ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal
Regular Issue, Vol. 13 (2024), e31602
eISSN: 2255-2863
DOI: https://doi.org/10.14201/adcaij.31602

A Systematic Analysis of Various Word Sense Disambiguation Approaches

Ganesh Chandra[a], Sanjay K. Dwivedi[b], Satya Bhushan Verma[c], Manish Dixit[a]

[a] Department of Computer Science & Engineering, Madhav Institute of Technology & Science (Deemed University), Gwalior, M.P, India
[b] Department of Computer Science, Babasaheb Bhimrao Ambedkar (A Central) University, Lucknow, UP, India
[c] Computer Science & Engineering, Shri Ramswaroop Memorial University Lucknow Deva Road, Barabanki, India, 225003
✉ ganesh.iiscgate@gmail.com, skd200@yahoo.com, satyabverma1@gmail.com, dixitmits@mitsgwalior.in

KEYWORDS

word sense disambiguation; knowledge based approach; supervised approach; unsupervised approach

ABSTRACT

The process of finding the correct sense of a word in context is known as word sense disambiguation (WSD). In the field of natural language processing, WSD has become a growing research area. Over the decades, so many researchers have proposed the many approaches to WSD. A development of this field has created the significant impact on several Web-based applications such as information retrieval and information extraction. This paper contains the description of various approaches such as knowledge-based, supervised, unsupervised and semi-supervised. This paper also describes the various applications of WSD, such as information retrieval, machine translation, speech recognition, computational advertising, text processing, classification of documents and biometrics.

1. Introduction

Language is a powerful communication medium for sharing of information. The word or phrases in any language having more than one meaning is known as ambiguous. For example, in the sentence «what is apple rate» here the word apple has two meanings, apple fruit or apple phone. Nowadays majority of people used web for searching of information. Searching may suffer if ambiguity occurs. Ambiguity is the problem of natural language processing (NLP) which can change the sense of word

used in the sentence or query in different ways (Kaplan, 1955; Zipf, 2013; Chandra et al., 2019; Chandra et al., 2020).

Word sense disambiguation (WSD) is the technique of identifying the correct sense (meaning) of word, if word have more than one meaning or sense in a sentence (Weiss, 1973; Madhu et al., 1965; Resnik et al., 1999; Brown, 2005; Chandra et al., 2014). WSD is highly demanding technology in various applications such as information retrieval (IR), information extraction, machine translation (MT) (Masterman, 1961), question answering system, cross lingual applications, and document classification. WSD uses corporate, training, and lexical databases to provide the correct meaning of particular word in a given context (Edmonds et al., 2007).

The general structure of WSD is shown in figure 1 where several meanings are obtained from various resources for ambiguous word. The disambiguation algorithm is applied on obtained results to fetch the correct meaning. WSD research began in the early days of machine translation in the late 1940s as a distinct computational task. In 1949, the problem of WSD in computational context was firstly introduced by Warren Weaver (1949). Later, Bar-Hillel (1960) stated that the problem of WSD cannot be easily solved by computer (Bar-Hillel, 1960). The problem of WSD was associated with AI in 1970s as subtask of semantic interpretation system. By the 1980s, WSD research reached at their turning point due to the development of lexical resources such as Oxford Advanced Learner's Dictionary of Current English (OALD). Lexical resources are used to extract knowledge using automatic method. In the 1990s, the development of Word Net, sensval and statistical revolution in NLP moves in the direction of computational linguistics. During this period, WSD research started to use the supervised learning technique. During 2000s, research moved towards coarse-grained senses, domain adaptation, semi-supervised, unsupervised, a combination of different methods or approaches, and knowledge based on graph-based methods.

The paper is organized as follows: Section 2 describes WSD approaches, Section 3 presents the resources of WSD, Section 4 describes the importance and applications of WSD, Section 5 describes the challenges of WSD, and finally the paper is concluded in Section 6.

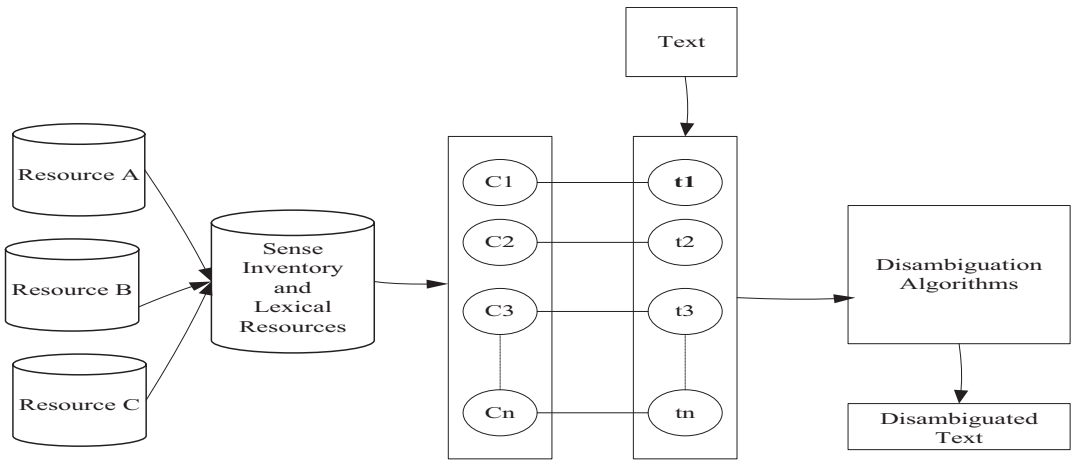

Figure 1. General model of WSD (Kwong, 2012)

2. WSD Approaches

In the field of information retrieval, WSD plays an important role. Firstly, improves the retrieval precision by identifying the correct meaning against an ambiguous word in both query & document. Secondly improves the recall results by improving the relationship between the query words and which are not in query. WSD is highly useful in improving the quality of retrieval results. In WSD, the matching is performed between sense knowledge vector and context vector for word disambiguation.

In order to solve the issue of ambiguity, the various approaches used for WSD are knowledge-based, supervised (Niu et al., 2005), unsupervised (Leacock et al., 1998) and semi-supervised. These are described in the subsections below. Table 1 shows the previous research work on WSD.

Table 1. Research work on WSD

No.	Name of Researcher	Research Title/Work	Year
1.	Weaver (1949)	Translation. Mimeographed.	1949
2.	Kelly et al., (1975)	Computer Recognition of English Word Senses.	1975
3.	Lesk (1986)	Automatic Sense Disambiguation Using Machine Readable Dictionaries.	1986
4.	Black (1988)	An Experiment in Computational Discrimination of English Word Senses.	1988
5.	Guthrie et al., (1991)	Subject Dependent Co-Occurrence and Word Sense Disambiguation.	1991
6.	Brown et al., (1991)	Word-Sense Disambiguation Using statistical Methods.	1991
7.	Yarowsky (1992)	Word Sense Disambiguation Using Statistical Models of Roget's Categories Trained on Large Corpora.	1992
8.	Pedersen (2007)	Unsupervised Corpus-Based Methods for WSD.	2006
9.	McCarthy (2009)	Word Sense Disambiguation: An Overview.	2009
10.	Navigli et al., (2009)	Word Sense Disambiguation: A Survey.	2009
11.	Jin et al., (2010)	A Knowledge Based Method for Chinese Word Sense Induction.	2010
12.	Navigli (2012)	Multilingual WSD with Just a Few Lines of Code: the BabelNet API.	2012
13.	Jin (2013)	A Word Sense Probabilistic Topic Model.	2013
14.	Adala et al., (2014)	A Novel Semantic Approach for Web Service Discovery Using Computational Linguistics Techniques.	2014
15.	Singh et al., (2015)	Role of Semantic Relations in Hindi Word Sense Disambiguation.	2015
16.	Duque et al., (2016)	Can Multilinguality Improve Biomedical Word Sense Disambiguation.	2016
17.	Raganato et al., (2017)	Word Sense Disambiguation: A Unified Evaluation Framework and Empirical Comparison.	2017
18.	Correa et al., (2018)	Word Sense Disambiguation: A Complex Network Approach.	2018
19.	Elayeb (2018)	Arabic Word Sense Disambiguation: A Review.	2019
20.	Singh et al., (2019)	Word Sense Disambiguation for Punjabi Language Using Deep Learning Techniques.	2020
21.	Kohli (2021)	Transfer Learning and Augmentation for Word Sense Disambiguation.	2021'
22.	Chandrika et al., (2022)	Word Sense Disambiguation for Indian Regional Language Using BERT Model.	2022

