also developed a system that labels the relations between clusters of entities pairs. In this system, entities that co-occur in the same sentence are first identified and then the similarities of entities are clustered. However, this approach depends on predefined entities and does not cover the whole concept for the domain.

The second approach is to extract concepts and concept pairs, which hold a given relationships such as part-of and cause-effect relations. In this approach, a set of patterns (i.e. two nouns in a sentence that hold the given relations), is extracted. [3], [5] and [6] are some examples of works which used this approach. Both [3] and [6] identified part-of relations while, [5] presented a method to identify cause-effect relations. However, this approach depends on pre-defined relations only and the number of relationships is limited and does not properly represent the domain texts.

The third approach is an approach to extract the concepts and relations and then assign appropriate relation between a pair of concepts. In this approach, there are two main tasks involved: (1) extracting the potential relation, and (2) relation labeling.

The extracting of the potential relation task is to discover which terms in the domain texts can be used as a potential relation to link between concepts. In this task, a term with a predicate (or verb) tag generally has been extracted as a potential relation for the domain. This is because in a sentence the predicate is used to describe a connection or a relation between two nouns (i.e. subject and object of a sentence). For example, a sentence "The student was reading a book". Based on this example, read is a predicate that describes the relation between student and book. Thus, read is selected as a potential relation for the domain text.

The relation labeling task is to determine which extracted potential relation is selected as the most appropriate relationships for the concepts. For example, predicates such as read, borrow, and write are extracted as the potential relations between the concept student and book. One of these potential relations could be the most appropriate relationship to show the relation between the concept student and concept book. In this task, the most appropriate potential relation with the concepts are computed by using several techniques such as association rule, log-likelihood, and chi-square.

The association rule is a technique to discover the association between data by computing the support and confidence value [1]. The support value shows the frequency of the item in the database. Meanwhile, the confidence value shows the weight of the rule. [12], [19] and [18] are some works that used the association rule technique to identify the association between predicate and concept pairs in domain texts. Chi-square is a statistical technique to determine the significant difference between the expected frequencies and the observed frequencies in one or more categories. [4] and

Manuscript received July 24, 2017; revised August 2, 2017. This work was supported in part by the Universiti Sains Islam Malaysia under Grant PPP/USG-0116/FST/30/11616.

N.F.Nabila and Nurlida Basir are with the Universiti Sains Islam Malaysia (USIM),71800 Nilai, Negeri Sembilan, Malaysia (e-mail: {fatin, nurlida} @usim.edu.my).

A.Mamat is with the Universiti Putra Malaysia, 433000 UPM Serdang, Selangor, Malaysia (e-mail: ali@upm.edu.my).

Mustafa Mat Deris is with Universiti Tun Hussein Onn Malaysia, Batu Pahat, 86400, Johor, Malaysia (e-mail: mmustafa@uthm.edu.my).

[9] have used chi-square techniques to determine the significant ranking of concepts and predicates in the domain texts against those in the general corpus and computed the co-occurrence of predicate and concepts. Log-likelihood is a statistical technique to measure two models. [14] and [15] have used the log-likelihood technique to identify a valid verb (i.e. independence or dependence of verb) that co-occurs with two concepts in a sentence.

All the techniques are used to determine the most suitable relationship for the concepts. A predicate that has the highest degree of occurrences between concepts is selected as the most suitable relationship for those concepts. Since the purpose of this current study was to determine the association between the relations and the concepts where the concepts do not only appear in a single sentence, but also appear in different sentences, we used the association rule technique in order to determine the appropriate extracted relations.

Since the first and second approaches of extracting non-taxonomic relations depend on pre-defined concepts and relations, the number of relationships is limited and does not properly represent the domain texts. Therefore, this research work was based on the third approach because it is more preferable for extracting non-taxonomic relations from domain texts. Works of [12], [10], [2], [9], [19], [14], [15], [18], [16] as well as [13], are some examples of works that use the third approach. [10] presented a technique to identify relationships between two concepts in a single sentence. This technique extracts the verb that frequently occurs with two concepts to fulfill the pattern of verb-concept-concept (VCC(n)). [2] developed a method called Wanderlust to find semantic relations between two entities using dependency grammar patterns. [9] proposed a technique to identify Noun Phrases (NPs) that occur before a verb is selected as the subject of the sentence and NPs that occurs after the verb is selected as the object of the sentence. The verb that holds the place between the subject and object of the sentence is then extracted. Then, the extracted verbs are expanded by gathering all verbs occurring between the same patterns of a concept pair.

