A MERANAO TRANSLATOR IMPLEMENTING STRING MATCHING ALGORITHM

AN UNDERGRADUATE THESIS

By:
JOHAIRA ALI SAMSODDEN
NIM. 11650108

DEPARTMENT OF INFORMATICS
FACULTY OF SCIENCE AND TECHNOLOGY
ISLAMIC STATE UNIVERSITY – MAULANA MALIK IBRAHIM MALANG
2018

A MERANAO TRANSLATOR IMPLEMENTING STRING MATCHING ALGORITHM

AN UNDERGRADUATE THESIS

Submitted to:
Islamic State University Maulana Malik Ibrahim Malang
In Partial Fulfillment of the Requirements of the Degree
Course Bachelor of Science in Computer (BS-Com)
/ Sarjana Komputer (S.Kom)

By:
JOHAIRA ALI SAMSODDEN
NIM. 11650108

DEPARTMENT OF INFORMATICS
FACULTY OF SCIENCE AND TECHNOLOGY
ISLAMIC STATE UNIVERSITY – MAULANA MALIK IBRAHIM MALANG
2018

LETTER OF APPROVAL

A MERANAO TRANSLATOR IMPLEMENTING STRING MATCHING ALGORTHM

AN UNDERGRADUATE THESIS

By:
JOHAIRA ALI SAMSODDEN
NIM. 11650108

Has been reviewed and approved for the examination
Dated January 5, 2018

Supervisor I	Supervisor II
Dr. Cahyo Crysdian M.CS	A'la Syauqi, M.Kom
NIP. 19740424 200901 1 008	NIP. 19771201 200801 1 107

Approved by,
The Head of the Department of Informatics

Dr. Cahyo Crysdian M.CS
NIP. 19740424 200901 1 008

LETTER OF ENDORSEMENT

A MERANAO TRANSLATOR IMPLEMENTING STRING MATCHING ALGORITHM

AN UNDERGRADUATE THESIS

By:

JOHAIRA ALI SAMSODDEN
NIM. 11650108

Has been conducted in the presence of the board of thesis examiners and stated accepted as one of the requirement to obtain the title Bachelor of Science in Computer / Sarjana Komputer (S.Kom)

Dated January 5, 2018

Board of the Examiner		Signature
Chairperson	: **Roro Inda Melani, S.Kom., M.Sc** NIP. 19780925 200501 2 008	( )
Head	: **Irwan Budi Santoso, M.Kom** NIP. 19700103 201101 1 004	( )
Secretary	: **Dr. Cahyo Crysdian, M.CS** NIP. 19740424 200901 1 008	( )
Member	: **A'la Syauqi, M.Kom** NIP. 19771201 20080S1 1 107	( )

Assessed and approved by,
The Head of the Department of Informatics

Dr. Cahyo Crysdian, M.CS
NIP. 19740424 200901 1 008

LETTER OF STATEMENT

The undersigned below

Name	: Johaira Ali Samsodden
Student's Identity Number (NIM)	: 11650108
Faculty	: Science and Technology
Department	: Informatics
Thesis Title	: A Meranao Translator Implementing String Matching Algorithm

States that the thesis mentioned above is my original work except for the quotations and statements whose resources are acknowledged on the references. Any shortcomings in this present work, therefore, are entirely my own responsibility.

Moreover, this work is not a plagiarism result and if it is found that this statement is false, my academic records will attest to invalidation and I will be responsible for that as well.

Malang, 9 January 2018
Signed by,

Johaira Ali Samsodden
11650108

DEDICATION

This thesis is specially dedicated to my beloved parents
(a.k.a. the inspiring source of my human super power)
whose affection, love, encouragement and prayers of days and nights
made me able to achieve favorable success in this phase of my life.
Thank you!
Alhamdulillah!

MOTTO

“Success is Never a Coincidence”

“Success is Never a Coincidence”

“Success is Never a Coincidence”

ACKNOWLEDGEMENT

بِسْمِ اللهِ الرَّحْمٰنِ الرَّحِيْمِ

I would like to express my greatest gratitude to Allah The Almighty for providing me enough spiritual and mental strength in pursuing this work of knowledge for the seek of His will and hopefully for the benefit of the rest of the humanity following same intent of knowledge interest in accumulating for the advancement of any related study.

Every challenging work needs self-efforts as well as guidance from those who have the relevant knowledge regarding the study and unconditional help, support, guidance, advice and actual hands-on contributions. With that being said, I would like to say thanks to the following people:

1. Mr. Cahyo Crysdian as the Head of the Department of Informatics for his moral support despite hectic schedule and his large extent of persistent in providing informative contribution for the compilation of this paper.
2. Mr. Ala Syauqi as the second advisor for his sincere advice and concern for this piece of work to be fulfilled in deserving delicate manner.
3. Informatics Department Professors (*Dosen*), Laboratory Staffs, and Administrator Staffs for the kindness treatment and informative guidance leading this attainable effort toward fulfilling not just the paper but the entire program study covering since day one to present.
4. My concern informatics colleagues for their untiring assessment and company expressing brotherhood and sisterhood throughout these years.

5. Mr. Mohammad Mualif as the Head of the International Office for his dynamic role and compassionate care for the international students' overall affairs and for his painstaking dealings with a diplomatic approach regarding our occasional difficulties.

Thus, I would like to indicate that this is a substantial knowledge of my own work, thus, I am in deep apology for any lacking provisions necessary for critics' sense and all that is not presented and hopefully be recognized and fulfilled for further accomplishments.

Malang, 9 January 2018

Johaira Ali Samsodden

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Johaira Ali Samsodden, 2018. A Meranao Translator Dictionary Implementing String Matching Algorithm. An Undergraduate Thesis. Department of Informatics. Faculty of Science and Technology. Islamic State University of Maulana Malik Ibrahim Malang. Supervisor**:** (I) Dr. Cahyo Crysdian, M.CS (II) A'la Syauqi, M.Kom

---

Keywords : *Meranao Language*, *Translator*, *Corpus*, *String Matching Algorithm*, *Text Analysis*.

One of the biggest challenges on preserving a language that is not documented is the lack of nurturing and retaining the use of it specially the minority of employing such language into expanding the use of it deliberately within the community. Discovering Meranao text without ascertain bases for exploration is likely resulting to deprivation of its conservation.

String Matching plays an important role in solving a problem regarding most text analysis related. The building of database corpus composed of Meranao words from an e-book and other paper-based dictionary is collectively usable for the implementation of String Matching algorithm. Experimental test evaluations has been made on establishing an application for the accuracy rate on finding the translated word from the database corpus in comparison with a Meranao language library source.

For this study, the performance of the application has statistically dependent on the number of input text being translated resulted from the highest performance at 80% accuracy rate deflated down to 35%. The larger number of text being translated, the more possibility for errors encountered due to grammatical exclusion on this study.

CHAPTER I

INTRODUCTION

1.1 Background of the Study

Philippines has a long history of struggles to free the country from colonizers. It even took many years before the country have the official language. To have an own language is a proof that the country is a free nation. Maranao is an Austronesian language of the Greater Central Philippines subgroup spoken primarily on the southern Philippine island of Mindanao in the provinces of Lanao del Sur and Lanao del Norte, although a considerable number of speakers have migrated throughout the entire Philippines as merchants.

Outside of the Philippines, there are a small number of Maranao who have migrated to the United states, and a larger number living in various Middle Eastern countries as Islamic scholars, students, and contract workers.

In terms of sociolinguistics and language attitudes, the Maranao language(and culture) is likely one of the least endangered in the Philippines. The Maranao are one of the few Philippine ethno-linguistic groups who in the twenty-first century still resolutely maintain use of their language and wear traditional clothing, even in Manila and other large cities.

Because the Maranao entries are stems without inflectional affixes, the user will not find in the dictionary every form he/she encounters in the language. Upon inspection of a form, the user must first locate the root much as is the case for classical languages such as Greek or Latin. With this belief of profound cases

behind one's own language regarding the word formations and without conservation can lead to losing heritage main medium of communication.

There's not much of the study regarding Maranao Verbs as it is not being taught under the Philippine government education and so, supporting the retention of the Maranao Verbs by implementing a Morphological Analyzer will provide a different level of reliability based on an automated word construction.

String matching is very important subject in the domain of text processing and it has been one of the most extensive problems in computer technologies during past two decades. It has applications such as DNA analysis, information retrieval systems and several other fields (Rajashekharaiah et al, 2012)

In a world of increasing intersectionality and diversity, it is no longer rare to find teenagers with several ethnic backgrounds or upbringings, nor is it necessarily a negative trait. Teenagers who have grown up in an environment where they've been exposed to several different languages have a unique pathway ahead of them when it comes to educational and social development.

