A New Word Sense Disambiguation System Based on Deduction

S.M. Fakhrahmad, A.R. Rezapour, M. Zolghadri Jahromi, and M.H. Sadreddini

Abstract— **Word sense ambiguity resolution is one of the major issues in the process of machine translation. Statistical and example-based methods are usually applied for this purpose. In statistical methods, ambiguity resolution is mostly carried out by making use of some statistics extracted from previously translated documents or dual corpora of source and target languages. In this paper, we look at the problem from a different viewpoint. The proposed system consists of two main parts. The first part includes a data mining algorithm which runs offline and extracts some useful knowledge about the co-occurrences of the words. The second part of the system is an expert system whose knowledge base includes the set of association rules generated by the first part. For the inference engine of the expert system, we propose an efficient algorithm based on forward chaining in order to deduce the correct senses of the words. The performance of the system in terms of applicability and precision will be analyzed and discussed through a set of experiments.**

Index Terms— **Machine Translation, Ambiguity Resolution, Word Sense Disambiguation, Association Rule Mining, Expert Systems, Forward Chaining**

I. Introduction

Machine Translation is one of the most attractive and applied fields in natural language processing (NLP). Machine translation (MT) is the process of automatically analyzing a text in a source language and producing a text in a target language. To date, machine translation has met with limited success. Conventional machine translation systems used to adopt *rule-based* methods, in which grammatical and linguistic restrictions are applied for translation. However, rule-based machine translation systems have many shortcomings. One of the major issues is ambiguity resolution and meaning interpretation. Rule-based systems suffer from inability to select the most suitable equivalent translation in many cases [1].

Word sense ambiguity can be thought of as the most serious problem in machine translation systems. The human mind is able to select the proper target equivalent of any source language word by comprehension of the context. A human being may also automatically consider a group of words, rather than just one word, in order to understand the meaning of a sentence, even if the words of the group are not relevant. In order to simulate this behavior in a machine, a huge amount of data will be required as input and the output may still not be free from errors.

There are two other categories of translation methods namely, *example-based* and *statistics-based* approaches proposed to overcome the shortcomings of rule-based methods.

In example-based translation methods, a large set of translation samples (i.e., pairs of source text and its translation) are stored and used for similar translations. Example-based methods are mostly used in order to detect and translate expressions.

Statistics-based machine translation was firstly proposed by Warren Weaver in 1949. It was then re-introduced in more details by researchers of IBM's Thomas J. Watson Research Center in 1991. Today, this category of machine translation methods is widely-studied and has attracted the attention of many other researchers in the field of machine translation.

In statistics-based Translation methods translations are generated on the basis of statistical or probabilistic models whose parameters are extracted from the analysis of a bilingual corpus.

Statistical translation is based on the study of frequencies of various linguistic units, including words, lexemes, morphemes, letters, etc., in a sample corpus in order to calculate a set of probabilities, so that various linguistic problems such as ambiguity can be solved.

Although example-based and statistics-based techniques outperform rule-based methods [1], they still have their own problems. For example, both methods require a huge bilingual corpus which is difficult to be collected, stored and processed [2, 3]. The other problem is that there is really no efficient algorithm to extract knowledge from this large-scale amount of data, which is required to be used for ambiguity resolution and other related purposes.

In this paper, we look at the problem of word sense disambiguation from a different viewpoint and propose a new ambiguity resolution system. The main advantage of the proposed system is that it does not essentially require a huge and complete corpus to obtain a good performance. Moreover, it is very efficient since it usually disambiguates other ambiguous words existing in the context rather than the main ambiguous word in just one pass through the knowledge base. The system consists of two major parts. The first part is a data mining system which first extracts a set of valuable knowledge and then represents this knowledge in a novel format. The proposed method in this part tries to avoid extraction of redundant knowledge.

The second part of the system is an expert system used to resolve any translation ambiguity of different words in a

S.M. Fakhrahmad is with the Dept. of Computer Eng., Islamic Azad University, Shiraz branch, (corresponding author, phone: +98-9177038028; e-mail: mfakhrahmad@ cse.shirazu.ac.ir).

