**Proceedings of the 2007 IEEE Symposium on**
**Computational Intelligence and Data Mining (CIDM 2007)**

# Extraction of Non-hierarchical Relations from Domain Texts

Janardhana Punuru and Jianhua Chen

***Abstract*— Ontology of a domain mainly consists of concepts, hierarchical relations, and non-hierarchical relations. Even though there exists a variety of methods for extracting concepts and hierarchical relations, very little concentration is on identification and labeling of non-hierarchical relations. In this paper, we present an unsupervised technique for the identification of non-hierarchical relations between the concepts using VF*ICF metric and log-likelihood ratios. The proposed approach is experimented with the electronic voting domain texts and is also compared with one of the existing approaches.**

## I. Introduction

Ontology of a domain mainly consists of concepts, taxonomy, and semantic relations among the concepts. Ontologies are wide spread in use for various research fields like information retrieval, information extraction, conceptual mapping, and knowledge management. Recently, Semantic Web [2], proposed by Tim Berners-Lee inventor of the WWW, has got a great attention to make a new kind of representation for the WEB using XML, RDF, and OWL. Semantic Web aims at representing the textual content into ontologies, which allow machines to comprehend semantics of documents and return more accurate results to the user queries. Though there exists a wide variety of applications for ontologies, as of now, ontologies for the domains are developed manually. Since the manual construction of ontologies is costly both in time and labor, now various research projects are focused in automating the ontology construction process [15]. Concepts of a domain are often identified by extracting domain relevant terms occurring in texts of the domain. Various techniques are presented in literature [23] [18] [19] for concept extraction task. Either to extend or to construct the thesaurus, custom dictionaries, or lexical knowledge bases, considerable attention has been given for taxonomy extraction [11] [3] [4]. Even though various methods exist for extracting the concepts and taxonomic relations between the concepts, very little attention is focused on identification and extraction of non-taxonomic relations. Most of the existing techniques are focused on extracting concept pairs for a given relation type such as part-whole [8] [1] or cause-effect [9]. Very few techniques exists for identification of relationship labels for a given set of concepts.

In addition, extraction of non-taxonomic relations is also useful in Question Answering systems for answering queries such as *Who manufactures X?*, *What is written by X?*, and etc. Along with the Semantic Web and Question Answering systems, Information Retrieval, Information Extraction, and Tex Summarization systems often need to identify semantic relations of the domain.

We consider, in this research, the relations of the form $C_i \rightarrow Rl \rightarrow C_j$ as instances for non-taxonomic relations where $Rl$ is a relation name different from "IS-A". If concepts $C_i$ and $C_j$ are related and $Rl$ indicates the relationship from $C_i$ to $C_j$ then the ordered triple ($C_i$, $Rl$, $C_j$) is considered as a valid non-taxonomic relation of the domain. For example, the triple (voter, cast, ballot) indicates a valid non-hierarchical relation. For simplicity from here on we use the word relations to refer to non-hierarchical relations. We consider the problem of identification of relations as two sub problems. One is identification of the concept pairs($C_i$, $C_j$) such that some relationship holds from $C_i$ to $C_j$. And the other is identification of labels for the relations from $C_i$ to $C_j$. Concept pairs are obtained based on the position of occurrence in domain texts. Candidate relationship labels are identified using the VF*ICF metric. And log-likelihood ratio method is used to assign the relation labels between the concepts. The main advantage of the proposed method is that it is completely an unsupervised technique. That is, it does not require any pre-labeled training data. Also, it is a domain independent approach and also does not use any external knowledge bases like WordNet [17].

The rest of the paper is organized as follows: the following section gives a brief review on the existing methods for the extraction of relations. In section III, we present an overview of the the proposed method and a description on extraction of concept pairs. Section IV discusses the method for extracting the candidate labels for the relations. The log-likelihood method for the relationship label assignment is presented in Section V. Section VI puts all the steps together for relations extraction. Section VII provides the experimental results of the proposed method. Sections VIII and IX gives future directions and conclusions of the paper respectively.