It is so inevitable how one's own language is so essential to human being existence. Allah created us all full of wisdom in every aspect of our own being. Language is so important that without them, communication would not be possible if this does not created as well. Learning one's own language is part of worship as it is appreciating what Allah has given toward us and as we nurture it, we are showing our creator a gratitude of what has He given us as a gift. Allah mention in the Quran (30:22):

"وَمِنْ آيَاتِهِ خَلْقُ السَّمَاوَاتِ وَالْأَرْضِ وَاخْتِلَافُ أَلْسِنَتِكُمْ وَأَلْوَانِكُمْ ۚ إِنَّ فِي ذَٰلِكَ لَآيَاتٍ لِلْعَالِمِينَ"

Meaning in translation:

"And among His Signs (is the) creation (of) the heavens and the earth, and the diversity (of) your languages and your colors. Indeed, in that surely (are) Signs for those of knowledge" Qur'an (30:22).

Our native language not only allows us to communicate and connect with one another, but it allows us to understand and appreciate the history of our ancestors and our upbringing. It cultivates an appreciation and understanding that is beyond beneficial for this new generation, especially those from diverse familial backgrounds.

1.2 Statement of the Problem

How accurate String Matching Algorithm to translate Meranao sentences into English Language.

1.3 Scope and Limitations of the Study

a. This study focuses on simple common Meranao Words used in today's language for Meranao people and linguistic students or language enthusiast.

b. The rules on finding the matching words applies only to the same exact spelling on the data corpus available.

c. This study does not include Meranao phonology in its scope.

1.4 Object of the Study

To measure the accuracy of String Matching Algorithm in search for the exact translation provided by the corpus collection.

1.5 Significance of the Study

a. This study will be of great contribution, not only to the Computer Science society, but also, to the Meranao community.

b. Meranao Translator will aid in any Meranao dictionary developer to produce the closest Meranao translation from Meranao word input into English Translation.

c. This study is relevant to areas such as Information Retrieval(IR), including classification, clustering or retrieval of documents.

CHAPTER II

LITERATURE REVIEW

2.1 Meranao Languge

a. Mera Parse

In Bagul (2006), a Mera Parser Research study, have developed a parsing algorithm for Meranao language taking the same parser used for analysis patterned similarly to TAGFWO stands for Tagalog Free Word Order. Bagul research is concentrated mainly on the result of the given input text broadly in a sentence form and process the result output syntactically correct or wrong in a "YES" and "NO" label. The Mera Parser, because of Lexical Fuctional Grammar use for the Tagalog Free Word Order, they have implemented it to Meranao Language formalism for the grammar check.

The Mera Parser have gone through tests and have successfully proven with every input sentences measuring time for execution. They have done a fifty(50) different free-word for the test required.

b. MELT: Meranao to English Language Translation

Pangandaman (2008) in further enhancement for Mera Parser by Bagul (2006), MELT which stands for Meranao to English Language Translator, in their research paper, they have explained that they have used the Transfer Approach technique to assist them in having a much clearer stages result, and analization result and transfer generation. They have stated that their study works only for unilingual translation, that is, from Meranao to English. In their study Mera Parser

plays a major role in the implementation process before any translation of words is done for MELT. They pointed out that initially, MELT parses the Meranao input sentence to be grammatically correct and that otherwise an error handler handles the input sentence on their study.

They have explained that MELT has tested ample inputs that translated 147 Meranao sentence in its VerbSubject-Object, Verb-Object-Subject, Verb-Subject, Verb-Object, and Verb patterns only. The MELT study as they have stated that they have successfully parsed input sentences in such patterns. They have had the translator first stores the words from the input sentence in tokens. So for every word in the tokens generate in MELT, the English equivalent and word attributes are obtained. Then, they have confirmed that it determines the subject of the equivalent English translation through the article "so" or "si" before the noun in the subjective voice. And then MELT determines the object of the Meranao input sentence by the article "sa" before noun. Finally, if no object is present in the input sentence, a pronoun is being supplied to replace it.

c. Phonemes: Learn the Maranao Language

The website provided the following information:

Maranao is a language spoken by nearly 1,142,000 people living in the island of Mindanao, Lanao del Norte and Lanao del Sur Provinces. It belongs to the Southern Philippine subbranch of the Western Austronesian language family, and is most closely related to the Ilanun language spoken in Sabah, Malaysia and Maguindanaon, spoken in Maguindanao, North Cotabato, South Cotabato, Sultan Kuderat, and Zamboanga del Sur provinces.

Phonology

Letters in Maranao alphabet (dictionary letters, not standardized): A B D AE E G H I K L M N NG O P R S T W Y. AE represents the pepet (central high) vowel, Philippine schwa. As the Meranao pronunciation differs massively from English pronunciation. The Latin alphabet equivalent above is what comes near regarding the most sounding-like result after the real Latin alphabet pronunciation.

Word Classes

Word classes in Meranao represent many categories as it goes to English equivalence depending on the part that they play on the sentence. The following tables are given to illustrate clear lists of each classes for further Meranao language understanding.

Pronouns

In this section, Table 2.1 presents the pronoun in English Language and its counterpart in Meranao Language afflictions. The sets of Meranao changes accordingly are illustrated depending on grammatical and formality use. We present the complete sets and changes on Meranao pronouns on Table 2.1 in columns Set I, Set II, Set III, and Set IV in a recognizable outline.

Table 2.1 Meranao Pronouns on Different Level of Use

	Set I	Set II	Set III	Set IV
I	ako	aken	raken	saken
you (sg.)	ka	ngka/ka	reka	seka
he/she/it	sekaniyan	niyan/iyan	rekaniyan	sekaniyan
we (excl)	kami	mi/ami	rekami	sekami

we (incl)	tano	tano	rektano	sektano
we dual	ta	ta	rekta	sekta
you (pl)	kano	niyo/iyo	rekano	sekano
they	siran	ran/iran	kiran	siran

Common Particles:

Traditionally, Meranao Language depending on the context has particles that may or may not inflect the grammar rulings. Here are some common particles that Meranao Language has and its English equivalent:

anda 'where', peman 'again', imanto 'now', den 'emphasis', di' 'no', ba 'interrogative marker', a go 'and', ka 'because', kagia 'when', o 'if', anday 'as soon as'.

Ligatures: a (links appositives); na (links subject and predicate). Ligatures in Meranao Language uses this to connect two words and becomes one. More elaboration sample on Meranao ligatures in a form of sentence are shown at the end of this segment.

Determiners: Common: so, o, sa, ko; Personal: si, I, ki. The determiner shows in Table 2.2 listing a detailed information about the expression to be used on each Verb types in regard to the Actor, Object, Referent, and Instrument context.

Table 2.2 Determiners for Meranao Verb Types

Verb Type	Actor	Object	Referent	Instrument
Active	so	sa	ko	
Objective	o	so	ko	ko

Referential	o	sa	so	ko
Instrumental	o	sa/ko	ko/sa	so

Verbs

Meranao Verbs are identified based on Voice, Mode, Aspect and Tenses. It is illustrated further below on how this types are inflected based on the grammatical rules and correctness in Meranao language use. The following are the afflictions changes and its corresponding contextual references;

Voice: active –om- (so phrase actor); objective –en; referential –an; instrumental I-.

Mode: indicative (unmarked); obligatory –a or –I; aptative ka-; causative paka-, paki-; aptative causative kapaki-

Aspect: distributive pang-

Tense: 4 tenses: past/completed –iy-; immediate future = replacive ae; present progressive pe- + immediate future; neutral (unmarked)

Verbal Paradigms

Verbal paradigms and the complications in the form of each grammar indicating verbal word form changes are occurred to be very multi-formal cases. The following tables are the grammatical categories in complex verbal paradigms and break down into table for categorical distinction.

Table 2.3, Verbal Paradigm in an Active form showing General, Aptative, and Distributive in corresponds to Neutral, Past, and Immediate Future on tense bases.

Table 2.3 Verbal Paradigm Active form

	General	Aptative	Distributive
Neutral	-om-	maka-	maN-
Past	-omiy-	miyaka-	miyaN-
Immediate Future	#-ae-	paeka-	#paeN-

Table 2.4, similarly to the table above, this table contains the Verb modal on Neutral, Past, Immediate Future and the addition of Present Progressive, and Obligatory in corresponds to its Emphatic Distributive, Causative, and Aptative Causative form depending on the Meranao Language grammatical context to be used.