[12] and [19] are examples of research that identified the relation between two concepts where the concepts are referred to as an existing ontology concept. [12] proposed a technique to identify the relation (i.e. predicate phrases) between two concepts, where the concepts are ontology concepts that appear in the same sentence. [19] proposed a technique that identifies nouns as concepts and verbs that hold a place between two nouns, that occur in a single sentence, as a relationship. All nouns and verbs are identified by using parts-of-speech (POS) taggers that were applied to each sentence from the documents collection to fulfill the pattern: <term><verb><term>, where the terms are identified nouns that also exist in ontology concepts. In both works, if the concepts do not exist in ontology concepts, then the relation is not identified.

[14] and [18] also used the predicate as a relation between two concepts. In contrast to [12] and [19], these works do not refer to ontology concepts to find the relevant concepts. [14] proposed the Subject-Verb-Object (SVO) Triples method to identify non-taxonomic relations between two concepts, where the concepts must appear as the subject and the object of a sentence. They used MINIPAR dependency parser to determine the appearance of concepts. Then, the verb that occurred together with the concept pair was identified. [18] proposed a technique that used an NLP approach and data mining technique to identify potential non-taxonomic relationships from textual sources. [18] proposed a semi-automatic method called PARNT to extract non-taxonomic relations from texts. [16] proposed a framework that used A Nearly New Information Extraction System (ANNIE), Stanford dependency parser and association technique, to enrich ontology relations. The framework extracted relations between two concepts, where the concepts must appear as the subject and the object of a sentence, as similar to [14]. This work extracted non-taxonomic relationships of the domain of tennis sport collected from various sources. [13] proposed a hybrid framework that used four features such as the bag of word model, NLP approach, Lexical and semantic based UMLS, to extract relation from biomedical datasets collected from MEDLINE database. This work extracted all verb phrases that occur between treatments and disease entities in the sentence. However, all works only identify two concepts that appear as subject and object of a sentence as a concept pair and then only the predicate that occurs together with the concepts are identified.

The model proposed in this paper should improve previous works by addressing many of the limitations. For example, most existing works in non-taxonomic relations extracting predicate or verb phrases that link concepts that appear as subject and concepts that appear as objects in the same sentence as potential relationships. However, a problem arises when either the object or subject of a sentence is “missing” or “uncertain”, which will cause the domain texts to be improperly presented as the relationship between concepts is not extracted. Although gathering all predicates that occur between the same patterns of subject and object in the same group may be sufficient to describe the knowledge of a domain. The relations obtained only in a general view or top level of relationships between concepts is then identified.

II. The Method

This section describes the proposed method for extracting concepts and relationship for constructing the ontology from domain texts. Figure 1 shows the flow of the approach. This approach involves the following steps: (1) Concept and Predicate Extraction, (2) Generating Subject-Object Pair and (3) Relation Labeling.

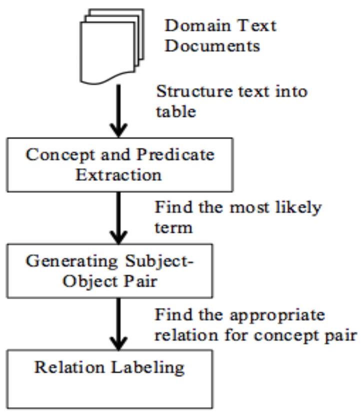

Fig. 1. Flow of the approach design

ISBN: 978-988-14047-5-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

A. *Concept and Predicate Extraction*

In this section, the pre-processing tasks and statistical analysis (i.e. term frequency-inverse document frequency metric, tf.idf) are used to extract relevant terms (as concepts) from the text. Then, the dependency pairs between these terms (i.e., grammatical relation between subject and/or object with the predicate) are identified using the Minipar shallow parser [11]. All identified terms and their relations are presented in the information table. In this table, all incomplete sentences with missing object or subject are highlighted by '*'. This "*" is known as an "uncertain" value of the sentences.

B. *Generating Subject-Object Pair*

This phase determines the most likely terms to replace the "uncertain" value of concept (i.e. subject or object) in order to complete the ontology component table. Two steps were involved to identify most-likely term: (1) Synonym Predicate Match, (2) probability of most likely term. First step is used when the predicates in the ontology component have synonym of predicates. Then, both predicate can be used as relations between the concepts. If the ontology component table still has uncertain value after Step 1, then second step is used to replace the uncertain subject or object with the most likely term suggested by probability theory.