2.1. Knowledge Based Approach

In this approach, the various resources such as dictionary, thesaurus, WordNet, SemCor, Wikipedia etc. are used to identify the appropriate meaning of word in a context. This approach came into existence in between 1970s to 1980s (McInnes, 2009).

Figure 2 describes the general structure of knowledge-based approach. For disambiguation of words, this approach uses hand coded rule or grammar rule. The algorithms used in this approach are selectional preference, Overlap of Sense Definitions.

This approach either uses hand coded rule or grammar rule for the disambiguation of words and uses the sense whose definition is most similar to the context of the ambiguous word. The various knowledge-based approaches are described as follows (Jin and Chen, 2013).

2.1.1. Selectional Preferences

The knowledge-based method of WSD frequently uses selectional preferences as a source of linguistic information (McCarthy et al., 2003). This approach finds the information related to the relationship between different types of word categories. Although, it is important process of disambiguation but its restricted semantic significance makes it less successful. This is not about the individual word form for disambiguation, rather it is specific to verb or adjective classes. The basic assumption made by selectional preferences is that the distinct senses of the same predicating word (verb or adjective) will have different selectional preferences with respect their arguments (noun). The simplest way of understanding the selectional preferences is to determine the semantic association provided by word-to-word relationship Verma, S. B., & Saravanan, C. (2018, September).

In Figure 3, selectional preferences are obtained from automatically preprocessed and parse text. The parser is applied to data to identify the grammatical relationship between nouns, verbs, and adjectives. Selectional preferences are acquired for the grammatical relationship between nouns, verbs, and adjectives using WordNet synsets to define the sense inventory Verma, S. B., Yadav, A. K. (2021).

2.1.2. Overlap of Sense Definitions

Overlap of sense definition is a knowledge-based approach depends on the calculation of word overlap between the sense definition of target words (two or more). This is also called as Lesk algorithm or

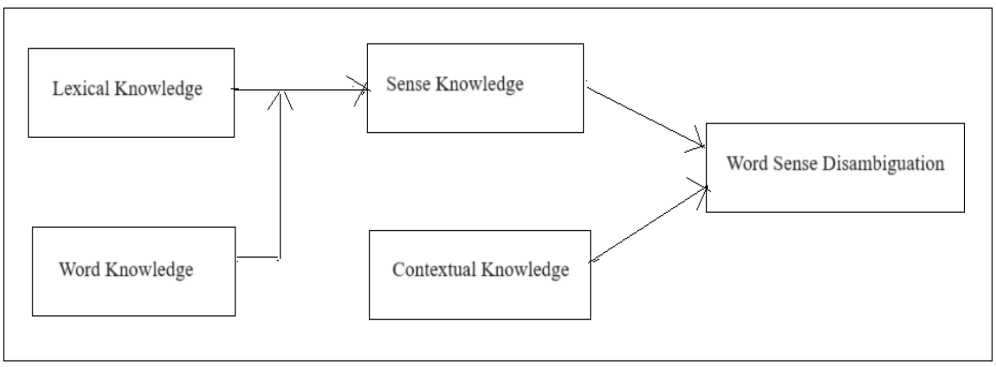

Figure 2. Knowledge-based WSD method

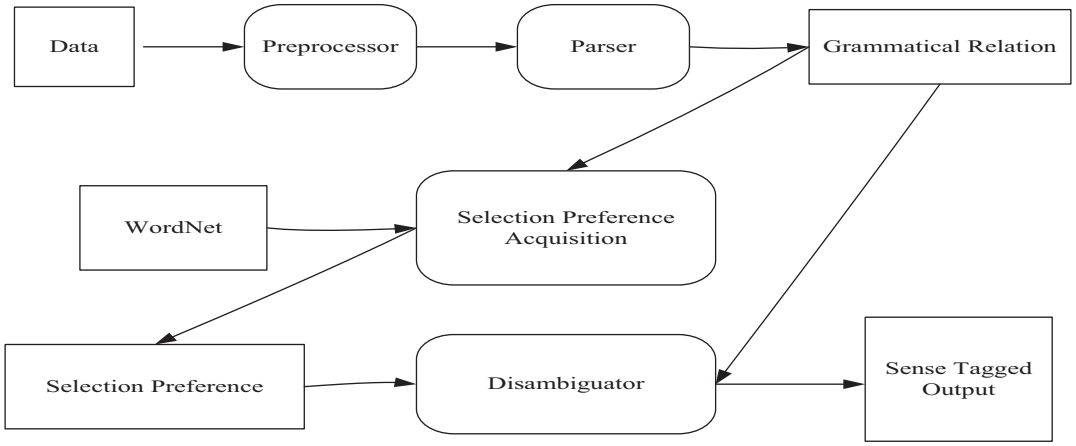

Figure 3. Process of selection preferences (McInnes et al., 2009)

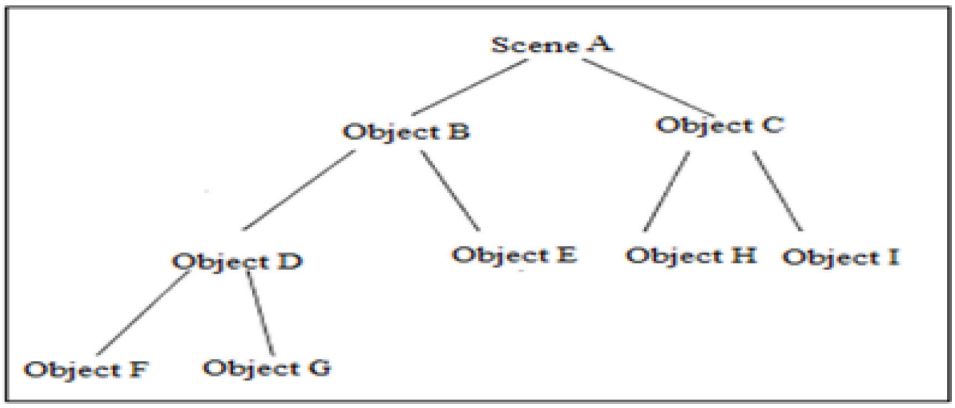

Figure 4. Hierarchical structural representation

Gloss overlap (Lesk, 1986) which utilizes two types of bags: sense and target bag. Sense bag contains the words in the definition of a candidate sense of the ambiguous word and context bag contains the words in the definition of each context word. A machine-readable dictionary (MRD) is used in this algorithm and its basic function is to identify an overlap between the features of different sense of an ambiguous word and the word found in its context. Here, weight is assigned to each sense and the sense that has the maximum overlap value is nominated as the contextually appropriate word (Pedersen, 2007).