Table 2.4 Verbal Paradigm Active form with Appends

	Emphatic Distributive	Causative	Aptative-Causative
Neutral	mamaN-	ma#ki-	ma#paki-
Past	miyamaN-	miya#ki-	miya#paki-
Immediate Future	#paemaN-	#paeki-	#mapaeki
Present Progressive	pepaemaN-	#pepaeki-	
Obligatory	pamaN-	paki-	

Table 2.5, the Verbal Paradigm Objective Form shows the form of Verb affixes in modal context Neutral, Past, Immediate Future, Present Progressive,

Obligatory, and Obligatory Immediate Future illustrating its word form in General, Aptative, Distributive, and Emphatic category.

Table 2.5 Verbal Paradigm Objective Form

	General	Aptative	Distributive	Emphatic Distributive
Neutral	-en	ma-#	paN- -en	pamaN- -en
Past	-iy-#	miya-#	piyaN-#	piyamaN-#
Immediate Future	-ae- -en	kae-#	paeN- -en	paemaN- -en
Present Progressive	pe- -ae- -en	pekae-#	pepaeN- -en	pepaemaN- -en
Obligatory	-a	ka- -a	paN- -a	pamaN- -a
Oblig. Immed. Future	-ae- -a	kae- -a	paeN- -a	paemaN- -a

Table 2.6 are lists of word forms categorizing from Causative, Causative2, and Aptative/Causative in the verbal mode Neutral, Past, Immediate Future, Present Progressive, Obligatory, Obligatory Immediate Future, and Obligatory Present Future.

Table 2.6 Verbal Paradigm Objective Form with Appends

	Causative	Causative 2	Aptative/Causative
Neutral	paki- -en	paka- -en	mapaka- #
Past	piyaki-#	piyaka-#	miyapaka-#
Immediate Future	paeki- -en	paeka- -en	mapaeka-#
Present Progressive	pepaeki- -en	pepaeka- -en	
Obligatory	paki- -a	paka- -a	

Oblig. Immed. Future	paeki- -a	paeka- -a	
Oblig. Pres. Prog.	pepaeki- -a	pepaeki- -a	

Table 2.7, a lists of verbal mode on Neutral, Past, Immediate Future, Present Progressive, and Obligatory and its corresponding category on General, Aptative, and Distributive form are illustrated.

Table 2.7 Verbal Paradigm Instrumental Form

	General	Aptative	Distributive
Neutral	i-	mi-	ipaN-
Past	ini-	mini-	inipaN-
Immediate Future	i- -ae-	k#i-	ipaeN-
Present Progressive	ipe- -ae-	pek#i-	ipepaeN-
Obligatory	-en		

Table 2.8, a lists of verbal mode on Neutral, Past, and Immediate Future and its corresponding category on Emphatic Distributive and Aptative-Causative word forms are illustrated.

Table 2.8 Verbal Paradigm Instrumental with Append

	Emphatic Distributive	Aptative-Causative
Neutral	ipamaN-	Kipaki-

Past	inipamaN-	Minipaki-
Immediate Future		K#ipaeki-

Table 2.9 shows the verbal mode on Neutral, Past, Immediate Future, Present Progressive, Obligatory, Obligatory Immediate Future, Immediate Present Progressive and its corresponding form on General, Aptative, and Distributive use.

Table 2.9 Verbal Paradigm Referential

	General	Aptative	Distributive
Neutral	-an	ka- -an	paN- -an
Past	-iy- -an	kiya- -an	piyaN- -an
Immediate Future	-ae- -an	kae- -an	paeN- -an
Present Progressive	pe- -ae- -an	pekae- -an	pepaeN- -an
Obligatory	-i	ka- -i	paN- -I
Oblig. Immed. Future	-ae- -i	kae- -i	paeN- -I
Oblig. Pres. Prog.	pe- -ae- -i	pekae- -i	pepaeN- -i

Table 2.10, illustrate verbal mode on Neutral, Past, Immediate Future, Present Progressive, Obligatory, Obligatory, Immediate Future, Present Progressive, Obligatory, Obligatory Immediate Future, and Obligatory Present Progressive corresponding to Emphatic Distributive, Causative, Aptative/Causative word form.

Table 2.10 Verbal Paradigm Referential with Append

	Emphatic Distributive	Causative	Aptative/Causative
Neutral	pamaN- -an	paki- -an	kapaki- -an
Past	piyamaN- -an	piyaki- -an	kiyapaki- -an
Immediate Future	paemaN- -an	paeki- -an	kaepaki- -an
Present Progressive	pepaemaN- -an	pepaeki- -an	pekaepaki- -an
Obligatory	pamaN- -i	paki- -i	kapaki- -I
Oblig. Immed. Future	paemaN- -i	paeki- -i	kaepaki- -I
Oblig. Pres. Prog.	pepaemaN- -i	pepaeki- -i	pekaepaki- -i

Enlists are example sentences from A Meranao Dictionary (McKaughan and Macaraya, 1996) and its valid grammatical approach:

1. *Patot a di' ka manalipenda'an dapay o kena' a ka pagari so tebangan*. It is proper for you not to neglect him even if it is not your brother who needs help.
2. *Tomininda' so manga baebay si'i ko pagapoyan*. The women cooked in the fireplace.
3. *Di'i ako iran paganta'an*. They are gossipping about me.
4. *Si'i gomegenek so mala' a seda' sa laod*. The big fish is living far from the shore.
5. *Saken i miyakailap ko pirak aken*. I myself lost my money.

6. *Balilisa ngka siran oto a manga rarata' i ongar*. Beat those who are misbehaving.
7. *So taw a malowig na di' paeka'awid sa mapened*. A weak person cannot carry something heavy.
8. *Aya antap iyan ko pirak iyan na dado*. He aims to use his money for a plow.
9. *Kamolo'a ngka a piyak anan ago ngka raken began*. Snatch that chick quickly and give it to me.
10. *Manginon ako den a kasasakitan ako na asar a di' kasakitan so wata' aken.* I will endure pain as long as my child does not suffer.
11. *Tingel aken sekaniyan ka kagiya piyagonaronar ako niyan*. I slapped him because he insulted me.
12. *Di' ako mata'o pikir.* I do not know how to think.
13. *Da'ir o lopa' aken so lopa' iyan*. His land is adjacent to mine.
14. *Miyakaleseb so garangan o kiping ko tolan.* The blade of the knife penetrated to the bone.
15. *Mata'an sekaniyan a paekaoma imanto*. It is certain that he will arrive today.

(Phonemes: Learn the Maranao Language, 2008)

2.2 Rule-Based Morphological Analysis

Klink and Kieninger (2001) in this study, researches begins by identifying the processes requirements which is dividing the phases of processing for the document understanding to initialize necessary data for document image processing. They tend to equip a study that will process a document structure

analysis in both layout and textual feature of a document provided. Through their research study they have experimented a document processing with a rule-based document structure understanding in a fuzzy combination of layout and textual features and have gone through phases that have been resulted to document understanding and proceeded to rule-based evaluation which they have found that the complexity of the rules applied adds to its enhancement but also a massive drawbacks on the study capability for further preferences.

Tamames (2005) pointed out that the study Text Detective: a rule-based system for gene annotation in biomedical texts, demands a particular procedure as it is a critical steps in the development on text mining applications in biosciences. Gene as an objective of the study makes one of the factor that this study may require rigid project consideration. In this study, Tamames did a total different kind of approach because of the gene name difficulties in their detection. The rules that is being implemented to meet the aim design and the several lexicons of biological concept on Text Detective System have gone through procedures occurring possible availability matching the names and gene symbols in a tagging lexicon contexts. the overall results of the performance of the above procedure with the method referencing on this study still displays dependency on the quality of the lexicon which is provided in the gene lists. As Tamames concluded that the observation on errors focuses on few difficult cases and so the study continues.

Tang (2006) tackled a paper study entitled English Morphological Analysis with Machine-learned Rules. Tang quoted that machine learning can avoid problems such as costly human labor, rule inconsistency and can provide additional statistical information which can be used in morphological analysis

procedure. The study on specifying a morphological rules acquiring by the statistical learning from lists of the words, in considering a particular morphological features on English language that has taken carried of through out the study. Under the procedure process on analyzing the morphological that has taken into consideration, two types of uncertain cases have found out such as intersectional ambiguity and combinatory ambiguity. As any regarding morphological analysis procedure, Stemming the algorithm given and morphological parsing, which along stages, Stemming on the de-suffixing step and the recoding step can be done in sequence or simultaneously. On the other hand, morphological parsing which Koskenniemi (1983) represented the Model of Two-level Morphology considering Morphological analysis and transformation in different phonological situations and encodes the correspondence between surface form and lexical form with finite-state transducer. On Tang conclusion, the segmentation in Chinese is more important than the morphological analysis in Indo-European languages and because Tang believe that the identification of internal structure of word forms provides very useful information for other language analysis tasks, then Tang build a motivation per se in pursuing an analyzer for English. At the end of his paper, he had stated that after the experiment with the satisfactory performance and the fairly higher result comparing to other algorithms, problems on the other hand still remain.

