

# Supervised Word Sense Disambiguation with Recurrent Neural Network Model



Chandrakant D. Kokane, Sachin D. Babar

**Abstract:** Disambiguating words is a branch of artificial intelligence that deals with natural language processing. The dissatisfaction of the motive of the word deals with the polysemy of the ambiguous word, processing a single word in natural language, having two or more meanings where the corresponding context discriminates the meaning. Humans are intelligent enough to derive the meaning of the word because they are a biological neural network. Computers can be trained in such a way that they should function similarly to biological neural networks. There are four different suggested approaches to clutter as the knowledge-dependent approach and the machine learning based models which are further classified as supervised, semi-supervised and unpublished learning models. The purpose of this research is to improve better communication between computers and humans. The discussed model used a supervised learning approach with recurrent neural networks.

**Keywords:** Supervised learning, recurrent neural network, word sense disambiguation..

## I. INTRODUCTION

The word-sensing discrepancy in natural language processing mainly focuses on important problems related with identifying the correct understanding of the word. Oblivion is related to which word is used to know the true and precise meaning of a sentence. A lot of words have been used in the English language which have different senses and meanings. WSD deals with the problem of choosing the correct meaning related to correct word perception. The solution to this problem improves the relevance of search engines with real-time information. The human brain (biological neural network) is quite smart and trained to decipher words. A simple reference is what a person needs to understand the true meaning or meaning of a word. Human language develops in a way that reflects the natural potential provided by the human brain i.e. biological neural network. In computer science and information technology, training computers for natural language processing and predictions by machine learning or artificial intelligence has been a

long-term challenge. Currently there are so many applications or technologies that they are working on solving this problem but still there is scope for improvement and accuracy. In today's modern world, people are heavily invested in the ability of computers to solve various problems in their daily lives. From finding directions through GPS to calculating their tax returns, most people rely on computer equipment in one way or another. For better user experience and better interfacing between user and machine, clear communication between them is required. One obstacle in the way is the problem of ambiguity in the word senses. To reduce this problem and increase computer intelligence, we propose our research for word sense disambiguation. For example, consider a word bank in English which has different meanings: various as commercial banks, blood banks and riverbanks. The word sense appropriately replaces the ambiguous word based on the context surrounding the sentence. Here we are focusing on the English language and in the same, there are a lot of ambiguous words whose meaning is stated by the surrounding and position of the word in the sentence.

There are two different approaches to WSD,

### 1. Knowledge-based approach

A knowledge-based algorithm uses various readable resources such as machine readable dictionaries (MRD) to identify correct understanding of WordNet words. These algorithms are easy to implement and first had to be developed while trying to solve the WSD problem. A knowledge-based system only needs access to commercial dictionary resources to begin the process of lexicography. The drawback of these algorithms is that their performance is limited to the speed of discovery and retrieval of these resources. As resources grow, latency increases and performance decreases.

Example LASK algorithm

## II. MACHINE LEARNING-BASED MODELS

The Machine Learning based approach consists of three different learning methods as,

### a. Supervised learning

Supervised methods depend on the hypothesis that the available reference can provide enough information by itself to disregard words. Because similar meanings and arguments are considered redundant. The learning set is designed primarily to infer the exact meaning of a fuzzy word using a few sentences for a fuzzy system, with a specific meaning for that particular word. A system finds the correct meaning of a vague term for a particular context based on that defined training set.

Revised Manuscript Received on December 30, 2019.

\* Correspondence Author

**Chandrakant D. Kokane**, Reseach Scholar, Department of Computer Engineering, Smt. Kashibai Navale College of Engineering, Pune (India).  
Email:cdkokane1992@gmail.com

**Dr. Sachin D. Babar**, PhD Department of Computer Engineering, Sinhgad Institute of Technology, Lonavala, Pune (India).  
Email:sdbabar@sinhgad.edu

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

The approach always generates over performance more than any other method.

Further, these supervised models are subjects of acquiring the new deep information as they rely on a considerable amount of manually meaning-indexed resources for learning purposes that are not difficult and inexpensive to generate. Examples are Na Method, Bayes Method, Decision Tree Method.

