

Word Sense Disambiguation Using Differential Evolution Algorithm

Elaf Nassir Abud, Suaad Mohsen Saber and Zuhair Hussein Ali

*Department of Computer Science, College of Education, Mustansiriyah University, 10052 Baghdad, Iraq
Elaf.n.sh@uomustansiriyah.edu.iq, suaad.m.saber@uomustansiriyah.edu.iq, zuhair72h@uomustansiriyah.edu.iq*

Keywords: Word Sense Disambiguation, Differential Evolution, Word Relatedness, Semcor.

Abstract: Word sense disambiguation (WSD) can be defined as the process of determining the meaning of a word as it is used in a specific context from a set of possible word senses. This task is of significant importance in many applications, such as text summarization, information retrieval, and information extraction. Many researchers have focused on the use of metaheuristic algorithms to find the most suitable sense that reflects the meaning of ambiguous words. However, the use of metaheuristic approaches remains limited and therefore requires efficient exploration of the problem space. In this paper, a new method is proposed to address the WSD problem based on differential evolution. The proposed method consists of two main phases. The first phase concerns estimating a numerical value for word relatedness based on a shortest-path measure. In the second phase, a differential evolution algorithm is applied to select the best sense corresponding to the ambiguous word. The proposed method was evaluated using the SemCor dataset, and performance was measured using three evaluation metrics: recall, precision, and F-score. The results demonstrate that the proposed method outperforms related approaches, particularly in terms of F-score and precision, achieving values of 85.6% and 89.2%, respectively.

1 INTRODUCTION

Dealing with ambiguity is one of the significant tasks of Natural Language Processing (NLP) aimed at facilitating effective communication between computer systems and humans. It is considered a key aspect of natural language understanding. In other words, a single word may have multiple meanings depending on the context in which it appears within a text [1]. Determining the precise meaning of a word in order to resolve ambiguity is regarded as one of the most complex problems in NLP. Humans naturally manage linguistic ambiguity by acquiring and refining semantic knowledge throughout their lives. However, in computational systems, the extraction, representation, and analysis of knowledge play a critical role in resolving ambiguity, as identifying the appropriate meaning often requires considerable time and effort [2].

In the English language, several types of ambiguity exist, including lexical, syntactic, scope, and structural ambiguity. Lexical ambiguity, particularly in written texts, arises from the presence of multiple meanings for a single word. For example, the word bank can refer to a financial institution or the side of a river [3].

The concept of Word Sense Disambiguation (WSD) refers to the process of determining the intended sense of a word that appears in a specific context when the word conveys multiple meanings [4]. It is a subfield of NLP that focuses on automatically identifying the correct meaning of a word by analyzing its surrounding context and assigning it a predefined sense that distinguishes it from alternative interpretations [5]. WSD is commonly defined as the computational task of specifying the correct sense of a word. In artificial intelligence, AI-complete problems are considered among the most challenging to solve [6]. Sense disambiguation is often described as an intermediate task rather than a final objective; however, it is fundamental for many applications that require deep language understanding, such as machine translation [7], text summarization [8], and question answering [9].

WSD approaches can be broadly classified into three main categories: supervised, unsupervised, and knowledge-based methods. The knowledge-based approach relies on external lexical resources and is considered a fundamental component of WSD [10]. Such resources provide rich semantic information for associating words with their possible senses and can

take various forms, including text corpora, machine-readable dictionaries, and ontologies. These resources often utilize grammatical and semantic information to support the disambiguation process [11].

The supervised approach employs sense-annotated corpora for training classifiers to distinguish between different word senses. It typically consists of two phases: a training phase and a testing phase. During the training phase, the classifier learns from labeled examples associated with a target word, while in the testing phase, it attempts to determine the correct sense based on the surrounding context in unseen sentences [12]. In contrast, the unsupervised approach derives contextual information directly from unannotated text, where word senses are inferred using similarity measures or clustering techniques without prior labeling [13].

2 RELATED WORKS

Great effort was exerted to solve the WSD problem since its appearance. The work done on WSD has been greatly expanded because of its wide-ranging applications [14]. There are various methods that have been used to solve the problem of ambiguous words. In 2021, the authors suggested a new method for WSD based on RoBERTa and a bi-encoder model. Sentence pairs are input to the system. The first sentence includes the ambiguous sentence, while the second sentence includes a set of candidate words that represent possible meanings of the ambiguous word. The two sentences are merged together to form a sentence pair that is tokenized and provided as input to the RoBERTa transformer. The proposed model was evaluated using the SemCor benchmark dataset. The system achieved an 82% F-score [15].