Majdi, Eric, and Mohammad (2013) in SALMA research, their paper study have put a review on the SALMA-Tools which stands for Standard Arabic Language Morphological Analysis that the researches made it an open-source standard collection for the availability of broadening the scope of Arabic word

structure. The complexity in the morphological formation of Arabic text concerning both vowelized and non-vowelized text makes the study more complicated than any other language studies. Tag-assignment in implementing in the Arabic language study has seen complexity because of the language itself. The SALMA as they have describe it as a fine grained morphological analyzer they have set a Tag for encoding standardizing morphological features of Arabic for the compact yet transparent notation format. As they have stated in their paper that Arabic is a living language that belongs to the Semitic group of languages, the evaluation on applying such algorithms are quite processed in another field of research studies as well and so, the comparison made it unable to be done because of this reason. The automation of morphological annotation of Arabic has seen problems as well through designing and computing a developed algorithms thus they may address the problem. They have done surveys in order for the Arabic morphological analyzer in discovering the theoretical current challenges on the morphological analysis for Arabic practice application. As the result of their research study, development of resources and development of linguistically-informed standards for Arabic morphological analysis. In their final statement, they have concluded that the information collected for the SALMA to be well-develop has been put into resources and standardize for different domains, formats and genres.

Khumbar, Braja, Dipankar and Sivaji (2012) the researchers define morphological analysis as it is concerned mainly on retrieving the syntactic and morphological properties or the meaning of morphologicall y complex word. In their paper study, the Morphological Analyzer for Kokborok introduces the design

and implementation in the purpose of a contraining resource and less computerized for Indian language. The researchers use a database driven affix stripping algorithm on designing the Morphological Analyzer which leads to accumulating grammatical information combining collection of words. They have undergone experiment in testing the Morphological Analyzer for Kokborok with morethan fifthy thousand words and confirmed the accuracy at 80 % in attaining manual check. There are exemptions on classifying the words within sentences as the information for the Morphological Analyzer Test and so the distinction between the noun class and verb classes is understandably relative but so to mention, that the distinction between the noun and adjective is yet unclear. In their conlusion they have confirmed that the tagging in morphology driven on Kokborok POS is certainly much dependent on the lexical rules based on the morphological analysis on each category. In their conclusion, they have proposed the need for more text corpus in order for more advance better transfer rules and techniques to accomplish a high value reference output for future improvise Morphological Analyzer.

Aswani and Sindul (2014) the researchers regarding their study pointed out that analyzing, generating and also understanding languages deals with the general use of Natural Language processing. Morphological Analyzer is implemented for this case as it is important procedure when it comes to language processing due to the scientific set of any natural language structural form. They have illustrated and implemented as well different techniques in morphological analyzer in their paper particularly for Malayalam language. Their study survey a morphological techniques and break down methods and advantage so that the

limitation can describe and elaborate how method works in its own advancement for further related research studies. In conclusion to that, the researches confirmed that the works are done in nouns and verbs which is also reach in inflections and the complexity is pursued to stem and its affixes study for the development of the language.

Mijit, Tatsuya, Akbar and Askar (2016) in this paper study, researchers studied an Uyghur language Morphological Analyzer and focused on concatenating suffixes as they have pointed out that the proper makes a large number of combinations of morphemes, and greatly increases the word-vocabulary size, causing out-of-vocabulary(OOV) and data sparseness problems for statistical models. Researchers put into consideration the provision on high coverage and smaller lexicon size and semantic and syntactic information. During the designing and implementing it supervised morpheme segmentation tool that they have standardized and manually segmented the Uyghur morphemes, and they have specialize the suffixes and summarized and classified morpho-phonetic rules and their implementations. They have stated that the general purpose for doing a morpho-phonetic analyzer is to that it can segment Uyghur text into phonemes, syllables, morphomes, and words with high accuracy. They have confirmed in their paper study that the implementation mention above can be apply into different research purposes. For the researchers to obtain reliable statistics for Uyghur language on different units, they have collected number of text and speech corpora as for the understanding corpus compiling process for Uyghur language. Although they may seem to understand and mentioned that Morpheme unit provides a small lexicon and better statistical properties, they also concluded

that it can provide a good foundation for data processing of natural language processing applications and finally the research result will set a good example for any research seekers as a resource reference languages that has the same class of morphology.

Sembok, Abu Ata, Abu Bakar (2011) This research paper focuses on stemming which the researches used as the main information retrieval system in reducing many kind of word forms and provide common roots for the enhancement of retrieval effectiveness. The paper also mention the need for an effective stemming algorithm for the indexing and retrieval of Arabic documents which is require in developing the Al-Omari in studying and establishing a new versions proposed for the performance advancement. In their tests for implementing the algorithm, they have proven that the stemming algorithm has successfully performed better than Al-Omari's algorithm and also, they have shown on their experiments an error resulting from stemming and may consider ways in changes for best applied in further research.

Rober and Gersam (2008) the researchers presented a study that Filipino verbs conjugated many different forms as inputs and analyzes them to produce the affixes used, the infinitive form, and the tenses of the original input verbs. The changes and removal of the letters for several verb form depending on the category has a corresponding algorithm. The analyzer also serves as category determiner for the right infinitive form whenever it is found. During the test, they have undergone almost one thousand Filipino verbs conjugated in three different tenses and some verbs used in everyday life. They have found out a less error for determining the tense and it is mainly because of irregular verb forms. In their conclusion, the high accuracy rate initial results is expected and

already implemented. They have proposed a future work inclusion on lexicon in an a more advanced on obtaining an appropriate meaning enhance and support context-driven machine for a more better translation system.

2.3 String Matching

String-matching is a very important subject in the wider domain of text processing. String-matching algorithms are basic components used in implementations of practical softwares existing under most operating systems. Moreover, they emphasize programming methods that serve as paradigms in other fields of computer science (system or software design). Finally, they also play an important role in theoretical computer science by providing challenging problems. Although data are memorized in various ways, text remains the main form to exchange information. This is particularly evident in literature or linguistics where data are composed of huge corpus and dictionaries. This apply as well to computer science where a large amount of data are stored in linear files. And this is also the case, for instance, in molecular biology because biological molecules can often be approximated as sequences of nucleotides or amino acids. Furthermore, the quantity of available data in these fields tend to double every eighteen months. This is the reason why algorithms should be efficient even if the speed and capacity of storage of computers increase regularly. String-matching consists in finding one, or more generally, all the occurrences of a string (more generally called a *pattern*) in a *text*. All the algorithms in this book output all occurrences of the pattern in the text. The pattern is denoted by x=x[0 .. m-1]; its length is equal to m. The text is denoted by y=y[0 .. n-1]; its length is equal to n.

Both strings are build over a finite set of character called an *alphabet* denoted by **Σ**with size is equal to σ ("Introduction" 2016).

Vidya, Prof. Akhtar and Dr. Nilay (2012) in this paper, the researchers define string as an important class of string algorithms that try to find a place where one or several strings are found within a larger string or text. It is a classical problem in computer science and a convenient problem solver for string matching issues on different kind of application designed for pattern based algorithm. The experiment on this paper mainly focuses on the importance of memory and time efficiency when it comes to using computers for excellency in performance. Specifically, exact and approximate string matching algorithm composes a several problems when it comes to solvable condition for the efficient performance.

Rajashekharaiah, MadhuBabu, and Raju (2012) emphasized that, String matching is very important subject in the domain of text processing and it has been one of the most extensive problems in computer technologies during past two decades. It has applications such as DNA analysis, information retrieval systems and several other fields. String matching is the process of finding the occurrence of a pattern P in a text T, where T is longer than P. This occurrence is either exactly matched or partially matched with the pattern, based on this; string matching algorithms are classified as Exact and Approximate string matching algorithms. Exact string matching algorithms are concerned with the number of occurrences of the pattern into a given text, while approximate string matching algorithms are concerned with the similarity percentage between the pattern and

the text or any part of the text.