2.1.3. Structural Approaches

The structural or syntactic approach is used to represent the hierarchical (tree-like) structure (as shown in Figure 4). Here, structural information is used for the recognition and measurements of patterns, both in a local and global context (Bunke et al.,1990). Its aim is to classify data based on the structural interrelationships of features.

2.1.4. Discussion

An advantage of the knowledge-based approach over the other approaches of WSD is that data is not required for each word that needs to be disambiguated. Another advantage of this approach is that it allows the systems to disambiguate the words in running text, which is referred to as all-words disambiguation. An all-words disambiguation method is more important than lexical-sample disambiguation methods because lexical-sample methods can only disambiguate words for which there is a large set of training data. All-word disambiguation methods are scalable and can be used in real-word practical applications. The disadvantage of the knowledge-based approach is that it is language- and domain-dependent because a knowledge source is required in the appropriate language and domain (Sammut and Webb, 2010).

2.2. Supervised Approach

The supervised learning approach of WSD uses a machine learning technique to induce a classifier from manually sense-annotated datasets. Classifiers (word experts) are used to assign the correct sense of each word for classification in the supervised approach. Supervised approaches have two phases: the training phase and the testing phase (Grozea, 2004). In the training phase, a sense annotated training corpus is used by classifiers to extract the semantic and syntactic features. In the testing phase, the classifiers aim to find the most suitable sense of each word based on surrounding words in the sentence. The supervised approach includes decision lists, naïve Bayes, decision tree and neural networks (Verma et al., 2023).

In Figure 5, the instance is the context in which the target word is used. The ambiguous word that occurs in a sentence is known as the target word. The different meanings of a target word are referred to as senses or concepts which are obtained from a sense inventory or concept inventory. Manually annotated training data is used as input for an evaluation module that divides the data into training and

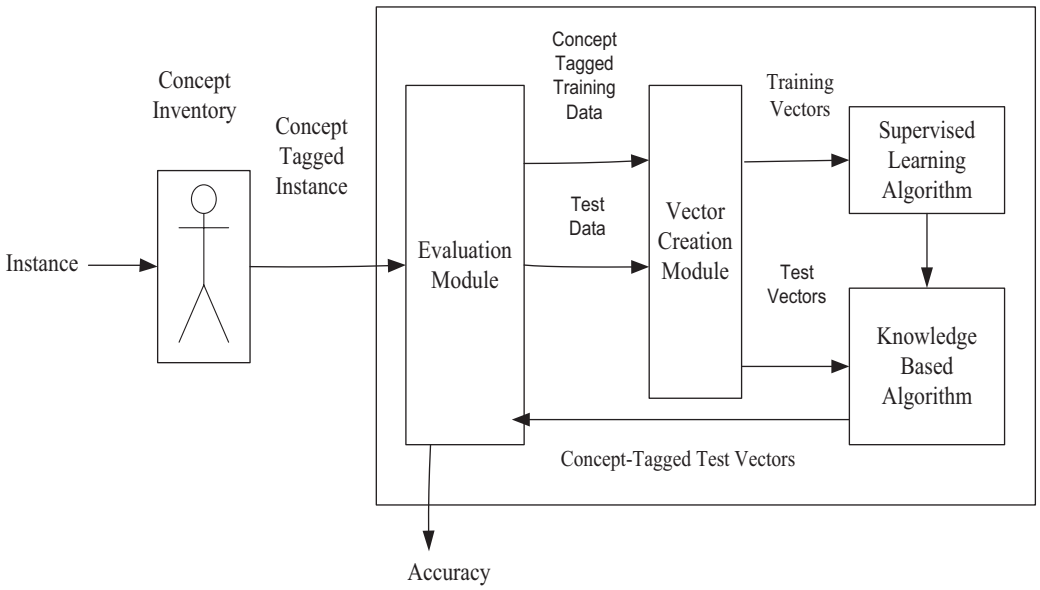

Figure 5. Supervised WSD (McInnes, 2009)

testing data. The vector creation module extracts the features from training data and produces vectors (training vector and test vector). The training vector is used by the supervised learning algorithm as input. The supervised learning model is created by the supervised learning algorithms which take the test vector as input to assign a suitable concept to each of the vectors. The result obtained from supervised learning becomes an input for the evaluation module, where the accuracy of the model is calculated (Chandra and Dwivedi, 2020a).

2.2.1. Decision Lists

The decision lists, first described by Rivest (1987), are an ordered set of rules in a list format which is based on a set of weighted if-then-else rules. In 1994, a decision list was first used by Yarowsky (1994) on the senseval corpus. Decision lists are based on a probabilistic measure where features are extracted from the set of the training corpus. The features that are extracted in the training phase of the WSD algorithm are: part-of-speech, semantic and syntactic features, collocation vector and co-occurrence vector. Once the features are selected, different kinds of rules (features-value, sense, score) are created, and applied in the decreasing order of score to form a decision list.

2.2.2. Naïve Bayes Method

The naïve Bayes method is considered under the probabilistic approach based on the application of Bayes theorem (Berrar, 2019). The conditional probability is calculated for each sense of word for the features given in context. The naïve Bayes classifier uses a large context and tests each word around the target word where each word provides some information for the identification of the sense of the target word. The naïve Bayes classifier mostly assigns the maximum probability to the correct class for accurate results. In this model, texts such as sentences or documents are represented in the form of bag (multiset) of words without using grammar rule and the order of words depends on multiplicity.

2.2.3. Decision Trees

Based on prediction model, decision trees were introduced by Quinlan (1987). In decision tree, a sense-tagged corpus is used as resource on which training is performed. In decision tree, a classification rule is applied in the form of a «yes-no» rule. With the help of these rules, the training dataset is partitioned recursively. Each internal node of a decision tree represents a feature on which experiments are conducted, each branch of a decision tree represents a feature value, and an external node represents the sense. The feature vectors of a decision tree are the same as the feature vectors of a decision list, for the disambiguation of a word in the testing phase, each feature vector is traversed through the tree to reach the leaf node. The sense contained in the leaf node is considered as the correct sense of word. The most popular algorithms used for the generation of a tree are: ID3, ID4, ID5 or ID5R and C4.5 algorithm (Quinlan, 1987).

An example of a decision tree for WSD is reported in Figure 6. For instance, if the noun bank must be classified in the sentence «we sat on a bank of sand», the tree is traversed and, after following the no-yes-no path, the choice of sense bank/RIVER is made. The leaf with empty value (-) indicates that no choice can be made on the basis of those specific feature values.