In 2021, the authors proposed a bi-encoder approach for Word Sense Disambiguation (WSD), leveraging transformer-based architectures and lexical knowledge from WordNet. The authors experimented with various optimization strategies, including contrastive and triplet loss, and introduced multi-stage pre-training using hypernym data. Their model achieved state-of-the-art results on standard benchmarks, outperforming previous methods such as GlossBERT. The study highlights the benefits of incorporating relational information and efficient embedding-based inference. Future work explores richer semantic relations and advanced sampling techniques [16].

In 2024, the authors suggested a BERT-PageRank technique for Word Sense Disambiguation (WSD)

that combines BERT embeddings with a Multiple Sense Graph. The method improves the accuracy of disambiguation by using PageRank to rank sense importance. It performs better than Word2Vec and standalone BERT, obtaining higher F1 scores, reaching up to 83% in experiments on common datasets (e.g., Senseval, SemEval). Through the integration of semantic graph structures and global context, the approach overcomes the drawbacks of conventional WSD. Richer graph-based representations for polysemous words may be investigated in future research [17].

In 2024, the authors suggested a method based on both Textual-WSD (choosing the correct word senses) and Visual-WSD (choosing relevant images). The authors proposed PolCLIP, a unified image-text model for Word Sense Disambiguation (WSD). In order to create multimodal representations for missing modalities, the model simulates stable diffusion and image captioning by utilizing the CLIP architecture and an image-text complementarity technique. BabelNet and BabelPic are used to align senses with images, creating a disambiguation-oriented dataset. With a 2.53% higher F1-score on Textual-WSD and a 2.22% higher HR@1 on Visual-WSD, PolCLIP outperforms state-of-the-art techniques. Using a refined GPT-3.5 model, the method demonstrates robustness across speech components and benefits from semantic augmentation [18].

In 2025, the authors suggested a sense representation approach. This study presents Learning to Rank Senses (LTRS), a novel approach for WSD that ranks a longer list of definitions. In contrast to traditional methods, LTRS improves disambiguation, especially for low-frequency senses (LFS), by utilizing similar senses across words (such as “wide” and “broad”). The approach incorporates BERT-based encoders for context and definition modeling and uses list-wise ranking objectives (ListNet and ListMLE). With significant improvements in few-shot and zero-shot scenarios, LTRS attains a state-of-the-art F1-score of 79.6% when tested on Chinese WSD. Additionally, it exhibits faster convergence than earlier techniques. Reliance on high-quality sentence embeddings and difficulties with fine-grained sense distinctions are among its drawbacks [19].

In 2025, the authors investigated hybrid transformer-based large language models (LLMs) for low-resource language Sesotho sa Leboa word sense disambiguation (WSD). The authors used hybrid architectures (BiGRU + transformer models) and optimization strategies (Adam, attention

mechanisms) to improve performance and assessed baseline models (RNN-LSTM, BiGRU, LSTMLM, DeBERTa, and DistilBERT). The efficiency of bidirectional context modeling was demonstrated by the maximum accuracy (85%) achieved by BiGRU with attention. The paper identifies domain-specific graph representations and annotated datasets as solutions to the problems of polysemy and data scarcity. According to the results, hybrid models – such as BiGRU + DeBERTa – perform better than standalone models, advancing NLP for underrepresented languages. The work fills gaps in low-resource language processing, thereby advancing computational linguistics [20].

3 MATHEMATICAL FORMULATION OF THE WSD PROBLEM

In NLP, WSD is the task of deciding which meaning of a word is activated by the use of the word in a specific context. This is a process that seems to be highly intuitive for people. WSD could be expressed as follows. Given a text document that can be defined as a set of words $C = \{w_1, w_2, \dots, w_n\}$, the main goal is to find the most suitable sense(s) for all words $w_i \in C$. That is, the goal is to find a relational map M from each word to its sense, such that $M(w_i) \subseteq SD(w_i)$, where $SD(w_i)$ is the set of senses that exist in a dictionary D and represent the word w_i , and $M(w_i)$ is a subset of appropriate senses of $w_i \in C$. There are two types of WSD problems: lexical sample and all-words. For the lexical sample, there is a set of specific words to be disambiguated, while the all-words task intends to disambiguate all words in the document.

4 PRELIMINARY CONCEPTS

In this section, the related background of differential evolution and word relatedness measures is described.