A.R.Rezapour, M.Zolghari Jahromi and M.H. Sadreddini are with the Dept. of Computer Eng. and IT, Shiraz University (e-mail: {Rezapour, zjahromi, sadredin} @Shirazu.ac.ir).

The partially financial support by the Iranian research institute for Information and Communication Technology (ICT) is kindly acknowledged.

sentence. The main advantages of the proposed system are its novelty, its efficiency and its high degree of accuracy.

The rest of this paper is organized as follows. Section 2 is devoted to the introduction of the related work in the literature. Sections 3 and 4, illustrate the two major parts of the proposed system. Experimental results are presented in Section 5. Finally, Section 6 concludes the paper.

II. Related Work

There are a lot of proposed methods for word sense disambiguation which follow supervised learning techniques, e.g., Naïve Bayesian [4], Decision List [5], Nearest Neighbor [6], Transformation Based Learning [7], Winnow [8], Boosting [9], and Naïve Bayesian Ensemble [10]. Among the mentioned methods, the method that uses Naïve Bayesian Ensemble has been reported to have the best performance for ambiguity resolution tasks with respect to data set used [10]. In order to determine the correct meaning of each ambiguous word, all of the above methods build a classifier, using features that represent the context of the ambiguous word.

Brown et al. (1991) proposed a corpora-based disambiguation method which can be applied in machine translation systems[11]. They use data from syntactically related words in the local context of the ambiguous word. In order to obtain statistical data, a word-aligned bilingual corpus is required.

Each occurrence of an ambiguous word should be labeled with a sense by asking a question about the context in which the word appears. The system was tested by translating 100 randomly selected Hansard sentences, each containing 10 words or less in length and obtained the accuracy of 45%.

In [12], Yarowsky et al. assumes that each word is located in a major category. In order to disambiguate word senses they have used the Roget's Thesaurus data set. By searching the hundred surrounding words as indicators of each category, the most probable category of a word can be determined. During the training phase, firstly, a stemming process is performed over all words in order to achieve more useful statistics. Then, by examining the hundred surrounding words for indicators of each category, the indicator words are obtained and weighted. The measure used as the weight of each indicator word is the log of word's salience as shown in (1)

$$\text{weight}(w \text{ for } cat) = \text{Log}(\Pr(w|cat)/\Pr(w)), \quad (1)$$

where *w* is an indicator word and *cat* stands for a category. Pr (w|cat) is the probability that *w* appears in the context of a word from the category *cat* and Pr(w) is the probability of the *w*'s occurrence in the corpus as a whole. For useful words, the computed weight, i.e., the log of salience will be greater than one.

The system proposed in [12] is not limited to particular word categories and works in a wide domain. This system achieves accuracy of between 72% and 99%. The first challenge of the system is that it cannot disambiguate topic-independent distinction words that occur in many topics. Another problem is that it does not consider the distance of words in the contexts it handles.

Another method for word sense disambiguation was proposed in [13] by Dagan et al. (1994). The method chooses the most probable sense of a word using frequencies of the related word combinations in a target language corpus. In this method, first of all, the system identifies syntactic relations between words using a source language parser and maps those relations to several possibilities in the target corpus using a bilingual lexicon. Two evaluations were carried out for this method, one using Hebrew sentences and the other using German sentences. The accuracy of the system was 91% and 78% for Hebrew and German sentences, respectively.

The other method of word sense disambiguation proposed in [14] by Justeson et al., uses syntactically or semantically relevant clues. This method disambiguates adjectives using only nouns that are combined by the adjectives. The system was evaluated on five of the most frequent ambiguous adjectives in English: 'right', 'hard', 'light', 'old', and 'short' on large sets of randomly selected sentences from the corpus that contained the adjectives and the accuracy of the system reached 97%. However, for adjectives which can be differently accompanied by the same noun, this method cannot be helpful in disambiguation.

