Enrichment of Semantic Relation Networks in a Knowledge Base

Avvaru Padmaja[1], Kotha Mahesh[2], Kumara Swamy[3], Mastan Vali[4]

[1]*Assistant Professor, CSE Dept, Sree Dattha Institute of Engineering and Science, Hyderabad*
[2]*Assistant Professor, CSE Department, Tirumala Engineering College, Hyderabad*
[3]*Assistant Professor, CSE Dept, Sree Dattha Institute of Engineering and Science, Hyderabad*
[4]*Assistant Professor, Dept of IT, J.B Institute of Engineering and Technology, Hyderabad*

Abstract: **Enrichment of semantic relations at human level is an important. In this regard WordNet has been built to be the most systematic and as close to the human level and is being applied actively in various works by using semantic information processing in various works. It is observed that semantic gaps I real world of WordNet. Hence we tried to build an enrichment of knowledge base for automatic semantic relation network. We adapted are rule based, inference methods to enrich WordNet. After comparative study a word sense disambiguation method is constructed with the state of art algorithm.**

KeyWords: **Knowledge Base, WordNet, Semantic networks, Semantic information Processing.**

1. INTRODUCTION

The exponential growth of the amount of data in the World Wide Web (WWW) requires the automatising of processes like searching, retrieving and maintaining information. One of the difficulties that prevents the complete automatising of those processes is the fact that the contents in the WWW are presented mainly in natural language, whose ambiguities are hard to be processed by a machine. The Semantic Web (SW) constitutes an initiative to extend the web with machine readable contents and automated services far beyond current capabilities. A common practice is the annotation of the contents of web pages using ontology. Sometimes people are facing problems in understanding correct meaning of the sentence. Since, sentence comprised of ambiguous words. In such case, correct meaning is taken by the context of the sentence. Usually, it is found in English language. In other words, we can say that context uniquely identifies meaning of the sentence. Based on this interpretation the ambiguity of word, known as lexical ambiguity is disambiguated; which is called as a process of WSD. Manual method of meaning extraction uses approach of searching words correct meaning in typical or online dictionaries which had several drawbacks.

2. WORD SENSE DISAMBIGUATION

Word ambiguity removal is a task of removing ambiguity from a word, i.e. correct sense of word is identified from ambiguous sentences. This paper describes a model that uses Part of Speech tagger and three categories for word sense disambiguation (WSD). Human Computer Interaction is very needful to improve interactions between users and computers. For this, the Supervised and Unsupervised methods are combined. The WSD algorithm is used to find the efficient and accurate sense of a word based on domain information. The accuracy of this work is evaluated with the aim of finding best suitable domain of word. To resolve an ambiguity in a sentence, natural language processing provides word sense disambiguation which governs a sentence in which the sense of a word or meaning is used, when the word has multiple meanings (polysemy). WSD is a process which identifies the correct sense of a word with the help of surrounding words in a sentence.

The correct sense of a word is obtained from the context of the sentence. a different meaning of the single word is associated in each sentence based on the context, the remaining sentence gives us. Thus, if the word imagination appears near the word play, we can say that it is related to free time and not related to a sport which is known as local context. Computers that read words, one at a time must use word sense disambiguation process for finding the correct meaning (sense) of a word. A disambiguation process requires a dictionary in which senses are to be specified and disambiguated. For identifying the correct sense of the word the 'WordNet' domain is used. A domain consists of different syntactic categories of synsets. It groups senses of the same word into uniform clusters, with the effect of reducing word polysemy in WordNet. WordNet domain provides semantic domain as a natural way to establish semantic relations among word senses. This functionality is used in creation of MySQL database. The system for disambiguation of ambiguity in a sentence aims to identify domain of intended sense of word. Basically, input provided to the system is a sentence with ambiguous words and the output is identified as domain of word.