Gong-shen and Sheng-jun (2009) pointed out in their study that about analyzing the factors that effects the algorithm for the time complexity. As they have performed and discussed their experiments as well for the proof purposes, they have figured out a best one when processing seven kinds of pattern in their paper. The smallest patter makes it a useful preference in text retrieval as they have confirmed in one of their experimental data result. In their paper, they have figured out and listed the factors that originally proposed throughout the paper and concluded that it was very useful for many application as the study for any related field will find it difficult for establishing new algorithm such as introducing a new searching method.

Yue, QingShi, Li & PeiFeng (2011) in their study Giant complete automaton for uncertain multiple string matching and its high speed construction algorithm, researchers put into an experiment the recognizing large numbers of strings with U-, V-, and U-V-uncertain-strings, and carefully unify large number of possible essential data in identifying any cause of harming the information. In proposing a complete automaton and its high-speed construction for the said algorithm in a large scale strings intercrossing multiple strings and uncertain strings, the whole system may set some regular expression that are not eliminated to a special point. The experiment and their test result for the maximum of the parallel in completing the automaton of the string and the methods for eliminating the intersection of the two sets of string are likely on a high comparison for permitting the methods on elimination during the process.

Soni, Vyas, and Sinhal (2014) in the Importance of String Matching in Real World Problems, researchers explained that the String Matching as the classical and existing problem, despite the fact that the real world aspects belonging to the research field of computer science. In their String Matching applications they have described a few applications for the real world problems. They have mentioned several study from the past that implement contraction on automata for pattern based and any related algorithm that also used each multiple patters. They have implemented a methodologies from the past that described the same concern. They have confirmed based on their discussion in the Information Retrieval that string matching plays very vital role in their study like as information extraction, topic tracking, question answering and so on. The text mining design as the main destination for their study, analysing large qualities of text. In their conclusion, they finally pointed out that string matching has a large proportion in the area of any innovative searching algorithm and it is expected to be very demanding for any application system on the future perspective.

Nsira and Ellouni (2017) in this research, the On-line String Matching in Highly Similar DNA Sequences, they consider the problem of on-line exact string matching of a pattern considering a highly similar sequences. The study is applicable in the areas where indexing in a particular related order for convenience feasibility. They have presented a two classical algorithms that is being adapted in considering to deal with the errors. In their experiments they have undergone tests, computing 100 times for each pattern length on average searching time. They have attempted to demonstrate that from a certain number of sequences, it is more efficient to use their solution than a classical exact string

matching solution. Finally, they have concluded that the searching time complexity of our algorithm depends on the size of the reference text. They also have pointed out that further research directions include to overcome this limitation in a certain gap between consecutive variations.

CHAPTER III

RESEARCH METHODOLOGY

In this chapter, research methodology on analyzing the processes will be discussed in the next section to elaborate the architectural system model to identify stages and necessary information for the target purpose on building an application for Meranao Translator using String Matching Algorithm.

3.1 Data Sources and Collection

The Term "Maranao", "Meranao", or "Mranao" derives from the word *ranao* which means "lake" Hence, "Meranao" means "people of the lake". The traditional home of the Meranao is the area surrounding Lake Lanao, the second largest in the Philippine archipelago next to Laguna de Bay in Luzon. Located in Lanao del Sur near the border of Lanao del Norte. The Meranao language spoken in the Lanao provinces is part of a subgroup of languages called the "Danao langauges".

As a native Meranao Speaker, a personal self-involvement regarding the Meranao verbs as the subject matter for this research is presented. Consultation with human expert on Meranao Language is also considered for word afflictions and authentication.

Data collection for this study as the Meranao Verbs are collectively taken from selected traditional paper dictionary and electronic dictionaries such as ebooks and developed websites.

Following are direct sources of the word collected required for the database corpus:

a. *A Maranao Dictionary* by Howard P. McKaughan and Batua Al-Macaraya, published 1996. Available source of e-book at http://www-01.sil.org/asia/Philippines/online-mrw.html and https://eric.ed.gov/?id=ED013450 for the entire book review. A Sample content of this source is shown in Figure 3.1.

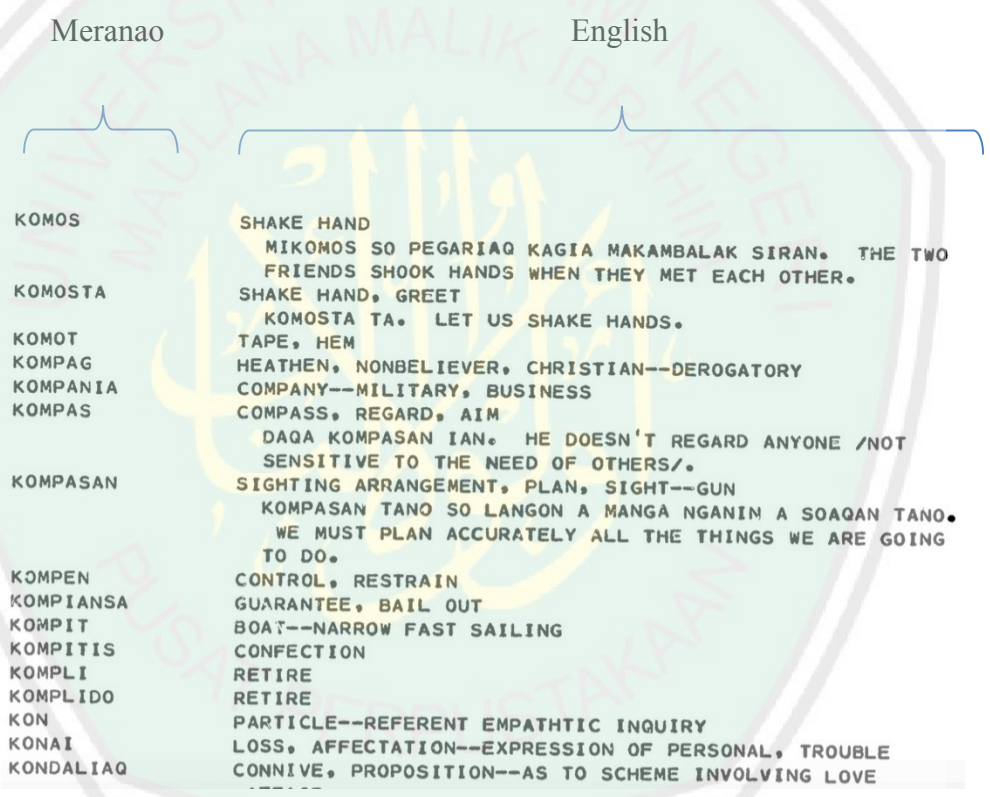

Figure 3.1 Screenshot from an E-book *A Meranao Dictionary*

Figure 3.1 A screenshot display taken from 'A Maranao Dictionary' e-book showing how listings of words in Meranao and its English meaning are put order in all capital form.

b. SEAlang's Maranao Online Dictionary

SEAlang's Maranao resources are based on three primary sources: the database underlying A Maranao Dictionary by Howard McKaughan and Batua Al-Macaraya conducted at De La Salle University and the Summer Institute of Linguistics, Manila, using electronic files supplied by Prof. McKaughan. Website access at http://www.sealang.net/maranao/dictionary.htm

c. MODERN ENGLISH-FILIPINO-MARANAO DICTIONARY by Mario "Guese" Tungol, Sari-manok publisher as supplementary source for incorporate further Meranao Words.

In the lists are additional information on understanding Meranao Language based on previews studies done by Language experts and Meranao Language researchers and has achieved consensus of the following set of rules:

- The high front sound will be represented by ***i*** as in *l**i**ma*** (five);
- The high back sound will be represented by ***o*** as in *s**o**lotan*** (sultan);
- The low central vowel sound will be represented by ***a*** as in *t**a**bas*;
- The ***i*** sound next to another vowel shall be written as ***iy***, if ***i*** comes under the two conditions: With words as in ***iy**an*** (him or her). When ***i*** sound comes second as *ma**i**to*** (small) it shall be as ***i***; except at the end of the word when then it shall be written ***y*** as in *wala**y*** (house); As added, ***i*** will be in the initial position in the word not ***y*** except in such borrowed words as in ***y**ahodi*, and ***y**o yo*;
- The ***o*** and ***w*** sound comes under the two conditions: Both ***o*** and ***w*** should be written in words which ***o*** precedes ***w*** as in *b**ow**is*** (tax). If ***o*** comes second after ***w*** in the medial position in words, the ***o*** alone should be

written and the ***w*** is discarded as in *ma**o**lad* (wide); As the recent update from Wikipedia, ***w*** should be used for terminal part of the words whenever appropriate except for improper names, otherwise as a common practice as in *ta**w*** (person);

- Single quotation mark (**'**) an apostrophe represents a glottal stop as in *maito'* (small), *wata'* (child);
- ***e*** should represent the simple unstressed pepet sound whenever it occurs as in *b**e**l* (smoke), *g**e**dam* (feeling). The consonant ***y*** and ***w*** should be introduced between vowel sound whenever appropriate as *i**y**op* (blow), and *o**w**ay* (yes);
- The special combination ***mb*** and ***nd*** should be combined as in ***mb**ama"* (to chew betel nut), ***mb**anog* (to find), ***nd**o"* (air), and ***nd**itar* (to wear);
- There is no silent letters in Meranao words as there in English;
- The long back vowel ***u*** differs from ***o***; signifies future tense as in *s**u**l**u**ten* (to wear) as added by Wikipedia; and
- ***h*** should represent immediate future tense when stressed sound as in mathay (too long), phatay (to die), and khan (to eat) as added very recent in Meranao language study.