## b. Semi-supervised methods

The learning approach is called a mixed approach. Many word perception exclamation specified methodology used a semi-supervised type of learning's that allows both manually published and unpublished data for training. Due to the lack of training data in this approach, the novel bootstrapping method starts with a based level small amount of root information for each word. The information is either a fixed number of fixed fire decision rules (eg, 'bank' in the context of 'transaction' almost always states a financial institution) or manual formally tagged training corpus. The root knowledge is used to train the initial classification, using any type of the supervised methods. This said classifier is further used on the untouched part of the tagged corpus to extract a large training set, which contains only the best assured classifications. This process repeats, with each new classifier being sequentially trained on large training data, until complete data is trained, or till a given maximum number of training iterations is remain.

Example Yarowsky bootstrapping method is label propagation algorithm.

## c. Unheard method

In the untrained learning approach, label information is not available for training, so the most important task is to derive patterns from fuzzy data. The unprocessed learning approach is the biggest challenge in word sense disambiguation. The bottommost assumption is that the words with same or similar meaning must and should occur in same contexts and thus the senses can be stimulated by the reduction of words using some measure of the context of similarity, a process as word perception discrimination or induction is referred to the performance of the unavailable learning approach has been observed to be minimum than the other teaching methods that are said above. But it has been observed that untrained education will remove the impediment to knowledge as they are independent of manual efforts.

Examples are using the co-event graphing / hyper-lex algorithm, WSD parallel corpora.

## III. PROCEDURE FOR PAPER SUBMISSION

### A. Submission of the paper

Author (s) can send paper in the given email address of the journal. There are two email address. It is compulsory to send paper in both email address.

### B. Plagiarism Check

Good quality plagiarism software/ tool (Turnitin / iThenticate) will be used to check similarity that would not be more than 20% including reference section. In the case of exclusion of references, it should be less than 5%.

### C. Quality Check

All submitted paper should be cutting edge, result oriented, original paper and under the scope of the journal that should belong to the engineering and technology area. In the paper title, there should not be word 'Overview/brief/ Introduction, Review, Case study/ Study, Survey, Approach, Comparative, Analysis, Comparative Investigation, Investigation'.

### D. Paper Submission Criteria

Any one author cannot submit more than 05 papers for the same volume/issue. The authors of the accepted manuscripts will be given a copyright form and the form should accompany your final submission. It is noted that:

- Each author profile along with photo (min 100 word) has been included in the final paper.
- Final paper is prepared as per journal the template.
- Contents of the paper are fine and satisfactory. Author (s) can make rectification in the final paper but after the final submission to the journal, rectification is not possible.

## IV. LITERATURE SURVEY

The literature survey is carried out with a consideration of four different words.

### A. Knowledge-based approach

Myung Yoo Kang, Tae Hong Min, Ja Sung Lee [8] (2018) The authors elaborate the term vector semantic model to reject more granular meanings in reference vectors by incorporating extended mean space. They have used a huge Korean meaning-tagged corpus and created an embedded emotion space together with knowledge-based model and evaluated effectiveness of the embedded meaning for word perception. The results of their experiment with a Korean corpus displayed that the proposed method, i.e., the embedded meaning space model, is more efficient than the word-space model [8]. The embedded sensation space model is not useful because the sense reference or the unrelated word reference are not available in the general context.

Uday Raj Dhungana et al. [12] (2015)

In this paper, the authors used a knowledge-based approach. They have used the adapted LASAK algorithm to dislike the word polysem in the Nepali language. They grouped each meaning of a polysim word based on the verb, noun, adverb and adjective, with which the word polysim can be used in a sentence. Experiments with test data consisting of 201 Nepali sentences have been performed on 348 words (including 59 polyseamy words and different senses of reference words), indicating their system accuracy of 88.05%.

Michael Lesk et al.[16] (1986)

A paper written in 1986 by Michael Lesk has been proven to be a revolutionary work in Word Sense Disambiguation (WSD). In this paper, he introduced his famous LASAK algorithm which has been a key algorithm for the knowledge-based approach WSD. The LESK algorithm uses various MRD's to calculate the correct senses of ambiguous words. The algorithm searches for overlap in different senses or signatures of a word.