The system presented by Ng and Lee (1996) in [15] is based on the Nearest Neighbor method. The prototypes are the instances of the ambiguous word in the training corpus, each containing the following features: singular/plural; POS tags of the current word; three words on either side; support for verbs, which have a different verbal morphological feature; a verb–object syntactic feature for nouns; and nine local collection features. These features are calculated for each instance of w in the sense-tagged training data. The results are stored as exemplars of their senses. By calculating the same feature vector for the current word and comparing by all the examples of that word, the given word is disambiguated choosing the closest matching instance. The accuracy of the system on test sets from Brown corpus and WSJ corpus was reported to be 58% and 75.2%, respectively. The results were calculated on a task including 121 nouns and 70 verbs, using fine-grained sense distinctions from WordNet.

The method presented by Brown et al. [11] requires a bilingual word-aligned corpus, which is costly to build. This is one of the challenges of this method, which makes difficult the applicability of the method to other pairs of languages.

The other method proposed by Mosavi et al. in [16] is somewhat the same as the method presented in [13] which uses a target language model. They use Persian as the target language and consider the co-occurrences of the multiple-meaning words in a monolingual corpus of the Persian language. By calculating the frequencies of these words in the corpus, the most probable sense for the multiple-meaning words is chosen. However, instead of considering syntactic tuples in the target language corpus, they consider just co-occurrences of certain words in that corpus without having a syntactic analysis for the corpus. In this method, no analysis is performed either for the source or the target language corpus from the syntactic viewpoint. The only task of the proposed algorithm, for gaining the required statistical information, is determining the nearest noun, pronoun, adjective, or verb to the ambiguous word, whether it is a noun, a verb, an adjective, or an adverb. When applying this method for the comparison of English and Persian, only a

ISBN: 978-988-19251-4-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

small portion of ambiguous words in English can be correctly translated into Persian.

III. Knowledge Extraction System

The first part of the proposed system is nothing but a two-step data mining algorithm by which a large-scale bilingual corpus is converted into a set of associations between words, phrases and concepts. In other words, the knowledge extracted from the corpus is represented by a set of association rules with different degrees of importance. The proposed algorithm is a statistics-based method. However, it is based on a novel idea and looks at the problem from a different viewpoint. Moreover, in this algorithm, we do not generate nor store all possible associations. We omit any discovered knowledge which can be deduced from others.

In this algorithm, we regard each sentence as a transaction in *Market Basket Data Analysis* problem [18]. Thus, the words included in a sentence play the role of purchased items in a transaction. The goal of the system is to discover the associations between the words in source language according to the meanings used for them in target language. The mining algorithm consists of two steps. As a preprocessing step, we have to find alignments between the words and meanings in the bilingual corpus. The result of this part is the set of all sentences (in source language) in which each word is coupled with a meaning (in target language). These sentences will then be used by the association rule mining algorithm to find co-occurrences of bilingual words and phrases.

A. Alignment

As a preprocessing step, each sentence in the parallel bilingual corpus is transformed to a sequence of *connection*s. A connection is defined as a word in the source language aligned with its current meaning in the target language. This is accomplished by comparing the sentences of the two parallel parts of the corpus and aligning each sentence to its translation.

Our method for sentence alignment is based on the Jaccard Similarity metric, which is defined as follows:

$$\text{Jaccard Similarity} = \frac{A \cap B}{A \cup B} \tag{2}$$

For each sentence in the source language corpus, using the Jaccard Similarity metric, we aim to find a sentence in the target language corpus which seems to have the most relevance and detect it as the translation of the sentence.

However, since we are going to measure the similarity between two sentences from two different languages, there is no intersection among the sets. In order to solve this problem, we substitute each word of the first sentence (i.e., the sentence which is in the source language) by the set of all of its possible senses in the target language. So, the sentence of the source language will be transformed into a set of words in the target language. The Jaccard Similarity can be computed for the sets, since they are now from the same language.

Finally, for each sentence, the most similar sentence is determined as the translation and the parallel corpus is converted to a set of connections, which is required for the next part.

