Comparative Analysis of Various Word Sense Disambiguation Techniques

Shweta Vikram and Sanjay K. Dwivedi
Babasaheb Bhimrao Ambedkar University, Lucknow_India

Abstract: Due to increase in digital content, new technologies and approaches are required to properly access the online documents. Due to language barrier, many of the relevant documents in a particular context become difficult to access and explore to the users and researchers. Machine translation tools are designed to bridge this language gap, However the ambiguity issue may cause the major hurdle in the translation process that might affect meaning of translated text. The Word Sense Disambiguation (WSD) aims to provide solution to this problem through various algorithms.

This paper tries to critically elaborate various features and the performance of popular WSD approaches under the two broad category of approaches i.e. knowledge based and machine learning based, that may provide solution to the ambiguity in natural text so that the translations between pair of languages can be improved.

Index Terms- **Supervised Classification, Text Mining, Word Sense disambiguation, Text Ontology.**

I. INTRODUCTION

Word Ambiguity is a challenging task in almost Natural Language Processing (NLP) based application and Word Sense Disambiguation (WSD) is a research area which helps in appropriately handling the ambiguity issue. WSD aims to automatically identify the correct sense of a word in a particular context by applying a suitable technique. This problem persists since a long in NLP which lead to many researchers to make Machine Translation (MT) projects development meaningfull

WSD can be at the level of coarse grained (homonymy) or can be fine grain (polysemy). The polysemous words usually require in-depth knowledge of the context to resolve ambiguity. Open-class words frequently have numerous implications, either because of polysemy or to homonymy. Among the example of such words is: "Fan", the fan has stop working. I am a devoted fan of super power hero films. Another popular example is "bank"; it can mean the land beside a river, or an economic organization. In Hindi language as well there are numerous examples of ambiguity, for example, □□ (hal) can mean □□□□ □□□□□ □□ □□ □□□□□ (an apparatus used to plough discipline) or □□□□□□/□□□□□□ (solution).

Thus, word senses provide an entry into world knowledge (in the shape of ontologies, for instance) that can be used to enrich the text and make it, to a certain extent, comprehensible to a machine. Such information is useful when establishing co-reference in texts, identifying lexical chains, etc. WSD is applicable to many other application areas including of word relations from source to target language. It is also a rich resource of information for building information extraction systems (rule-based or statistical), for information retrieval, question answering, etc.

Whole paper is organized as follows. The second section provides various researchers works in this area, while third section discusses various techniques of WSD, fourth section provides a comprehensive discussion followed by summary in section five.

II. RELATED WORK

In [36] authors proposed one model which is first to incorporate the glosses into an end-to-end neural WSD model. In this way, model can benefit from not only massive labeled data but also rich lexical knowledge. In order to handle semantic relationship of context and glosses, author proposes a glossaugmented neural network (GAS) in an improved memory network paradigm. Further expand the gloss through its semantic relations to enrich the gloss information and better infer the context. Finally author extend the gloss module in GAS to a hierarchical framework in order to mirror the hierarchies of word senses in WordNet.

In [37] Basile et al. (2014) utilize a distributional way to deal with definitions and the setting of the objective word. They make semantic vectors for sparkles and settings to process closeness of the objective word and the setting of an objective word, while this work likewise figure the comparability of a sense and its setting specifically utilizing sense embeddings.

In [38] paper, present data sense, an unsupervised framework for word sense disambiguation. In this work, given a sentence, the framework picks the most significant feeling of each information word regarding the semantic comparability between the given sentence and the synset comprising the feeling of the objective word. data sense has two methods of activity. The inadequate mode utilizes the customary vector space model to evaluate the most comparable word sense relating to its specific circumstance. The thick mode, rather, utilizes synset embeddings to adapt to the sparsity issue. We portray the design of the present framework and furthermore direct its assessment on three diverse lexical semantic assets for Russian.