Unlike the English language, the Meranao languge is rich in affixations to inflect a given stem. There are 16 types of affixations considered in the study, which are illustrated in the table 3.2 below. The inflection symbols represent the position of the affixes in the inflected word. The stem is represented by the symbol R, while the prefix, infix and suffix are represented by x, y and z, respectively.

3.3 System Design

The Meranao Translator system design application is mainly focusing on the accuracy of the translation of the word Meranao into English. Considering the String Matching as the algorithm implemented in the system, the process of this particular technique is assumed as the qualified equipment in achieving the system main goal.

The general overview of the flow of the system design is going to be illustrated in Figure 3.2. The system design is to build a Meranao Translator that will accurately translate a data acquired from the user into English meaning. The illustration on Figure 3.2. shows the step by step sequence on how the system periodically take the progress activity on acquiring the main task assigned for the attainment of the system design. The System phase consists of the *input Text Collection*, *Tokenization*, *String Matching*, *Fetching* and *Conversions of Data*, *joining of extracted Data* and then the *Result Output*.

The flow of the system particular order performs at acquiring a data from the user into *Text Collection*, then *Tokenization* takes place for the preprocessing series, then the output tokens will then perform *Data Formation into String Array*. On the *String Matching* process, the string elements will fetch the data on the database and turns back to matching before the next process will occur. If it matches the data, it will now perform the *Data Conversion into Cell Array,* which is in a form of an extracted elements and so *Combining/Joining Extracted Data* will now do its task before finalizing the *Result Output*.

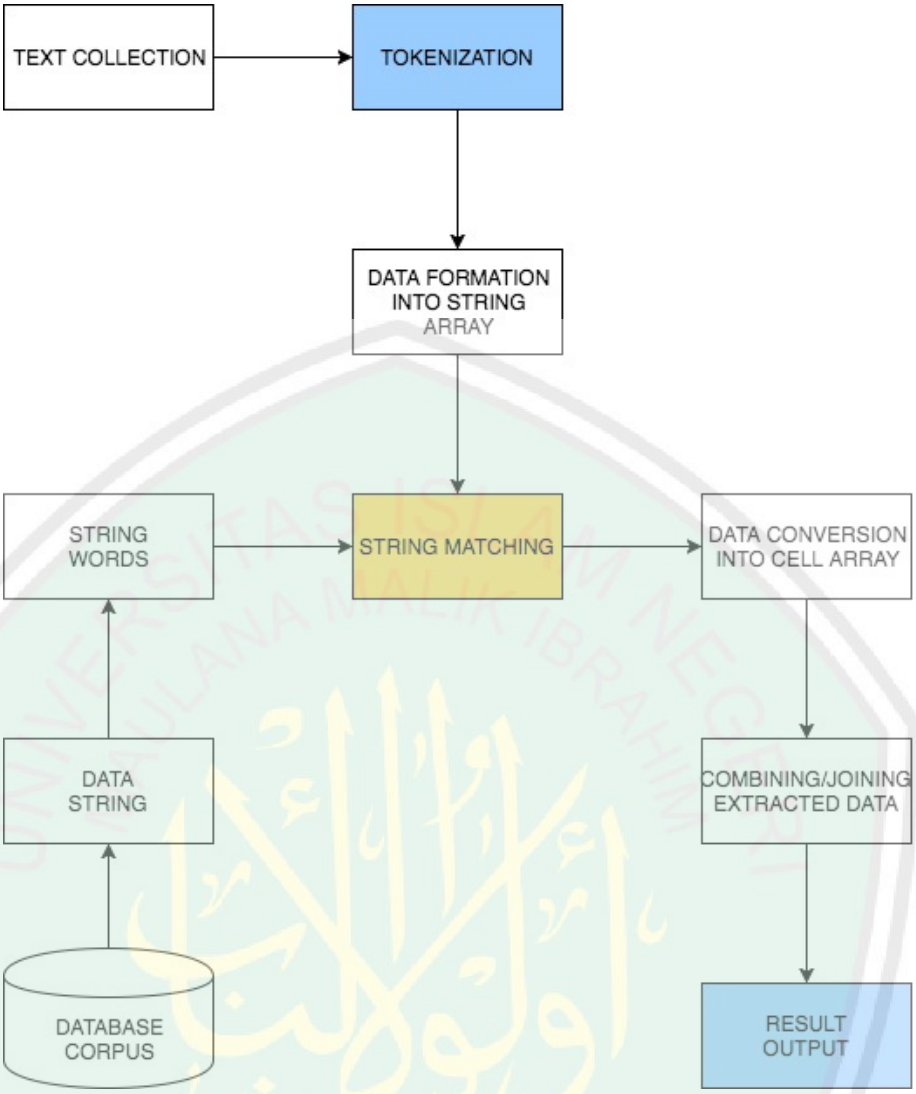

Figure 3.2 System Design

On the color coded *Tokenization*, *String Matching*, and *Result Output* are the three(3) main points in the system design shown in Figure 3.2. which is going to be the main areas of this section. Further details about the sequence of the general system design is emphasized in the following subsections.

3.3.1 Tokenization

In this process, input text of the words to be translated occurs in preprocessing before any steps for the translation. Naturally, before any real text processing is to be done, text needs to be segmented into linguistic units such as words, punctuation, numbers, alpha-numerics, etc. This process is called tokenization.

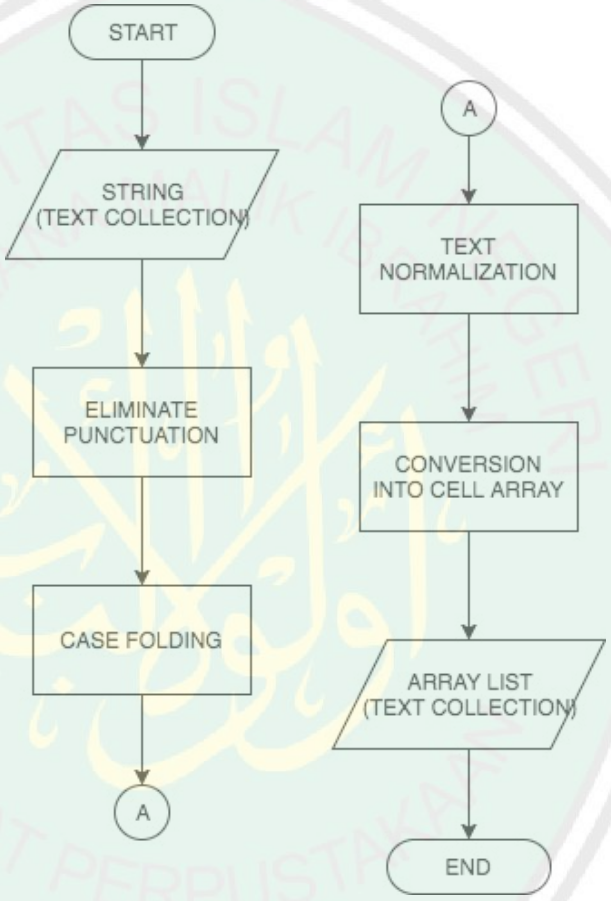

Figure 3.3 Flowchart on Input Text and Tokenizing

In Figure 3.3, After the collection of text has been entered, the next process is to *eliminate* the *punctuation* of the text then convert capital letters into lowercase for *case folding* then proceed to *text normalization* for *converting text* data to *cell array* and then collect the tokenized text into an array list before the

next process. In the Figure 3.4 shown below is a screenshot of the code listing for the Tokenization process.

```
str = get(handles.edit1,'String');

% Erase the punctuation
a = erasePunctuation(str);

% Convert to lowercase
b = lower(a);

% Tokenize the text.
documents = tokenizedDocument(b);
```

Figure 3.4 Coding for the Tokenization Process

Figure 3.4, a screenshot figure of the code program for the text tokenization within a Matlab platform.