3. LITERATURE SURVEY

For Word sense disambiguation, the first attempt effectively used by Michael E. Lesk was based on the Dictionary approach. The problem with this algorithm is that, it defines context in a more complex way which is overcome by Simplified Lesk algorithm. It can be effectively used with the WordNet lexical database. Such an attempt is made at Indian Institute of Technology, Bombay [3] and the results are promising. Navigili had found that the right sense for a given word amounts to identifying the most “important” node among the set of graph nodes representing its senses. Ling Che Yangsen and Zhang described a general framework for domain adaptation which contained instance pruning and weighting and the training instance augmentation. Agirre described a thorough overview of the current WSD techniques and performance of systems on data sets, as well as a brief history of the field and some truly insightful discussions on potential developments. In we find the most general and well-known attempt to utilize information in machine-readable dictionaries for WSD, that of Lesk , which computes a degree of overlap that is, number of shared words--in definition texts of words that appear in a ten-word window of context.

4. SYSTEM MODEL

This paper focuses mostly on the first point. WordNet is the knowledge base (KB) that is most widely used for Semantic Information Processing (SIP). WordNet developed at Princeton University is a lexical database of English based on psycholinguistics and it has been continuously expanded since 1985. WordNet has been put to use for SIP such as semantic document indexing, semantic document topic detection, query expansion, ontology extension information retrieval by semantic similarity, and knowledge integration.

In this paper, an enrichment method to build a KB that can maximally reflect the semantic relation network used in the real world is proposed in order to minimize the semantic gap between humans and machines and an enriched WordNet (E-WordNet) is created as the result. Because KB is the criteria when a machine analyzes information and makes a decision, the research on automatic construction, extension, or enrichment of the KB must be 100 percent accurate. E-WordNet is enriched through two methods to maximally reflect such a characteristic.

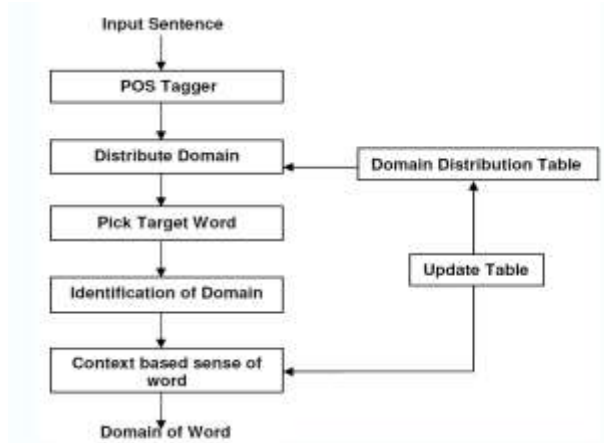

Fig. 1. An example of system model for WSD.

The enrichment of semantic relation networks in a Knowledge base must be carried out using accurate information. Methods that rely on probability and patterns may include false information. To obtain a KB without false information, the base information must be accurate and the method for grasping the senses of terms must not rely on probability or patterns. In order to achieve 100 percent accuracy, this research is divided into two; 1) a method that uses the WordNet’s glossary and 2) an inference method that uses the WordNet relation type axioms. Among the enrichments, the result of the second method can be applied to the first method but, it will be independently enriched because the objective of this research is to get the independent results from both enrichments. According to the experimental result, the direction of our future research will be determined. Fig. 2 schematizes the entire process for this research.

5. RELATED WORK

WordNet 2.1 defines 143,847 noun words using 81,426 senses. Between those senses, semantic network is formed with 203,762 relations. In addition, each sense has a glossary with natural language. This section describes the method for forming the glossary relation (“;gr”) and the inverse glossary relation (“-gr”) and the result. First, Part-Of-Speech tagging is needed to extract the noun words included inside the glossary. For this research, we used the pos-tagger2 2008-05-19 version that was developed by the Stanford NLP Group and recently updated [8], [9].