In [39] present an abstract of 110 Statistical Machine Translation frameworks worked from parallel corpora of 11 Indian dialects having a place with the Indo-Aryan and Dravidian families. The authors break down the connection between interpretation precision and the dialect families included. For their investigations, they constructed express based frameworks and a few expansions. Over various dialects, they demonstrate enhancements for the benchmark expression based frameworks utilizing these augmentations: (1) source side reordering for English-Indian dialect interpretation, and (2) transliteration of untranslated words for Indian dialect Indian dialect interpretation. These improvements outfit shared attributes of Indian dialects. To invigorate comparable advancement broadly in the NLP people group, they have made the prepared models for these dialect combines openly accessible.

In another work[40], authors exhibit a graphical UI to peruse and investigate the IndoWordnet lexical database for different Indian dialects. IndoWordnet visualizer extracts the related ideas for a given word and shows a sub chart containing those ideas. The interface is upgraded with different includes with the end goal to give adaptability to the client. IndoWordnet visualizer is made publically accessible. In spite of the fact that it was at first built for making the wordnet validation process less demanding, it is turned out to be extremely valuable in investigating different Natural Language Processing errands.

In [41] propose an inventive strategy to do the sentiment processing for news sentences. All the more uniquely, in view of the online networking information (i.e., words and emojis) of a news sentence, a word feeling affiliation organize is worked to mutually express its semantic and feeling, which establishes the framework for the news sentence assumption calculation. In view of WEAN, a word feeling calculation is proposed to get the underlying words feeling, which are additionally refined through the standard feeling vocabulary. With the words feeling close by, work can figure each sentence's sentiment.

III. TECHNIQUES OF WSD

As discuss in the previous section, a number of researches have explored this area for various natural languages and have shown varying degree of improvement in the MT by using WSD algorithm based on the work carried by researches including some classic a most popular researches in this area, this section classifies these approaches into different categories.

3.1. Knowledge Based Approaches

It relies on knowledge resources of Machine Readable Dictionaries (MRD) in form of WordNet, and Thesaurus *etc.* They may use grammar rules and hand coded rules for disambiguation. In recent years, most dictionaries made available in Machine Readable Dictionaries format (MRD) like that of Oxford English Dictionary, Collins, Longman Dictionary of Ordinary Contemporary English (LDOCE);Thesauruses which add synonymy information like Roget Thesaurus ; and Semantic networks which add more semantic relations like WordNet, EuroWordNet. These are for English [6].

The knowledge based approaches can be of two types

A) Selectional Preferences based approaches

This approach also called selectional restrictions requires exhaustive enumeration of argument-structure of verbs. They usually combine statistical linguistics and knowledge based approaches.

B) Using Overlap Based Approaches

These require a Machine Readable Dictionary (MRD) [12].These machine readable dictionaries may include WorldNet, Thesaurus *etc.* Thesaurus based disambiguation makes use of the semantic categorization provided by a thesaurus or a dictionary with subject categories. Roget's International Thesaurus (Roget, 1946) has been used a one of the most popular thesaurus which was put into machine-tractable form in 1950s. This approach is base on finding the features of ambiguous word alongwith in context, in this way such algorithms are basically overlap based algorithm.

Among major algorithms widely discussed and cited under overlapped based approaches are as follows.

(i) Lesk's Algorithm

The Lesk algorithm for disambiguation proposed in 1986 has opened the way for researches to use MRDs, many researchers had since then started using MRD as structured source for lexical Knowledge for WSD. The underlying idea of the algorithm is that that word senses that are related to each other, are often defined in a dictionary using many of the same words. To selects a meaning for a particular target word its dictionary definitions of possible senses are compared with those of the other content words in the surrounding window of context. Lesk's algorithm treats glosses as unordered bags of words, and simply counts the number of words that overlap between each sense of the target word and the senses of the other words in the sentence [14].