Sessions with maximum number of overlaps are selected as the exact senses of the ambiguous word in given context. Lesk concludes that the algorithm yields an accuracy of about 50–70% depending on the use of MRD.

### B. Supervised Learning Approach

Edilson A. Corsa, Alneau A. Lopes, Diego R. Amancio [1] (May 2010)

In this system the supervised learning model is used for disambiguation, the words of the documents are represented as nodes. If two nodes are connected if and only if they are semantically similar. The major challenge the authors have given is to form networks and explain the word by the presence and context of the word. The learning algorithm adopted in bipartisan networks mostly proves better results, if and only if grounded features are correctly mapped to the data context. The challenge of this method is edge failure when mapping word sense. The said model gives best results even though a small or moderate data is available for training.

Tsinghua Wang, Junyang Rao, Qi Hu [3] (2014)

In this system, the supervised learning approach is used for word comprehension, with the formation of a BoW of reference. There are three different proposed stages in the form of pre-processing, kernel, and classification. In the creation of the word bag (BoW), the semantic diffusion kernel in which the SDK is applied before classification to improve the performance of the SVM. In the third and final phase, the Support Vector Machine (SVM) model is used for classification purposes. The limitations of supervised training here are class labels[6] and co-occurrence information is required. The limitation of BoW is that cluttered grammar and order will lead to incorrect reference to the word. Abdulgabar Saif, Nazalia Omar, Umami Zakia Zanodin, Mohammad Zuzadin Ab Aziz [5] (2018)

In this approach, the authors address the first phase of WSD which is constructing the meaning tag. The construction of comprehension tag data is very important for a supervised learning approach that has yielded promising results in word comprehension. Manual knowledge-based approach is labor and time consuming task. The proposed method by authors starts with mapping dataset of Arabic language Wordnet to the Wikipedia. To select a Wikipedia article for the same meaning in Wordnet. In a very first step of the mapping, the cross-language model is used to identify the similarity between the features of Wikipedia and WordNet meanings. The proposed approach works as candidate extraction model, candidate alignment method and example extraction in three different stages. Ali Alkhatlan, Jugal Kalita, Ahmed Alhadad [6] (2018) The proposed approach and the paperwork are limited in Arabic, even though near about half a billion are native Arabic speakers. In this work, authors have presented several solutions to the problem of disambiguation in Arabic language and use breakthroughs in word embedding learning with two different approaches like as GloVe and Word2vec. The primary drawback of word embedding is the unvalued vector representation of the meaning of a word, even though so many words are polysemous words. The author's main contribution to this work is to computationally obtain an embedding for each meaning using Arabic WordNet (AWN) to overcome the WSD problem. They also calculate word meaning similarity giving consideration to several Arabic stemming

approaches. Finally, the authors provide a large pre-processed corpus to be prepared for single experiment and a WSD test data based on AWN to fill gaps in Arabic NLP (ANLP) as compared to English language[9]. P. of Tamil Nadu, s. Srivatsa [10] (2012) the authors in this research have applied a supervised model to word perception using neural networks with minimum feature sets. As we know on a neural network with a hidden layer there are three layers in which hidden neurons from 5 to 20 are formed for clutter, with an increasing of 5-9 neurons at a time. At least two features, Bigra and a discount is tried with a maximum of trice features, trigram. The number of inputs to the neural network is depending on the total number of features that are taken for the training process. According to authors Bigrams considers only two features including ambiguous words and trigram considers only three features (Including ambiguous terms). The total Performance is calculated by using four different error functions. By considering 60 different network architectures, 20 network architecture generates trigram-based pattern recognition network with cognitive 85.72% accuracy.