## II. Related work

One of the least tackled problems in ontology learning is extraction of non-hierarchical relations from domain texts. Very few works are concentrated in finding the relations between concepts. Most of the existing works are focused on finding relations between named entities [10] [22] [24] [20]. Identification of relations between named entities concentrates on only a fixed set of predefined entities such as *person*, *location*, *organization*, and etc. These entities are fixed irrespective of the domain. Since the actual concepts vary based on the domain, we believe the techniques developed for learning

The authors are with the Computer Science Dept, Louisiana State University, Baton Rouge, LA-70803, USA. Email: {jpunur1, jianhua}@csc.lsu.edu
This research was partially supported by National Science Foundation grant: IIS-0326387 and AFOSR grant: FAA9550-05-1-0454.

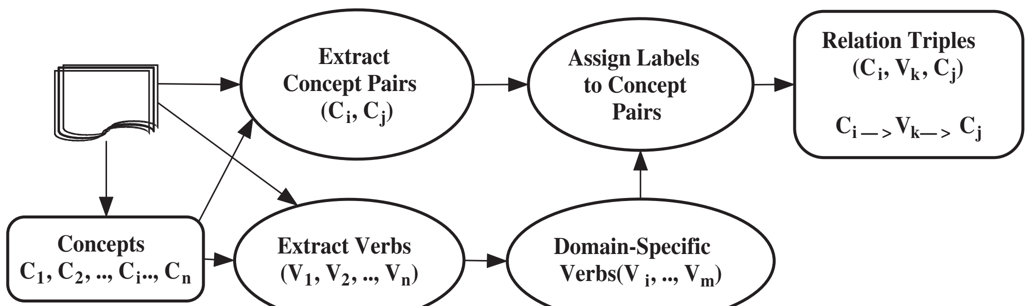


Fig. 1. **Framework for Relations Extraction**

ontological relations should not rely on the predefined entities or concepts. Very few methods are exist for identification of relations among the non-predefined concepts [7] [5] [13] [21]. In [7], Faure et al considers relations extraction problem as learning of selection restrictions for verbs. In this method all terms occurred along with a verb are clustered and each of the clusters are manually labeled. Whereas methods presented in [5], [13] and [21] exploit the syntactic structure and dependencies between the words for relations extraction. Both [5] and [21] extracts concept pairs which are in pre-specified dependency relations and use the chi-square test to verify the statistical significance of the occurrence of concept pairs together. Alexander et al's [21] technique builds the relation triples by extracting relevant pairs(Predicate and concept pairs). This technique used football domain texts for the experimentation. This Ciaramita et al's work [5] is experimented with the Molecular Biology domain texts. In [5], chi-square test is employed to learn the patterns such as $SUBJ \rightarrow$ bind $\rightarrow DIR\_OBJ$. And learned patterns are used to to extract semantic relations. Kavalec et al's [13] approach, initially, forms candidate triples($C_1$, $V$, $C_2$) such that concepts $C_1$ and $C_2$ occur with in the predefined distance from $V$ in the domain text. Using the triples constructed, labels for the relations between the concepts are identified based on the above expectation measure defined in equation 1. This measure emphasizes that if the occurrence of a verb, $V$, with a given pair of concepts($C_1$, $C_2$) is greater than the its occurrence with the individual concepts then the verb $V$ is considered as the candidate label for the relationship between the concepts. Here the authors used the tourism domain texts for their experiments.

$$AE((C_1 \bigwedge C_2)|V) = \frac{P((C_1 \bigwedge C_2)|V)}{P(C_1|V).P(C_2|V)} \qquad (1)$$

Main constraint of the above approach is that AE measure method does not suggest the direction of the relationship. That is, it is not known whether the verb($V$) indicates the relationship $C_1 \rightarrow C_2$ or $C_1 \leftarrow C_2$. To overcome the problem of relationship direction identification, we considered concept pairs($C_1$, $C_2$) such that the relationship $C_1 \rightarrow C_2$ is ensured before assigning the label for the relationship. In this paper, we also compared the results of the above measure with our presented approach.

## III. Overview

Our approach for extraction of relations from texts involves the following steps.

1) Extraction of domain specific concepts.
2) Identification concept pairs($C_i$, $C_j$) such that $C_i$ and $C_j$ are related.
3) Extraction of the candidate labels, $Rl$, for the relations.
4) Assignment of labels, $Rl$, for the relations between the concepts.

In our approach, we initially extract all relevant concepts of the domain using information retrieval measures. Concept pairs are identified based on the position of occurrence of concepts in texts. Candidate labels for relations are identified from texts and are assigned to concept pairs to obtain the final relations. The overall system architecture for obtaining the relations is shown in Figure 1.