In the Lesk's description of algorithm included various ideas for future research, and in fact several of the issues he raised continue to be topics of research even today. Though it opened the path for knowledge based WSD research, it also has few criticism and limitations towards its performance, for example since dictionary glosses are very short they often fail to provide the fine grained senses, in such situations the disambiguation may be drastically affected. Therefore it has hypothesized that the length of the glosses is likely to be the most important issue in determining the success or failure of this method [13].

(ii) Walker's Algorithm

Walker (1987) proposed an algorithm based on thesaurus in which each word is assigned to one or more subject categories in the thesaurus to which the word belongs. Then the sore for each sense is computed using the word context. If the word is assigned to several subjects, then it is assumed that they correspond to different senses of the word. Black applied this approach to five different words and achieved accuracies around 50% [15].

(iii) Wilks' Approach

Wilks observed that dictionary glosses are too short to result in a proper disambiguation. Motivated from the observations in the Lesk's approach, they expanded the glosses using context vector approach with related words, by doing so, a wider and more relevant matching became possible that resulted in finer grained distinctions in meaning than is possible with short glosses. To achieve this, they used Longman's Dictionary of Contemporary English (LDOCE). Since the vocabulary of LDOCE for gloss matching is much larger, it increased the likelihood of finding overlaps among word senses [18].

(iv) Cowie's Approach

Cowie et.al. after analyzing the Lesk's approach found that the despite it is capable disambiguation, the only issue is the computational complexity which could be enormous for practical purposes. In order to search for senses simultaneously for all content word in a sentence, they used simulated annealing. They further analyses that if the sense assignment is appropriate done complexity may be reduced. The simulated annealing can be better used as solution that globally optimizes the assignment of senses among the words in the sentence to further minimize search effort [19].

(v) Veronis & Ide's Approach

Apart from the Lesk's work, Quillian's spreading activation networks has also been used by researchers. One important among them is Veronis and Ide who represented the senses of words in a dictionary in a semantic network in a way that word nodes are connected to sense nodes which are then connected to the words that are used to define that sense. Disambiguation is performed via spreading activation, that is, word that appears in the context is assigned the sense associated with a node that is located in the most heavily activated part of the network [20].

(vi) Kozima & Furugori's Approach

Kozima and Furugori [21] used LDOC glosses to construct a network consisting of nodes to represent the controlled vocabulary, and links in order to know the co-occurrence of these words in glosses.

(vii) Niwa & Nitta's Approach

In their work, Niwa and Nitta getting inspired by the Quillian network used and compared two vectors i.e. context vectors derived from co–occurrence statistics of large corpora and the vectors derived from the path lengths in a network that represent their co–occurrence in dictionary definitions. They further explored Wilk's context vector method of disambiguation, to conclude that dictionary contents are better source of co–occurrence information than the corpora [22].

(viii) Sussna's Approach

Sussna attempt of disambiguation is based on minimizing a semantic distance function to assigns a sense to each noun in a window of context among their possible senses, it was a measure of relatedness among nouns introduces by him. He utilized the WordNet noun hierarchy where in a single link provides a better conceptual distance compared to the links lower in the hierarchy. [23]. The comparisons of the some KB approaches are shown in Table 1.

Though knowledge based approaches are widely used for disambiguation, there are some underlying issues that may affect the disambiguation accuracy. These include

- The dictionary definitions present in MRD are generally very small.
- The dictionary entries rarely take into account the distributional constraints of different word senses *e.g.* selectional preferences, kinds of prepositions, *etc.*
- They suffer from the problem of sparse match.[33], it occurs in NLP problems wherein many events occur rarely, even when large quantities of data are available
- The proper nouns are not present in a MRD. Hence these approaches fail to capture the strong clues provided by proper nouns *e.g.* 'Ricky Ponting' strongly refers to the category 'sports' as Ricky Ponting plays cricket.

Table 1: Below shows a comparative analysis of knowledge based approaches [12].