Ignacio Icobaciak, Mohammad Tahir Pilhwar, Roberto Navigali [11] (2014) The main focus of this paper is on the word embedding. This means collecting semantic information from a collection of datasets. This is an example of a knowledge-based approach. Word embedding is usually a collection of names for advanced learning techniques in language modelling and natural language processing. In this, the results are evaluated using two methods as,

1. Lexical samples for WSD experiment.
2. All word for WSD usage.

The main interests were on the embedding and training parameters of WSD facilities which were influencing WSD performance. The maximum accuracy observed during this experiment was 69.9%.

Niladri Chatterjee and Rohit Mishra [15] (2009)

In this paper, team members have represented a trainable approach for Word Sense Disambiguation (WSD). The model mainly uses the concepts of context information theory to get the exact and appropriate meaning of a word if and only if the context information of the word is provided or available. Looking at the training lesson, the model learns to classify each and every occurrence of the target ambiguous word in the true sense. The presented model uses the principle of maximum penetration, looking at the sense of the ambiguous word as a random initialized variable with different results. The model then estimates the probability of each meaning by measuring the 'bias' of the surrounding words. The emotion with the highest probability is chosen as the true meaning of the target word. The model has been proven to provide accuracy up to 85%.

### C. Untrained teaching approach

Yoán Gutierrez, Sonia Vezquez, Andres Montoyo [2] (2011é) In this paper, he addressed the first step of the word rank algorithm by the page-rank algorithm to specify the term as the term or formation of a co-occurrence network. . Two different approaches proposed for a knowledge-based approach that use machine-readable dictionaries such as WordNet and a second corpus-based approach that use co-occurrence to measure similarity between words.

K.P. Shruti Shankar, et al. [8] (2014)





In this paper, the authors are mainly focusing on the word sensation discrepancy system for Malayalam, a language mainly used as a communication interface in Kerala, India. The system proposed by the authors uses a data that is collected from distributed Malayalam documents that are distributed in nature[6]. For every possible meaning of the target word, a unique set of training models called seed sets are listed and represent the valuable meaning. According to authors the collected and most related terms are counted as training models.

Alok Pal, Anupam Munshi and Digant Saha [13] (2013)

The main focus of this paper is to speed up the process of disambiguation using a filtering approach that detects the specific senses of a given ambiguous word through part-of-speech tagging. The exact fraction of the ambiguous word is obtained at that particular instance. In the next method, online dictionaries are referred to as wordnets, etc., which are related to part-to-speech to refute the true meaning of that particular ambiguous word.

In the training data phase, Brown Corpus is used for online speech as part of speech tagging and WordNet. In this method to speed up the process of WSD, some relevant glosses (words) are filtered and accuracy is increased.

### D. Semi-supervised learning approach

Bridget T. McInnes, Mark Stevenson [4] (2014)

Methods suggested in this paper include both supervised and unhelpful learning approaches. The use of supervised learning information or labels creates training data, where an unsupervised approach relies on UMLS (Joint Medical Language). The authors propose two different scenarios as mean\_similarity and max\_similarity in a supervised approach. Meaning in similarity, the combined means of calculating the similarity and degree of similarity in between the concepts of each and every ambiguous words with all possible occurred senses. The authors conclude that the supervised learning approach produced 98% accurate results for the Abrive dataset (dataset of summarization only).

Bartosz Baroda, Mackeyz Piaseki [9] (2009)

In this paper, the word meaning ambiguity is resolved using a semi-supervised approach and the results described that the proposed approach is very close to its accuracy for the supervised approach. The drawbacks of using a supervised and unhelpful approach are that the supervised approach requires lots of human efforts to prepare training data. On the other hand, the unused approach expresses significantly minimum accuracy and the outcomes are not satisfied to solve the problem. The main function of this model is to reduce human involvement, but to manually assign senses using a lexical-semantic resource known as WordNet. Here, Lexicographer (LexCSD) is used to collect the corpus from the given keyword. This keyword is then divided into groups and some common keywords are found from the given word. It is analyzed to find common features or senses in each cluster. The evaluation is done later by crossing MRDs.