As mentioned before, there exists various statistical [18] [23] and syntactic based techniques [12] for the extraction of domain specific concepts. For this part of the work, high relevant terms occurred in domain texts are considered as candidate concepts of the domain. To extract high relevant terms, TF*IDF values are computed for each of the terms. Among the terms with high TF*IDF values, terms with fewer senses are considered as the candidate concepts of the domain. Detailed discussion on the above mentioned approach for concept extraction is presented in [19].

To identify concept pairs($C_i$, $C_j$) such that there exists a relationship from $C_i$ to $C_j$, we maintain two sets $CS$ and $CO$ of concepts. Here, set $CS$ consists of concepts which occur as subjects in sentences. Similarly, $CO$ consists of set of concepts which occur as objects in sentences. To determine the subject and object(s) of sentences, MINIPAR [14] shallow parser is used. MINIPAR[1] parser produces dependency relations with 88% precision and 80% recall. Dependency relations produced by MINIPAR are analyzed to identify the subject and object(s)

[1]www.cs.ualberta.ca/ lindek/minipar.htm

of sentences. From $CS$ and $CO$ sets, concept pairs of the form ($C_i$, $C_j$) are constructed using the following conditions:

1) $C_i \in CS$ and $C_j \in CO$.
2) There exists a sentence $S$ such that $C_i$ is subject and $C_j$ is an object of $S$.

Here, condition 1 ensures the direction of the relationship from $C_i$ to $C_j$. Condition 2 reduces irrelevant pairs from getting added to candidate pairs.

Extraction of candidate labels for the relationships and assignment of the labels for the relationships are described in the following sections.

## IV. Finding Candidate Relation Labels

To find the relations between concepts in concept pairs, we first identify candidate labels for the relations and then map the labels to concept pairs. This section describes the method employed to identify candidate labels for relations. It is quite intuitive to believe that verbs which occur along with concepts in sentences could be useful for labeling relationships. Thus, it is reasonable to consider frequent verbs as candidates for labeling the relations. But most of the high frequency verbs are of the form *do*, *is*, *have*,..etc; which do not signify much semantic information of the domain. To find the domain-specific verbs, we defined VF*ICF metric, similar to the TF*IDF used in information retrieval, for finding the domain-specific verbs as in shown in equation 2. Informally, VF*ICF metric can be explained as follows. Verbs which occur with only a few set of concepts are more significant compared to the verbs which occur along with all of the concepts.

$$VF * ICF(V) = (1 + logVF(V)) * log(\frac{|C|}{CF(V)}) \quad (2)$$

TABLE I
Top 10 verbs with high VF*ICF value

| Verb(V) | VF*ICF(V) |
|---|---|
| produce | 25.010 |
| check | 24.674 |
| ensure | 23.971 |
| purge | 23.863 |
| create | 23.160 |
| include | 23.160 |
| say | 23.151 |
| restore | 23.088 |
| certify | 23.047 |
| pass | 23.047 |

In equation 2, $|C|$ is the total number of concepts, $VF(V)$ is the count of the occurrence of verb $V$ in domain texts and $CF(V)$ is the count of the concepts with which the verb $V$ is associated. A verb $V$ is considered to be associated with concept $C$, if both of them occur in a sentence. Table I shows top 10 verbs with corresponding VF*ICF values. Evaluation of VF*ICF metric for identification of domain specific verbs is presented in the experiments section.

## V. Assignment of Relation Labels

Another component in identification of relations from domain texts is the assignment of labels for the relationships between concepts. Here we use domain specific verbs extracted using VF*ICF metric as candidates to label the relations for concept pairs. Assignment of labels for relationships between the concepts is performed using log-likelihood ratios. Before going to the details on the formulation for computing the log-likelihood ratios, here, we describe the terms used. Let $S(C_1, C_2)$ be the set of sentences in which both $C_1$ and $C_2$ occur. Similarly, let $S(V)$ as the set of sentences in which verb $V$ occurs. Let $n_C = |S(C_1, C_2)|$, $n_V = |S(V)|$, $n_{CV} = |S(V) \cap S(C_1, C_2)|$, and $N = \sum_{i=1}^{n} \sum_{j,k=1}^{|C|} |S(V_i) \cap S(C_j, C_k)|$. Where $n$ is the count of domain-specific verbs and $|C|$ is the count of concepts in relevant concept pairs.