Algorithm	Accuracy
Selectional Restriction Brown Corpus Algorithm	44%
Lesk's algorithm	50-60%
WSD using conceptual	54%

density	
Walker's algorithm	50%

3.2. Machine Learning Based Approaches

The main features of machine learning based approaches are that these basically rely on corpus evidence. The training of the model can be done using tagged or untagged corpus.

It can be classified under following categories

- Supervised approaches- It is based on a labeled training set. The system uses training set of 'feature-vectors' along with sense labeling
- Semi-supervised algorithms- It is based on unlabeled corpora. The system uses training set of 'feature-vectors' without their appropriate sense label
- Unsupervised Algorithms- They combine the advantages of both supervised and unsupervised approaches. These are like knowledge based approaches in that these do not need tagged corpora but like supervised approaches in extracting the evidence from corpus. Connections between words in a sentence can help in disambiguation. The graph is a natural way to capture connections between entities, which utilize relations between senses of various words [34].

3.2.1 Supervised Learning

Supervised learning techniques collects a set of training data with known labels in order to classify new set or data items. These identify patterns in the dataset associated with each particular class, and generalize those patterns into rules which are they added to classify new set. In this way, they are class of methods that induces a classifier from manually sense-tagged text using machine learning techniques. Such techniques use any form of sense tagged resources, Syntactic Analysis (POS tagger, Chunker, Parser) [6]. Its scope is typically one target word per context; part of speech of target word resolved or lends itself to 'targeted word' formulation. The WSD therefore becomes a classification problem wherein a target word is assigned the most appropriate sense based on the context in which it occurs.

A generalized approach of supervised learning is as follows [6]

- A sense-annotated trained corpora is created
- Built classifiers using machine learning techniques
- recognize the appropriate senses depending on context of surrounding sentence

We discuss below some popular supervised algorithms used for word sense disambiguation.

A) Naïve Bayesian Classifiers

Naïve Bayesian Classifier is a popular supervised machine learning algorithm and has been widely used for WSD. It uses classifiers based on Bayes theorems for computation of conditional probability for each sense of a word. It has usually thousands of binary features that indicate if a word is present in the context of the target word (or not). This algorithm may however, suffers from the problem of data sparseness. It requires a large number of parameters to be trained [6]. Intestinally, since the scores are based on a product of probabilities, it is possible that some weak features might pull down the overall score for a sense causing poor performance.

B) Decision Lists and Trees

Decision trees have become popular to be used since very early years of WSD research. It is a word-specific classifier and a separate classifier needs to be trained for each word. It uses the single most predictive feature which eliminates the drawback of Naïve Bayes. It is based on 'One sense per collocation' property. The training labeled data set is used to train the classifiers for the first time to identify the main features. The nearby words provide strong and consistent clues as to the sense of a target word. Decision List for WSD is given by Yarowsky, 1994.

***C)* Exemplar Based WSD (K-NN)**

It is a word-specific classifier algorithm. In this, an exemplar based classifier is constructed for each word to be disambiguated; it uses a diverse set of features (including morphological and noun- subject-verb pairs). For a sentence containing ambiguous word a test example is constructed which is then compared with training sets to select few closest set. The most prevalent amongst these is then selected as the correct sense.

D) WSD Using SVM (Support Vector Machines)

It is a word-sense specific classifier. It's a binary classifier that separates positive samples from negative samples It gives the highest improvement over the baseline accuracy. It uses a tagged corpus, the training for a sense of a word is done using a variety of rich features.

E) WSD Using Perceptron trained HMM (Hidden Markov Model)

It users corpus such as WordNet super senses rather than actual senses. A broad coverage classifier as the same knowledge sources can be used for all words belonging to super sense. A discriminative Hidden Markov Model is trained using the feature such as; POS of neighboring words, Local collocations, Shape of the word and neighboring words.

Table 2: Below presents a comprehensive analysis of various supervised approaches [42].