B. Association Rule Mining

The algorithm applied in this section for mining frequent itemsets and generating association rules is FastARM, which was proposed in [19]. In this algorithm, two well-known metrics, namely *Confidence* and *Support* are used in order to filter the generated rules. The definitions of these metrics for a typical rule A→ B are as follow:

$$\text{Confidence}\ (A \rightarrow B) = P(B|A) \tag{3}$$

$$\text{Support}\ (A \rightarrow B) = P(A \cap B) \tag{4}$$

In this work, we indicate the fitness of any generated rule by a rule-weight whose value is simply calculated as the product of the evaluation measures, i.e., Confidence and Support:

$$W(A \rightarrow B) = \text{Confidence}\ (A \rightarrow B) * \text{Support}(A \rightarrow B) \tag{5}$$

This part of the system discovers association rules. These rules show the connections between the words from the source language and their meaning in the target language.

As an example, assume A and B are two words in source language, where *A* has three different senses in target language, namely m_{A-1}, m_{A-2} and m_{A-3} and the set of possible meanings for B are $m_{B-1,}$ m_{AB-2} and m_{B-3}. Now assume that we have a small bilingual corpus containing 5000 sentences. The word A occurs in 100 sentences with the meaning of m_{A-1}. Among these 100 sentences, 40 sentences contain the word B. In 35 cases of these 40 sentences, B means m_{B-1} and in 5 others it has the meaning of m_{B-2}. In this situation, the following pair of rules can be generated:

Rule_1: $A = m_{A-1} \rightarrow B = m_{B-1}$
Confidence= 35% , Support= 35/5000 = 7% , W_1= 0.0245
Rule_2: $A = m_{A-1} \rightarrow B = m_{B-2}$
Confidence= 5%, Support= 5/5000= 1% , W_2 = 0.0005
(W_i stands for the weight assigned to Rule_i)

In this part of the system, if for example the source language consists of 2000 words each having two possible meanings (on average), then there are totally 4000 possible items, for which we are going to mine the existing inter-associations.

IV. The Sense Disambiguating Expert System

In this section, the central engine of the proposed system which is an expert system is described. The goal of this expert system is to resolve the semantic ambiguity of multi-sense words. The knowledge base of the system contains the association rules generated from the previous part. The main strategy used by this system for ambiguity resolution is based on a forward chaining approach. The main advantage of the proposed method is its ability to find the correct meanings of several words in a single pass. Moreover, it can deduce the conceptual relationships among two or more words (which is needed for word sense disambiguation), even if the words have never co-occurred in the training corpus.

Assume that the system has been asked to translate a sentence that includes a set of ambiguous words. The process starts with the word which is not ambiguous (if any) or the word having the minimum ambiguity (i.e., the word having the fewest number of different senses). If the selected word has more than one meaning, we assume one of the meanings to be correct for the current context. The pair of the selected word and its meaning is then inserted into the working memory of the expert system as a new fact. By creation of the first fact in the working memory, the process of disambiguation starts following the forward chaining method.

In the chaining method, every generated fact (in the working memory) is compared with the left-hand side of all the rules (i.e., the rule premise) included in the knowledge base of the system. If there is a rule that can be fired by the current fact, the right-hand side of the rule is concluded and inserted as a new fact into the working memory. The concluded fact suggests a meaning for another word which has a conceptual relationship with the selected word and may be present in the sentence under investigation. This chaining process continues until we reach the desired knowledge or there are no more rules that can be fired by the current facts.

Since the rules included in the knowledge base have different weights, the new facts deduced by the chaining process do not have the same values. Some results are more precise or more valuable. In order to assign a score to each deduced piece of knowledge (showing how valuable it is), we propose and apply the following heuristic rule:

"For a fact deduced from the chaining process over a set of rules, compute the product of the weights of all the rules in the chain and assign the result as the score of the fact."