2.2.4. Neural Networks

A neural network model was proposed by McCulloch and Pitts (1943). It is an interconnected group of artificial neurons that uses a computational model, based on a connectionist approach, for

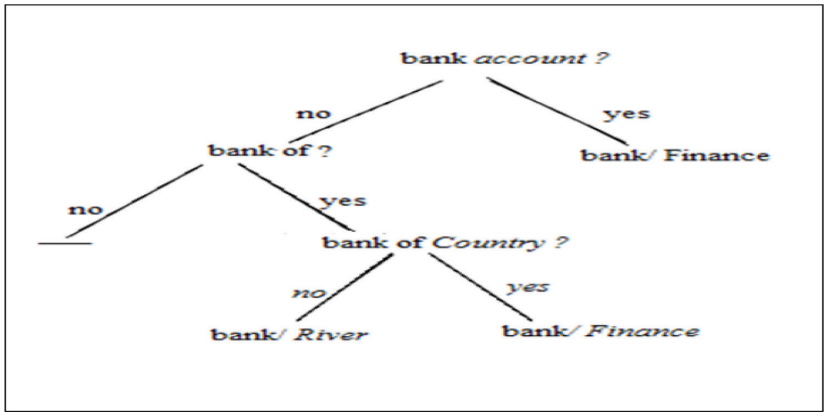

Figure 6. An example of a decision tree

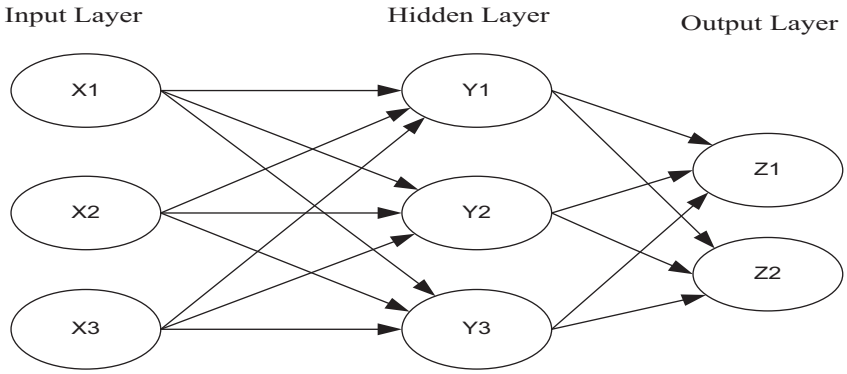

Figure 7. A feed forward neural network WSD with 3 features and 2 responses

the processing of data. In this approach, pairs of input features and desired responses are used as input to the learning program. The different types of neural network models include the perceptron network model, the feed-forward network model and the recurrent network model. Out of these models, the perceptron model is used for the purpose of WSD. The perceptron model is based on the hidden Markov model (HMM) and the backpropagation feed-forward networks. The purpose of this approach is to partition the training contexts into non-overlapping sets for the desired responses using input features. Major problems occur in neural networks, namely, difficulties in interpreting the results, the need for a large quantity of training data, and the tuning of parameters such as thresholds, decay, etc.

Figure 7 represents the multilayer perceptron neural network (input layer, hidden layer, and output layer) which has three features and outputs the two senses of a target word in context.

2.2.5. Support Vector Machines

Support vector machine (SVM) was proposed by Boser et al., (1992) and it plays an important role in classification problems. SVM are based on the structural risk minimization principle. SVMs are a linear

classifier that produces the hyperplane for separating the positive and negative training examples (called support vector) with largest margin. Nowadays, SVM is also used to solve the problems of pattern recognition, image recognition and NLP-related problems such as chunking, parsing and WSD etc.

2.2.6. Discussion

Word sense ambiguity is considered a central problem among NLP researchers across different languages. To solve this problem, many international research groups are actively working on WSD approaches. The supervised approach uses semantically annotated corpora to train machine learning algorithms to decide which sense of word fits which contexts. The words in these annotated corpora are manually tagged with semantic classes taken from a particular lexical semantic resource. To improve the performance of current supervised WSD systems, the set of features with linguistic knowledge must be enriched, however, this is currently not available in wide-coverage lexical knowledge bases. In general, supervised learning methods produce very high disambiguation accuracy in comparison to other WSD methods. The disadvantage of supervised learning is that it requires manually annotated training data for each word that needs to be disambiguated. This is a labor intensive and time-consuming process. The major challenges found in this approach are data quality and data selection (Ng, H.T, 1997).

2.3. Unsupervised Approach

In the supervised approach, the training data must be manually created, which is very expensive (Wang, 2006). This problem is known as knowledge acquisition bottleneck. The unsupervised approach solves this problem by introducing the concept that the sense of a particular word, depends on its neighboring words. The unsupervised WSD approach divides the occurrence of a specific word into a number of classes in order to decide whether the occurrence of that word has the same sense or not. The important task of this approach is to identify sense clusters. The different types of methods used in the unsupervised approach are context clustering, word clustering and co-occurrence graph (Chen et al., 2012). In the context clustering method, clusters are created by finding the co-occurrence of a word with the target word. In the word clustering method, words are created according to their semantic nature (e.g., subject-verb, adjective-noun etc.). In the graph-based method, a graph is created with the help of a grammatical relationship (Navigli, 2010).

2.3.1. Context Clustering

In the context clustering method, every occurrence of a target word is represented as a context vector in corpus. These vectors are grouped into clusters for the identification of the sense of the target word. The most famous context clustering algorithm was proposed by Schutze, where the occurrences of an ambiguous word are grouped into clusters of senses based on the contextual similarity of their occurrences.

Figure 8 represents the general model of the clustering method. An evaluation module takes the test data as input for evaluation and then sends this data to the vector creation module. The test data and unannotated training data is used by the vector creation module as input. A vector is created by the vector creation module for each instance of testing data and training data. The training vector is grouped together into clusters using clustering algorithms in discrimination module. The test vectors and clusters are used by the disambiguation module as input to assign a concept for each cluster

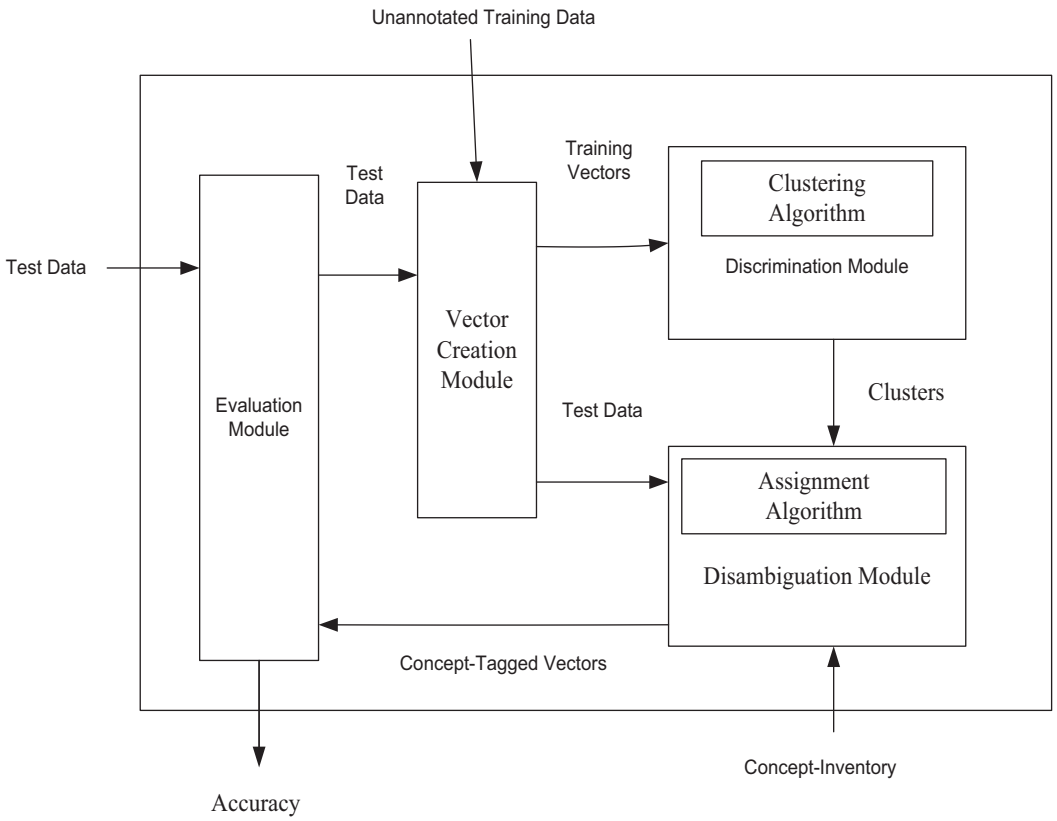

Figure 8. WSD clustering method (McInnes, 2009)

using assignment algorithms. The concept whose vector is closest to the test vector is assigned to the target word. After that, the accuracy of the system is calculated by evaluation module SBV, BP, and BKG (2022).