As an example, the voting machine text was used as a case study. All sentences were extracted and presented in the ontology component table as shown in Table I. In Table I, u1, u2, ….., u10 represent sentences. C is an attribute of sentences that consist of subject, object and predicate of sentence.

TABLE I
A PART OF SENTENCES IN VOTING MACHINE DATASET

SentenceID	C		
	Subject	object	Predicate
u1	voter	machine	trust
u2	company	*	supply
u3	machine	paper	produce
u4	voter	*	check
u5	machine	record	produce
u6	*	record	produce
u7	governme	machine	provide
u8	voter	*	trust
u9	*	name	verify
u10	machine	record	evaluate

Step 1: Synonym predicate match

This phase identified the most-likely terms by using synonymous predicate as defined in Definition 4.1. Here, synonym predicate consists of different predicates, but have similar meaning. The synonymous of predicate is referred to WordNet.

Definition 4.1. Let $C_i = \{s_i, o_i, p_i\}$ be a complete regular sentence and $C_j = \{s_j, *, p_j\}$ be a irregular sentence in text. If (p_i **is equivalent to** p_j) then $* = o_i$.

Based on Definition 4.1, if p_i is equivalent to p_j, then this work handles two scenarios as follows:

(1) Scenario 1

For example, in Table I, the predicate supply in u2 is synonymous with the predicate provide in u7. Since the predicate are synonym, then in this work, both predicate can be considered as relations between the subject and object even when the subject and object appear in different sentences. Thus, the object in u7 (i.e. machine) is selected as most-likely term to replace the uncertain value of object in u2. Since the uncertain value of object in u2 is replaced with machine, then u2 will has one set of triples.

u2 = {*company, * , supply*}, u7= {*government, machine, provide*},
If (*supply* **is equivalent to** *provide*), then * = *machine*.
Thus, u2 = {*company, machine, supply*}.

(2) Scenario 2

For example, in Table I, both u4 and u9 are irregular sentences where u4 has uncertain value of object, while u9 has uncertain value of subject. Since the predicate check in u4 is synonymous with the predicate verify in u9, the object name in u9 is selected as most-likely term to replace the uncertain value of object in u4. While, the subject voter in u4 is selected as most-likely term to replace the uncertain value of subject in u9.

u4 = {*voter, * , check*}, u9= *{*, name, verify*}.
If (*check* **is equivalent to** *verify*), then * in u9 = *voter*, * in u4 = *name*
Thus, u4 = {*voter, name, check*}, u9 = {*voter, name, verify}*

Table II shows the updated incomplete ontology component table after step one. Based on table II, the table is incomplete ontology component table because the uncertain values still exist. Therefore, the second step, i.e. probability of most likely term, will be used in the following section.

TABLE II
INCOMPLETE ONTOLOGY COMPONENT TABLE AFTER PHASE 1

SentenceID	Subject	object	Predicate
u1	voter	machine	trust
u2	company	machine	supply
u3	machine	paper	produce
u4	voter	name	check
u5	machine	record	produce
u6	*	record	produce
u7	government	machine	provide
u8	voter	*	trust
u9	voter	name	verify
u10	Machine	Record	evaluate

Step 2: Probability of most likely term

In this step, the uncertain value can be calculated using the probability theory defined as follows:

ISBN: 978-988-14047-5-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Definition 4.2. The probability of most-likely term is denoted as $P(D_m)$. This can be defined as

$$P(D_m) = D/ |U|$$

where,

-D is probability that the same subject and object with predicate occur

-|U| is total number of sentences in ontology component table

The highest probability value was selected as most-likely term to replace the uncertain value.

The algorithm for selecting most-likely term using probability is given in Figure 2. In Figure 2, the algorithm used incomplete ontology component table as an input. The determination of most-likely term algorithm consists of several main steps. Each step details in the algorithm are described below.

Algorithm: Determination of most-likely term
Input: incomplete ontology component table
Output: most-likely term
Begin
Step 1. Calculate the probability for each attribute (i.e. subject, object, predicate)
Step 2. Calculate the probability of set C where C = {subject, object, predicate}
Step 3. Determine the probability of same subject and object pair
Step 4. Determine the most-likely term
End

Fig. 2 The determination of most-likely term algorithm

a) Calculating the probability of each attribute of set C.

