Ilorin Journal of Science
Volume 10, Number 1, 2023, pp. 62 – 81 (Printed in Nigeria)
ISSN: 2408 – 4840 © 2023 Faculty of Physical Sciences, University of Ilorin
https://doi.org/10.54908/iljs.2023.10.01.005

ILORIN
JOURNAL OF SCEINCE

ILJS-23- 012

Sentiment Lexicon Disambiguation: A Systematic Literature Review

Adeniji[1], A. B., Kolajo*[1], T. and Agbogun[2], J.B.

[1]Department of Computer Science, Federal University Lokoja, Lokoja, Kogi State, Nigeria
[2]Department of Computer Science and Mathematics, Godfrey Okoye University, Ugwuomu-Nike, Enugu State, Nigeria

Abstract

In our society, different channels are used to express one's opinion or feelings about a situation, event, product, or service. The languages used also differ from one person or region to another. Sentiment analysis is carried out to gain a good understanding of the writer's opinion or feelings. Still, ambiguity in languages affects the result when conducting sentiment analysis. The available sentiment lexicon is also negatively impacted due to ambiguity problems. In this research, a systematic review of sentiment lexicon disambiguation was carried out. The databases used are ScienceDirect, Scopus, Semantic Scholar, and SpringerLink. The first string search that was carried out resulted in 89,605 journals, conference papers, and book chapters. An inclusive and exclusive criterion was applied, and 61 papers were finally selected for the research. Findings show that ambiguity occurs in most available languages and negatively affects the reader's or audience's understanding when conducting sentiment analysis. We also look at the techniques used to carry out disambiguation and consider their strengths and weaknesses. Future research can focus on developing more sentiment lexicons free of ambiguity in low-resource languages, which will further help researchers conducting sentiment analysis.

Keywords: Sentiment analysis, Ambiguity, Lexicon, Lexicon disambiguation, Natural language processing

1. Introduction

Sentiment analysis is one area that needs more research in our world. Every minute and second, people share their opinions on the internet, books, video, audio, web, and social media such as WhatsApp, Facebook, Twitter, Telegram, and many more. The amount of opinion data on the internet or in documents is increasing daily, and writers express their likes or dislikes differently. Sentiment analysis helps explore feelings, attitudes, and opinions regarding products, topics, or discussion issues (Balshetwar *et al.*, 2019).

One of the important steps in sentiment analysis is the building of a sentiment lexicon, which contains words that have different meanings (Chang *et al.*, 2021). Generally, we have many lexicons, and these lexicons can contain ambiguous words. Not only that, people in the education sector, business sector, social media, and so on express their opinions and feelings through text, and this text also contains ambiguous words that affect the meaning of what the writer is saying. So, it became so important to disambiguate this text as well as the lexicon that contains ambiguous words for the result of our analysis to be accurate (Dang *et al.*, 2020; Kumar, 2018).

Corresponding Author: Kolajo, T.
Email: taiwo.kolajo@fulokoja.edu.ng

Just as sentiment lexicons can be in different languages, ambiguity also occurs in different languages, and some researchers have worked on the issue of ambiguity in some languages, but this is still an area of concern. Hence, there is a need for more research in this area because the ambiguity that exists in these different languages can harm the efficiency, accuracy, and robustness of the few available lexica and end up affecting the analysis that the researcher or individual is carrying out. Therefore, the process of resolving ambiguous phrases or words that exist in a sentiment lexicon is known as "sentiment lexicon disambiguation." Various natural language processing techniques, such as word sense disambiguation, part of speech tagging, and others, have been used to disambiguate the ambiguous words that exist in the sentiment lexicon (Yin *et al.*, 2020).

The structure of the paper is as follows: Section 1 presents an introduction to what the paper is all about. Section 2 presents the background and related works. In Section 3, the research methodology is discussed. The findings from our research are presented in Section 4. Section 5 presents a discussion based on the results from Section 4. Section 6 briefly presents the limitations of the review work. Finally, the conclusion and further research work are provided in Section 7.

2. Background of Study and Related Works

This section presents some concepts in sentiment analysis and the related works by different researchers. Some of these concepts are discussed subsequently.

2.1 Sentiment Analysis

Sentiment analysis is a subdomain of natural language processing and can detect positive, negative, or neutral sentiments in a sentence or text (Mukherjee *et al.*, 2021). Sentiment analysis is very important when it comes to decision-making. It helps the decision-makers better understand the expression in the text, feedback, and available information. It is also very useful in different fields of life, especially when making calculations or identifying and expressing sentiment (Amrani *et al.*, 2018). Sentiment analysis can be carried out using three methods: machine learning, lexicon-based, and hybrid approaches. Machine learning can be supervised, semi-supervised, or unsupervised (Sadia *et al.*, 2018).

2.2 Lexicon disambiguation

A lexicon may be thought of as a word dictionary that comprises many different words in a particular language or as a collection of people's vocabularies. These lexicons may include words that repeat themselves and take on many meanings depending on the context in which they are used. Various words in various languages, despite having the same spelling, have multiple meanings, depending on how they are used. A statement that, depending on how it is used, may have two or more interpretations is considered ambiguous in language. Some of the types of ambiguity are lexical, pragmatic, anaphoric, semantic, structural, and syntactic. Lexical ambiguity deals with a single word's ambiguity. On the other hand, syntactic ambiguity is a sort of ambiguity that happens when a certain sentence can be understood in several distinct ways. When a word has multiple meanings depending on its context, it is said to be pragmatically ambiguous. Anaphoric ambiguity occurs in sentences when words like "it," "I," and "so on" are repeated at the beginning of a sentence. (Priya *et al.*, 2021; Kumar, 2018). Contrarily, structural ambiguity occurs when a text is unclear due to the way it is organized or written (Adenike *et al.*, 2018). Disambiguation is the process of getting rid of ambiguity, and when we talk about lexicon disambiguation, we mean getting rid of ambiguity in a lexicon.

2.3 Related Work

Some of the literature review and systematic literature review papers related to sentiment lexicon disambiguation over the years are discussed in this section.

Alian *et al.* (2017) researched the disambiguation of Arabic word senses. Focusing on the English and Arabic languages, a review of word disambiguation was carried out. In the review, the knowledge-based, unsupervised, and hybrid approaches for Arabic word sense disambiguation were discussed, and it was found that the majority of researchers in Arabic used different kinds of datasets, e.g., dictionaries, hand-labeled data, etc., of limited size to do their research and It was also found that using Arabic WordNet had some issues when compared to English WordNet, some of which included noise, precision, and limited coverage.