The log-likelihood ratios are computed with the assumption of hypotheses $H_1$ and $H_2$ separately. Here hypothesis $H_1$ formalizes that the occurrence of a verb $V$ is independent of the occurrence concept pair($C_1$, $C_2$). Whereas $H_2$ formalizes that the occurrence of $V$ is dependent on the occurrence of ($C_1$, $C_2$).

- **Hypothesis1 ($H_1$).** $P(V|(C_1, C_2)) = P(V|\neg(C_1, C_2))$
- **Hypothesis2 ($H_2$).** $P(V|(C_1, C_2)) \neq P(V|\neg(C_1, C_2))$

Now the log-likelihood ratio is computed using the equation 3.

$$log\lambda = log\frac{L(H_1)}{L(H_2)} \quad (3)$$

Assuming $H_1$ is true, $P(V|(C_1, C_2)) = P(V|\neg(C_1, C_2)) = p = \frac{n_V}{N}$. The likelihood of $H_1$ is

$$L(H_1) = b(n_{CV}; n_C, p)b(n_V - n_{CV}; N - n_C, p). \quad (4)$$

In the same way, assuming $H_2$ is true, $P(V|(C_1, C_2)) = p_1 = \frac{n_{CV}}{n_C}$ and $P(V|\neg(C_1, C_2)) = p_2 = \frac{n_V - n_{CV}}{N - n_C}$. The likelihood of $H_2$ is

$$L(H_2) = b(n_{CV}; n_C, p_1)b(n_V - n_{CV}; N - n_C, p_2) \quad (5)$$

In equations 4 and 5, $b(k; n, x) = \binom{n}{k} x^k (1 - x)^(n - k)$. $L(H_1)$ and $L(H_2)$ are computed assuming binomial distribution of the observed frequencies.

Similar formulation for collocation discovery using log-likelihood ratios is described in [16](§5.3.4§) and [6]. Since we want triples with high $L(H_2)$ and low $L(H_1)$ scores, we multiplied $\log \lambda$ with $-2$. It is also mentioned in [16] that if $\lambda$ is the likelihood ratio then the quantity $-2log\lambda$ is asymptotically $\chi^2$ distributed. For our purposes, we consider the triples($C_1, V, C_2$) with high $-2log\lambda$ score as valid non-taxonomic relations of the domain.

## VI. Putting It All Together

Subject-Verb-Object Triples (SVO Triples) method is developed, combining individual components described in the previous sections, to extract non-hierarchical relations from

Input: Candidate concept pairs($C_i$, $C_j$) of CP
For each pair ($C_i$, $C_j$) in CP
    maxLambda =0; relLbl = "";
    Extract set $L$ of labels associated with $C_i$ and $C_j$
    For each $v$ in L
        vLambda = -2*$log\lambda$ of ($C_i$, $v$, $C_j$);
        if maxLambda $<$ vLambda
            maxLambda = vLambda;
            relLbl = $v$;
    output ($C_i$, $v$, $C_j$);

Fig. 2. Procedure for Relationship Labeling

domain texts. As presented in Figure 2, the algorithm labels the relationship between the concepts for each of the concept pairs. For each concept pair($C_i$, $C_j$), set of candidate labels(L) are extracted. Among candidate labels, the label($v$) with highest log likelihood ratio is determined and assigned to the concept pair to output the triple ($C_i$, $v$, $C_j$).

## VII. Experiments

The presented approach for extraction of related concepts and the identification of relation labels is experimented with the *Electronic Voting* domain texts collected from *New York Times* website. From the voting domain texts, a total of 164 concepts are extracted from domain texts. Experimental results of VF*ICF metric for extraction of domain specific verbs and relationships assignment method(SVO) for concept pairs are shown as follows.

### A. Evaluation of VF*ICF Metric

Using the extracted concepts and the domain texts, for each of the verbs in the text VF*ICF scores are computed. We initially removed the stop words from the extracted verbs. From the remaining verbs, top 20% of them with high VF*ICF scores are considered as candidate labels for the relationships. To evaluate the performance of VF*ICF metric, each of the verbs are manually classified as either relevant or not. Whether a given verb is considered as relevant or not is determined based on the authors knowledge about of the domain. After the manual classification, the precision score for top 20% of verbs is 57%. To give an intuition on what kind of verbs we considered as relevant, some of the relevant and irrelevant verbs for relation labeling in *Electronic Voting* domain are shown in Table II.

