

**FILIPINO LANGUAGE PROCESSOR WITH GRAMMAR
CHECKER AND SUMMARIZER APPLICATION**

THESIS

**JAYSON CABANGLAN
JOHN MARK B. JAUDIAN**

College of Engineering and Information Technology

CAVITE STATE UNIVERSITY

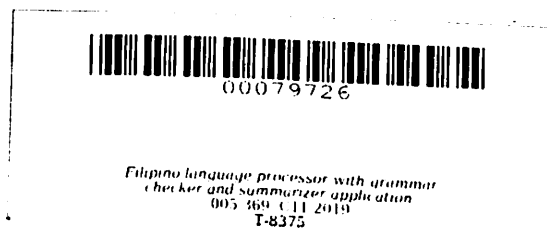
Indang, Cavite

June 2019

FILIPINO LANGUAGE PROCESSOR WITH GRAMMAR CHECKER AND SUMMARIZER APPLICATION

Undergraduate Thesis
Submitted to the Faculty of the
College of Engineering and Information Technology
Cavite State University
Indang, Cavite

In partial fulfillment
of the requirements of the degree
Bachelor of Science in Computer Science



JAYSON CABANGLAN
JOHN MARK B. JAUDIAN
June 2019

ABSTRACT

CABANGLAN, JAYSON and JAUDIAN, JOHN MARK B. FILIPINO LANGUAGE PROCESSOR WITH GRAMMAR CHECKER AND SUMMARIZER APPLICATION. Undergraduate Thesis. Bachelor of Science in Computer Science. Cavite State University, Indang, Cavite. June 2019. Adviser: Ms. Julie Ann C. Lontoc.

This study was conducted from March 2018 to April 2019 at Cavite State University - Main Campus. The purpose of the study was to develop a system that can create and manipulate Filipino text language that enables to establish a mechanism for leveraging natural language processing for Filipino language.

Extreme Programming was used as the methodology of the system. Four (4) instructors and 36 students from Cavite State University – Main majoring in Filipino and 10 faculty members of the Department of Information Technology evaluated the system. The results were collected through an evaluation form based on ISO9126.

According to the overall evaluation results, the system was judged to be excellent in all software quality criteria such as functionality, reliability, usability, efficiency, maintainability, portability, and user-friendliness. Upon the completion of the study, the proponents concluded that this would be a solution for the common problems encountered by Filipinos that uses word processor.

TABLE OF CONTENTS

	Page
BIOGRAPHICAL DATA	iii
ACKNOWLEDGMENT	iv
ABSTRACT	vi
LIST OF FIGURES	ix
LIST OF TABLES	x
LIST OF APPENDIX FIGURES	xi
LIST OF APPENDIX TABLES	xii
LIST OF APPENDICES	xiii
INTRODUCTION	1
Statement of the Problem	2
Conceptual Framework	3
Objectives of the Study	4
Significance of the Study.....	5
Time and Place of the Study	6
Scope and Limitations of the Study.....	6
Definition of Terms	8
REVIEW OF RELATED LITERATURE.....	10
Related Studies	21
Related Commercial Software	30
METHODOLOGY	34
Materials	34

Methods	34
System Architecture.....	38
RESULTS AND DISCUSSION.....	41
System Development	41
System Overview	42
Software Testing	45
Software Evaluation	46
SUMMARY, CONCLUSION, AND RECOMMENDATIONS	52
Summary	52
Conclusion	53
Recommendations	54
REFERENCES	55
APPENDICES	56

LIST OF FIGURES

Figure		Page
1	Conceptual framework of the study	4
2	First stage of Levenshtein algorithm	13
3	Final stage of Levenshtein algorithm.....	15
4	Extreme programming methodology (Auer & Miller, 2005).....	35
5	System architecture	40
6	Screenshot of the main page of the application	42
7	Screenshot of the File menu.....	43
8	Screenshot of Edit menu.....	43
9	Screenshot of suggestion for the underlined word.....	44
10	Screenshot of suggestion applied to the underlined word.....	44
11	Screenshot of the text and summarize button.....	45
12	Screenshot of the result of clicking the summarize button.....	45

LIST OF TABLES

Tables	Page
1 How to calculate costs for cells.....	14
2 Example of tokenized lexeme.....	16
3 Contribution of related local studies to the system.....	29
4 Comparison of related studies to the proposed system.....	30
5 Comparison of commercial system to the system.....	33
6 ISO 9126 characteristics and sub-characteristics	37
7 Rating scale for software evaluation	47
8 Participants' evaluation of the software in terms of its functionality.....	47
9 Participants' evaluation of the software in terms of its reliability	48
10 Participants' evaluation of the software in terms of its usability.....	48
11 Participants' evaluation of the software in terms of its efficiency.....	49
12 Participants' evaluation of the software in terms of its maintainability.....	49
13 Participants' evaluation of the software in terms of its portability.....	50
14 Participants' evaluation of the software in terms of its user-friendliness ...	50
15 Summary results of the evaluation.....	51

LIST OF APPENDIX FIGURES

Appendix Figures	Page
1 Fishbone diagram (Absence in handling Filipino Language in a word processor)	58
2 Fishbone diagram (Perplexion of the user in today's processor)	58
3 Fishbone diagram (Unable to summarize Filipino language in existing word processor)	59
4 Gantt chart of the development	59
5 Use case diagram of the system	60
6 Survey result of question no. 5	61
7 Survey result of question no. 3	61
8 Survey result of question no. 9	62
9 Survey 2 result of question no. 3	62

LIST OF APPENDIX TABLES

Appendix Table		Page
1	Frequency distribution of scores of the functionality indicators.....	64
2	Frequency distribution of scores of the reliability indicators.....	64
3	Frequency distribution of scores of the usability indicators.....	64
4	Frequency distribution of scores of the efficiency indicators.....	64
5	Frequency distribution of scores of the maintainability indicators.....	64
6	Frequency distribution of scores of the portability indicators.....	65
7	Frequency distribution of scores of the user-friendliness indicators.....	65

LIST OF APPENDICES

Appendix		Page
1	Survey Questionnaire	67
2	Sample Software Evaluation Sheet	71
3	Sample Accomplished Software Evaluation Sheet	75
4	Unit Testing.....	79
5	Integration Testing	86
6	System Testing	90
7	Sample Source Code	94
8	Letters, Forms, and Certificates	100

FILIPINO LANGUAGE PROCESSOR WITH GRAMMAR CHECKER AND SUMMARIZER APPLICATION

Jayson Cabanglan
John Mark B. Jaudian

An undergraduate thesis submitted to the faculty of Department of Information Technology. College of Engineering and Information Technology, Cavite State University, Indang, Cavite in partial fulfillment of the requirements for the degree Bachelor of Science in Computer Science with Contribution No. CEIT 2018-19-2-143. Prepared under the supervision of Ms. Julie Ann C. Lontoc

INTRODUCTION

Word Processing is an application program for manipulating text-based documents. It utilizes a word processor or editor as an electronic equivalent of paper, pen, typewriter, eraser and most likely, dictionary and thesaurus (Kunde, 2008).

Natural Language Processing (NLP) is a field that covers computer understanding and manipulating of human language, and it's ripe with possibilities for newsgathering (Pesce, 2017). NLP is a way for computers to analyze, understand, and derive meaning from human language in a smart and useful way. By utilizing NLP, developers can organize and structure knowledge to perform tasks such as automatic summarization, translation, named entity recognition, relationship extraction, sentiment analysis, speech recognition, and topic segmentation.

This study aimed to develop a system that can create and manipulate Filipino text language that enables to establish a mechanism for leveraging NLP for Filipino language.