Algorithm Class	Method	Test Data	Performance
Naïve Bayesian Classifier	Naïve Bayesian Algorithm (Le and Shimazu, 2004)	Small dataset of four words, Large dataset extracted from DSO corpus	On small dataset 92.3% accuracy and on DSO corpus accuracy is 66.4% for verbs and 72.7% for nouns.
Examplar Based Classifier	Examplar Based Learning Algorithm (Ng and Lee, 1996)	Manually Sense Tagged Data Set of about 192.800 words	Improve performance in comparison to Miller 1994, Yarowsky, 1993 etc.
Decision List Classifier	Using Decision List (Yarowsky)	Spanish Test Data	99% accuracy in general and 90% accuracy for most difficult ambiguities
Maximum Entrophy Classifier	Maximum Entrophy approach with rich feature sets (Tratz et. Al. 2007)	SemCor and Example Sentences	Results are better than the baseline
Lazy Boosting Algorithm	Based on Lazy Boosting Algorithm (Escudero et. Al. 2001)	TALP test data (TALP is a research center)	Fine grained accuracy 61.51% and Coarse Grained accuracy of 69.00%

3.2.2 Semi-Supervised Algorithms

As discussed in section 3.2 above, the Semi-Supervised Algorithm uses the strategy of its supervised version even though it needs significantly fewer amounts of tagged data. It expands applicability of supervised WSD; therefore it usually has all the advantages and disadvantaged of its supervised version.

The algorithms that come under this category use bootstrapping approaches. The common features of bootstrapping approach are – use of some labeled data, large amounts of unlabelled data and One or more basic classifiers. The output by this approach is a new classifier that improves over the basic classifiers.

The bootstrapping is an example of Yarowsky's algorithm that uses Decision Lists. It relies on two heuristics and a decision list

- One sense per collocation :
  The neighborhood words have strong connection to get the correct sense of a target word.
- One sense per discourse :
  Given a document, there is a strong possibility of getting the sense of a target word within it.

The two popular algorithms under semi supervised category of WSD are decision list (bootstrapping) and monosemous. The performance of decision list algorithm is usually found better that that of the second one. The bootstrapping approach starts with a small size of seed data for each word. This seed is taken as initial classifier and trained using any supervised algorithm to get a bigger trained dataset and the process is repeated until the entire corpus is trained. Other approaches used co-occurrence information as supplement to tagged corpora.

Table 3: Below presents a comprehensive analysis of various supervised approaches [42].

Algorithm Class	Method	Test Data	Performance
Bootstrapping Approach	Yarowshky Algorithm	Test data was extracted from a 460 million word corpus containing new articles, scientific abstracts, novel etc.	96.1% in comparison to the Schutze, 1998 92.2% accuracy.
	Self-Training (Rada	Test Data from the Senseval-2 and a large	Performanced Improved by error

	Mihalcea, 2004)	new corpus of unlabeled examples	reduction of 25.5%
	Co-Training (Rada Mihalcea, 2004)	Test Data from the Senseval-2 and a large new corpus of unlabeled examples	Performanced Improved by additional error reduction of 9.8% with global parameters.

3.2.3 Unsupervised Algorithms

These approaches are among the toughest of all other WSD methods. The task of unsupervised WSD is challenging because there is no manually labeled data present in this case. The underlying assumption is that if the context is same /similar then the words appearing in these contexts will also have similar senses and measure of similarity of context may identify the correct sense. If Sense tagged text is available, it can be used for evaluation. The performances of unsupervised approaches are good for only a limited set of target words.

Some of the prominent algorithms under the unsupervised category are-

A) Lin's Algorithm

It is a universally useful wide inclusion approach. It can even work for words which do not appear in the corpus.

B) Hyperlex

In this algorithm instead of using 'lexicon characterized senses, uses the senses from the corpus' itself. It faces difficulty in identifying fine grain senses of a word.

C) Yarowsky's Algorithm

It is a broad coverage classifier. It can be used for words which do not appear in the corpus but it was not tested on an 'all word corpus'.