The probability of each attribute of set C (i.e. subject, object and predicate) that occurs in a sentence will be calculated. Definition 4.3 is to calculate the probability of attributes that exist in a sentence. Definition 4.4 is used to calculate the attribute if the attribute has uncertain value.

Definition 4.3. Let C= {s, o, p}. If an attributes x has value other than '*', then P(x) = 1, i.e., the probability of x is 1.

Example 1. From Table II the probability of subject voter, P(voter) in u1 is

P(voter) = 1,

Similarly for probability of object machine, P(machine) in u1 is 1.

Definition 4.4. Let C= {s, o, p}. If there exists '*' value (i.e. uncertain value) for attributes, then the probability of not '*' attributes (i.e. subject/ object) in the table is calculated based on the formula below.

$$P(x) = (|(x)|/U, \; x \in C)$$

where

-x is not '*' value that exist in the correspondence attribute (subject/ object),

-(| |) represents the cardinality of the sets,

-U represents number of sentences in ontology component table.

Example 2. From Table II, the object (i.e. '*') of u8 is a missing value, thus the probability of other terms that could appear as object (i.e. machine, paper, name and record) are calculated as follows:

For the set of attribute object = {*, machine, paper, name, record},

P(*) = 1/10,
P(machine) = 3/10,
P(paper) = 1/10,
P(name) = 2/10
P(record) = 3/10

In this example, attribute object has uncertain value and other terms that appear as object in Table 4.4 are machine, paper, name and record. Thus, the probability of other terms that could appear as object was calculated.

b) Calculate the probability of set C

The probability that three independent events (i.e. subject, object and predicate) will occurs in a sentence, $P(C_i)$, is calculated by using a formula defined below.

Definition 4.5. The events s, o and p are independent, then

$$P(C_i) = P(s_i) * P(o_i) * P(p_i)$$

where

- *c* is the three independent attribute (i.e. subject, object, predicate)
- *i* is referring to number of sentence the triplet occurs
- Event *s* is referring to term that appear as subject in a sentence
- Event *o* is referring to term that appear as object of a sentence
- Event *p* is referring to term that appear as predicate of the subject

Example 3 Based on Table II, the probability of C for u1 and u8 are calculated as follows:

-For u1,

C = {voter, machine, trust}, $P(C_{u1}) = 1*1*1=1$

-For u8,

C = {voter, machine, trust}, $P(C_{u8}) = 1*3/10*1=3/10$
C = {voter, paper, trust}, $P(C_{u8}) = 1*1/10*1=1/10$
C = {voter, name, trust}, $P(C_{u8}) = 1*2/10*1=2/10$
C = {voter, record, trust}, $P(C_{u8}) = 1*3/10*1=3/10$

In this example, the decision value for u1 is 1 and u8 has four values since there exists four probabilities of terms (i.e. machine, paper, name and record) that appear as object in Table II to replace the object is '*'.

c) Determine the probability of same subject and object pair

The probability that has the same subject and object with predicate in step b was identified using the formula below:

$$D = \sum_{\text{same } s,o} P(C_i)$$

ISBN: 978-988-14047-5-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Figure 3 show the result of step c for table II.

SentenceID	Subject-object	Predicate	D
u6	(Voter, record)	produce	4/10
	(machine, record)		13/10
	(company, record)		1/10
	(government, record)		1/10
u8	(voter, machine)	trust	13/10
	(voter, paper)		1/10
	(voter, name)		2/10
	(voter, record)		3/10

Fig. 3 The result of step 3

d) Determine the most-likely term

The most likely term for uncertain value was calculated by using Definition 4.3.

Example 4. Based on Fig. 3, the results of probability of most likely term for u6 are

(voter, record) ⇒ produce, P (D_m) = 0.4/10 = 0.04
(machine, record) ⇒ produce, P (D_m) = 1.3/10 = 0.13
(company, record) ⇒ produce, P (D_m) = 0.1/10 = 0.01
(government, record) ⇒ produce, P (D_m) = 0.1/10 = 0.01

Here, the probability of machine and record with predicate produce is the highest probability value. Therefore, machine is considered as most-likely term for uncertain value of object in u6. Table III shows the complete ontology component table for table I. In this table, all the uncertain values are replaced with the most likely terms for the uncertain value.

