

**UNIVERSITI TEKNOLOGI MARA**

**CONTEXT ENRICHMENT  
FRAMEWORK FOR SENTIMENT  
ANALYSIS IN HANDLING WORD  
AMBIGUITY RESOLUTION**

**NOR NADIAH B INTI YUSOF**

Thesis submitted in fulfilment  
of the requirements for the degree of  
**Doctor of Philosophy**  
**(Computer Science)**

**College of Computing, Informatics and Mathematics**

**May 2024**

## ABSTRACT

The rise of social networking platforms has created a surge in online opinionated data. Despite the challenges posed by this wealth of data, analysing and assessing this data presents an opportunity to gain valuable insights through sentiment analysis. However, ambiguous words pose a challenge for sentiment analysis algorithms as they require the identification of the correct meaning and word polarity within a specific context. Word sense disambiguation is a critical problem to be addressed in sentiment analysis due to the challenge of context-dependant nature, word ambiguity and negation present in sentiment texts. An enhancement of a generic sentiment analysis framework is introduced to address these issues, ensuring accurate texts classification in predicting the sentiment orientation. The proposed framework integrates enriched semantic information and a word ambiguity resolution strategy into sentiment analysis which is called context enrichment framework and aims to classify the sentiment texts by incorporating both context and semantic information. This is further elaborated in two novel components which are a word ambiguity resolution model (WAR) and a negation analysis model (NAM). The movie review dataset, considered as the baseline in sentiment analysis, is utilized, and the validation approach employs the SemCor 3.0 dataset. Word prior-polarity extraction is facilitated by the utilization of dictionary lexicons with predefined sentiment scores. The enhancement of this process utilises a hybrid lexicon, combining SentiWordNet and So-Cal, exploiting on the strengths of both resources. The WAR model identifies ambiguous words and generates context terms with a window size of three to resolve its ambiguity. The similarity between ambiguous words and their context words is evaluated using the cosine similarity approach. A rule-based method is introduced to select context words based on their similarity. The resolution of word ambiguity is addressed by introducing a formula that adjusts the polarity of ambiguous words. The evaluation of WAR's performance includes a comparison with the baseline model and evaluating the accuracy through the summation approach, which is based on aligning word polarity values with the document's assigned class label. A NAM is implemented within the negation scope to accurately address negation. The model incorporates a proposed negation rule-based method that considers the threshold value condition of negative words in the documents. NAM is evaluated by comparing its performance with the baseline negation rule-based, through accuracy assessment using the summation approach. Eventually, the assessed models of WAR and NAM, along with the evaluated word polarity extraction from dictionary lexicons, are integrated into the proposed CEF. Machine learning algorithms are deployed to perform sentiment classification. CEF is evaluated through a series of experiments, showcasing its superior effectiveness over the baseline approach. It achieves an accuracy of 75%, outperforming the baseline approach's accuracy of 67.75%. Additionally, the validation of CEF demonstrated its superiority with an accuracy of 78.95%, compared to the baseline approach's accuracy of 63.16%. The results show CEF through the implementation of WAR and NAM effectively addresses word ambiguity issues and enhances the performance of sentiment analysis. Future work in sentiment analysis could involve the incorporation of diverse datasets, investigation of additional techniques for resolving word ambiguity, and exploration of alternative dictionary lexicons. For classification performances, optimization of machine learning parameters and exploration of deep learning approaches can be applied for further enhancement.

## ACKNOWLEDGEMENTS

First and foremost, I express my gratitude to Allah for granting me the opportunity to undertake and successfully complete this extensive and challenging PhD journey. I extend my appreciation to the Ministry of Higher Education Malaysia and Universiti Teknologi MARA for their financial support throughout my studies.

I am profoundly thankful to my dedicated supervisors, Assoc. Prof. Dr. Shuzlina Abdul Rahman and Prof. Dr. Azlinah Mohamed for their unwavering support, guidance, motivation, and patience. Their continuous encouragement has been a source of inspiration and strength throughout this journey.

The support of my family has been invaluable, and I am deeply grateful to my husband, Mohd Hanafi Nasri@Masri, and my children, Puteri Addeena Safiyya, Muhammad Hakim, and Muhammad Helmy, for their understanding and encouragement during this journey. Heartfelt thanks to my parents, Hj. Yusof Ali and my siblings, extended family, and in-laws for their boundless love and prayers.

I extend special appreciation to my fellow postgraduate friends for their encouragement and moral support, which played a significant role in making this journey more manageable.

Lastly, I want to express my gratitude to all those who directly and indirectly contributed to this research. Thank you immensely.

## **TABLE OF CONTENT**

	<b>Page</b>
<b>CONFIRMATION BY PANEL OF EXAMINERS</b>	<b>ii</b>
<b>AUTHOR'S DECLARATION</b>	<b>iii</b>
<b>ABSTRACT</b>	<b>iv</b>
<b>ACKNOWLEDGEMENTS</b>	<b>v</b>
<b>TABLE OF CONTENT</b>	<b>vi</b>
<b>LIST OF TABLES</b>	<b>x</b>
<b>LIST OF FIGURES</b>	<b>xiii</b>
<b>LIST OF ABBREVIATIONS</b>	<b>xvi</b>
<b>CHAPTER ONE: INTRODUCTION</b>	<b>1</b>
1.1 Research Background	1
1.2 Problem Statement	7
1.3 Research Questions	11
1.4 Research Objectives	11
1.5 Research Scope	12
1.6 Research Significance and Contribution	13
1.7 Overview of the Research Methodology	15
1.8 Thesis Organisation	16
<b>CHAPTER TWO: LITERATURE REVIEW</b>	<b>18</b>
2.1 Introduction	18
2.2 Natural Language Processing	18
2.3 Sentiment Analysis	21
2.3.1 Generic Framework of Sentiment Analysis	23
2.3.2 Trends and Applications of Sentiment Analysis	25
2.4 Approaches in Sentiment Analysis	28
2.4.1 Machine Learning Approach	28

# **CHAPTER ONE**

## **INTRODUCTION**

The substantial increase in online opinionated data has resulted in an abundance of information and knowledge, which can be quite overwhelming. Analysing these opinions becomes crucial for extracting meaningful insights, identifying trends, and making informed decisions. Sentiment analysis plays an essential role in extracting valuable knowledge from the sea of data, making it a valuable tool in the context of the online environment. This research aims to address the challenge of word ambiguity within sentiment analysis. The primary objective is to explore how the inclusion of semantic context can reduce the word ambiguity in the sentiment texts. Additionally, addressing negation is considered crucial, and this effort contributes to a more refined understanding of sentiments expressed in textual data. Ultimately, integrating enriched semantic information is seen to enhance the accuracy and efficacy of sentiment analysis. In this chapter, the strength of this thesis is explained concisely. The research background is explained in section 1.1. Section 1.2 deliberates the research problem. Research questions and research objectives are stated in section 1.3 and 1.4, respectively. Consequently, Section 1.5 discusses the research scope, and the significant and contribution of this research is explained in section 1.6. Section 1.7 elaborates the overview of the research methodology. This chapter ends by describing the organisation of the thesis structure in section 1.8.

### **1.1 Research Background**

A massive use of internet has created a huge volume of information available on the web. Users are able to post content via blogs, forums and social network and they actively generate online information by expressing their thought, opinions or sentiments about all kinds of topics. This situation creates a great opportunity to understand the opinion of users by analysing the rich content of opinionated data. Due to the overloaded online opinions and reviews, sentiment analysis has gained much attention to researchers especially in the field of natural language processing (NLP). The main function of natural language is for communication and information exchange. The computer systems need to understand natural language and analysing it