D) WSD using Parallel Corpora

It overcomes the issue of hyperlex in that it can distinguish even between finer senses of a word as the fine grain senses of a word get translated as distinct words. Such algorithms usually needs a word aligned parallel corpora and require large number of parameters for training.

Table 4: presents a comparative summary of Un-supervised approaches [42].

Algorithm Class	Method	Test Data	Performance
Latent Semantic Analysis	Phil Katz and Pau method	Test data from Senseval-3	Only Slight improvement in performance. However not better than the Naïve Bays classifier
	Jason Blind method	Data set derived from SemCor-2.0 corpus	No major improvement in performance.
Parallel Text	Parallel Corpora approach (Diab and Resnik, 2001)	Pseudo Translated Corpus (English-French)	Performance improved in comparison to other unsupervised systems
	Nancy Ide	Parallel Corpora based on Orwell's Novel	Outperforms the Monolingual Bootstrapping process.
	Bilingual Bootstrapping (Li and Li, 2004)	Dataset from Wall Street Journal, Few words of Yarowsky study	Outstanding improvement in Bootstrapping in comparison to monolingual Bootstrapping
Spreading Activation Networks (SAN)	SAN method (Tsatsaronis et.al. 2007)	Senseval-2 data using word thesaurus	Bootstrapping improvement in comparison to the unsupervised approach of (veronis and Ide, 1998)

IV. DISCUSSION

Broadly the three approaches discussed have their own advantages and disadvantages. A number of specific approaches under each category have shown improvement in disambiguation of texts. The knowledge based approaches are usually good but the MRD used in these approaches are usually small. The Walker's algorithm has accuracy50% when tested on 10 highly polysemous English words.The lesk algorithms is a famous example of Knowledge base approach and set a milestone in the use of MRD, it has however, also hypothesized that the length of the glosses is likely to be the most important issue in determining the success or failure of this method. The wilks approach considered the observations of Lesk algorithm, and treat the LDOCE glosses as a corpus, and build a co–occurrence matrix for the defining vocabulary for enhancing the chances of better overlapping of words.

Machine learning based approaches basically rely on corpus evidence. The training of the model can be done using tagged or untagged corpus they can be supervised, unsupervised or semisupervised.

Naïve Bayesian is a famous supervised algorithm with good performance; it may however suffer from the problem of data sparseness. The Decision List algorithm uses the single most predictive feature which eliminates the drawback of Naïve Bayes and achieves the highest precision among other algorithms.

The unsupervised approaches take the advantages of both supervised and unsupervised approaches. These are like knowledge based approaches in that these do not need tagged corpora but like supervised approaches in extracting the evidence from corpus. Connections between words in a sentence can help in disambiguation. Among the unsupervised approaches, the Hyperlex approach has shown slightly better performance.

The semi supervised algorithms eases the need of annotated corpora so the knowledge acquision bottleneck is minimized, despite this minimal requirements, these algorithms work at par with supervised approaches.

V. CONCLUSIONS

Based on the study of word sense ambiguity in the field on linguistic, it is found that a large number of attempts have been made to resolve the ambiguities in various languages. Many attempts that have been carried in this area have shown the potential as indicated by the experiments and results. Among the three main categories of the approaches for WSD that have been explored are Knowledge Based Approaches, Machine Learning Based Approaches (Supervised, Semi-supervised and Un-supervised). The comparative analysis of various popular techniques involving various parameters has also been discussed. The performances of the algorithms which have been shown in the comparison have been obtained from the literatures of various research works carried by researchers to show the overall scenario of progress of research in this area.