Singh and Saraswat (2019) reviewed different techniques used for the disambiguation of polysemy words in languages like English, Nepalese, Telugu, Hindi, Sinhala, Tamil, German, Kannada, and Malayalam. Various approaches used over the years were also reviewed, and approaches like supervised and unsupervised were discussed, as well as how they are used with the different languages. Some of their findings are that Hindi and English word sense disambiguation started with the advent of the LESK algorithm, and the techniques used evolved from information content-based and rule-based algorithms to graph-based and machine learning-based techniques. For Malayalam word sense disambiguation tasks, they later evolved using Naive Bayes classifiers.

Singh (2021) conducted a review on word sense disambiguation using WordNet, IndoWordNet, and other corpora as the primary data resources for different languages. According to the review, the majority of word sense disambiguation systems for foreign languages use supervised approaches that only support nouns, verbs, and adjectives. Another point is that word-sense disambiguation is implemented using knowledge-based and supervised algorithms in the majority of Indian languages. According to the review, the supervised algorithm outperforms all other techniques when carrying out sentiment disambiguation. It was also shown that applying the same technique used to disambiguate to different categories of languages causes variations in performance in terms of accuracy.

In a study of the most recent developments in word sense disambiguation, (Bevilacqua *et al.*, 2021) reviewed the various word meaning disambiguation approaches. The two different types of data utilized in word sense disambiguation, such as sense inventories and annotated corpora, were explored along with the knowledge-based and supervised approaches. It was also discovered that, despite being computationally convenient, when using a discrete sense inventory, it prevents scaling to newer and more inventive word uses, which served as a limitation.

Yadav *et al.* (2021) reviewed how ambiguities are resolved in natural language processing. In the review, different disambiguation tools and techniques, such as style guides, controlled natural language (CNL), knowledge-based approaches, and transfer learning techniques, were compared, evaluated, and analyzed. The findings show that some disambiguation tools fail to eliminate ambiguities. And another thing that the review shows is that some tools are still undergoing further development, and they might be capable of eliminating ambiguities in the future.

Mente *et al.* (2022) reviewed word sense disambiguation and IT approaches to showing ambiguities in natural languages. The review showed that WSD is applied in natural language processing such as dialogue, speech synthesis, machine translation, question answering, and information retrieval. The review examined different strategies for carrying out WSD, such as dictionary and knowledge-based, supervised, semi-supervised, and unsupervised approaches. The findings showed that the supervised approach outperformed all or any other approaches.

Kaddoura and Ahmed (2022) also reviewed Arabic word sense disambiguation when it comes to natural language processing applications. It was reviewed that Arabic lacks standards and diacritics, in addition to a severe lack of resources, which results in ambiguity in some of the available Arabic lexicon. A detailed review of studies that addressed word ambiguity using various strategies and methods was covered. Some of the word sense disambiguation methods used in word sense disambiguation, like supervised, unsupervised, semi-supervised, hybrid, knowledge-based, and metaheuristic, have been exploited over the years by researchers to carry out word sense disambiguation and have produced good results over the years.

The related works mentioned above concern ambiguity in different languages and disambiguation. Our review work will look at different sentiment lexicon disambiguation. It has been observed that one of the key factors that affect sentiment analysis and understanding of a word or sentence is the ambiguity that exists in the language used to communicate. We will further review the effect of this ambiguity, and we will also present the techniques used to carry out lexicon disambiguation.

3. Research Methodology

A systematic literature review was carried out to examine different aspects of sentiment lexicon disambiguation. This includes the effects of ambiguity on sentiment analysis, the different sentiment lexicons available, the approaches used for lexicon disambiguation, and the limitations of the existing methods.

3.1 Research Question

The following research questions are answered:
Research Question 1: What are the effects of ambiguity in carrying out sentiment analysis?
Research Question 2: What are the different sentiment lexicons available?
Research Question 3: What approaches or techniques are used for lexicon disambiguation?

3.2 Search String

The search string process that is good enough for research work involves the combination of words or statements that are carried out in a search engine or database to discover related work. In this paper, we considered the following search strings: “sentiment analysis” OR “building a sentiment lexicon” OR “resolving ambiguities in sentiment analysis” OR “disambiguation in sentiment lexicon” OR” approaches to lexicon disambiguation” OR “ambiguities in low-resource language” OR “ambiguities in natural processing language." These search strings were inputted into the ScienceDirect, Semantic Scholar, SpringerLink, and Scopus databases.

3.3 Data Sources

Getting research information from a rich database becomes so important because research work in different disciplines is increasing daily, and there is a need to produce quality work to meet expectations or stand out. The research databases used were:

ScienceDirect: This is a good database that provides access to open-access or subscription-based papers where students, researchers, teachers, lecturers, and others who need information can easily access it. About 1.4 million open-access articles, more than 600 open-access publications, 363K topic pages, 43K eBooks, 2,650 peer-reviewed journals, and 19M articles and chapters are available in the database. Articles in fields like physical science and engineering, health sciences, life sciences, social sciences, and humanities are available.

Scopus: This is also an Elsevier expertly abstract and citation database. It has enriched data and was designed to meet the needs of educators, students, researchers, librarians, and administrators across the academic community. It covers about 87 million+ documents, 1.8 billion+ cited references, 17 million+ author profiles, 335K+ books, 7K+ publishers, and 94K+ affiliation profiles. Journals in different fields, such as social sciences, life sciences, health sciences, physical sciences, etc., are available on the database.

Semantic Scholar: This is a search engine that provides open resources. They are AI-driven, free, and content discovery tools that help researchers properly search for the paper they need. It has over 207 million academic papers from all fields of science that are useful for researchers to use in their work. They collaborate with other publishers such as IEEE, ACM, Scientific.Net, Springer Nature, Frontiers, and many more.

SpringerLink: This database provides a platform for researchers to access and discover scientific documents in proceedings, journals, series, books, reference works, and protocols. It covers various topics from disciplines such as biomedicine, business and management, chemistry, computer science, earth sciences, economics, education, engineering, environment, geography, history, law, life sciences, literature, materials science, mathematics, medicine and public health, pharmacy, philosophy, physics, political science and international relations, psychology, social sciences, and statistics. A researcher can browse about 15,022,349 resources that contain 8,318,054 articles, 5,050,620 chapters, 1,378,050 conference papers, 701,224 reference work entries, 70,482 protocols, and 365 videos.

3.4 Data Retrieval

The fact that the most impactful articles can be found in ScienceDirect, Scopus, Semantic Scholar, and SpringerLink informed our choices. Seven search strings were used with the conjunction of an ‘OR’ boolean between the years 2017 and 2023. The first search string from ScienceDirect, SpringerLink, Semantic Scholar, and Scopus gave an output of 6,766, 21,671, 58,601, and 2,567, respectively. We had a total of 89,605 articles in the first search.