Example1: Suppose that we are going to translate a sentence containing three words A, B and C, all of which involve sense ambiguity. Assume the word A has 5 different senses (namely m_{A-1} to m_{A-5}), B has 2 senses (m_{B-1} and m_{B-2}) and there are 4 different meanings for C (namely m_{C-1} to m_{C-4}). Consider the set of rules shown in Figure 2 as the knowledge base of the system:

$A = m_{A-1} \rightarrow D = m_{D-1}$,	W= 0.6
$B = m_{B-1} \rightarrow C = m_{C-3}$,	W= 0.7
$B = m_{B-2} \rightarrow E = m_{E-2}$,	W = 0.4
$C = m_{C-3} \rightarrow D = m_{D-2}$,	W = 0.55
$B = m_{B-1} \rightarrow E = m_{E-1}$,	W= 0.82
$B = m_{D-2} \rightarrow D = m_{E-2}$,	W= 0.9
$C = m_{C-2} \rightarrow D = m_{D-3}$,	W = 0.65
$D = m_{E-2} \rightarrow A = m_{A-3}$,	W = 0.72
$E = m_{E-3} \rightarrow D = m_{D-1}$,	W = 0.8
$C = m_{C-2}$ AND $D = m_{D-2} \rightarrow A = m_{A-4}$,	W = 0.85

Fig.2. The knowledge base of the expert system used in Example1

The process of disambiguation starts by selecting the least ambiguous word (the word B in this example) and assuming one of its senses (m_{B-1}) to be correct. The forward chaining process will continue as shown in Figure 3 and the deduced word senses (of some other words) will be inserted into the working memory of the expert system.

$A = m_{A-1} \rightarrow D= m_{D-1}$
$B= m_{B-1} \rightarrow C = m_{C-3}$
$B=m_{B-2} \rightarrow E = m_{E-2}$ 1
$C=m_{C-3} \rightarrow D = m_{D-2}$
$B = m_{B-1} \rightarrow E = m_{E-1}$ 2
$D=m_{D-2} \rightarrow E = m_{E-2}$
$C = m_{C-2} \rightarrow D = m_{D-3}$ 3
$E = m_{E-2} \rightarrow A = m_{A-3}$
$E = m_{E-3} \rightarrow D = m_{D-1}$
$C=m_{C-2}$ AND $D =m_{D-2} \rightarrow A= m_{A-4}$

Fig.3. A view of the forward chaining process, supposing m_{B-1} as the correct sense for the word B

The deduced results in this example (i.e., m_{A-3} and m_{C-3}) are based on the initial hypothesis, where m_{B-1} was assumed to be the correct meaning of B. Using the proposed heuristic, the following values are computed for these two results:

$A = m_{A-3}$ value = 0.7 * 0.55 * 0.9 * 0.72 = 0.25
$C = m_{C-3}$ value = 0.7

The above process has to be redone for the next meaning of B (i.e., m_{B-2}). The chining process for this case is shown in Figure 4. The results obtained here (the meanings of two words, D and E) are relevant to the current context, because the input sentence includes none of the words D and E. Therefore, the results based upon the first assumption (the first meaning of B) are accepted.

$A = m_{A-1} \rightarrow D = m_{D-1}$
$B= m_{B-1} \rightarrow C = m_{C-3}$
$B=m_{B-2} \rightarrow E = m_{E-3}$
$C=m_{C-3} \rightarrow D = m_{D-2}$ 1
$B = m_{B-1} \rightarrow E = m_{E-1}$
$D=m_{D-2} \rightarrow E = m_{E-2}$
$C = m_{C-2} \rightarrow D = m_{D-3}$
$E = m_{E-2} \rightarrow A = m_{A-3}$
$E = m_{E-3} \rightarrow D = m_{D-1}$
$C = m_{C-2}$ AND $D = m_{D-2} \rightarrow A = m_{A-4}$

Fig.4. A view of the forward chaining process, supposing m_{B-2} as the correct sense for the word B

V. Experimental Results

One of the main challenges in evaluation of translation systems is the lack of a parallel (bilingual) aligned corpus. Thus, as a preprocessing step, we first designed a crawler which is able to browse through the world-wide web and

ISBN: 978-988-19251-4-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

download a set of English documents as well as their translation in Persian (mostly from some special sites such as Wikipedia). Using the algorithm proposed in Section 3-1, we align the English and Persian related documents at the sentence level.

In order to evaluate our system, we focused on 8 nouns, namely *palm*, *bass*, *crane*, *plant*, *motion*, *tank* and *match* all of which involve sense ambiguity. The reason of choosing these words is that we also have a benchmark corpus named TWA which includes a set of texts including the mentioned words in different senses. These ambiguous words have higher frequencies compared with other words in the texts and appear in different situations in the training data. Thus, we combined TWA with our constructed corpus.