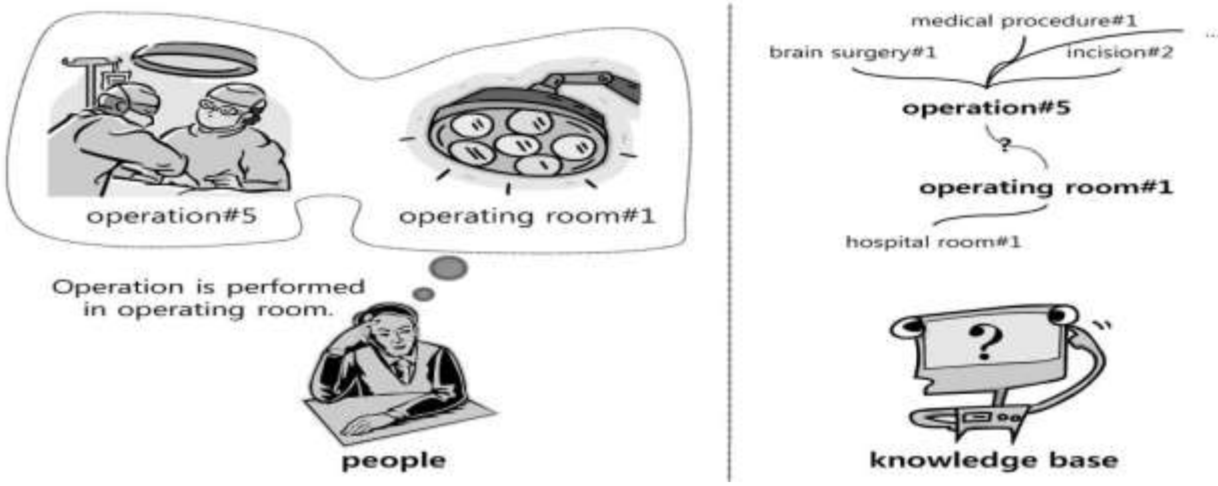

Fig. 2. An example of semantic gap between people and knowledge base.

To extract such nouns, we input glossary for each sense into the pos-tagger. This subchapter describes a method to accurately grasp the senses for the nouns in each glossary noun list that was extracted in the preceding part. Much existing research use the Word Sense Disambiguation (WSD) method to grasp the While the KB-based WSD and the maximum entropy-based WSD algorithm were applied for this research, an error rate of 15 percent or slightly higher were encountered. Applying these methods to semantic relation network enrichment was unreasonable because KB must guarantee purity.

Therefore, in this research, we design the method which grasps senses of a noun list based on pure clues. Concepts that have semantically close distance in relation network defined in WordNet can also have similar glossaries. In other words, words in glossaries can also have close relations in terms of the meaning. Therefore, word sense can be more easily grasped when integrating the glossary noun lists for neighboring concepts. For example, "operation#5" forms a semantic relation network in WordNet like in Fig. 3.