TABLE III
COMPLETE ONTOLOGY COMPONENT TABLE

SentenceID	Subject	object	Predicate
u1	voter	machine	trust
u2	company	machine	offer
u3	machine	Paper	produce
u4	voter	machine	check
u5	machine	Record	produce
u6	machine	record	produce
u7	government	machine	provide
u8	voter	machine	trust
u9	voter	name	verify
u10	machine	record	evaluate

C. Relation Labeling

In this phase, the support and confidence metric in association rule have been used to identify the most appropriate relations among concepts (i.e. subject-object pair). In this paper, the predicate that has the highest value of confidence for the subject-object pair is selected as the most appropriate relation for the pair. The confidence is high if the subject-object pair co-occurred frequently with the predicate in domain texts.

III. EXPERIMENT

For conducting the experimental evaluation, a tourism datasets was used. Tourism corpus was collected from Wikipedia websites and the Los Angeles Time website and consisted of 65 texts and over 29,000 words describing tourism.

To evaluate our method, the prototype based on the proposed method was developed using Java and Javascript. Two existing works, namely methods by [14] and [18], were developed and tested by using the same texts.

In this paper, all domain texts were given to the experts for them to identify all relevant relations of the domain texts manually and the results were used as benchmarks for the system. The results produced by the proposed solution, [14] and [18], were compared with the results produced by domain experts. The experiments' results were then analyzed using precision, recall, and F-measure metric to measure the relevancy and quality of the extracted relations.

Table IV shows the results for tourism domain texts. In this table, the domain expert 1 has identified 368 valid relations and the domain expert 2 has identified 305 valid relations. Based on Expert 1, the recall value for the SVO method [14] and [18] method were 31.79% and 34.51%. The proposed solution obtained 96.74% which is higher than the SVO and Serra method [18]. Meanwhile, the recall value for the proposed solution, SVO method and Serra method based on Expert 2 were 94.10%, 37.05% and 40.33%, respectively. Meanwhile, for precision value, the proposed method has achieved 78.07% (based on Expert 1) and 62.94% (based on Expert 2), which is slightly higher than the SVO and Serra method.

TABLE IV
THE EXPERIMENT RESULTS

Method	# of extracted relations	Expert 1: 368 relevance relations				Expert 2: 305 relevance relations			
		# of correct relations	P	R	F	# of correct relations	P	R	F
SVO Method	205	117	57.07%	31.79%	40.83%	113	55.12%	37.05%	44.31%
Serra Method	232	127	54.74%	34.51%	42.33%	123	53.02%	40.33%	45.81%
Proposed Method	456	356	78.07%	96.74%	86.41%	287	62.94%	94.10%	75.43%

Notes: P:precision, R: recall, F:f-score

Figure 4 shows the average of the F-score values based on Expert 1 and Expert 2. In this figure, the graph shows that the proposed solution is sufficient and capable to extract non-taxonomic relationships, even when the concepts (i.e. subject and object) occur in different sentences and for irregular sentences. Therefore, based on the evaluation results, it shows that the proposed method helps in the enrichment of the domain ontology to represent the corpus.

ISBN: 978-988-14047-5-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

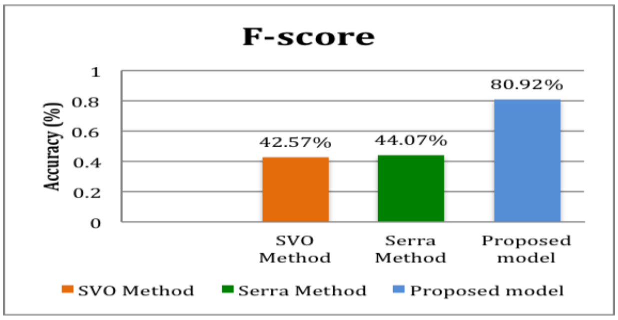

Fig. 4. The percentage of F-score by three methods

The existing works have used the predicate as the potential relation to representing relationships between concepts. However, most of the existing methods [14][15][18][19] focus on extracting potential relations between concepts (i.e. subject and object), and are only within the scope of a single sentence. Moreover, if the sentence did not fulfill the subject-predicate-object (S-P-O) pattern i.e., denoted as an irregular pattern, the existing extraction techniques will ignore it. Our work in contrast, focuses on extracting potential relations by using synonyms of the predicate. The synonyms of predicates are used as the reference to relate the concepts that occur not only in the same sentence, but also in different sentences. In addition, the proposed method is able to identify terms that it are more likely to replace the uncertain value by using probability theory and can be used to solve the issue of an irregular sentence pattern. The reliability of the extracted relations is evaluated based on the experts' judgment.