Table 1: Result of the First Search String

Paper Source	ScienceDirect	Semantic Scholar	SpringerLink	Scopus	Total
Number of Papers	**6,766**	58,601	21,671	2,567	89,605

Additional refinement was carried out on the papers; we limited the subject area to computer science and engineering. The content type was limited to an article, conference paper, conference proceedings, and book. The journal and conference papers were limited by IEEE Access, ACL, SEMEVAL, and SSRN. In Scopus, we further filtered it by limiting it to the 1st and 2nd quartiles. The total number of 24,449 papers remained, excluding 65,156.

Table 2: Result of the Second Search String

Paper Source	ScienceDirect	Semantic Scholar	SpringerLink	Scopus	Total
Number of Papers	3,717	2,932	16,592	1,208	24,449

We decided to review the titles to check the relevance of the papers. We excluded 22,523 irrelevant items at this stage, leaving 1,926.

Table 3: Result of the Third Search String

Paper Source	ScienceDirect	Semantic Scholar	SpringerLink	Scopus	Total
Number of Papers	575	347	654	352	1,926

The abstract, introduction, and conclusion of the papers were read. A quick overview was carried out for some papers to ascertain whether the paper was relevant or irrelevant. In the end, we had a total of 61 relevant papers.

Table 4: Final Selection

Paper Source	ScienceDirect	Semantic Scholar	SpringerLink	Scopus	Total
Number of Papers	13	27	12	9	61

3.4.1 Inclusion Criteria

Papers that were published in conferences, workshops, and journals between the years 2017 and 2023 were included in the search.

3.4.2 Exclusion Criteria

The excluded papers were those that were not written in English, those with titles that were irrelevant to the research domain, and papers whose introduction and conclusion did not support our research topic.

4. Results

The research questions served as a guide for carrying out the systematic review. The findings from the 61 selected final papers are now presented as answers to the research questions.

Research Question 1: What are the effects of ambiguity in carrying out sentiment analysis?

Ambiguity is a statement that can be understood in two or more meanings depending on how it is used. Some types of ambiguity are lexical, pragmatic, anaphoric, semantic, structural, and syntactic. Lexical ambiguity deals with the ambiguity of a single word, which can be interpreted in many ways. Syntactic ambiguity, on the other hand, occurs when a particular sentence can be interpreted in many different ways. A word is said to be semantically ambiguous if that word can be misinterpreted. Anaphoric ambiguity occurs when a phrase entity is repeated in a sentence, for instance, the repetition of ‘I’, ‘it’, or any other phase in a sentence. Pragmatic ambiguity occurs when a phrase's context gives it multiple meanings (Priya *et al.,* 2021). Structural ambiguity, on the other hand, is when a text becomes ambiguous as a result of the arrangement or syntax of the text (Adenike *et al.,* 2018).

Orimaye *et al.* (2012) pointed out that sentiment analysis still has some constraints due to ambiguities in the language, thereby affecting the performance of the sentiment or subjectivity analysis algorithms. The semantic ambiguity of a single word in different languages harms the robustness of the sentiment analysis. Ambiguity occurs in different areas. For example, according to Abdel-Hafiz (2017), some newspaper headlines are ambiguous, affecting readers' understanding. Data used for sentiment analysis can also be collected from different sources, such as blogs, online reviews, Facebook, Twitter, YouTube, email, and news forums. Some of the data collected from these sources contains ambiguity (Kolajo *et al.,* 2020), which negatively affects the results of sentiment analysis algorithms. Also, we need to know that different things cause ambiguity; for example, in Arabic, the lack of diacritics in most digital documents causes word ambiguity because the same word can appear in different places with different senses (Elayeb, 2019).

Ambiguity in language negatively affects the audience's understanding and the result acquired when conducting sentiment analysis. Hence, it is very important to handle the issue of ambiguities. Ambiguity occurs in most of the available languages and harms the reader's or audience's understanding. Unhandled ambiguities result in miscommunication or communication failure.

Research Question 2: What are the different sentiment lexicons available?

A sentiment lexicon, called an opinion lexicon, is a lexical resource used for sentiment analysis. A sentiment lexicon is a database or collection of lexical units (phrases, words, word senses, sentences, etc.) for a particular language, along with the lexical unit sentiment orientations. Another way to see the sentiment lexicon is as a collection of positive or negative words, which is called polar. A sentiment lexicon can be constructed manually, although it is time-consuming and requires a lot of effort. This construction has four main approaches: crowdsourcing and gamification, bootstrapping, machine learning or probabilistic learning approaches, and human sentiment coding (Khoo & Johnkhan, 2018). The quality of your lexicon will also go a long way in determining the quality of your sentiment analysis (Chang *et al.,* 2021). Some of the available lexicons are briefly discussed subsequently.

WordNet: WordNet is a lexical database that was first created in English and therefore contains semantic relations between different words in more than two hundred languages. It is the most commonly used lexicon resource, and many other lexicons are generated from this lexicon (Sadia et al., 2018). For instance, the African WordNet project, which aims to develop wordnets for indigenous languages such as isiZulu, Setswana, isiXhosa, Sesotho, Xitsonga, Tshivena, Siswati, and isiNdebele (Griesel & Bosch, 2020), and Indo-Aryan WordNet for the Gujarati language were derived from WordNet (Vaishnav & Sajja, 2019). WordNets are used in different natural language processing applications (Kanojia *et al.,* 2022).
SentiWordNet: SentiWordNet is hosted inside WordNet and available to the public. According to (Mukherjee *et al.,* 2021), this lexicon provides positive, negative, and objective sentiment scores for each synset of WordNet, that is, the words in the WordNet. According to Feng *et al.* (2018), SentiWordNet is the most widely used lexicon resource.

BabelNet: The BabelNet lexicon was created by merging Wikipedia, WordNet, and some other multilingual lexical resources. This lexicon is also available to the public, and it can be edited by any volunteers, which helps the lexicon provide a wider encyclopedia of knowledge (Scarlini *et al.,* 2020; Ayetiran & Agbele, 2018).

SemCor: The SemCor lexicon is another example of a vocabulary resource containing 362 texts with over 200,000 words manually tagged with senses from WordNet and part of speech (Ayetiran & Agbele, 2018; Mente *et al.,* 2022).