From the observation of the results for VF*ICF metric, we believe that further research is needed in finding the labels for relations. We also think that using only verbs for labeling the relations may not be sufficient.

### B. Evaluation of SVO Method

Because of the lack of gold standard for identification of the conceptual relationships of the domain, it is difficult to verify the performance of the SVO method. To compute the recall for presented method, it is required to have all possible

TABLE II

Some of the relevant and irrelevant labels

| Relevant | Irrelevant |
|---|---|
| make | say |
| vote | try |
| produce | ensure |
| cast | know |
| certify | tell |
| install | help |
| count | believe |
| elect | want |

TABLE III

Examples of Concept Pairs

| Concept Pairs($C_i, C_j$) |
|---|
| (election, official) |
| (company, voting machine) |
| (ballot, voter) |
| (manufacturer, voting machine) |
| (polling place, worker) |
| (polling place, precinct) |
| (poll, security) |

relations of the domain. In this experiment we evaluate the performance of the methods using the accuracy of the results produced. Here accuracy is defined as the percentage of the relations obtained are correct. Further more, accuracy of the method is evaluated based on the following three constraints.

- **Constraint 1**. In a concept pair($C_1$, $C_2$), $C_1$ and $C_2$ are non-hierarchically related.
- **Constraint 2**. In a triple($C_1$, $V$, $C_2$), $V$ is the label for relation either $C_1 \rightarrow C_2$ or $C_1 \leftarrow C_2$.
- **Constraint 3**. In a triple($C_1 \rightarrow V \rightarrow C_2$), $V$ is a label for the relation from $C_1 \rightarrow C_2$ only.

Constraint **1** verifies whether the concepts in the concept pair are non-taxonomically related. Since SVO Triples method extracts concept pairs initially and then assigns the label for the relationship between the concepts, this evaluation is useful to verify the accuracy in extraction of concept pairs. Constraint **2** is useful for identification of whether the assigned label is a valid one for the relationship between the concepts in the concept pair without considering the direction of the relationship. Similarly constraint **3** verifies whether the direction of the relationship is maintained. Verification with respect to constraint **3** is also required because the direction of the relationships should also be maintained by the methods developed for automating the ontological relations extraction process. For example in the triple `(voter, cast, ballot)`, the label `cast` indicates the relationship from `voter` to `ballot` but not in reverse.

As mentioned in Section III, two concepts which occur together at least once in a sentence are considered as valid pairs. With the above notion, a total of 184 concept pairs are resulted. Of these pairs, top 20% pairs with high log-likelihood score are considered as candidate pairs. For illustration, some of the concept pairs such that their constituents are non-taxonomically related are shown in Table III.

TABLE IV
SVO TRIPLES METHOD EXAMPLE RESULTS

| Concept($C_1$) | Label($V$) | Concept($C_2$) |
|---|---|---|
| machine | produce | paper |
| voter | cast | ballot |
| voter | record | vote |
| official | tell | voter |
| voter | Trust | machine |
| worker | direct | voter |
| county | adopt | machine |
| company | provide | machine |
| machine | record | ballot |

TABLE V
EVALUATION OF AE AND SVO METHODS

| Method | ($C_1$,$C_2$) | ($C_1$, $V$, $C_2$) | ($C_1 \rightarrow V \rightarrow C_2$) |
|---|---|---|---|
| AE Measure | 89.00 | 6.00 | 4.00 |
| SVO Triples | 89.47 | 68.42 | 68.42 |

Now the resultant verbs with VF*ICF metric are used to determine the relation label for the candidate pairs. For each pair of concepts extracted, verbs which occur in at least one sentence along with the concepts in the pair and having high VF*ICF value are considered as candidate labels. Among all the candidate verbs, the verb with highest likelihood score is considered as the label for the relationship between the concepts.

In each of the candidate triples($C_1$, $V$, $C_2$) obtained using SVO method, $C_1$ has to be subject and $C_2$ has to be of object of verb $V$. And also $V$ has to have high VF*ICF score. Because of the above restrictions, very few(only 19) triples are resulted. Among the triples obtained, most of them are valid semantic relations. In SVO triples approach, even though very few relations are obtained, most of them satisfied the constraint **3** i.e. direction of the relationship maintained. For illustration, some of the obtained triples with SVO Triples approach are shown in Table IV.

According to each of the above constraints, we evaluated the SVO Triples approach. Its accuracy with respect to three constraints is shown in Table V.