In the last line of the Figure 3.4, the code "`documents = tokenizedDocument (b);` " indicates an array of tokenized document, which is later set as a representation of collected token words for later text analysis. This is also the basic function of the MATLAB platform for featuring toolboxes on word preprocessing text data for more precise and short coding and put it together in specified syntaxes.

3.3.2 Rule-based

Considering as one of the effective ways in any complicated decision making systems, this study is unable to implement this approach due to insufficiency of complete required data for the advancement of the system level of complexity.

As for the relevant basis for the standard system design, we arguably provide the information necessarily for the assurance of literature based on what we have conducted in the effort of providing essential material in the case of this study.

The Rule-based Databases consist of two phase, the dictionary and rule tables. They are primarily filled with defined rules and lexicon under the Meranao language. As the system goes, the databases can learn new rules and words as the user approves. It can learn a new similar language by mounting the language's lexicon and rules to the Rule-based Database.

a. The Dictionary database contains three tables: *roots*, *candidate stem* and *language*. The roots table contains only a list of Meranao root words (verbs) that literally contains forms of affixes by which the algorithm may read as affix and eliminate it during stemming the affixes. This in effect, decreases the size of the dictionary because only a few stems are being stored. Also, it is intended to avoid over stemming a word. For instance the root word "*inom*" (drink) is stored in the roots table because it is formed by the syllables *in-* and *om-*, which can be found in the affix tables. If no restriction is specified for this kind of words, this may lead to stemming the prefix *in-* (PT marker) and suffix *-om-* (SUBJ marker).

Another table in the dictionary, the *candidate stem*, stores the candidate stems which contain forms of affixes that the algorithm removes during the stemming process but in the end results to over stemming. For example, the word *malang* (block), from the stem *alang*, constitute an entry to the candidate stems because it literally contains the prefix *ma-* and prefix *m-* (the correct prefix of this word). In the affix stemmer, the longer prefix *ma-* will be removed instead of the correct

prefix, *m*-. This will result to an invalid stem *lang* once *ma-* is removed. An entry to the candidate stem table is of the form:

<candidate_stem>/<stem>/<prefix>/<infix>/<suffix>/<language>

The backslash operator represents the column separator in the database. The prefix, infix and suffix attribute state the corresponding affixes that are attached to the stem to form the candidate stem. The language attribute indicates to which language the candidate stem belongs. For the example word *alang*, it is stored in the table as

<malang>/<alang>/<m>/<NULL>/<NULL>/<Meranao>

This way, the word *malang* is now determined by the algorithm to have the prefix *m-* instead of *ma-*. The language must be specified for every word in the dictionary. Currently, the language specified in the database is Meranao, which is stored in the language table. As stated earlier, the database can be added with a new similar language to enhance the system's capabilities.

b. The Rule Tables play an important role in this study. It contains the tables which store the rules from which the algorithm refers to in order to analyze correctly the morphology of Meranao verbs. These rules can later on be expanded as per the user decides. Adding new rules make the system more capable and reliable with no bad effect to the algorithm's performance.

The tables that comprise the database of Rule Tables include the affix tables (prefix, infix, and suffix) and the inflections table. Each of the affix tables stores the affixes and their corresponding tense and voice. An entry in each table follows the format:

<affix>/<tense>/<voice>/<language>

The *affix* represents any defined affix, e.g., *pe*, *in*, *om*, etc. It comes with the *tense* and *voice* attributes that mark the grammatical function of the affix indicated. Generally, these attributes do not come together in a row because one affix either marks the tense or voice of a word. That is, when the tense has a value, the voice is null, and vice versa. The language must be specified for every item in the tables, e.g., *Meranao*. This is because new rules can be added for a new language. To illustrate the contents of the prefix table, refer to Table 3.1. The table shows that the tense and voice attributes do not come together.

Table 3.1 Sample entries in the Prefix Table

Affix	Tense	Voice	Language
pe	PPT		Meranao
in	PT		Meranao
om		SUBJ	Meranao
i		INS	Meranao

Another table that makes up the rule tables is the inflections tables. This table contains the tense, voice, category and affixes correspondence. This table is deliberately designed in order to support the dynamism of the whole system and its future expansion to adopt new rules in inflecting a stem in different language aside from Meranao. Items in the inflections table follow the format:

<tense>/<voice>/<category>/<prefix>/<infix>/<suffix>/<initial>/<language>

The *tense* value can be any from {NT, PT, IFT, and PPT}; whereas *voice* can be any from {REF, OBJ, INS, and SUBJ}. The *category* determines the type of inflection, such as voice *(V)*, aptative *(A)*, or causative with ki- *(C)*. In every

category, there is a 4x4 correspondence between the tense and voice. That is, every tense has four voice inflections. This means that a single stem can have 3 x (4x4) inflections generated, that is 48 in total. These items can be null to indicate zero inflection in some cases. The sample entries in the inflections table are listed in Table 3.2.

Table 3.2 Sample Entries in the Inflection Table

Tense	Voice	Category	Prefix	Infix	Suffix	Initial	language
NT	REF	V	Null	Null	An	Null	Meranao
PT	OBJ	V	Null	I	An	V	Meranao
PT	OBJ	V	Null	In	Null	C	Meranao
IFT	INS	A	Ki	null	Null	Null	Meranao
NT	SUJB	C	Maki	Null	Null	Null	Meranao

3.3.3 String Matching

String Matching in this process has the main role in the system design operation. We put the string matching as the algorithm for the system as it is particularly appropriate for basic text mining systems. The process of String Matching is finding the equivalent set given in a program assuming the approximate result will provide the expected output during the process of probation.

In this process, We Assume that the text is an array. T[1..n] of length n and that the patters is an array of length[1..m] of length m and that m<=n. The character arrays T and P are often called strings of characters.

We say that pattern P occurs with shift s in text T (or equivalently that the pattern P occurs beginning at position s+1 in text T) if 0<=s<=n-m and T[s+1….s+m]=P[1..m] if P occurs with shift s in T then we calls a valid shift otherwise we call s an invalid shift. The string matching algorithm is the problem of finding all valid shift with which a pattern P occurs in given text.

String Matching on Matlab Function Reference later on the subsection of this segment emphasize the direct keyword on finding the match string based on Matlab syntaxes for the logical clarity of each step within the flowchart.

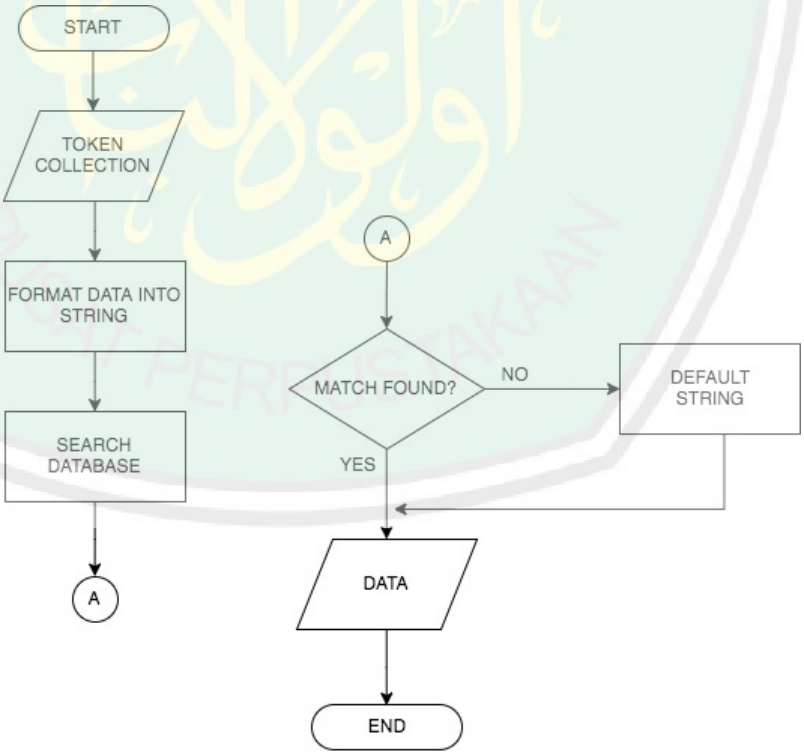

Figure 3.5 Flowchart for String Matching Process

In Figure 3.5, the tokenized text collection 'TOKEN COLLECTION' is being formatted into array of data string 'FORMAT DATA INTO STRING' for the search execution on 'SEARCH DATABASE'. Then, if the element is 'YES' on 'MATCH FOUND' then it will proceed to the next operation. If it is 'NO' then the entered string will display the same as the default entered text 'DEFAULT STRING'. The rules apply when the string entered has matched the exact same spelling from the database corpus provided.