Valence-Aware Dictionary and Sentiment Reasoner (VADER): The VADER English lexicon contains English words and their sentiments. VADER can be translated into other languages. Oyewusi *et al.* (2020) used VADER to calculate compound sentiment scores for about 14,000 Nigerian Pidgin tweets in their research work. Abiola *et al.* (2023) used VADER to calculate text sentiment by collecting the lexical features (words) categorized as positive or negative based on their semantic orientation in the text. Bhavsar *et al.* (2022) used the VADER library to evaluate sentimental data from new headlines that were scraped from Google News. Different researchers have also constructed sentiment lexicons for low-level language and used them during their research. For instance, the DUTIR sentiment lexicon contains 27446 common Chinese words with sentiment polarity labeled either positive, negative, or neutral (Feng *et al.,* 2018). The Urdu sentiment lexicon was created by translating existing sentiment lexicons in the English language into the Urdu language. Bilingual dictionaries were used to translate the English words. Khattak *et al.* (2021) also acquired an Urdu corpus of 89,000 tweets in Pakistan on the political situation, which they used to build the lexicon.

A few sentiment lexica exist in the Persian language, some of which are the National Research Council (NRC) word-emotion association lexicon, Senti_Str, CNRC (customized NRC), etc. The NRC word-emotion association lexicon was constructed by directly translating the English lexicon into Persian and comprised 9450 words. The Senti_Str lexicon was also derived by translating English into the Persian language. The CNRC was constructed from the NRC by removing words that do not convey sentiment in Persian, words that correspond to long phrases in Persian, and finally, filtering words that were not correct. The CNRC has 2,698 term words (Basiri & Kabiri, 2017).

IgboSentilex: IgboSentilex was developed by Ogbuju and Onyesolu (2019). It contains 7,000 words, of which 2,100 were positive and 4,900 were negative. The words in the Igbo language were gotten by translating Liu's lexicon into the Igbo language. The Igbo language Bible was also used to get more positive or negative words in the research.

HindiSentiWordnet: HindiSentiWordnet was developed by translating words in SentiWordNet to the Hindi language, and these translated words were arranged based on their parts of speech (Kulkarni & Rodd, 2021).

Arabic Root-Based Lemmatiser (ARBL): The Arabic Root-Based Lemmatiser lexicon consists of 69 patterns, 3,829 roots, and a set of 346 Arabic words. It consists of 28,760 words with sentiment scores of positivity, negativity, or objectivity. The root words are categorized into 16 groups, and some of the categories are nouns, prepositions, verbs, numerals, and conjunctions. The lexicon was developed based on WordNet, SentiWordNet, and an Arabic morphological analyzer (Obiedat *et al.,* 2021).

Senti-Foclóir: Senti-Foclóir Lexicon is regarded by Afli et al. (2017) as the first-ever Irish language sentiment lexicon. The AFINN-111 lexicon was manually translated from the English language to the Irish language to form the Senti-Foclóir Lexicon. It contained 5,571 words when the authors published their research work.

BengSentiLex: BengSentiLex is one of the sentiment lexicons available in the Bengali language. The corpus-based approach was used to develop this lexicon, and it also contains opinion words that people use in informal communication (Sazzed, 2021).

Marathi WorldNet: Marathi WorldNet was constructed using paradigmatic relations such as entailment, hyponymy, synonymy, antonymy, etc. Synsets, or synonym sets, are for verbs, adverbs, adjectives, and nouns. Marathi WorldNet is mostly used for natural processing languages in the Marathi language (Patil *et al.,* 2020).

SenZi: The SenZi lexicon used existing resources to generate a list of Arabizi sentiment words by translating the English words into Arabizi words. The Facebook corpus was used to add more words to the lexicon (Tobaili *et al.*, 2019).

SauDiSenti: Al-Thubaity *et al.* (2018) worked on the SauDiSenti lexicon. This lexicon comprises 4,431 words and phrases manually extracted from the Twitter trending hashtags in Saudi Arabia. 24% of the words and phrases were made of positive words, while 76% comprised negative words and phrases.

According to the research, WordNet was the first lexicon created in the English language and is also the one that is most frequently used for sentiment analysis and disambiguation. Many lexicons in English and other languages have been derived from WordNet, and researchers still require more lexicons in local resource languages.

Research Question 3: What approaches or techniques are used for word disambiguation?

Disambiguating the context of a word is very important in the sense that it helps to improve the accuracy of sentiment analysis. After all, one word can mean different things depending on the usage domain. One method that has been used to handle ambiguity issues is the addition of the word-sense disambiguation process. The word sense disambiguation process is used to disambiguate text so the computer can deduce the correct sense of the given words (Abid *et al.,* 2017). Word sense disambiguation replaces the ambiguous word based on the sentence’s context. Word Sense Disambiguation is applied in many areas, such as sentiment analysis, machine translation, information retrieval (IR), and knowledge graph construction (Wang, *et al.,* 2020). There are two different approaches to word sense disambiguation: knowledge-based approaches and machine learning-based models. The dictionary and knowledge-based approaches include the LESK algorithm, selectional preferences, semantic similarity, the heuristic method, Walker’s algorithm, overlay of sense definitions, structural approaches, genetic algorithms, and more. The machine learning models can be supervised learning (Naïve Bayes Method, Decision Tree Method, Logistic Regression, K-Nearest Neighbors, Ensemble Methods, Support Vector Machine (SVM) Method, Neural Network Method, Exemplar-Based or Instance-Based Learning, AdaBoost, etc.), semi-supervised methods (Bootstrapping, Monosemous Relatives, etc.), and unsupervised methods (K-means, Co-event Graphing / Hyper-lex Algorithm, Word Clustering, Context Clustering, Co-occurrence Graphs, etc.) (Kolajo *et al.*, 2020; Kokane & Babar, 2019; Mente *et al.*, 2022; Patil *et al.*, 2020). Table 5 lists some of the techniques, their underlying concepts, strengths, and weaknesses that were employed to carry out disambiguation.

