

Universiti Teknologi MARA

**Sentiment Analysis for Malay Newspaper
(SAMNews) Using Negative Selection
Algorithm**

Nur Amalina Binti Redzuan

**Thesis submitted in fulfillment of the requirements for
Bachelor of Computer Science (Hons)
Faculty of Computer & Mathematical Sciences**

January 2013

DECLARATION

I certify that this thesis and the research to which it refers are the product of my own work and that any ideas or quotation from the work of other people, published or otherwise are fully acknowledged in accordance with the standard referring practices of the discipline.



.....
NUR AMALINA BINTI REDZUAN

2010453932

JANUARY 23, 2013

ABSTRACT

Newspapers express sentiments during reporting on recent events every day. It also can be known as a new domain in textual type for sentiment analysis that deals with many suggestions. The newspaper is documented in long sentences but the contents that express the sentiment is compressed and clearly understood by human. However, for the machine learning text representation it causes some problems because of the noisy text. As the solution, this project is conducted on the purpose to determine the polarity of the sentiment in the newspapers' sentences. This project is implemented based on five phases in methodology part which consists of background study, data collection and preparation, prototype design, prototype development and evaluation and documentation. Sentiment Analysis for Malay Newspaper (SAMNews) is constructed using the negative selection algorithm which is able to classify the sentiment in newspaper's sentences into the polarity (positive, negative or neutral) automatically based on detectors' words. The sentiment analysis in this project utilized 1000 newspaper's sentences for the training and classification phase and testing data to evaluate the average of accurateness. The evaluation is made on three experiments which in Experiment I used 700 newspaper's sentences as the training data and 300 newspaper's sentences as the testing data. The accuracy of this experiment is about 59.99%. In Experiment II, 800 newspaper's sentences and 200 newspaper's sentences are used as the training data and testing data. The accuracy of this experiment is increased about 58.58%. While in Experiment III used 900 newspaper's sentences as the training data and 100 newspaper's sentences as the testing data and the accuracy is improved to 65.81%. In future, a comparative study on Artificial Immune System and other techniques or algorithms can be carried out to enhance the performance of the classification model.

Keywords – negative selection algorithm (NSA), text mining, sentiment analysis, newspaper.

TABLE OF CONTENTS

CONTENT	PAGE
ACKNOWLEDGEMENT	iv
ABSTRACT	v
TABLE OF CONTENTS	vi

CHAPTER ONE: INTRODUCTION

1.1	Research Background.....	1
1.2	Research Problem.....	2
1.3	Scope of Research.....	3
1.4	Objective of Research.....	4
1.5	Significance of Research.....	4
1.6	Research Framework.....	5
1.7	Conclusion.....	6

CHAPTER TWO: LITERATURE REVIEW

2.1	Background of Study.....	8
2.1.1	What is Sentiment?.....	8
2.1.2	Variety of Sentiment.....	9
2.2	Text Mining and Sentiment Analysis.....	10
2.2.1	Text Mining.....	10
2.2.2	Sentiment Analysis.....	11
2.2.3	Subtask in Sentiment Analysis.....	11
2.2.4	Related Works in Sentiment Analysis.....	12
2.3	Preprocessing Technique.....	18
2.4	Artificial Immune System (AIS).....	21
2.4.1	Clonal Selection.....	22
2.4.2	Negative Selection.....	23

2.4.3	Immune Network	23
2.5	Negative Selection Algorithm (NSA)	23
2.5.1	Related Research on Negative Selection Algorithm	25
2.6	Sentiment Analysis Using Newspaper	26
2.7	Conclusion.....	26

CHAPTER THREE: METHODOLOGY

3.1	Research Framework.....	27
3.2	Preliminary Study.....	28
3.2.1	Knowledge Acquisition and Comprehension	29
3.3	Data Collection and Preparation	30
3.3.1	Data Preparation.....	30
3.3.2	Data Collection	30
3.3	Prototype Design.....	31
3.3.1	Engine Design.....	31
3.3.2	Preprocessing Engine Design	32
3.3.3	Negative Selection Engine Design.....	36
3.3.4	Testing Phase in classification model	39
3.4	Prototype Development and Evaluation.....	42
3.5	Documentation	44
3.5.1	Software and Hardware Requirement	44
(a)	Hardware Requirement Description.....	44
(b)	Software Requirement Description	44
3.6	Conclusion.....	45

CHAPTER FOUR: RESULT AND ANALYSIS

4.1	Data Description.....	46
4.1	List of detectors for Sentiment category	47