After preparation of the corpus, we used the algorithm proposed in Section 3-1 and selected a set of sentences in English coupled with their translations in Persian to be stored as the training data. Then, using the rule mining algorithm presented in Section 3-2, a set of association rules was generated and stored as the knowledge base of the expert system. The algorithm was learned using more than 8000 parallel sentences.

The minimum support threshold (denoted by *MinSupp*) used in the mining algorithm in order to extract frequent itemsets was set at 0.05 after trying different values from 0.01 to 0.2. The best value of disambiguation precision (88%) was obtained for *MinSupp* values of 0.05 and 0.06.

In order to evaluate our approach, we selected 24 English sentences semi-randomly from the corpus as test set. We ensured that these sentences contained all of the mentioned ambiguous words (with multiple Persian translations). The senses (multiple translations) of the ambiguous words were obtained from the dictionary. The number of senses per test word ranged from 2 to 4, and the average was 3.

For our evaluation, we used two measures; *applicability* and *precision*. The *applicability* is the proportion of the ambiguous words that the algorithm could disambiguate. The precision is the proportion of the correctly disambiguated senses for the ambiguous word. For conflict resolution in the process of disambiguation, we evaluated all the three approaches introduced in Section 4.1, in turn. In order to compare the results with other disambiguation methods, we executed some of the existing corpora-based methods (the methods proposed in [11, 12, 13, 15, 16]) over the same data. The results are shown in Table1.

TABLE I
APPLICABILITY AND PRECISION RESULTS OF DIFFERENT WSD METHODS USING A 8000 SENTENCE CORPUS

	Degan method	Ng method	Mosavi method	The proposed method
Applicability (%)	90	83.3	90	91.3
Precision (%)	86.4	75.7	89.9	84.5

In order to evaluate the sensitivity of our approach to the size of the training corpus (in comparison with other approaches), we repeated our experiment twice more times by reducing the number of sentences in the training corpus to 6000 and 4000 sentences, respectively. In both experiments, we used the same test set as used in the first experiment. The results are shown in Tables 2 and 3.

TABLE II
APPLICABILITY AND PRECISION RESULTS OF DIFFERENT WSD METHODS USING A 6000 SENTENCE CORPUS

	Degan method	Ng method	Mosavi method	The proposed method
Applicability (%)	86.3	79.5	85.2	92.9
Precision (%)	80.8	73.2	90.1	79.4

TABLE III
APPLICABILITY AND PRECISION RESULTS OF DIFFERENT WSD METHODS USING A 8000 SENTENCE CORPUS

	Degan method	Ng method	Mosavi method	The proposed method
Applicability (%)	81.2	78.5	82.6	91.3
Precision (%)	77.4	73.7	84.9	72.8

VI. Conclusion

In this paper, we proposed a word sense disambiguation system. The proposed system consists of two main parts. As a preprocessing method, we constructed a bi-lingual corpus including English texts coupled with the translation in Persian. In the first part of the system, by applying our association rule mining algorithm, we extracted the required knowledge about the co-occurrence of the words in the form of association rules. In the second part of the system we developed an expert system and for its knowledge base we used the set of association rules generated by the first part. We used a forward chaining approach in the inference engine of the expert system in order to discover the correct senses of the ambiguous words in an inductive manner. The performance of the proposed system in terms of Applicability and Precision was encouraging compared to its counterparts. However, the main advantage of the proposed system as shown via the experiments is its relative independence on the size of the training corpus. Due to the inductive process followed in the disambiguation engine of the system, a huge and complete corpus is not essentially required for obtaining a good performance. In the proposed system, the semantic association among two typical words which have a semantic relevance can be deduced, even if they have never co-occurred in the training corpus. That's why the applicability of the method is promising even when the training corpus is not very large.