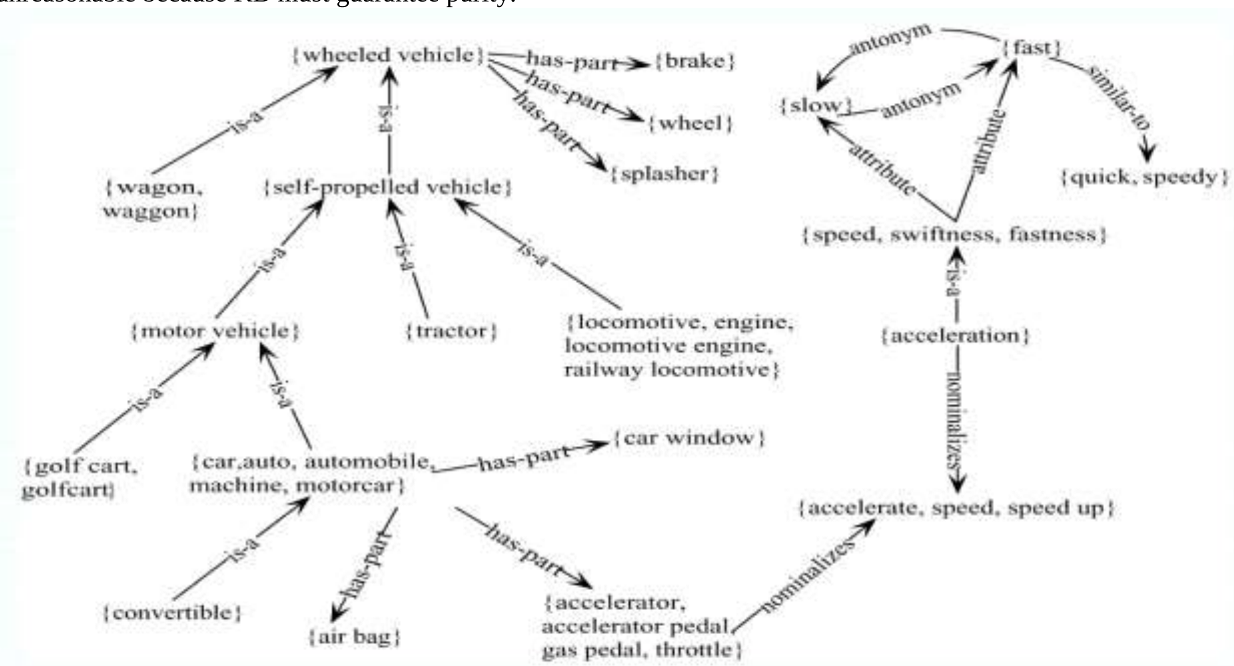

Fig. 3. An excerpt of the WordNet semantic network

6. CONCLUSION

The system improves the self-learning process by obtaining correct sense of a sentence by resolving ambiguity from a word with full automation. This paper describes the research to enrich the semantic relation network of WordNet and the WSD method using It proved possible to enrich WordNet's semantic network higher. For the coverage of human level knowledge, concept pairs that have general relations in the real world were selected and input to the three types of KB and to the KB of SSI. In addition, WSD-SemNet, a new WSD algorithm, was proposed and evaluated by applying each of the three types of KB and comparing the results with the SSI algorithm under identical conditions.

In the evaluation, E-WordNet which was enriched through this research could cover human level knowledge most satisfactorily and WSD-SemNet algorithm also exhibited better performance. The E-WordNet and the WSD-SemNet are expected to be used in a diverse manner in areas of knowledge-based SIP. These areas include semantic document indexing, document topic detection, query expansion, ontology extension, semantic information retrieval and knowledge integration and so on. Based on the E-WordNet and the WSD-SemNet, we are carrying out research on social network information mining, named entity definition and semantic document interconnection. In this system, when the number of target word is correctly disambiguated system gives 100% accuracy. Else, the accuracy may be 66% or 50%. Hence, the overall 80% accuracy is evaluated. These results generated by the system are beneficial for Human Computer Interaction as it is motivating people to learn the language by themselves using computer in the absence of teacher.

7. REFERENCES

[1]. Automatic Enrichment of Semantic Relation Network and Its Application to Word Sense Disambiguation Myunggwon Hwang, Chang Choi, and Pankoo Kim, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 23, NO. 6, JUNE 2011,PP-845-858.

[2]. An Improved Approach for Word Ambiguity Removal, Priti Saktel & Urmila Shrawankar, International Journal of Human Computer Interaction (IJHCI), Volume (3) : Issue (3) : 2012, pp-71-82 and respective references of this paper.

[3]. Automatic extraction of semantic relationships for WordNet by means of pattern learning fromWikipedia, Maria Ruiz-Casado, Enrique Alfonseca and Pablo Castells

[4]. Word Sense Disambiguation: A Survey, ROBERTO NAVIGLI, *Universit `a di Roma La Sapienza*, ACM Computing Surveys, Vol. 41, No. 2, Article 10, Publication date: February 2009.pp-10:1-10:69 .

Ms.Avvaru Padmaja, received M.Tech (CSE) from Jawaharlal Nehru Technological University (JNTUH),Hyderabad. Her research interest includes Data Mining, Network Security & Cloud Computing. Now she is working as an Assistant Professor in CSE Dept, Sree Dattha Institute of Engineering and Science**,** Hyderabad.

Mr. Kumara Swamy, received M.Tech (CSE) from Jawaharlal Nehru Technological University (JNTUH), Hyderabad. His research interest includes Cloud Computing, Data Mining & Network Security. Now he is working as an Assistant Professor in CSE Dept, Sree Dattha Institute of Engineering and Science**,** Hyderabad.

Mr. Kotha Mahesh, received M.Tech (CSE) from Jawaharlal Nehru Technological University (JNTUH), Hyderabad. His research interest includes Data Mining, Big Data, Pure Mining & Cloud Computing. Now he is working as an Assistant Professor in CSE Dept, Tirumala Engineering College, Hyderabad.

Mr. Mastan Vali, received M.Tech (CSE) from Jawaharlal Nehru Technological University (JNTUH), Hyderabad. His research interest includes Computer Networks & Security, Cloud Computing and Data Mining. Now he is working as an Assistant Professor in Dept of IT, J.B Institute of Engineering and Technology,Hyderabad.