Pratibha Rani, Vikram Pudi, Deepti Mishra Sharma [10] (2017)

In this paper, the authors present a general Word Sense Disambiguation (WSD) method using a semi-supervised approach. They state that current WSD systems use extensive domain resources and require advanced linguistic knowledge. Therefore, to improve these factors, they propose a system that extracts seed-based data from a small amount of reference data that contains emotion tags and

unpublished training data. Their experiments in the Hindi and Marathi language domains show that the system gives good performance without language-specific information, with the exception of the Sense ID present in the training set with approximately 60–70% accuracy[10].

Lokesh Nandanwar, Kalyani Mamulkar [14] (2013)

This is a survey paper that describes three approaches used in word comprehension:

1. Supervised Approach.
2. Semi-supervised approach.
3. Unproven approach.

These approaches are found to be very useful and successful in the field of word perception. They are classified on the basis of the main source of knowledge used for different senses and the amount of annotated corpora required [2].

The following approaches described above the semi-supervised approach require lower amounts of annotated corpora than the supervised approach. Here, annotate means that it is necessary to add some opinions to the text and corpora [8]. By observing and testing the approaches, the supervised approach gives better performance than other approaches.

## V. METHODOLOGY

The proposed methodology is divided into Training and Testing as,

### 1. Training

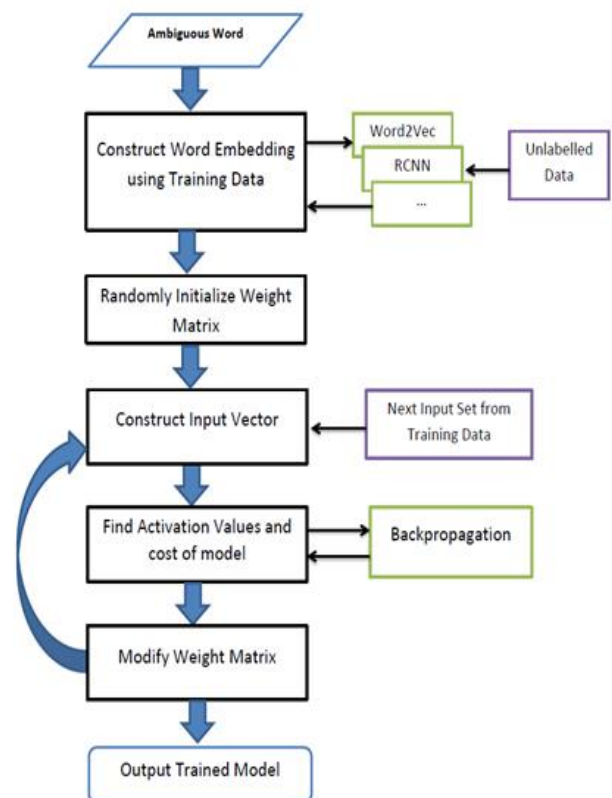


Figure 1: Method for Training

2. Testing

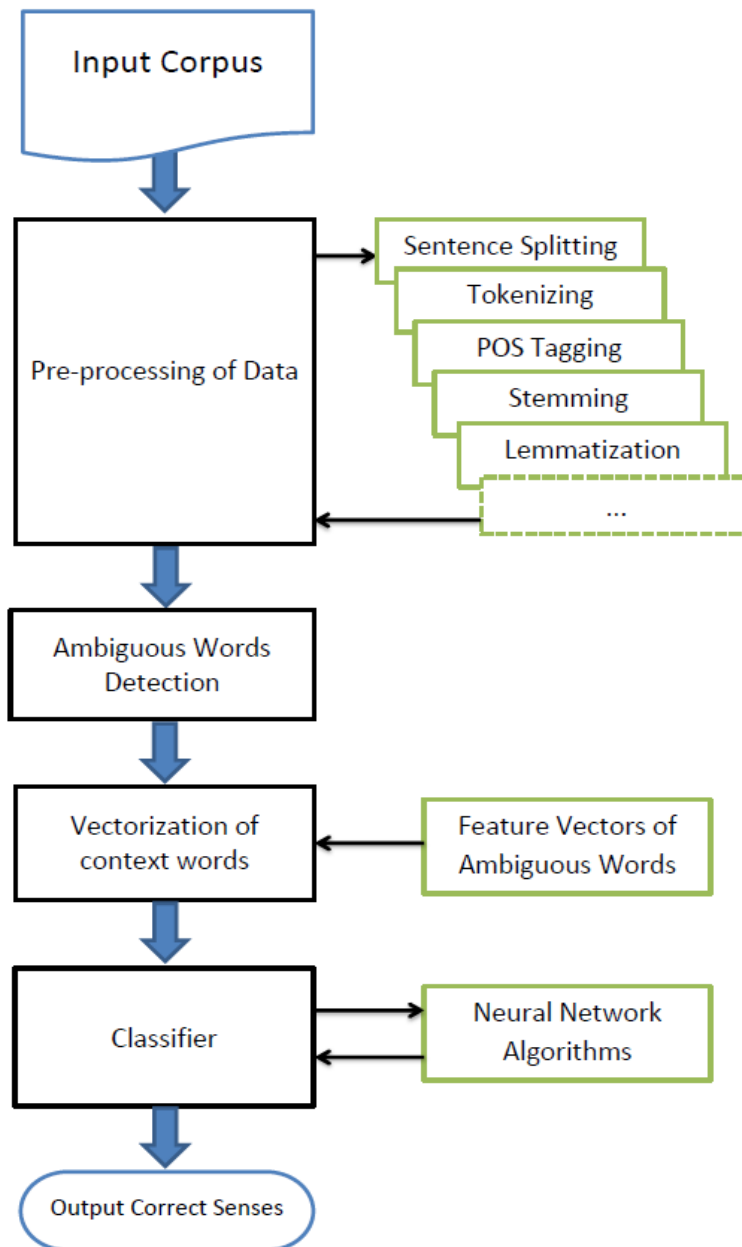


Fig.2 Method for Testing

VI. RESULTS

The Supervised model generates more accurate results and gives 96.3 percent of accuracy for polysemy words. The dataset Wordnet gives the final sense and word embedding technique word2vec will generate word embedding vector which will be provided as an input to the recurrent neural network with some randomly initialized weights.

Table 1 Result for Wordnet

Sr. No.	Sentence	Ambiguous Word	Sense	Meaning
1	Let's go to the bank	bank	3	Financial Institution

2	Let's go to the bank	bank	2	Blood storage Unit
3	Let's go to the bank	bank	4	Unit of Rivers
4	Lets meet in may	may	6	Month
5	May I Come in?	may	2	Asking for Permission

## VII. CONCLUSION

In a supervised learning approach, the main issue is to tag the words with the right senses and the knowledge-based approach is the most appropriate MRD (machine readable dictionary). In this research, I observed the need for Word Sense Disambiguation and saw that the accuracy of the

systems working on this task needed improvement. I have noted the need to increase the accuracy of existing WSD systems, and to advance our hypothetical model based on a supervised approach. The model uses a two-tier algorithm that uses our improvisation at the first level and the LASK algorithm at the second level to maximize accuracy.