Table 5: Technique for lexicon disambiguation

TECHNIQUE	IDEA	STRENGTH	WEAKNESS	REFERENCES
LESK Algorithm	The sense of the target word that most closely matches the context is chosen by comparing it to the senses of the other words in the sentence.	The ability to consider the target word's context is one of its strengths and is easy to implement.	For terms that are poorly represented in dictionaries and other knowledge sources, it might not be as useful.	Kolajo et al. (2019), Tripathi *et al.* (2020), Mente *et al.* (2022)
Selectional Preferences	The idea is to leverage source knowledge to identify meanings and learn more about potential connections between word kinds.	Word-to-word link	The restricted number of participant roles	Mente *et al.* (2022), Choi *et al.* (2017)
Semantic Similarity	Examine how the text's words and sentences relate to one another semantically.	It can accurately and nuancedly disambiguate words by capturing their context-specific, rich, and complicated meanings.	The semantic similarity might not be able to differentiate between several meanings of a polysemous word, determine the proper interpretation of an idiomatic expression, or distinguish between various named things that have the same name, all of which could have an impact on disambiguation.	Zhu and Iglesias (2018), AlMousa *et al.* (2022)

Heuristic Method	The heuristic approach is based on the presumption that words that frequently appear in similar circumstances have related meanings.	This method can be applied when the rules-based strategy is insufficient.	Since heuristics are not based on precise rules but rather on experience, there is a chance of producing an incorrect result.	Purohit and Yogi (2022), Bhattacharjee *et al.* (2020), and AlMousa *et al.* (2022)
Walker's Algorithm	The idea is to decide the result for each sense by choosing the synonyms that match the current definition.	Because it is built on synonyms, it provides great resolution.	It is difficult to find synonyms that can handle the word's ambiguity.	Gidhe and Ragha (2018), Purohit, and Yogi (2022)
Glossy Overlap	identify ambiguous words' semantic duplicates and, as a result, identify words that appear in that context.	The particular context and the words being taken into consideration determine the strength of glossy overlap.	The technique employed uses glosses of a word's several senses, which might not be adequate to capture the entire range of meanings and complexity that a word has.	Mente *et al.* (2022)
Genetic Algorithm	The idea behind this algorithm is to identify the word's most relevant sense in light of its context as efficiently as possible.	It can swiftly and effectively search a large search space.	They may become trapped in local optimal states, which would prevent them from discovering the overall optimum.	Vaishnav and Sajja (2019), Bhatia *et al.* (2022)
Naïve Bayes	It is a probabilistic method in which the likelihood of correct outcomes is	The capacity to effectively manage vast volumes of data. Large datasets may be processed	The presumption of feature independence may not always be accurate, and overfitting can	Pal *et al.* (2021), Maitra *et al.* (2018)

	higher for an accurate class.	with the Nave Bayes method.	happen when the algorithm is trained on a small dataset.	
Decision Tree	Based on a collection of labeled training data, it constructs a tree-like model of decisions and their potential outcomes.	Has the ability to manage a variety of features and difficult decision-making processes.	Tend to capture simple interactions between variables but not complex ones, and the data tends to be over- or under-fitted.	Abid *et al.* (2017), Pal *et al.* (2021), and Bhattacharjee *et al.* (2020)
K-Nearest Neighbour	The fundamental principle of KNN is to categorize a new instance using the K nearest neighbors' class labels in the training set.	It can handle high-dimensional data and perform well with both linear and non-linear data.	It needs enough labeled data to function properly.	Daeli and Adiwijaya (2020)
Ensemble Methods	The idea is to integrate many models to create a single, more precise prediction.	They can deal with situations where certain models may have biases or flaws.	The ensemble may not be able to considerably increase performance if the basic models are very similar or have a strong correlation.	Patel *et al.* (2021), Canale *et al.* (2018), Krawczyk and McInnes (2018)
Support Vector Machine	It creates a hyperplane to divide the training samples into positive and negative ones.	It is possible to apply it to regression or classification; in fact, it plays a crucial role in classification issues.	Noise in the data will limit its effectiveness.	Gunawan *et al.* (2023), Pal *et al.* (2021)
Random forest	builds several decision trees as part of its ensemble learning	It can manage feature spaces of many dimensions	When the number of trees increases, the algorithm	Amrani *et al.* (2018), Fauzi (2018)

	approach to classification and regression.	and complexity.	works sluggishly.	
Neural network	According to the intended response, employ input features to divide training instances into separate sections.	It can learn complex patterns, is adaptable, and has non-linear mapping capabilities.	gradient exploding or vanishing problems	Mahadevaswamy and Swathi (2023), Ullah *et al.* (2022), Kumar (2020)
Exemplar-based or instance-based learning	To generate predictions, the algorithms are trained on a set of particular samples.	The method can handle large and varied data sets and provide more precise predictions.	More training examples are required, and it is computationally expensive.	Jones (2019), Kumar, (2018)
AdaBoost	The idea is to create a strong classifier; therefore, numerous weak classifiers are combined.	It can efficiently integrate multiple features to generate more accurate predictions, and it can handle noisy and confusing data.	Even though it can handle noise, if the noise in the train data is excessive, it may end up being overfitted to the noise.	Chengsheng *et al.* (2017)
Bootstrapping	The idea is to automatically identify and label a larger collection of unlabeled data using a small set of labeled data.	It can learn from incomplete data and manage ambiguity and uncertainty in language.	The quality and representation of the first labeled data set have a major effect on determining the bootstrapping technique's accuracy.	Mente *et al.* (2022), Li (2022), Almousa *et al.* (2022)
Monosemous Relatives	The idea is to employ a group of unambiguous terms or	Its capacity to use the context in which a word occurs to clear up any	The approach may not be successful in determining the meaning of the	Bolshina and Loukachevitch (2020), Mente *et al.* (2022)

	phrases to help explain the meaning of ambiguous words or phrases.	ambiguity about its meaning.	statement if the context is vague or unclear.	
K-means	The method determines which cluster center is closest to each data point before recalculating each cluster's center using the newly allocated points.	In terms of computation speed, ease of use, and effectiveness	One problem is its sensitivity to the original cluster centers and its inclination to converge to the local optimum instead of the global optimum.	Zul *et al.* (2018), Bhattacharjee *et al.* (2020) s
Co-event Graphing	It included examining a collection of texts to find related and concurrent occurrences.	It can identify both the implicit and explicit connections between events in a group of texts.	The approach depends on the availability of a significant corpus of text data, which may not always be the case.	Kokane and Babar (2019)
Word Clustering	Words are grouped according to how semantically related they are.	It enables the recognition of word senses that conventional dictionary definitions could miss.	Words with numerous meanings or those used in a variety of settings might not work well with them.	Bhatia *et al.* (2022), Maurya, and Bahadurv (2022), Kumar, (2018)
Context Clustering	The concept is to classify comparable words or phrases according to the context in which they are used.	It's capacity to determine the most likely meaning of a word in a specific context using contextual information.	Larger contexts might not be taken into consideration by context clustering.	Kumar, (2018), Bhatia *et al.* (2022), Maurya, and Bahadurv (2022)
Co-occurrence Graphs	Co-occurrence graphs can be used to infer a word's most likely	It captures the associations between words that are contextual and can help	The nuanced nature of language usage might not be captured by co-	Duque *et al.* (2018), Kumar (2018)

	meaning depending on the context in which it appears.	disambiguate their meaning.	occurrence graphs.	
Bag of Concepts	Identifies and modifies ambiguous concepts by generating a source of ambiguous sentiment concepts based on SenticNet.	Commonsense knowledge can increase accuracy using ConceptNet handling to overcome lost knowledge.	Do not properly handle domain-specific knowledge.	Rajabi et al. (2020)

We can see from Table 5 that while several strategies have been used to disambiguate, they each have their limitations. To get the most out of the disambiguation algorithms or techniques, we may need to combine them.