In Table V, the initial column shows the method applied. Second column shows accuracy of the methods according to the constraint **1**. Similarly, columns 2 and 3 indicate accuracies of the corresponding methods with respect to constraints **2** and **3** respectively.

In Table V, first row shows the results of AE measure presented in [13]. It is also discussed briefly in section II. AE measure identifies the candidate triples($C_1$, $V$, $C_2$) such that $C_1$ and $C_2$ appear within a pre-defined distance(8 words) from $V$. We also implemented the AE measure and applied to our domain texts. From Table V, the results of AE measure indicate that even though it is able to extract related concepts with high accuracy, it performed very poorly in identification of the labels for the relations and in maintaining the direction of the relationship as well. From the results in column 2 of Table V, it is clear that VF*ICF measure useful for filtering some of the irrelevant relation labels.

We believe main reasons for such a low accuracy on finding the labels in AE measure are as follows. Concepts in some of the concept pairs occur more as part of compound terms in texts rather than connected by some verb. For example, the compound term `voting machine` occurred more often on its own than the concepts `voting` and `machine` are connected by some verb. Another reason is some of the concepts which occur together more often are connected by a preposition or a conjunction rather than a verb showing the relationship between them. For example, in the sentence `there were constant problems with the hardware and software`, the occurrence of concepts, `hardware` and `software`, does not signify semantic relation between them to label. At the same time, using the verbs occurred along with the concepts in the concept pair in a sentence may indicate the relationship between some other concepts rather than concepts in the pair.

Enforcing the conditions mentioned in III, most of the concept pairs obtained are indeed related. Among the obtained concept pairs, very few of them got invalid labels. The few invalid labels might have been obtained due to parse errors. Further more, all of the valid relations obtained using SVO Triples method maintained the direction of the relationship ($C_1 \rightarrow C_2$). Even though most of the relations obtained by SVO method are valid, SVO Triples method extracts only a small fraction of the total relations from domain texts. Hence SVO Triples method gives poor coverage. From the experiments, we believe that even though SVO method is useful for extracting semantic relations, it is not sufficient to find all of the relations of the domain. To further confirm the accuracy of SVO Triples method, experiment with larger domain texts(TREC data) is under progress. We believe further research is needed to find the relations and relation labels between concepts which does not occur as subject and object(s) in the texts.

## VIII. DISCUSSION AND FUTURE WORK

Even though SVO Triples method is able to identify relations with high accuracy, the count of relations obtained does not represent the whole domain. To improve the coverage of non-taxonomic relations, we believe even better methods are required to find the relations. Relationship labels obtained using VF*ICF measure has only 57% accuracy. this measure shows the importance of better methods for identification of domain specific verbs. We believe considering only verbs as candidates for relational labels is not sufficient. As part of the future work, we are interested in using words with other parts of speech tags also as the candidates for labeling the relations. In addition to using the Subject-Verb-Object triples for finding the relations between the concepts, we also want to consider the prepositional phrases as candidates for finding the relations. The main idea of extracting relations from prepositional phrases is as follows. In a given prepositional phrase, two concepts are linked by a preposition, the relationship between the concepts is labeled based on the semantic classes of the concepts and the associated preposition. For example, in the phrase "`hand recount of paper ballots`", semantic class of `hand recount` is "Action" and of `paper ballot` is "object". Using pre-learned patterns based on the semantic classes and the prepositions involved, relationship between the concepts is automatically labeled as one of the predefined relation labels.

## IX. CONCLUSIONS

In this paper, we presented SVO Triples technique based on the log-likelihood ratios for finding the non-taxonomic relations between the concepts. Domain specific verbs occurred along with concepts are considered as the candidates for labeling relations between the concepts. We defined VF*ICF metric to find the domain-specific labels for relations. Empirical evaluation of the SVO Triples method with respect to three different constraints is performed. Our method for finding the relations is also compared with one of the existing methods. From the experimental results, we conclude that SVO Triples method is useful for finding the non-taxonomic relations between the concepts. But using only SVO Triples method may not be sufficient to find all relations exists in texts.

## REFERENCES


[1] Berland, M. and Charniak, E., *Finding Parts in Very Large Corpora* In Proc. $37^{th}$ Annual meeting of Association for Computational Linguistics. pp 57-64.

[2] Berners-Lee, T., Hendler, J, & Lassila, O., "The semantic web", Scientific American 2001, pp.30-37.