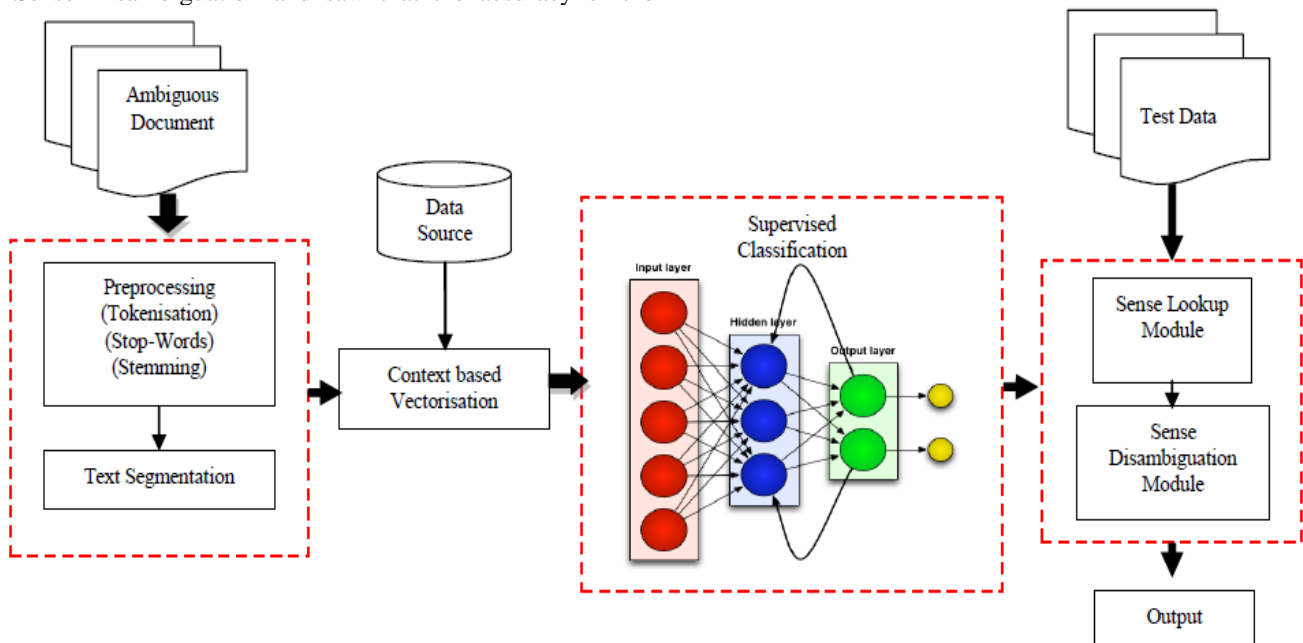


Figure3: System Architecture

## REFERENCES

- Edilson A. Corrêa, Alneu A. Lopes, Diego R. Amancio, "Word Sense Disambiguation: A Complex Network Approach", Information Sciences, Volumes 442-443, 2018, Pages 103-113, Issn 0020-0255.
- Yoan Gutiérrez, Sonia Vázquez, Andrés Montoyo, "Spreading Semantic Information By Word Sense Disambiguation", Knowledge-Based Systems, Volume 132, 2017, Pages 47-61, Issn 0950-7051.
- Tinghua Wang, Junyang Rao, Qi Hu, "Supervised Word Sense Disambiguation Using Semantic Diffusion Kernel", Engineering Applications of Artificial Intelligence, Volume 27, 2014, Pages 167-174, Issn 0952-1976.
- Bridget T. McInnes, Mark Stevenson, "Determining The Difficulty Of Word Sense Disambiguation", Journal Of Biomedical Informatics, Volume 47, 2014, Pages 83-90, Issn 1532-0464.
- Abdulgabar Saif, Nazlia Omar, Umami Zakiah Zainodin, Mohd Juziaddin Ab Aziz, "Building Sense Tagged Corpus Using Wikipedia For Supervised Word Sense Disambiguation", Procedia Computer Science, Volume 123, 2018, Pages 403-412, Issn 1877-0509.
- Ali Alkhatlan, Jugal Kalita, Ahmed Alhaddad, "Word Sense Disambiguation For Arabic Exploiting Arabic Wordnet And Word Embedding", Procedia Computer Science, Volume 142, 2018, Pages 50-60, Issn 1877-0509.
- Tamilselvi P, S.K. Srivatsa, "Optimal Word Sense Disambiguation With Minimal Feature Set Using Neural Network", Procedia Engineering, Volume 30, 2012, Pages 546-553, Issn 1877-7058.
- K.P. Sruthi Sankar, P.C. Reghu Raj, V. Jayan, "Unsupervised Approach To Word Sense Disambiguation In Malayalam", Procedia Technology, Volume 24, 2016, Pages 1507-1513, Issn 2212-0173.
- Myung Yun Kang, Tae Hong Min, Jae Sung Lee, "Sense Space For Word Sense Disambiguation", 2018 Ieee International Conference On Big Data And Smart Computing.
- Bartosz Broda And Maciej Piasecki, "Semi-Supervised Word Sense Disambiguation Based On Weakly Controlled Sense Induction". 2009 Proceedings Of The International Multiconference On Computer Science And Information Technology.
- Pratibha Rani, Vikram Pudi, Dipti M. Sharma "Semi-Supervised Data-Driven Word Sense Disambiguation For Resourcepoor Languages" 2017 14th International Conference On Natural Language Processing (Icon).
- Ignacio Iacobacci, Mohammad Taher Pilehvar, Roberto Navigli "Embedding For Word Sense Disambiguation: An Evaluation Study" 2016 Proceedings Of The 54th Annual Meeting Of The Association For Computational Linguistics, Pages 897-907
- Udaya Raj Dhungana, Subarna Shakya, Kabita Baral And Bharat Sharma "Word Sense Disambiguation Using Wsd Specic Wordnet Of Polysemy Words" 2015 Proceedings Of The 2015 Ieee 9th International Conference On Semantic Computing