4.1 Basics of Differential Evolution Algorithm

Finding a good near-optimal solution to a problem that is hard to solve is the aim of evolutionary algorithms (EA), which are typically used in optimization approaches. Many optimization problems can be classified under this class [21]. Differential Evolution (DE) belongs to the EA family

proposed by Storn and Price in 1997 [22]. DE is a population-based evolutionary optimization meta-heuristic. DE is a stochastic optimizer that uses a population of real-valued vectors that represent candidate solutions to the problem being solved [23]. This optimization algorithm is considered efficient, simple, and global. Due to its succinct structure and strong search ability, DE has been widely used in different applications [24].

DE has basically four operators: initialization, mutation, crossover, and selection. All population-based algorithms (including DE) consider the generation of the initial population as the first processing step. Initialization begins by using a set of candidate decision variables, which are commonly selected randomly using a uniform distribution within the solution space. Each individual in the population is then subjected to DE operators to produce a new population [25]. The cornerstone of DE is the mutation process, which uses scaled differences of individuals randomly selected from the population to generate new solutions. An important aspect that arises here is how to choose the most suitable mutation strategy, which is the basis for the success of a DE algorithm [26]. The generation of a new trial vector uses a crossover operator that combines both the current and mutant vectors. This process governs the number of components that are mutated and which components are mutated in each individual of the current population [27]. The final step of the DE algorithm is the selection operator. To decide whether or not the trial vector will become a member of the next generation, its fitness value is compared with that of the target vector. The trial vector replaces the target vector if it has a better fitness value; otherwise, the target vector remains, and the trial vector is discarded [28].

4.2 Word Relatedness Measure

In many fields of artificial intelligence and computational linguistics, word relatedness measures are considered an essential component, particularly in areas such as information retrieval and word sense disambiguation. There are many word relatedness techniques. Recently, WordNet has been widely used for word relatedness measurement due to its high efficiency. WordNet is used as a lexical database for the English language developed at Princeton University. In general, WordNet works as a dictionary by covering a variety of terms from different domains and relating them to each other. It contains nearly 100,000 terms arranged in a hierarchical structure. Each word's noun, verb,

adjective, and adverb forms are organized into sets of synonyms. Additionally, all synonym sets are arranged according to senses. Hyponym/Hypernym and Meronym/Holonym relations are also connected to synonym sets [29].

A path-based measure is used to compute relatedness similarity. The basic idea of path-based measures depends on the principle that the similarity between any two concepts is a function of the path length linking these two concepts and the location of the concepts in the taxonomy [30].

5 METHODOLOGY FOR WSD

This section introduces the proposed method for addressing WSD. Here, DE is used to solve the WSD problem by finding a sequence of senses or meanings that maximizes semantic similarity among the target words. This type of problem can be considered an optimization problem, in which the exact meanings of words represent the solution variables, and the relatedness similarity measures between two meanings represent the fitness function used to evaluate the quality of each solution. In the proposed method, a path-based measure is used to assign a score to each word sense, and then DE is employed to maximize the overall semantic similarity of the set of parsed words. The following subsections illustrate the details of each component.

5.1 Shortest-Path for Relatedness Measure

To compute the relatedness between two sets of words a shortest-path based was used, the similarity degree between two sets of words was given as numerical value. The Wordnet is represented as a directed graph $G=\{V,E\}$, where a set of synsets represented as a set of vertices V , and the semantic link can be represented as a set of edge E between vertices. The bidirectional approach used to represent the link between S which is the start set of vertices represent synset of the first word and the goal[target] set of vertices T that represent synset of the second word, where each length is represented by 1. The distance between S and T can be computed as in (1).

$$\text{Dis}(S,T)=\min_{\substack{s \in S, \\ t \in T}} \text{Path_length}(s,t), \quad (1)$$

The method based on computing the shortest path from S to T in breadth first search as in [31].