5. Discussion

This review shows that ambiguity in language has a great negative impact on effective sentiment analysis. This problem affects the audience's understanding when listening to or reading a write-up or text. One of the factors that causes ambiguity in the language, especially in some low-resource languages, is the lack of diacritics in the text document in the English language and other languages. So there is still a need for more research to resolve the ambiguity problem in low-resource languages in a situation where diacritics are absent.

One factor that affects sentiment analysis is the quality of the lexicon used. Improper building of the lexicon has affected sentiment analysis. WordNet is one of the proper lexicons used in sentiment analysis, and many other quality lexicons have been built over the years. Some of these lexicons are SentiWordNet, BabelNet, HindiSentiWordnet, and many others, but we still lack inadequate lexicons in many low-resource languages. Out of over 7,000 languages spoken in different parts of the world, there are very few available lexicons in these low-resource languages, so more research efforts are encouraged to generate lexicons in low-resource languages.

A quality lexicon is very important for many reasons. The word sense disambiguation (WSD) process has been one of the ways many researchers have used to resolve the issue of ambiguity in the sentiment lexicon. WSD is applied in sentiment analysis, machine translation, information retrieval (IR), and knowledge graph construction. Many other low-resource languages have also adopted WSD to resolve ambiguity in their language and the sentiment lexicon built into that language. For example, Hindi, Marathi, and many others have effectively used WSD to resolve ambiguity.

Many researchers have used different techniques to carry out disambiguation, and even though these techniques are effective, they still have limitations that in some way compromise the precision of the results. It is important to note that many other approaches can be used to carry out disambiguation apart from the one listed in this review, and we also need to manage these limitations well to achieve efficiency and accuracy. It is recommended that more research focus on these areas. This review has shown us different techniques used for disambiguation and some of the lexicons used in carrying out sentiment lexicon. Also, we have come to know that the majority of the lexicon that we have was developed for WordNet.

6. Limitation of the Research Work

The papers used for this systematic review were selected based on inclusive and exclusive criteria. Some papers might have been useful, but they were not written in English, making it difficult for us to review them. We used ScienceDirect, Scopus, Semantic Scholar, and SpringerLink in searching for papers, and there are possibilities that some other useful papers are available but not in these databases.

7. Conclusion and Further Works

The fact that sentiment analysis is a growing area of research in the natural processing language field makes it a field that we encourage more researchers to explore. One key area that needs more attention is solving ambiguity problems. More attention is needed to disambiguation because one of the major problems affecting sentiment analysis is ambiguity in most of our languages. Disambiguation in the English language has witnessed much research that is sometimes worth commending. Still, it is also necessary to conduct more research to resolve disambiguation in low-resource languages. The methods or techniques currently used to handle ambiguity must also improve to generate more accurate results.

This systematic review has examined some available sentiment lexicons, the effect of ambiguity in languages, techniques for lexicon disambiguation, and the strengths and weaknesses of the existing techniques used in carrying out disambiguation. In the paper, we answered research questions that focused on the effects of ambiguity in carrying out sentiment analysis, some of the available sentiment lexicons, and the approaches used in carrying out lexicon disambiguation. Future research can focus on developing more sentiment lexicons in low-resource languages and solving the ambiguity problems in these low-resource languages.

References

Abdel-Hafiz, A. S. (2017): Ambiguity in Egyptian Newspaper Headlines. *African Journalism Studies*, **38**(3–4), 1–25.

Abid, M., Habib, A., Ashraf, J., and Shahid, A. (2017): Urdu word sense disambiguation using machine learning approach. *Cluster Computing*, **21**(1), 515–522.

Abiola, O., Abayomi-Alli, A., Tale, O. A., Misra, S., and Abayomi-Alli, O. (2023): Sentiment analysis of COVID-19 tweets from selected hashtags in Nigeria using VADER and Text Blob analyser. *Journal of Electrical Systems and Information Technology*, **10**(1), 1-20.

Adenike, A., Odetunji, A., and Akeem, S.s (2018): Lexical ambiguity resolution system for standard yorùbá verbs. *American Journal of Engineering Research (AJER)*, **7**(6), 170-176.

Afli, H., Maguire, S., and Way, A. (2017): Sentiment translation for low-resourced languages: Experiments on Irish general election tweets. *In: 18th International Conference on Computational Linguistics and Intelligent Text Processing, Budapest Hungry*, **P**p. 17-21.

Alian, M., Awajan, A., and Al-Kouz, A. (2017): Arabic word sense disambiguation-survey. In *2017 international conference on new trends in computing sciences (ICTCS)*, **P**p. 236-240.

AlMousa, M., Benlamri, R., and Khoury, R. (2022): A novel word sense disambiguation approach using WordNet knowledge graph. *Computer Speech & Language,* **74**, 101337.

Al-Thubaity, A., Alqahtani, Q., and Aljandal, A. (2018): Sentiment lexicon for sentiment analysis of Saudi dialect tweets. *Procedia computer science,* **142**, 301-307.

Amrani, A., Lazaar, M., and Kadiri, K. (2018): Random forest and support vector machine based hybrid approach to sentiment analysis. *The First International Conference on Intelligent Computing in Data Sciences*, **127**, 511–520.

Ayetiran, F., and Agbele, K. (2018): An optimised Lesk-based algorithm for word sense disambiguation. *Open Computer Science*, **8**, 165-172.

Balshetwar, S., Tuganayat, R., and Regulwar, G. (2019): Frame tone and sentiment analysis. *International Journal of Computer Sciences and Engineering*, **7**(6), 24-40.

Basiri, M., and Kabiri, A., (2017): Sentence-level sentiment analysis in Persian. *In: 2017 3rd International Conference on Pattern Recognition and Image Analysis (IPRIA) IEEE*, **P**p. 84-89.

Bevilacqua, M., Pasini, T., Raganato, A., and Navigli, R. (2021): Recent trends in word sense disambiguation: A survey. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*. International Joint Conference on Artificial Intelligence, Inc.

Bhatia, S., Kumar, A., and Khan, M. M. (2022): Role of Genetic Algorithm in Optimization of Hindi Word Sense Disambiguation. *IEEE Access,* **10**, 75693-75707.

Bhattacharjee, K., ShivaKarthik, S., Mehta, S., Kumar, A., Phatangare, S., Pawar, K., ... and Verma, D. (2020): Survey and gap analysis of word sense disambiguation approaches on unstructured texts. *In 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC),* **P**p. 323-327.