REFERENCES

[1] Ide, N., & Véronis, J. (1998). Introduction to the special issue on word sense disambiguation: the state of the art. Computational linguistics, 24(1), 2-40.
[2] McCarthy, J. (1960). Programs with common sense (pp. 300-307). RLE and MIT computation center.
[3] ALPAC, L. (1966). Machines: Computers in Translation and Linguistics, Report by the Automatic Language Processing Advisory Committee, Division of Behavioral Sciences. National Academy of Sciences, National Research Council, Publication, 1416.
[4] Stevenson, M., & Wilks, Y. (2003). Word sense disambiguation. The Oxford Handbook of Comp. Linguistics, 249-265.
[5] Wielemaker, J., Schrijvers, T., Triska, M., & Lager, T. (2012). Swi-prolog. Theory and Practice of Logic Programming, 12(1-2), 67-96.
[6] Mihalcea, R., & Pedersen, T. (2005). Advances in word sense disambiguation. Tutorial in ACL, 5.
[7] Gaizauskas, R., & Wilks, Y. (1998). Information extraction: Beyond document retrieval. Journal of documentation, 54(1), 70-105.
[8] Specia, L., Srinivasan, A., Ramakrishnan, G., & Nunes, M. D. G. V. (2006, August). Word sense disambiguation using inductive logic programming. In International Conference on Inductive Logic Programming (pp. 409-423). Springer, Berlin, Heidelberg.
[9] Dalal, A., Nagaraj, K., Sawant, U., & Shelke, S. (2006). Hindi part-of-speech tagging and chunking: A maximum entropy approach. Proceeding of the NLPAI Machine Learning Competition.
[10] Sinha, M., Reddy, M., & Bhattacharyya, P. (2006, January). An approach towards construction and application of multilingual indo-wordnet. In 3rd Global Wordnet Conference (GWC 06), Jeju Island, Korea.
[11] Fellbaum, C., Grabowski, J., & Land, S. (1997). Analysis of a hand-tagging task. Tagging Text with Lexical Semantics: Why, What, and How?
[12] Khapra, M. M., Shah, S., Kedia, P., & Bhattacharyya, P. (2009, August). Projecting parameters for multilingual word sense disambiguation. In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1 (pp. 459-467). Association for Computational Linguistics.
[13] Pedersen, T., Banerjee, S., & Patwardhan, S. (2005). Maximizing semantic relatedness to perform word sense disambiguation. Research Report UMSI 2005/25, University of Minnesota Supercomputing Institute.
[14] Haroon, R. P. (2010, December). Malayalam word sense disambiguation. In Computational Intelligence and Computing Research (ICCIC), 2010 IEEE International Conference on(pp. 1-4). IEEE.
[15] Mutlum, B. (2005). Word Sense Disambiguation Based on Sense Similarity and Syntactic Context (Doctoral dissertation, Master Thesis of Science in Computer Engineering. Koc University September).
[16] Jones, K. S. (1964). Synonymy and semantic classification.
[17] Quillian, M. (1968). Semantic memory in M. Minsky (ed.), Semantic Information Processing, 227–270.
[18] Wilks, Y., Fass, D., Guo, C. M., McDonald, J. E., Plate, T., & Slator, B. M. (2007). Providing Machine Tractable Dictionary Tools. In Words and Intelligence I (pp. 167-216). Springer, Dordrecht.
[19] Cowie, J., Guthrie, J., & Guthrie, L. (1992, August). Lexical disambiguation using simulated annealing. In Proceedings of the 14th conference on Computational linguistics-Volume 1(pp. 359-365). Association for Computational Linguistics.
[20] Veronis, J., & Ide, N. M. (1990, August). Word sense disambiguation with very large neural networks extracted from machine readable dictionaries. In Proceedings of the 13th conference on Computational linguistics-Volume 2 (pp. 389-394). Association for Computational Linguistics.