IV. Conclusions

In conclusion, this paper has shown that this proposed method provides improved knowledge extraction from domain texts. The benefit of this solution is that it may be used in the conceptualization process of the ontology engineering process to assist ontology engineers in extracting knowledge from domain texts. Therefore, the solution may be a useful and beneficial to obtain valuable information from the variety of sources of natural language text description such as journal articles, manuscripts, structured databases of any domain, which enable to facilitate big data analysis.

Acknowledgment

Universiti Sains Islam Malaysia (USIM).

References

[1] Agrawal, R., Imieliński, T., & Swami, A., "Mining association rules between sets of items in large databases," In ACM SIGMOD Record, 1993, Vol. 22, No. 2, pp. 207-216.
[2] Akbik, A., & Brob, J., "Wanderlust: Extracting semantic relations from natural language text using dependency grammar patterns," *In WWW Workshop*, 2009.
[3] Berland, M., & Charniak, E., "Finding parts in very large corpora," In Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics, Association for Computational Linguistics, 1999, pp. 57-64.
[4] Ciaramita, M., Gangemi, A., Ratsch, E., Saric, J., & Rojas, I., "Unsupervised Learning of Semantic Relations between Concepts of a Molecular Biology Ontology," In *IJCAI*, 2005, pp. 659-664.
[5] Girju, R., & Moldovan, D. I., "Text Mining for Causal Relations," In*FLAIRS* Conference, 2002, pp. 360-364.
[6] Girju, R., Badulescu, A., & Moldovan, D., "Learning semantic constraints for the automatic discovery of part-whole relations," In Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1, Association for Computational Linguistics, 2003, pp. 1-8.
[7] Greenwood, M. A., & Stevenson, M., "Improving semi-supervised acquisition of relation extraction patterns," In Proceedings of the Workshop on Information Extraction Beyond The Document, Association for Computational Linguistics, 2006, pp. 29-35.
[8] Hasegawa, T., Sekine, S., & Grishman, R., "Discovering relations among named entities from large corpora," In Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, Association for Computational Linguistics, 2004, pp. 415.
[9] Imsombut, A., "A statistical approach for semantic relation extraction," In Natural Language Processing, SNLP'09. Eighth International Symposium, IEEE, 2009, pp. 54-58.
[10] Kavalec, M., Maedche, A., &Svátek, V., "Discovery of lexical entries for non-taxonomic relations in ontology learning," In SOFSEM 2004: Theory and Practice of Computer Science, Springer Berlin Heidelberg, 2004, pp. 249-256.
[11] Lin, D., "Dependency-based evaluation of MINIPAR," In Treebanks, Springer Netherlands, 2003, pp. 317-329.
[12] Maedche, A., &Staab, S., "Semi-automatic engineering of ontologies from text," In Proceedings of the 12th international conference on software engineering and knowledge engineering, 2000, pp. 231-239.
[13] Muzaffar, A. W., Azam, F., & Qamar, U., "A Relation Extraction Framework for Biomedical Text Using Hybrid Feature Set," Computational and mathematical methods in medicine, 2015.
[14] Punuru, J., & Chen, J., "Extracting of non-hierarchical relations from domain texts," In Computational Intelligence and data mining, CIDM, IEEE Symposium, 2007, pp. 444-449.
[15] Punuru, J., & Chen, J. "Learning non-taxonomical semantic relations from domain texts," In Journal of Intelligent Information Systems, 2012, pp. 191-207.
[16] Ribeiro, "Extraction of non-taxonomic relations from texts to enrich a basic ontology", 2014.
[17] Riloff, E., "Automatically generating extraction patterns from untagged text," In Proceedings of the national conference on artificial intelligence, 1996, pp. 1044-1049.
[18] Serra, I., & Girardi, R., "A Process for Extracting Non-Taxonomic Relationships of Ontologies from Text," in Intelligent Information Management, 2011, 3(4).
[19] Villaverde, J., Persson, A., Godoy, D., &Amandi, A., "Supporting the discovery and labeling of non-taxonomic relationships in ontology learning," Expert Systems with Applications, 36(7), 2009, pp.10288-10294.
[20] Yangarber, R., Grishman, R., Tapanainen, P., & Huttunen, S., "Automatic acquisition of domain knowledge for information extraction. In Proceedings of the 18th conference on *Computational linguistics-Volume 2*, 2000, pp. 940-946.

ISBN: 978-988-14047-5-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)