Bhavsar, H., Jivani, A., Amesara, S., Shah, S., Gindani, P., and Patel, S. (2022): Stock Price Prediction Using Sentiment Analysis on News Headlines. In *ICT with Intelligent Applications: Proceedings of ICTIS 2022, Volume 1* (**P**p. 25-34). Singapore: Springer Nature Singapore.

Bolshina, A., and Loukachevitch, N. (2020): Monosemous relatives approach to automatic data labelling for word sense disambiguation in russian1. In *linguistic forum 2020: language and artificial intelligence,* **P**. 12.

Canale, L., Lisena, P., and Troncy, R. (2018): A novel ensemble method for named entity recognition and disambiguation based on neural network. In *The Semantic Web–ISWC 2018: 17th International Semantic Web Conference, Monterey, CA, USA, October 8–12, 2018, Proceedings, Part I 17* (**P**p. 91-107). Springer International Publishing.

Chang, C.H., Hwang, S.Y., and Wu, M.L. (2021): Learning bilingual sentiment lexicon for online reviews. *Electronic Commerce Research and Applications,* **47**, 101037.

Chengsheng, T., Huacheng, L., and Bing, X. (2017): AdaBoost typical Algorithm and its application research. In *MATEC Web of Conferences* (Vol. **139**, p. 00222). EDP Sciences.

Choi, Y., Wiebe, J., and Mihalcea, R. (2017): Coarse-grained+/-effect word sense disambiguation for implicit sentiment analysis. *IEEE Transactions on Affective Computing*, **8**(4), 471-479.

Daeli, N.O.F., and Adiwijaya, A. (2020): Sentiment analysis on movie reviews using Information gain and K-nearest neighbor. *Journal of Data Science and Its Applications*, **3**(1), 1-7.

Dang, N. C., Moreno-García, M. N., and De la Prieta, F. (2020): Sentiment analysis based on deep learning: A comparative study. *Electronics*, **9**(3), 483.

Duque, A., Stevenson, M., Martinez-Romo, J., and Araujo, L. (2018): Co-occurrence graphs for word sense disambiguation in the biomedical domain. *Artificial intelligence in medicine*, **87**, 9-19.

Elayeb, B. (2019): Arabic word sense disambiguation: a review. *Artificial Intelligence Review,* **52**(4), 2475-2532.

Fauzi, M.A. (2018): Random Forest Approach for Sentiment Analysis in Indonesian. *Indones. J. Electr. Eng. Comput. Sci.*, **12**, 46-50.

Feng, J., Gong, C., Li, X., and Lau, Y. (2018): Automatic approach of sentiment lexicon generation for mobile shopping reviews. *Wireless Communications and Mobile Computing*, **2018**, 9839432.

Gidhe, P., and Ragha, L. (2018): Comprehensive Method of Knowledge-Based Approach for Word-Sense Disambiguation. In *Proceedings of International Conference on Recent Advancement on Computer and Communication: ICRAC 2017* (**P**p. 63-71). Springer Singapore.

Griesel, M., and Bosch, S. (2020): Navigating challenges of multilingual resource development for under-resourced languages: The case of the African Wordnet project. *In: Proceedings of the first workshop on Resources for African Indigenous Languages, European Language Resources Association,* Marseille, France, **P**p. 45-50.

Gunawan, L., Anggreainy, M. S., Wihan, L., Lesmana, G. Y., and Yusuf, S. (2023): Support vector machine based emotional analysis of restaurant reviews. *Procedia Computer Science*, **216**, 479-484.

Jones, M. N. (2019): When does abstraction occur in semantic memory: Insights from distributional models. *Language, Cognition and Neuroscience*, **34**(10), 1338-1346.

Kaddoura, S., and D. Ahmed, R. (2022): A comprehensive review on Arabic word sense disambiguation for natural language processing applications. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, **12**(4), e1447.

Kanojia, D., Patel, K., and Bhattacharyya, P. (2022): Indian language WordNets and their linkages with Princeton WordNet. *arXiv preprint arXiv*:**2201**.02977.

Khattak, A., Asghar, M., Saeed, A., Hameed, I., Hassan, S., and Ahmad, S. (2021): A survey on sentiment analysis in Urdu: A resource-poor language. *Egyptian Informatics Journal*, **22**, 53–74.

Khoo, C.S., and Johnkhan, S.B. (2018): Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons. *Journal of Information Science,* **44**(4), 491-511.

Kokane, C.D., and Babar, S.D. (2019): Supervised word sense disambiguation with recurrent neural network model. *Int J Eng Adv Technol (IJEAT),* **9**(2), 1447-1453.

Kolajo, T., Daramola, O., and Adebiyi, A. (2019): Sentiment analysis on Naija-tweets. *In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, ACL*, Florence, Italy, **P**p. 338-343.

Kolajo, T., Daramola, O., and Adebiyi, A. Seth, A. (2020): A framework for pre-processing of social media feeds based on integrated local knowledge base. *Information Processing and Management*, **57**, 102348.

Krawczyk, B., and McInnes, B. T. (2018): Local ensemble learning from imbalanced and noisy data for word sense disambiguation. *Pattern recognition*, **78**, 103-119.

Kulkarni, D.S., and Rodd, S.F. (2022): Word sense disambiguation for lexicon-based sentiment analysis in Hindi. *Webology,* **19**(1), 592-600.

Kumar, B. M. (2018): A Survey on Word Sense Disambiguation. *Language India*, **18**(2).

Kumar, P. (2020): Word sense disambiguation for Punjabi language using deep learning techniques. *Neural Computing and Applications,* **32**(8), 2963-2973.

Li, Q. (2022): Co-attention-based pairwise learning for author name disambiguation (Master's thesis, University of Twente).

Mahadevaswamy, U. B., and Swathi, P. (2023): Sentiment Analysis using Bidirectional LSTM Network. *Procedia Computer Science*, **218**, 45-56.

Maitra, S., Madan, S., Kandwal, R., and Mahajan, P. (2018): Mining authentic student feedback for faculty using Naïve Bayes classifier. *Procedia computer science*, **132**, 1171-1183.

Maurya, A. S., and Bahadur, P. (2022): A detailed analysis of word sense disambiguation algorithms and approaches for indian languages. *In Proceedings of Second Doctoral Symposium on Computational Intelligence: DoSCI 2021* (**P**p. 693-710). Springer Singapore.