Advantages:

- A large amount of manually annotated training data is not required.
- Language and domain independent.

Disadvantages:

- Training data is required for each word that needs to be disambiguated.

2.3.2. Word Clustering

In this method, the words that are semantically similar are clustered to form a specific meaning. The word clustering approach was introduced by Lin (1998). This approach consists in the identification of words that are similar to the target word. The relationship between the identified word and target word depends on the information content for single features, given by syntactic dependencies in

a corpus (e.g., subject-verb, verb-object, adjective-noun, etc.). The higher the information content, the more dependencies occur between words.

2.3.3. Co-Occurrence Graphs

In this method, a graph is created on the basis of the grammatical relationship between words. Co-occurrence graphs were introduced by Widdows and Dorow (2002). In a co-occurrence graph, every word in the text is called a vertex and, the syntactic relationship is called an edge. In this method, weight is assigned to the edge on the basis of the relationship. An iterative algorithm is applied to the graph to find the word that has highest degree node and, at last, minimum spanning tree is used to disambiguate the target word.

2.3.4. Discussion

Unsupervised learning is the greatest challenge for WSD researchers. One of the assumptions of this technique is that similar senses occur in similar contexts. In this technique, the senses can be induced from text by clustering word occurrences using measures of similarity of context. A task in which senses occur is referred to as word sense induction or discrimination. The new occurrences of the word can be classified into the closest induced clusters/senses. The performance of unsupervised methods is lower than that of other WSD methods. It is hoped that unsupervised learning will overcome the knowledge acquisition bottleneck because it is not dependent on manual effort. The limitations of unsupervised learning include its unsuitability for large-scale situations, the instances in training data may not assign the correct sense, clusters are heterogeneous, and the number of clusters may differ from the number of senses of a target word to be disambiguated.

2.4. Semi-Supervised Approach

The semi-supervised approach is also referred to as the weakly supervised approach or the minimally supervised approach. This approach uses both types of data: labeled data and unlabeled data. Mostly, fewer amounts of labeled data are used. In this approach, the bootstrapping algorithm is commonly used for WSD. This approach addresses the problem of how unlabeled data can be used with labeled data, to build better classifiers. The primary focus of the semi-supervised approach is to improve the performance of WSD. This approach is gaining popularity in the WSD field because small amounts of labelled data and unlabeled data are used. The bootstrapping method was the first semi-supervised learning approach for WSD and an advantage of bootstrapping is its simplicity. The semi-supervised approach of WSD is inexpensive and less time-consuming in comparison to other approaches.

3. Resources for WSD

The resources for WSD are described in the subsections that follow.

3.1. Machine Readable Dictionaries

Machine-readable dictionaries (MRDs) came into existence in the 1980s to provide a knowledge source for human-language modeling (Walker et al., 1986). The Longman Dictionary of Contemporary English (LDOCE) is frequently used in WSD. Many researchers such as Lesk have used

the MRDs as a structured source of lexical knowledge in their work for WSD systems. MRDs contain many inconsistencies and are created for human use only. They are not available for machine exploitation.

3.2. Thesaurus

Thesaurus provides information about the relationship between words, such as their synonyms and antonyms etc. In 1911, «Roget's International Thesaurus» came into existence in the field of WSD. Many researchers have used thesaurus in their research on the disambiguation of words (Miller et al., 1995; Resnik, 1999). WordNet, which was compiled at Princeton University, is the most popular thesaurus. Each of WordNet's 90 000 words is assigned to one or more synsets. A synset is a set of synonym words. WordNet's semantic network of noun synsets was used by the author for disambiguation. The author assigned a weight to all the synset relations to calculate the semantic distance between any two words in the semantic network. Thesaurus provides a rich network of word associations and a set of semantic categories which are potentially valuable for large-scale language processing (Grefenstette, 1993).

3.3. Wikipedia

Wikipedia is a collaborative Web encyclopedia composed of pages. A Wikipedia page provides knowledge about a specific concept [e.g., Soda (soft drink) or named entity e.g., food standard agency]. Each page typically contains hypertext linked to other relevant Wikipages. For instance, Soda (soft drink) is linked to Cola, flavored water and many others. Some corpora listed in following Table 2.

Table 2. List of some corpora

No.	Corpora	Description	Language	Year
	Monolingual Raw/POS-Tagged Corpora			
1.	Brown Corpus (Francis, 1979)	W.N. Francis &H. Kucera at Brown University	American Language	1979
2.	Wall Street Journal (WSJ) (Charniak et al., 2000)	Eugene Charniak at Philadelhpia University	English language	2000
3.	British National Corpus (Burnard, 2007)	Lou Burnard at Oxford University	British English language	2007
4.	WaCky (Web as Corpora Kool Yntiative; Baroni et al., 2009)	Macro Baroni at University of Trento	English (ukWaC), German (deWaC), Italian (itWaC)	2009
5.	Wikipedia dumps (Li, C, 2011)	Chenliang Li	English	2011
	Monolingual Sense-Tagged Corpora			
1.	SemCor (Miller et al., 1994)	George A. Miller at Princeton University	English	1994

(continued)

Table 2. List of some corpora (continued)

No.	Corpora	Description	Language	Year
2.	Defense Science Organization of Singapore (DSO) corpus (Ng et al., 1996)	Hwee Tou Ng &Hian Beng Lee at National University of Singapore	English	1996
3.	Open Mind Word Expert corpus (Chklovski et al., 2002)	Chklovski, Timothy & Rada Mihalcea at University of Texas	English	2002
4.	Senseval and SemEval data (Navigli, 2009)	Roberto Navigli at Herstmonceux Castle, Sussex, England	English, French, Italian, Basque, Chinese, Czech, Danish, Dutch, Estonian, Italian Japanese, Korean, Spanish & Swedish	2009
Parallel Raw/Aligned Corpora				
1.	Tatoeba Corpus (Breen, 2003)	Prof. Yasuh ito Tanaka at Hyogo Univeristy	Japanese, English, French, Esperanto, German, Polish, Russian, Spanish, Chinese, Ukrainian	2001
2.	Europarl Corpus (Koehn, 2005)	Philipp Koehn at Phuket Thailand	Danish, German, Greek, English, Spanish, Finnish, French, Italian, Dutch, Portuguese, and Swedish	2005
Parallel Sense-Tagged Corpora				
1.	MultiSemCor (Bentivogli, 2004)	Luisa Bentivogli & Emanuele Pianta at Princeton University	English, Italian	2004
2.	NTU-Multilingual Corpus (Tan et al., 2011)	Tan & Bond at Singapore	Arabic, Chinese, English, Indonesian, Japanese, Korean & Vietnamese	2011
3.	Japanese SemCor (JSemCor; Bond et al., 2012)	Francis Bond at Nanyang Technological University Singapore	English, Japanese	2012

4. WSD Importance and Applications

The global information exchange between different languages increases as the size and the use of Internet expands. WSD is beneficial for cross lingual information retrieval (CLIR) (Dwivedi et al., 2016) or monolingual information retrieval (IR). WSD helps to solve the ambiguity problem in query translation and document translation (Chandra et al., 2017). Thus, the most suitable information can be retrieved easily between different languages. WSD can also play an important role in the following areas: multilingual document retrieval, scientific data retrieval (domain-specific), interactive cross-language retrieval (ICLR), multiple language question answering (MLQA), cross-language image retrieval (ImageCLEF), CLEF web track (WebCLEF), cross-language geographical information retrieval (CLGIR), cross-language video retrieval (CLVR), multilingual information filtering (MIF).