REFERENCES

[1] Statistical Post-Editing of a Rule-Based Machine Translation System, Proceedings of NAACL HLT 2009: Short Papers, pages 217–220, Boulder, Colorado, June 2009. c 2009 Association for Computational Linguistics, pp 217-224

[2] Arabic/English Word Translation Disambiguation using Parallel Corpora and Matching Schemes, th EAMT conference, 22-23 September 2008, Hamburg, Germany, pp 6-11.

[3] On the use of Comparable Corpora to improve SMT performance Sadaf Abdul-Rauf and Holger Schwenk, Proceedings of the 12th Conference of the European Chapter of the ACL, pages 16–23, Athens, Greece, 30 March – 3 April 2009. c 2009 Association for Computational Linguistics, pp 16-23

[4] Gale, K. Church, and D. Yarowsky.: A Method for Disambiguating Word Senses in a Large Corpus. Computers and Humanities, vol. 26, pp. 415-439 (1992).

[5] Yarowsky.: Decision Lists for Lexical Ambiguity Resolution: Application to Accent Restoration in Spanish and French. In Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics, pp. 88-95 (1994).

[6] T. Ng and H. B. Lee.: Integrating Multiple Knowledge Sources to Disambiguate Word Sense: An Exemplar-based Approach. In Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics, pp. 40-47 (1996)

[7] Mangu and E. Brill.: Automatic rule acquisition for spelling correction. In Proceedings of the 14th International Conference on Machine Learning pp. 187-194.H. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1985, ch. 4.

[8] R. Golding and D. Roth. A Winnow-Based Approach to Context-Sensitive Spelling Correction. Machine Learning, vol. 34, pp. 107-130 (1999).

[9] Escudero, Gerard, Lluís Màrquez & German Rigau.: Boosting applied to word sense disambiguation. Proceedings of the 12th European Conference on Machine Learning (ECML), Barcelona, Spain, 129-141 (2000).

[10] T. Pedersen.: A simple approach to building ensembles of Naive Bayesian classifiers for word sense disambiguation. In Proceedings of the 1st Annual Meeting of the North American Chapter of the Association for Computation Linguistics, pp. 63–69, Seattle, WA (2000)

[11] Brown, P. F., DellaPietra, S. A., DellaPietra, V. J., and Mercer, R. L. (1991). Word Sense Disambiguation Using Statistical Methods. Proceedings. Annual Meeting of the Association for Computational Linguistics, pp. 264–70.

[12] Yarowsky, D. (1992). Word Sense Disambiguation Using Statistical Models of Roget's Categories Trained on Large Corpora. Proceedings of 15th International Conference on Computational Linguistics, pp.454–60.

[13] Dagan, I. and Itai, A. (1994). Word sense disambiguation using a second language monolingual corpus. Association for Computational Linguistics, 20(4): 563–96.

[14] Justeson, J. J. and Katz, S. M. (1995). Principled disambiguation: discriminating adjective senses with modified nouns. Computational Linguistics, 21(1): 1–28.

[15] Ng, H. T. and Lee, H. B. (1996). Integrating Multiple Knowledge Sources to Disambiguate Word Sense: An Exemplar-Based Approach. Proceeding of the 34th Annual Meeting of the Association for Computational Linguistics (ACL-96), Santa Cruz.

[16] T. M. Miangah and A. D. Khalafi, Word Sense Disambiguation Using Target Language Corpus in a Machine Translation Systems, Literary and Linguistic Computing, Vol. 20, No. 2, 2005

[17] Schütze, H.: Automatic WS discrimination. Computational Linguistics, 24(1):97-124 (1998)

[18] Brin, S., Motwani, R., Silverstein, C., Beyond market baskets: generalizing association rules to correlations, In Proc. of the 1997 ACM SIGMOD international conference on Management of data, pp.265-276, Tucson, Arizona, United States, 1997.

[19] S.M. Fakhrahmad, M.H. Sadreddini and M. Zolghadri Jahromi, " Mining Frequent Itemsets in Large Data Warehouses: A Novel Approach Proposed for Sparse Data Sets", The 8th International Conference on Intelligent Data Engineering and Automated Learning (IDEAL'07), Birmingham, U.K , pp. 517-526, 16-19 December 2007.