Mente, R., Aland, S., and Chendage, B. (2022): Review of word sense disambiguation and its approaches. *Available at SSRN*: http://dx.doi.org/10.2139/ssrn.4097221

Mukherjee, P., Badr, Y., Doppalapudi, S., Srinivasan, S.M., Sangwan, R.S., and Sharma, R., 2021. Effect of negation in sentences on sentiment analysis and polarity detection. *Procedia Computer Science*, **185**, 370-379.

Obiedat, R., Al-Darras, D., Alzaghoul, E., and Harfoushi, O. (2021): Arabic aspect-based sentiment analysis: a systematic literature review. *IEEE Access*, **9**, 152628-152645.

Ogbuju, E., and Onyesolu, M. (2019): Development of a general purpose sentiment lexicon for Igbo language. *Widening Natural Language Processing (WiNLP) Workshop, Association for Computational Linguistics (ACL) Conference*, Florence, Italy, P. 1.

Orimaye, S.O., Alhashmi, M., and Eu-gene, S. (2012): Sentiment analysis amidst ambiguities in YouTube comments on Yoruba language (Nollywood) movies. *In: Proceedings of the 21st International Conference on World Wide Web*, Lyon, France, **P**p. 583-584.

Oyewusi, W., Adekanmbi, O., and Akinsande, O. (2020): Semantic enrichment of Nigerian Pidgin English for contextual sentiment classification. Conference paper at ICLR **2020**, *arXiv preprint* arXiv:2003.12450.

Pal, A. R., Saha, D., Naskar, S. K., and Dash, N. S. (2021): In search of a suitable method for disambiguation of word senses in Bengali. *International Journal of Speech Technology*, **24**, 439-454.

Patel, N., Hale, J., Jindal, K., Sharma, A., and Yu, Y. (2021): Building on Huang et al. GlossBERT for Word Sense Disambiguation. *arXiv preprint arXiv:2112.07089*.

Patil, A., Patli, C., Ramteke, R., RP, B., and Darbari, H. (2020): Exploring Resources in word sense disambiguation for Marathi Language. *International Research Journal on Advanced Science Hub*, **2**, 108-111.

Priya, B., Nandhini, J.M., and Gnanasekaran, T. (2021): An analysis of the applications of natural language processing in various sectors. *In V.D.A. Kumar et al. (Eds.), Smart Intelligent Computing and Communication Technology*, *IOS Press*, **P**p. 598-602.

Purohit, A., and Yogi, K. K. (2022): A Comparative Study of Existing Knowledge Based Techniques for Word Sense Disambiguation. In *Proceedings of International Joint Conference on Advances in Computational Intelligence: IJCACI 2021* (**P**p. 167-182). Singapore: Springer Nature Singapore.

Rajabi, Z., Valavi, M.R., and Hourali, M. (2020): A context-based disambiguation model for sentiment concepts using a bag-of-concepts approach. *Cognitive Computation*, **12**(6), 1299-1312.

Sadia, A., Khan, F., and Bashir, F. (2018): An overview of lexicon-based approach for sentiment analysis. *In: 2018 3rd International Electrical Engineering Conference (IEEC 2018), IEP Centre*, Karachi, Pakistan, **P**p. 1-6.

Sazzed, S. (2021): BengSentiLex and BengSwearLex: creating lexicons for sentiment analysis and profanity detection in low-resource Bengali language. *PeerJ Computer Science*, **7**, e681.

Scarlini, B., Pasini, T., and Navigli, R. (2020): Sensembert: context-enhanced sense embeddings for multilingual word sense disambiguation. *In: Proceedings of the AAAI Conference on Artificial Intelligence*, **34**(5), 8758-8765.

Singh, C. P. (2021): A review on word sense disambiguation emphasizing the data resources on Wordnet and Corpus. *Information technology in industry*, **9**(2), 996-1016.

Singh, V., and Saraswat, K. (2019): Techniques for Disambiguation of Polysemy Words: A. *International Journal of Computer Applications*, **975**, 8887.

Tobaili, T., Fernandez, M., Alani, H., Sharafeddine, S., Hajj, H., and Glavas, G. (2019): SenZi: A sentiment analysis lexicon for the Latinized Arabic (Arabizi). *In: Proceedings of Recent Advances in Natural Language Processing*, Varna, Bulgaria, **P**p. 1203-1211.

Tripathi, P., Mukherjee, P., Hendre, M., Godse, M., and Chakraborty, B. (2020): Word sense disambiguation in Hindi Language using score based modified Lesk algorithm. *International Journal of Computing and Digital Systems*, **10**, 2-20.

Ullah, K., Rashad, A., Khan, M., Ghadi, Y., Aljuaid, H., and Nawaz, Z. (2022): A deep neural network-based approach for sentiment analysis of movie reviews. *Complexity*, **2022**, 5217491.

Vaishnav, Z. B., and Sajja, P. S. (2019): Knowledge-based approach for word sense disambiguation using genetic algorithm for Gujarati. In *Information and Communication Technology for Intelligent Systems: Proceedings of ICTIS 2018, Volume 1* (**P**p. 485-494). Springer Singapore.

Vaishnav, Z.B., and Sajja, P.S. (2019): Knowledge-based approach for word sense disambiguation using genetic algorithm for Gujarati. *In information and communication technology for intelligent systems*, Springer, Singapore, **P**p. 485-494.

Vedantam, V.K., (2021): The survey: advances in natural language processing using deep learning. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, **12**(4), 1035-1040.

Wang, S., Lv, G., Mazumder, S., and Liu, B. (2020): Detecting domain polarity-changes of words in a sentiment lexicon. *arXiv preprint* arXiv:2004.14357.

Yadav, A., Patel, A., and Shah, M. (2021): A comprehensive review on resolving ambiguities in natural language processing. *AI Open*, **2**, 85-92.

Yin, F., Wang, Y., Liu, J., and Lin, L. (2020): The construction of sentiment lexicon based on context-dependent part-of-speech chunks for semantic disambiguation. *IEEE Access*, **8**, 63359-63367.

Zhu, G., and Iglesias, C. A. (2018): Exploiting semantic similarity for named entity disambiguation in knowledge graphs. *Expert Systems with Applications*, **101**, 8-24.

Zul, M.I., Yulia, F., and Nurmalasari, D. (2018): Social media sentiment analysis using K-means and Naïve Bayes algorithm. *In: 2018 2nd International Conference on Electrical Engineering and Informatics (ICon EEI), IEEE*, Batam, Indonesia, **P**p. 24-29.

Author Contributions

Adedapo Bolaji Adeniji: Resources, Data Curation, Methodology, Writing – Original draft preparation. **Taiwo Kolajo:** Conceptualization, Investigation, Validation, Writing – Reviewing and Editing, Supervision. **Joshua Babatunde Agbogun:** Writing – Reviewing and Editing, Supervision.