5.2 Differential Evolution for WSD

The main goal of WSD is to select one sense for each word, one sense that is more suitable to represent the meaning of the word than the others senses in the given text. The DE used to select the sense that has maximum relatedness similarity as in (2).

$$\text{score}(S) = \frac{1}{\text{Dis}(S,T)}. \quad (2)$$

The algorithm based on finding the sense that maximize the score of relatedness similarity. There exist three versions of DE called DE/rand/1, jDE/rand/1, jDE/best/1. The proposed method used DE/rand/1. There are many steps to implement the DE algorithm. The first step concern with individual representation. The individual was a sequence of natural number between 1 to N , where N is the number of senses for word w_i . The algorithm ignores any word that doesn't have a sense. DE begins with a population with N solutions. That can be expressed as follows:

$$P = \{X_1, X_2, \dots, X_N\}.$$

The second step in the DE algorithm concern with computing the objective function, which based on maximizing the score(S) as in (2). The mutation is the third step in DE algorithm. DE algorithm relies heavily on a mutation operation to maintain the population diversity. It consists of adding a scaled difference vector $F \cdot (X_{r2} - X_{r3})$ between two randomly selected individuals X_{r2} and X_{r3} to a third individual X_{r1} , which is also randomly chosen, called the base vector (3).

$$V_i = X_{r1} + F \cdot (X_{r2} - X_{r3}). \quad (3)$$

Where $r1, r2$ and $r3$ are random number in $[1, N]$. F parameter called the scaling factor. This is found for each individual and obtained as a realization of a log-normally distributed random variable.

$$F = \exp(\beta). \quad (4)$$

Where β in $[0, 1]$.

Crossover is the next step of DE algorithm. In this step the trial vector is defined by combining the mutant vector V_i with a DE individual X_i given as (5).

$$U_{i,j} = \begin{cases} V_{i,j} & \text{if } \text{rand}_i, j \leq CP \text{ or } j = j_{\text{rand}} \\ X_{i,j} & \text{otherwise} \end{cases}. \quad (5)$$

CP is chosen randomly in $[0, 1]$.

Selection is the last step in DE algorithm. Depending on the WSD problem, maximizing the relatedness similarity is essential for the construction of the objective function. The objective function as in (2) is used to evaluate each individual. After the evaluation of all individuals, DE selection operation is performed as shown in (6). The resulted vector from this step have higher objective function score and will be passed to the next generated population.

$$u_{i,j}(G+1) = \begin{cases} u_{i,j}(G) & \text{if } Objfun(u_{i,j}(G)) > Objfun(x_{i,j}(G)) \\ x_{i,j}(G) & \text{otherwise} \end{cases} \quad (6)$$

6 EXPERIMENTAL RESULTS

In this section, the proposed method is evaluated using the SemCor benchmark dataset and compared with baseline approaches to assess its effectiveness.

6.1 Benchmark Dataset

The proposed method was tested on the Semantic Concordance (SemCor) dataset, which is a part of the Brown Corpus and represents one of the largest publicly available sense-annotated resources for Word Sense Disambiguation. SemCor is widely used for evaluating WSD systems due to its high-quality manual annotations. Out of the 352 documents included in the dataset, 186 are fully annotated for all parts of speech, including nouns, verbs, adjectives, and adverbs, while the remaining 166 documents are annotated for verbs only. The dataset contains a total of 802,443 words, among which 226,040 words are annotated with word senses [32].

6.2 Evaluation Metrics

The performance of the proposed WSD system is evaluated using three standard metrics: precision, recall, and F1-score [33]. Precision reflects the proportion of correctly disambiguated words among all words processed by the system. Recall measures the proportion of correctly disambiguated words relative to the total number of test words in the dataset. The F1-score provides a balanced evaluation by combining both precision and recall into a single performance indicator.

6.3 Results and Discussion

This section presents the experimental results obtained by the proposed approach and discusses its performance in comparison with other state-of-the-art

WSD methods. The parameter settings used for the Differential Evolution algorithm are summarized in Table 1.

Table 1: DE parameters.

Parameter	Value
Population size, N	100
Maximum number of Iteration	150
Crossover probability, CP	0.8
Mutation Rate, F	0.6

The parameters of the DE algorithm are obtained using a trial-and-error basis. The optimal values are determined based on the outputs obtained from the SemCor dataset for the input passages.

Table 2 clarifies the results of the proposed method compared with three methods mentioned in Section 2.

Table 2: Comparison of the proposed method with other methods.

Technique	Recall	Precision	F-score
ESR[15]	81	83	82
LTRS [19]	75.9	83.	79.6
BiGRU [20]	83	87	85
Proposed method	82.4	89.2	85.6

As shown in Table 2, the proposed method outperforms other methods in terms of precision, while BiGRU achieves the best performance in terms of recall. The trade-off between recall and precision is reflected by the F-score, which considers both metrics, and in this case, the proposed method achieves the best overall result.