4.1. Information Retrieval

Current search engines do not provide the most relevant documents to the user query. WSD can be useful for improving the accuracy of existing IR systems through a better understanding of queries. Krovetz and Croft (1992) proved that word senses do improve retrieval performance if senses are included as index terms. Apart from indexing, WSD can also be used in query expansion to fetch better results than those obtained by using the original query.

4.2. Machine Translation

WSD is required in machine translation (MT) for different senses of words that are potentially ambiguous within a given domain. Dagan et al., (1994) as well as Vickrey et al., (2005) showed in their research that the performance of MT systems increases with the proper utilization of WSD.

4.3. Speech Recognition

WSD is also used in speech recognition to provide accurate results. For instance, in the processing of homophones (words which are spelled differently but pronounced the same) for example: «sun» and «son» (Jain et al., 2012).

4.4. Computational Advertising

Computational advertising is a new sub-discipline of science and its central challenge is to find the best advertisement for user in a given context, such as querying a search engine or reading a web page. WSD identifies the appropriate meanings of the main terms in the given context and provide the best advertisement to match the given query or page.

4.5. Text Processing

The goal of text processing is to produce a set of indexing terms that make the best use of resources and produce an accurate matching of user query (Chandra et al., 2020). Text processing also helps to automatically detect the language of text i.e., when words are pronounced in more than one way depending on their meaning. For example: «lead» can be «in front of» or «type of metal». In such a case, WSD helps to detect the proper meaning of the words.

4.6. Classification of Documents

Document classification has been used to enhance information retrieval. Document classification is based on the clustering concept, which states that documents which have similar contents are also relevant to the same query. WSD provides a way to cluster documents that have a similar meaning.

4.7. Biometrics

Bioinformatics study relationships between genes and gene products, however, genes and their proteins often have the same name. WSD solves the problem of ambiguity where genes have different senses or meaning.

5. Challenges in WSD

5.1. Differences Between Dictionaries

One of the open problems of WSD is to decide on the correct sense of words. Different dictionaries define words differently. The solution to this problem is to select a particular dictionary when defining the set of senses. In some cases, the many senses are closely related to each other so the division of words into senses becomes much more difficult. Different dictionaries and thesauruses will provide different divisions of words into senses. To solve such problems, some researchers select a particular dictionary, and limit themselves to the use its set of senses (Pradhan et al., 2007). The research results obtained using broad distinctions in senses are better than the result obtained using narrow ones. However, given the lack of a full-fledged coarse-grained sense inventory, most researchers continue to work on fine-grained WSD (Sammut and Webb, 2010).

5.2. Part-of-Speech Tagging

In any real test, part-of-speech tagging, and sense tagging are closely related to each other. Now one of the questions is whether these tasks should be kept together or decoupled is still not unanimously resolved. Another question is whether part-of-speech tagging, and sense tagging are same or different because part-of-speech of a word is determined by immediate one to three words whereas the sense tagging of a word may be determined by words further way.

5.3. Human Interaction

A human provides a better result than a computer when searching for the sense of words. Thus, WSD systems generally tested their results on a task by comparing against a human. However, it is difficult for a human to memorize all the possible senses of words (Navigli et al., 2006).

5.4. Common Sense

Sometimes it is difficult to parse the meaning of words without common sense. Let's take the following two sentences as an example:

- «Ram and Shyam are fathers.»
- «Ram and Shyam are brothers.»

In first sentence each is independently a father, whereas in the second sentence they are siblings. Thus, for the proper identification of word senses, common sense is sometimes needed.

5.5. Sense Inventory and Algorithms' Task-Dependency

It is difficult to find the accurate sense of a word when multiple languages are at play. For example, the ambiguity of «mouse» (animal or device) is not relevant in English-French machine translation.

6. Conclusion

WSD is the toughest open problem of NLP with roots as old as that of machine translation. The different problems that occur in WSD are caused by WSD's dependance on knowledge drawn from different resources. WSD also deals with the complexities of languages. This paper surveyed the field of WSD and discussed the range of WSD approaches that are useful in disambiguating the senses of words. Knowledge-based, supervised, unsupervised and semi-supervised approaches are used in WSD. The supervised approach performs best in comparison to all the other approaches because the training data is fully dependent on a specific domain.

References

Adala, A., Tabbane, N., & Tabbane, S. (2014, March). A novel semantic approach for Web service discovery using computational linguistics techniques. In *Fourth International Conference on Communications and Networking, ComNet-2014* (pp. 1-6). IEEE. https://doi.org/10.1109/ComNet.2014.6840909

Bar-Hillel, Y. (1960). The Present Status of Automatic Translation of Languages. *Advances in computers, 1*, 91–163. https://doi.org/10.1016/s0065-2458(08)60607-5

Baroni, M., Bernardini, S., Ferraresi, A., & Zanchetta, E. (2009). The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation*, *43*(3), 209–226. https://doi.org/10.1007/s10579-009-9081-4

Bentivogli, L., Forner, P., & Pianta, E. (2004). Evaluating cross-language annotation transfer in the multisemcor corpus. In *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics* (pp. 364-370). https://doi.org/10.3115/1220355.1220408

Berrar, D. (2019). Bayes' theorem and naive Bayes classifier. *Encyclopedia of Bioinformatics and Computational Biology, 1*, 403–412. https://doi.org/10.1016/b978-0-12-809633-8.20473-1

Black, E. (1988). An experiment in computational discrimination of English word senses. *IBM Journal of research and development*, *32*(2), 185-194. https://doi.org/10.1147/rd.322.0185

Bond, F., Baldwin, T., Fothergill, R., & Uchimoto, K. (2012, January). Japanese SemCor: A sense-tagged corpus of Japanese. In *Proceedings of the 6th global WordNet conference (GWC 2012)* (pp. 56-63).

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992, July). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory* (pp. 144-152). https://doi.org/10.1145/130385.130401

Breen, J. W. (2003, July). Word usage examples in an electronic dictionary. In *Papillon (Multi-lingual Dictionary) Project Workshop*.

Brown, K. (2005). *Encyclopedia of language and linguistics* (Vol. 1). Elsevier.

Brown, P. F., Della Pietra, S. A., Della Pietra, V. J., & Mercer, R. L. (1991, June). Word-sense disambiguation using statistical methods. In *29th Annual meeting of the Association for Computational Linguistics* (pp. 264-270). https://doi.org/10.3115/981344.981378

Bunke, H., & Sanfeliu, A. (1990). Syntactic and structural pattern recognition: theory and applications (Vol. 7). In *World Scientific eBooks*. https://doi.org/10.1142/0580

Burnard, L. (2007). *Reference Guide for the British National Corpus (XML Edition)*. Published for the British National Corpus Consortium by the Research Technologies Service at Oxford University Computing Services.