[3] Caraballo, S. A., *Automatic Construction of a Hypernym-labeled Noun Hierarchy from Text* In Proc. $37^{th}$ Annual meeting of Association for Computational Linguistics.

[4] Cederberg, S. and Widdows, D., *Using LSA and Noun Coordination Information to Improve the Precision and Recall of Automatic Hyponymy Extraction* In Proc. of Conference on Natural Language Learning Edmonton, Canada 2003.

[5] Ciaramita, M., Gangemi, A., Ratsch, E., Jasmin, S., Isabel, R., *Unsupervised Learning of Semantic Relations between Concepts of Molecular Biology Ontology*, In Proc. of $19^{th}$ International Joint Conference on Artificial Intelligence, 2005.

[6] Dunning, T, *Accurate methods for the statistics of surprise and coincidence* in Computational Linguistics 19:61-74, 1993.

[7] Faure, D. and Nedellec, C., *A Corpus-based Conceptual Clustering Method for Verb Frames and Ontology Acquisition* In Proceedings of LREC Workshop on Adapting Lexical and Corpus resources to sublangauges and applications 1998 pp. 5-12.

[8] Girju, R., Badulescu, A., and Moldovan, D., *Learning Semantic Constraints for the Automatic Discovery of Part-Whole Relations* In Proc. HLT/NAACL-03 pp. 80-87 Edmonton Canada.

[9] Girju, R. and Moldovan, D. *Text Mining for Causal Relations* In Proc. of FLAIRS Conference 2002 pp. 360-364.

[10] Hasegawa, T., Sekine, S., Grishman, R., *Discovering Relations among Named Entities from Large Corpora* In Proc. of Association of Computational Linguistics 2004.

[11] Hearst, M. A., *Automatic Acquisition of Hyponyms from Large Text Corpora* in Proc. $14^{th}$ International Conference Computational Linguistics, Nantes France, 1992.

[12] Jacquemin, C., *A Symbolic and Surgical Acquisition of terms through variation*, in Symbolic Approaches to Learning for Natural Language Processing, pp. 425-438, January 1995.

[13] Kavalec, M., Maedche, A., Svatek, V., *Discovery of Lexical Entries for Non-taxonomic Relations in Ontology Learning*, in SOFSEM - Theory and Practice of Computer Science, Springer LNCS 2932 (2004).

[14] Lin, D., *MINIPAR: a minimalist parser*, In Maryland Linguistics Colloquium. University of Maryland, College Park. 1999.

[15] Maedche, A., Volz, R., *The Text-to-Ontology Extraction and Maintenance System*, in ICDM-Workshop on Integrating Data Mining and Knowledge Management, San Jose, California, USA 2001.

[16] Manning, C. D. and Schütze, H, *Foundations of Statistical Natural Language Processing*, MIT Press 1999.

[17] Miller, G., *WordNet: An on-line lexical database*, International Journal of Lexicography, vol 3(4) Special Issue 1994.

[18] Pantel P. and Lin, D., *A statistical Corpus Based Term Extractor*, In Stroulia, E. and Martwin, S. (Eds) AI 2001 Lecture Notes in Artificial Intelligence. pp. 35-46.

[19] Punuru, J. and Chen, J., *Automatic Acquisition of Concepts from Domain Texts* In Proc. of IEEE Granular Computing 2006.

[20] Riloff, E., *Automatically Generating Extraction Patterns from Untagged Text* In Proc. of $13^{th}$ National Conference on Artificial Intelligence, 1996, pp. 1044-1049.

[21] Schutz, A., Buitelaar, P., *RelExt: A Tool for Relation Extraction from Text in Ontology Extension*, In Proc. of $4^{th}$ International Semantic Web Conference (ISWC-2005) Galway, Ireland.

[22] Stevenson, M., *An Unsupervised WordNet-based Algorithm for Relation Extraction*, in Fourth International Conference on Language Resources(LREC-04), Lisbon, Portugal.

[23] Tomokiyo, T. and Hurst, M., *A Language Model Approach for Keyphrase Extraction* In Proc. of ACL 2003 workshop on Multiword Expressions: Analysis, Acquisition, and Treatment. PP 33-40.

[24] Yangarber, R., Grishman, R., Tapanainen P., Huttunen, S., *Unsupervised Discovery of Scenario-Level Patterns for Information Extraction*, In Proc. of Applied Natural Language Processing Conference 2000 Seattle WA.