- Alok Pal, Anupam Munshi And Diganta Saha "An Approach To Speed-Up The Word Sense Disambiguation Procedure Through Sense Filtering" 2013 International Journal Of Instrumentation And Control Systems (Ijics) Vol.3, No.4, October 2013 26.
- Lokesh Nandanwar And Kalyani Mamulkar "Supervised, Semi-Supervised And Unsupervised Wsd Approaches: An Overview" 2013 International Journal Of Science And Research (Ijsr)
- Niladri Chatterjee And Rohit Misra "Word-Sense Disambiguation Using Maximum Entropy Model" International Conference On Methods And Models In Computer Science, 2009
- Michael Lesk "Automatic Sense Disambiguation Using Machine Readable Dictionaries: How To Tell A Pine Cone From An Ice Cream Cone" Bell Communications Research Morristown, R. Navigli and P. Velardi, "Structural semantic interconnections: a knowledge-based approach to word sense disambiguation," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 7, pp. 1075-1086, July 2005. doi: 10.1109/TPAMI.2005.149
- Y. Choi, J. Wiebe and R. Mihalcea, "Coarse-Grained +/-Effect Word Sense Disambiguation for Implicit Sentiment Analysis," in IEEE Transactions on Affective Computing, vol. 8, no. 4, pp. 471-479, 1 Oct.-Dec. 2017. doi: 10.1109/TAFFC.2017.2734085
- R. Navigli and M. Lapata, "An Experimental Study of Graph Connectivity for Unsupervised Word Sense Disambiguation," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 4, pp. 678-692, April 2010. doi: 10.1109/TPAMI.2009.36
- A. Halioui, P. Valtchev and A. B. Diallo, "Bioinformatic Workflow Extraction from Scientific Texts based on Word Sense Disambiguation," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 15, no. 6, pp. 1979-1990, 1 Nov.-Dec. 2018. doi: 10.1109/TCBB.2018.2847336

21. Myunggwon Hwang, Chang Choi and Pan Koo Kim, "Automatic Enrichment of Semantic Relation Network and Its Application to Word Sense Disambiguation," in IEEE Transactions on Knowledge and Data Engineering, vol. 23, no. 6, pp. 845-858, June 2011. doi: 10.1109/TKDE.2010.163
22. F. Rojas Lopez, I. Lopez Arevalo, D. Pinto and V. J. Sosa Sosa, "Context Expansion for Domain-Specific Word Sense Disambiguation," in IEEE Latin America Transactions, vol. 13, no. 3, pp. 784-789, March 2015. doi: 10.1109/TLA.2015.7069105

### AUTHORS PROFILE



**Chandrakant D. Kokane**, Reseach Scholar,  
Affiliation: Savitribai Phule Pune University  
Department of Computer Engineering, Smt.  
Kashibai Navale College of Engineering, Pune  
(India).  
Email:cdkokane1992@gmail.com



**Dr. Sachin D. Babar**, PhD Affiliation: Savitribai  
Phule Pune University Department of Computer  
Engineering, Sinhgad Institute of Technology,  
Lonavala, Pune (India).  
Email:sdbabar@sinhgad.edu