Chandra, G., & Dwivedi, S. K. (2014, December). A literature survey on various approaches of word sense disambiguation. In *2014 2nd International Symposium on Computational and Business Intelligence* (pp. 106-109). IEEE. https://doi.org/10.1109/ISCBI.2014.30

Chandra, G., & Dwivedi, S. K. (2017). Assessing query translation quality using back translation in hindi-english CLIR. *International Journal of Intelligent Systems and Applications*, *9*(3), 51–59. https://doi.org/10.5815/ijisa.2017.03.07

Chandra, G., & Dwivedi, S. K. (2019). Query expansion for effective retrieval results of Hindi–English Cross-Lingual IR. *Applied Artificial Intelligence*, *33*(7), 567–593. https://doi.org/10.1080/08839514.2019.1577018

Chandra, G., & Dwivedi, S. K. (2020). Query expansion based on term selection for Hindi – English cross lingual IR. *Journal of King Saud University - Computer and Information Sciences*, *32*(3), 310–319. https://doi.org/10.1016/j.jksuci.2017.09.002

Chandra, G., & Dwivedi, S. K. (2020a). Term Ordering-Based Query Expansion Technique for Hindi-English CLIR System. In *Advances in data mining and database management book series* (pp. 283–302). https://doi.org/10.4018/978-1-7998-2491-6.ch016

Chandrika, C. P., & Kallimani, J. S. (2022, May). Word Sense Disambiguation for Indian Regional Language Using BERT Model. In *Smart Intelligent Computing and Applications*, Volume 2: Proceedings of Fifth International Conference on Smart Computing and Informatics (SCI 2021) (pp. 127-137). Singapore: Springer Nature Singapore.

Charniak, E., Blaheta, D., Ge, N., Hall, K., Hale, J., & Johnson, M. (2000). Bllip 1987-89 wsj corpus release 1. *Linguistic Data Consortium, Philadelphia, 36*.

Chen, P., Bowes, C., Ding, W., & Choly, M. (2012). Word Sense Disambiguation with Automatically Acquired Knowledge. *IEEE Intelligent Systems*, *27*(4), 46–55. https://doi.org/10.1109/mis.2010.134

Chklovski, T., & Mihalcea, R. (2002, July). Building a sense tagged corpus with open mind word expert. *WSD '02: Proceedings of the ACL-02 Workshop on Word Sense Disambiguation: Recent Successes and Future Directions*, *8*, 116–122. https://doi.org/10.3115/1118675.1118692

Correa Jr, E. A., Lopes, A. A., & Amancio, D. R. (2018). Word sense disambiguation: A complex network approach. *Information Sciences*, *442*, 103-113. https://doi.org/10.1016/j.ins.2018.02.047

Dagan, I., & Itai, A. (1994). Word sense disambiguation using a second language monolingual corpus. *Computational linguistics*, *20*(4), 563-596.

Duque, A., Martinez-Romo, J., & Araujo, L. (2016). Can multilinguality improve Biomedical Word Sense Disambiguation? *Journal of Biomedical Informatics*, *64*, 320–332. https://doi.org/10.1016/j.jbi.2016.10.020

Dwivedi, S. K., & Chandra, G. (2016). A Survey on Cross Language Information Retrieval. *International Journal On Cybernetics & Informatics*, *5*(1), 127-142. https://doi.org/10.5121/ijci.2016.5113

Edmonds, P., & Agirre, E. (2007). *Word Sense Disambiguation: Algorithms and Applications*. Springer Dordrecht. https://doi.org/10.1007/978-1-4020-4809-8

Elayeb, B. (2018). Arabic word sense disambiguation: a review. *Artificial Intelligence Review*, *52*(4), 2475–2532. https://doi.org/10.1007/s10462-018-9622-6

Francis, W. N., & Kucera, H. (1979). Brown Corpus Manual, Department of Linguistics, Brown University. Brown corpus manual. Department of Liguistics Brown University.

Grefenstette, G. (1993). Evaluation techniques for automatic semantic extraction: comparing syntactic and window-based approaches. In *Acquisition of Lexical Knowledge from Text*.

Grozea, C. (2004, July). Finding optimal parameter settings for high performance word sense disambiguation. In *Proceedings of SENSEVAL-3, the third international workshop on the evaluation of systems for the semantic analysis of text* (pp. 125-128).

Guthriee, J. A., Guthrie, L., Aidinejad, H., & Wilks, Y. (1991, June). Subject-dependent co-occurrence and word sense disambiguation. In *29th Annual meeting of the Association for Computational Linguistics* (pp. 146-152).

Jain, R., & Sulochana Nathawat, D. G. (2012). Sense Disambiguation Techniques: A Survey. *International Journal*, *1*(1).

Jin, P., & Chen, X. (2013, December). A Word Sense Probabilistic Topic Model. In *2013 Ninth International Conference on Computational Intelligence and Security* (pp. 401-404). IEEE. https://doi.org/10.1109/CIS.2013.91

Jin, P., Sui, R., & Zhang, Y. (2010, December). A Knowledge based Method for Chinese Word Sense Induction. In 2010 *Fourth International Conference on Genetic and Evolutionary Computing* (pp. 248-251). IEEE.

Kaplan, A. (1955). An experiment study of ambiguity and context. *Mechanical Translation*, *2*, 39-46.

Kelly, E. & Stone, P. (1975). *Computer Recognition of English Word Senses*, Amsterdam: North-Holland.

Koehn, P. (2005). Europarl: A parallel corpus for statistical machine translation. In *Proceedings of machine translation summit x*: papers (pp. 79-86).

Kohli, H. (2021, March). Transfer learning and augmentation for word sense disambiguation. In *European Conference on Information Retrieval* (pp. 303-311). Cham: Springer International Publishing.

Krovetz, R., & Croft, W. B. (1992). Lexical ambiguity and information retrieval. *ACM Transactions on Office Information Systems*, *10*(2), 115–141. https://doi.org/10.1145/146802.146810

Kwong, O. Y. (2012). New perspectives on computational and cognitive strategies for word sense disambiguation. Springer Science & Business Media.

Leacock, C., Chodorow, M., & Miller, G. A. (1998). Using corpus statistics and WordNet relations for sense identification. *Computational Linguistics*, *24*(1), 147-165.

Lesk, M. (1986, June). Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. In *Proceedings of the 5th annual international conference on Systems documentation* (pp. 24-26). https://doi.org/10.1145/318723.318728

Li, C., Sun, A., & Datta, A. (2011). A generalized method for word sense disambiguation based on wikipedia. In *Advances in Information Retrieval: 33rd European Conference on IR Research, ECIR 2011*, Dublin, Ireland, April 18-21, 2011. Proceedings 33 (pp. 653-664). Springer Berlin Heidelberg.

Lin, D. (1998, August). Automatic retrieval and clustering of similar words. In *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*, Volume 2 (pp. 768-774). https://doi.org/10.3115/980691.980696

Madhu, S., & Lytle, D. W. (1965). A figure of merit technique for the resolution of non-grammatical ambiguity. *Me Mechanical Translation and Computational Linguistics*, *8*(2), 9-13.

Masterman, M. (1961). Semantic message detection for machine translation, using an interlingua. In *Proceedings of the International Conference on Machine Translation and Applied Language Analysis*.

McCarthy, D. (2009). Word sense disambiguation: An overview. *Language and Linguistics compass*, *3*(2), 537-558. https://doi.org/10.1111/j.1749-818x.2009.00131.x

McCarthy, D., & Carroll, J. (2003). Disambiguating nouns, verbs, and adjectives using automatically acquired selectional preferences. *Computational Linguistics*, *29*(4), 639–654. https://doi.org/10.1162/089120103322753365

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, *5*(4), 115–133. https://doi.org/10.1007/bf02478259

McInnes, B. T. (2009). *Supervised and knowledge-based methods for disambiguating terms in biomedical text using the umls and metamap*. University of Minnesota.