The proposed method performs more effectively than the other methods for several reasons, which can be briefly explained. First, DE can avoid local optima, unlike BiGRU, which is based on gradient-based training and may converge to local optima. Second, DE provides a better balance between exploration and exploitation compared to ESR and LTRS, which are primarily rule-based methods. Finally, DE is a population-based search algorithm, meaning that multiple candidate solutions are evaluated independently, which increases the likelihood of finding an optimal or near-optimal solution.

7 CONCLUSIONS

Word Sense Disambiguation (WSD) is the process of identifying the correct sense of an ambiguous word within a given context. In this paper, a knowledge-

based approach was adapted to address this problem. The textual document is introduced as input to the proposed method in order to disambiguate ambiguous words within the text. First, the semantic relatedness between the context of an ambiguous word and the definitions of its possible senses is computed. Second, the Differential Evolution algorithm is applied to select the most appropriate sense among all candidate senses. The proposed method was evaluated on the SemCor dataset. The results demonstrate the effectiveness of the proposed approach compared with other related methods. Experimental findings show that the proposed method outperforms several existing approaches, highlighting its ability to accurately disambiguate word meanings. This improvement indicates that combining semantic similarity measures with an adaptive optimization algorithm such as DE provides a robust and efficient solution for knowledge-based WSD tasks.

REFERENCES

- [1] Y. Wang, M. Wang, and H. Fujita, "Word sense disambiguation: A comprehensive knowledge exploitation framework," *Knowledge-Based Systems*, vol. 129, art. no. 105030, 2019.
- [2] M. Menai, "Word sense disambiguation using evolutionary algorithms: Application to Arabic language," *Computers in Human Behavior*, vol. 41, pp. 92–103, 2014.
- [3] K. Abdalgader and A. Al Shibli, "Context expansion approach for graph-based word sense disambiguation," *Expert Systems with Applications*, vol. 168, art. no. 114313, 2021.
- [4] F. B. Goularte, D. Sorato, S. M. Nassar, R. Fileto, and H. Saggion, "MSC+: Language pattern learning for word sense induction and disambiguation," *Knowledge-Based Systems*, vol. 188, art. no. 105017, 2020.
- [5] D. O., S. Kim, and K. Kim, "Word sense disambiguation based on word similarity calculation using word vector representation from a knowledge-based graph," in *Proc. 27th Int. Conf. Computational Linguistics*, 2018, pp. 2704–2714.
- [6] H. Singh and P. Bhattacharyya, "A survey on word sense disambiguation," *White Paper*, 2019.
- [7] D. Vickrey, L. Biewald, M. Teyssier, and D. Koller, "Word-sense disambiguation for machine translation," in *Proc. HLT/EMNLP*, 2005, pp. 771–778.
- [8] Z. H. Ali, M. R. Ghanim, and R. N. Saleh, "Extractive multi-document text summarization based on ResNet and SALAS algorithm," in *AIP Conf. Proc.*, vol. 3282, no. 1, Apr. 2025.
- [9] H. Soudani, E. Kanoulas, and F. Hasibi, "Fine-tuning vs. retrieval augmented generation for less popular knowledge," in *Proc. 2024 Annu. Int. ACM SIGIR Conf., Asia Pacific*, Dec. 2024, pp. 12–22.
- [10] A. R. Pal and D. Saha, "Word sense disambiguation: A survey," *Int. J. Control Theory Comput. Model.*, vol. 5, no. 3, 2015.
- [11] P. Bala, "Knowledge-based approach for word sense disambiguation using Hindi WordNet," *Int. J. Engineering Science (IJES)*, vol. 2, no. 2, pp. 36–41, 2013.
- [12] L. Nandanwar and K. Mamulkar, "Supervised, semi-supervised and unsupervised WSD approaches: An overview," *Int. J. Scientific Research (IJSR)*, vol. 4, no. 2, 2013.
- [13] S. Singh, H. Kaur, R. Kanozia, and G. Kaur, "Empirical analysis of supervised and unsupervised machine learning algorithms with aspect-based sentiment analysis," *Applied Computer Systems*, vol. 28, no. 1, pp. 125–136, 2023.
- [14] A. Duque, J. Martinez-Romo, and L. Araujo, "Can multilinguality improve biomedical word sense disambiguation?" *Journal of Biomedical Informatics*, vol. 64, pp. 320–332, 2016.
- [15] Y. Song, X. C. Ong, H. T. Ng, and Q. Lin, "Improved word sense disambiguation with enhanced sense representations," in *Findings of the Association for Computational Linguistics: EMNLP 2021*, Nov. 2021, pp. 4311–4320.
- [16] H. Kohli, "Training bi-encoders for word sense disambiguation," in *Proc. Int. Conf. Document Analysis and Recognition*, Cham, Switzerland: Springer, Sept. 2021, pp. 823–837.
- [17] W. Li, H. Wang, X. Wang, and M. Hu, "A word sense disambiguation method based on multiple sense graph," in *Machine Learning and Intelligent Computing*, PMLR, July 2024, pp. 40–47.
- [18] Q. Yang, Y. Li, X. Wang, F. L. Wang, and T. Hao, "PolCLIP: A unified image-text word sense disambiguation model via generating multimodal complementary representations," in *Proc. 62nd Annu. Meeting of the Association for Computational Linguistics*, vol. 1, Aug. 2024, pp. 10676–10690.
- [19] H. Wang, Y. Wang, Q. Liang, and Y. Liu, "LTRS: Improving word sense disambiguation via learning to rank senses," in *Proc. 31st Int. Conf. Computational Linguistics*, Jan. 2025, pp. 1934–1942.
- [20] H. D. Masethe, M. A. Masethe, S. O. Ojo, P. A. Owolawi, and F. Giunchiglia, "Hybrid transformer-based large language models for word sense disambiguation in the low-resource Sesotho sa Leboa language," *Applied Sciences*, vol. 15, no. 7, p. 3608, 2025.
- [21] F. Zou, D. Chen, Q. Xu, and R. Lu, "A new prediction strategy combining T-S fuzzy nonlinear regression prediction and multi-step prediction for dynamic multi-objective optimization," *Swarm and Evolutionary Computation*, vol. 59, art. no. 100749, 2020.
- [22] R. Storn and K. Price, "Differential evolution: A simple and efficient heuristic for global optimization over continuous spaces," *Journal of Global Optimization*, vol. 11, no. 4, pp. 341–359, 1997.
- [23] P. Krömer, J. Platoš, and V. Snášel, "Differential evolution for the optimization of low-discrepancy generalized Halton sequences," *Swarm and Evolutionary Computation*, vol. 54, art. no. 100649, 2020.