[21] Kozima, H., & Furugori, T. (1993, April). Similarity between words computed by spreading activation on an English dictionary. In Proceedings of the sixth conference on European chapter of the Association for Computational Linguistics (pp. 232-239). Association for Computational Linguistics.
[22] Niwa, Y., & Nitta, Y. (1994, August). Co-occurrence vectors from corpora vs. distance vectors from dictionaries. In Proceedings of the 15th conference on Computational linguistics-Volume 1 (pp. 304-309). Association for Computational Linguistics.
[23] Sussna, M. (1993, December). Word sense disambiguation for free-text indexing using a massive semantic network. In Proceedings of the second international conference on Information and knowledge management (pp. 67-74). ACM.
[24] Agirre, E., & Rigau, G. (1996, August). Word sense disambiguation using conceptual density. In Proceedings of the 16th conference on Computational linguistics-Volume 1(pp. 16-22). Association for Computational Linguistics.
[25] Banerjee, S., & Pedersen, T. (2003, August). Extended gloss overlaps as a measure of semantic relatedness. In Ijcai (Vol. 3, pp. 805-810).
[26] Sharma, R., & Bhatia, P. G. (2008). Word Sense Disambiguation for Hindi Language (Doctoral dissertation).
[27] Badodekar, S. (2003). Translation resources, services and tools for Indian languages. Computer Science and Engineering Department, Indian Institute of Technology, Mumbai, 400019.
[28] Ramakrishnan, G., Prithviraj, B., & Bhattacharya, P. (2004). A gloss-centered algorithm for disambiguation. In Proceedings of SENSEVAL-3, the Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text.
[29] Dave, S., Parikh, J., & Bhattacharyya, P. (2001). Interlingua-based English–Hindi machine translation and language divergence. Machine Translation, 16(4), 251-304.
[30] English WordNet http://WordNet.princeton.edu
[31] Resnik, P. (1997). Selectional preference and sense disambiguation. Tagging Text with Lexical Semantics: Why, What, and How?.
[32] Kang, S. J., & Lee, J. H. (2001, July). Semi-automatic practical ontology construction by using a thesaurus, computational dictionaries, and large corpora. In Proceedings of the workshop on Human Language Technology and Knowledge Management-Volume 2001 (p. 6). Association for Computational Linguistics.
[33] Martinez, D. (2005). Supervised Word Sense Disambiguation: Facing Current Challenges. Procesamiento del Lenguaje Natural, 34.
[34] Barzilay, R., & McKeown, K. R. (2001, July). Extracting paraphrases from a parallel corpus. In Proceedings of the 39th annual meeting on Association for Computational Linguistics(pp. 50-57). Association for Computational Linguistics.
[35] Abdelaali, B., Tlili-Guiassa, Y., Schwab, D., & Tchechmedjiev, A. (2015). Ant colony algorithm for Arabic word sense disambiguation through English lexical information. International Journal of Metadata, Semantics and Ontologies.
[36] Luo, F., Liu, T., Xia, Q., Chang, B., & Sui, Z. (2018). Incorporating Glosses into Neural Word Sense Disambiguation. arXiv preprint arXiv:1805.08028.
[37] Basile, P., Caputo, A., & Semeraro, G. (2014). An enhanced lesk word sense disambiguation algorithm through a distributional semantic model. In Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers (pp. 1591-1600).
[38] Ustalov, D., Teslenko, D., Panchenko, A., Chernoskutov, M., Biemann, C., & Ponzetto, S. P. (2018). An Unsupervised Word Sense Disambiguation System for Under-Resourced Languages. arXiv preprint arXiv:1804.10686.
[39] Kunchukuttan, A., Mishra, A., Chatterjee, R., Shah, R., & Bhattacharyya, P. (2014). Sata-anuvadak: Tackling multiway translation of indian languages. pan, 841(54,570), 4-135.
[40] Chaplot, D. S., Bhingardive, S., & Bhattacharyya, P. (2014). IndoWordnet visualizer: A graphical user interface for browsing and exploring wordnets of Indian languages. In Proceedings of the Seventh Global Wordnet Conference (pp. 338-345).
[41] Jiang, D., Luo, X., Xuan, J., & Xu, Z. (2017). Sentiment computing for the news event based on the social media big data. IEEE Access, 5, 2373-2382.
[42] Dwivedi, S. K., & Rastogi, P. (2009). Critical analysis of WSD algorithms, In Proceedings of the International Conference on Advances in Computing, Communication and Control (pp. 62-67). ACM.