Miller, G. A. (1995). WordNet. *Communications of the ACM*, *38*(11), 39–41. https://doi.org/10.1145/219717.219748

Miller, G. A., Chodorow, M., Landes, S., Leacock, C., & Thomas, R. G. (1994). Using a semantic concordance for sense identification. In *Human Language Technology: Proceedings of a Workshop held at Plainsboro, New Jersey*, March 8-11, 1994. https://doi.org/10.3115/1075812.1075866

Navigli, R. (2009). Word sense disambiguation. *ACM Computing Surveys*, *41*(2), 1–69. https://doi.org/10.1145/1459352.1459355

Navigli, R. (2006). Meaningful Clustering of Senses Helps Boost Word Sense Disambiguation Performance. Proc. of the *44th Annual Meeting of the Association for Computational Linguistics joint with the 21st International Conference on Computational Linguistics (COLING-ACL 2006)*, Sydney, Australia. https://doi.org/10.3115/1220175.1220189

Navigli, R., & Lapata, M. (2010). An experimental study of graph connectivity for unsupervised word sense disambiguation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(4), 678–692. https://doi.org/10.1109/tpami.2009.36

Navigli, R., & Ponzetto, S. P. (2012, July). Multilingual WSD with just a few lines of code: the BabelNet API. In *Proceedings of the ACL 2012 System Demonstrations* (pp. 67-72).

Ng, H. T. (1997). Getting serious about word sense disambiguation. In *Tagging Text with Lexical Semantics: Why, What, and How?*

Ng, H. T., & Lee, H. B. (1996). Integrating multiple knowledge sources to disambiguate word sense: An exemplar-based approach. *arXiv preprint* cmp-lg/9606032. https://doi.org/10.3115/981863.981869

Niu, Z. Y., Ji, D., & Tan, C. L. (2005, June). Word sense disambiguation using label propagation based semi-supervised learning. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)* (pp. 395-402). https://doi.org/10.3115/1219840.1219889

Pedersen, T. (2007). Unsupervised Corpus-Based Methods for WSD. In *Text, speech and language technology* (pp. 133–166). https://doi.org/10.1007/978-1-4020-4809-8_6

Pradhan, S., Loper, E., Dligach, D., & Palmer, M. (2007, June). Semeval-2007 task-17: English lexical sample, srl and all words. In *Proceedings of the fourth international workshop on semantic evaluations* (SemEval-2007) (pp. 87-92). https://doi.org/10.3115/1621474.1621490

Quinlan, J. R. (1987, January). *Decision trees as probabilistic classifiers. In Proceedings of the Fourth International Workshop on Machine Learning* (pp. 31-37). Morgan Kaufmann. https://doi.org/10.1016/B978-0-934613-41-5.50007-6

Raganato, A., Camacho-Collados, J., & Navigli, R. (2017). Word sense disambiguation: a uinified evaluation framework and empirical comparison. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*: Volume 1, Long Papers (Vol. 1, pp. 99-110). https://doi.org/10.18653/v1/E17-1010

Resnik, P., & Yarowsky, D. (1999). Distinguishing systems and distinguishing senses: new evaluation methods for Word Sense Disambiguation. *Natural Language Engineering*, *5*(2), 113–133. https://doi.org/10.1017/s1351324999002211

Rivest, R. L. (1987). Learning decision lists. *Machine Learning*, *2*(3), 229–246. https://doi.org/10.1007/bf00058680

Sammut, C., & Webb, G. I. (2010). Encyclopedia of Machine Learning. Springer Science & Business Media. https://doi.org/10.1007/978-0-387-30164-8

Schutze, H. (1998). Automatic word sense discrimination. *Computational linguistics*, *24*(1), 97-123.

Singh, S., & Siddiqui, T. J. (2015). Role of semantic relations in Hindi word sense disambiguation. *Procedia Computer Science*, *46*, 240–248. https://doi.org/10.1016/j.procs.2015.02.017

Singh, V. P., & Kumar, P. (2019). Word sense disambiguation for Punjabi language using deep learning techniques. *Neural Computing and Applications*, *32*(8), 2963–2973. https://doi.org/10.1007/s00521-019-04581-3

Tan, L., & Bond, F. (2011, December). Building and annotating the linguistically diverse NTU-MC (NTU-multilingual corpus). In *Proceedings of the 25th Pacific Asia Conference on Language, Information and Computation* (pp. 362-371).

Verma, S. B., Pandey, B., & Kumar Gupta, B. (2023). Containerization and its Architectures: A Study. *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, *11*(4), 395–409. https://doi.org/10.14201/adcaij.28351

Verma, S. B., & Saravanan, C. (2018, September). Performance analysis of various fusion methods in multimodal biometric. In *2018 International Conference on Computational and Characterization Techniques in Engineering & Sciences* (pp. 5–8). IEEE. https://doi.org/10.1109/CCTES.2018.8674156

Verma, S. B., Yadav, A. K. (2021). Hard Exudates Detection: A Review., Emerging Technologies in Data Mining and Information Security. *Advances in Intelligent Systems and Computing*, vol 1286. Springer, Singapore. https://doi.org/10.1007/978-981-15-9927-9_12

Vickrey, D., Biewald, L., Teyssier, M., & Koller, D. (2005, October). Word-sense disambiguation for machine translation. In *Proceedings of human language technology conference and conference on empirical methods in natural language processing* (pp. 771-778). https://doi.org/10.3115/1220575.1220672

Walker, D., & Amsler, R. (1986). The use of machine-readable dictionaries in sublanguage analysis. *Analyzing Language in Restricted Domains*, 69-83.

Wang, Y. F., Zhang, Y. J., Xu, Z. T., & Zhang, T. (2006, August). Research on dual pattern of unsupervised and supervised word sense disambiguation. In *2006 International Conference on Machine Learning and Cybernetics* (pp. 2665-2669). IEEE. https://doi.org/10.1109/ICMLC.2006.258922

Weaver, W. (1949). Translation. Mimeographed, 12pp. July15, 1949. Reprinted in Locke, W. N. & Booth, A.D. (1955), *Machine Translation of Languages* (pp. 15-23). New York: John Wiley & Sons.

Weiss, S. F. (1973). Learning to disambiguate. *Information Storage and Retrieval*, *9*(1), 33-41. https://doi.org/10.1016/0020-0271(73)90005-3

Widdows, D., & Dorow, B. (2002). A graph model for unsupervised lexical acquisition. In *COLING 2002: The 19th International Conference on Computational Linguistics*. https://doi.org/10.3115/1072228.1072342

Yarowsky, D. (1992). Word-sense disambiguation using statistical models of Roget's categories trained on large corpora. In *COLING 1992 Volume 2: The 14th International Conference on Computational Linguistics*. https://doi.org/10.3115/992133.992140

Yarowsky, D. (1994). Decision lists for lexical ambiguity resolution: Application to accent restoration in Spanish and French. *arXiv preprint* cmp-lg/9406034. https://doi.org/10.3115/981732.981745

Zipf, G. K. (2013). 3. Relative Frequency and Dynamic Equilibrium in Phonology and Morphology. In *Eight Decades of General Linguistics* (pp. 57-75). Brill. https://doi.org/10.1163/9789004242050_005