- [24] J. Cheng, Z. Pan, H. Liang, Z. Gao, and J. Gao, "Differential evolution algorithm with fitness and diversity ranking-based mutation operator," *Swarm and Evolutionary Computation*, art. no. 100816, 2020, doi: 10.1016/j.swevo.2020.100816.
- [25] B. Kazimipour, X. Li, and A. K. Qin, "Effects of population initialization on differential evolution for large scale optimization," in *Proc. IEEE Congress on Evolutionary Computation (CEC)*, 2014, pp. 2404–2411.
- [26] K. Opara and J. Arabas, "Comparison of mutation strategies in differential evolution: A probabilistic perspective," *Swarm and Evolutionary Computation*, vol. 39, pp. 53–69, 2018, doi: 10.1016/j.swevo.2017.12.007.
- [27] D. Zaharie, "Influence of crossover on the behavior of differential evolution algorithms," *Applied Soft Computing*, vol. 9, no. 3, pp. 1126–1138, 2009, doi: 10.1016/j.asoc.2009.02.012.
- [28] L. Fajfar, J. Puhan, S. Tomažič, and A. Bürmen, "On selection in differential evolution," *Elektrotehniški Vestnik*, vol. 17, no. 5, pp. 275–280, 2011.
- [29] G. Varelas, E. Voutsakis, E. G. M. Petrakis, E. Milios, and P. Raftopoulou, "Semantic similarity methods in WordNet and their application to information retrieval on the Web," in *Proc. 7th ACM Int. Workshop on Web Information and Data Management*, 2005, pp. 10–16.
- [30] R. Rada, H. Mili, E. Bicknell, and M. Blettner, "Development and application of a metric on semantic nets," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 19, no. 1, pp. 17–30, 1989.
- [31] J. Ferlež and M. Gams, "Shortest-path semantic distance measure in WordNet v2.0," *Informatica*, vol. 28, pp. 385–390, 2004.
- [32] G. A. Miller, C. Leacock, R. Teng, and R. T. Bunker, "A semantic concordance," in *Proc. Workshop on Human Language Technology*, Association for Computational Linguistics, 1993, pp. 303–308.
- [33] N. H. Baqer, Z. H. Ali, and A. T. Sadiq, "Sentimental analysis for depression tweets using deep belief network," in *AIP Conf. Proc.*, vol. 3036, no. 1, Mar. 2024.