3.3.3.1. String Matching on Matlab Function Reference

On the process of finding the match word, we have illustrated a flowchart on string matching based on MATLAB command shown in Figure 3.6.

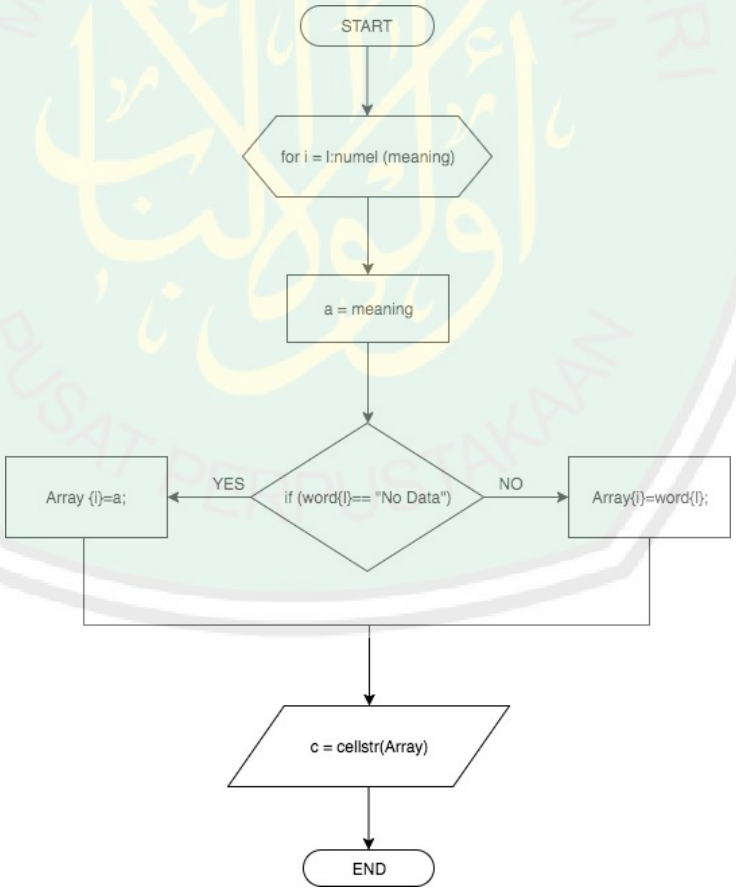

Figure 3.6 Flowchart on String Matching based on MATLAB syntaxes.

The flowchart started on declaring the function reference with the syntax for `i = 1:numel (meaning)` and proceed to the casting the `a = meaning` for the one-on-one ratio match. Then if the `word{1} == "No Data"` which is TRUE, then it will indicate the `array {i}` as the meaning on a, if FALSE, then it will print out the same as entered. The last flow of the process will declare all the match into a cell array for later output.

A screenshot of the code listing for converting document process is shown in Figure 3.7 and Figure 3.8.

```
%convert to cell array
C = doc2cell(documents);
meaning = C{1};

%Show the result to WordList
set(handles.listbox1, 'String', meaning);

% connect to database
dbname = 'meranao';
username = 'root';
password = 'root';

conn = database(dbname, username, password);

Array = {};

% initialize your matrix/vector
for i = 1:numel(meaning)

a = meaning(i);
```

Figure 3.7 A Screenshot Coding Image for Array Conversion Process

Figure 3.7, a screenshot figure of the code program for Array Conversion Process within a MATLAB platform. The database connection is strictly required to operate the array listing for the matching process. The data type as they have only one kind of character will put all the elements in the same field of array.

The initialization code iteration on above Figure 3.7 "`for i = 1:numel(meaning)`" is a number of array elements that returns in the '`(meaning)`' declaration inside this syntax which later, on syntax '`a = meaning(i)`' will be collected for data extraction.

```
% Matching word

sqlquery = sprintf('SELECT meaning FROM translate WHERE word LIKE "%s"',a);

curs = exec(conn,sqlquery);

curs = fetch(curs,1);
word = curs.Data;
curs.Data


if (word{1} == "No Data")
  Array{i} = a;

else
  Array{i} = word{1};
end

end

c = cellstr(Array);

close(curs)
close(conn)
```

Figure 3.8 A Screenshot Coding Image for Executing Word into Cell Array

Figure 3.8, a screenshot figure of the code program executing word into cell array within a Matlab platform. In this code listings, matching the string on converting data into text from any format, the operation on conversion functions is taken in the keyword '`sprintf`' as displayed in Figure 3.8. It controls the operation on notation, alignment, significant digits of the string specified parameter for text display and for output files.

3.3.4 Result Output

To finalize the translation process, the input array proceeds to construct the text together by joining each element on an array that has been matched from the previous segment. The Database section of the System Design specifies the matching process provided by the corpus collection to be search upon with the string values stored as a reference for text translation.

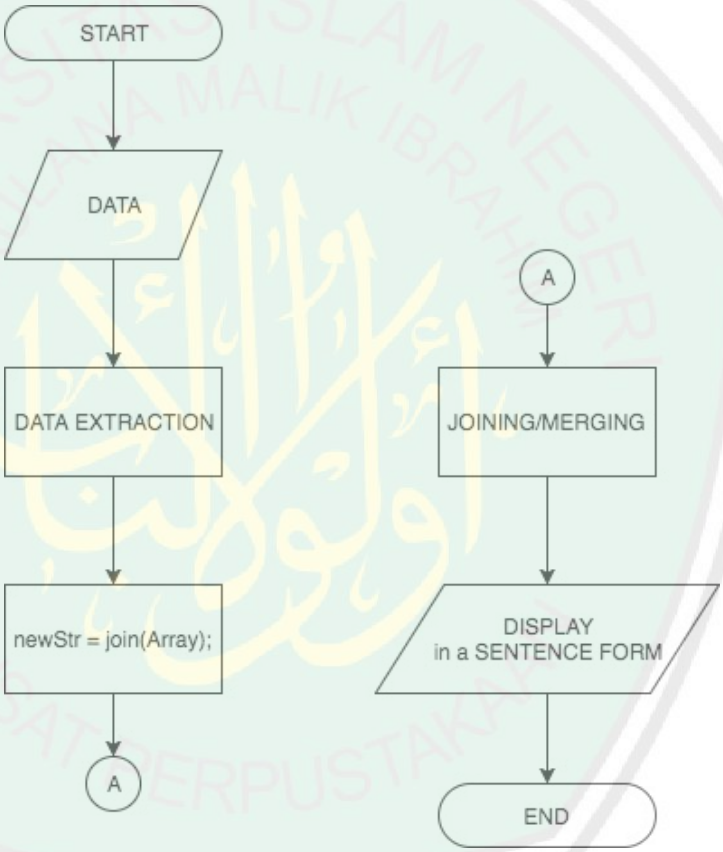

Figure 3.9 Flowchart on Translation Process

In Figure 3.9, a flowchart on input the array text taken from matching process then enjoined for the translation output result which is shown in Figure 3.10 as a screenshot code listing for the translation process.

```
set(handles.listbox3, 'String', c);

% change the array to sentence
newStr = join(c);

set(handles.edit2,'String', newStr);
```

Figure 3.10 A Screenshot Image on Joining an Extracted String

Figure 3.10, a screenshot figure of the code program for joining an extracted string within a Matlab platform. In this syntax '`newStr = join(c)`' is the process where input array combines the text by joining consecutive elements placing space character between them. The final translation displays as the texts combined together in one line as a sentence.

CHAPTER IV

EXPERIMENT AND DISCUSSION

This chapter will present the implementation and result based on the theoretical concept from previews chapter. The implementation stands as an evaluation for testing the accuracy of the application to obtain the anticipated objective result.

4.1 EXPERIMENTAL PROCEDURE

Steps to accomplish the objective result for this study emphasizes the main core on conducting the experiment which is the performance on how accurate is the designed system after the implementation of the method being used in this study. The experimental procedure as illustrated in Figure 4.1 shows a precise particular order on describing an Input-Process-Output pattern for the basic structure of the system operation.

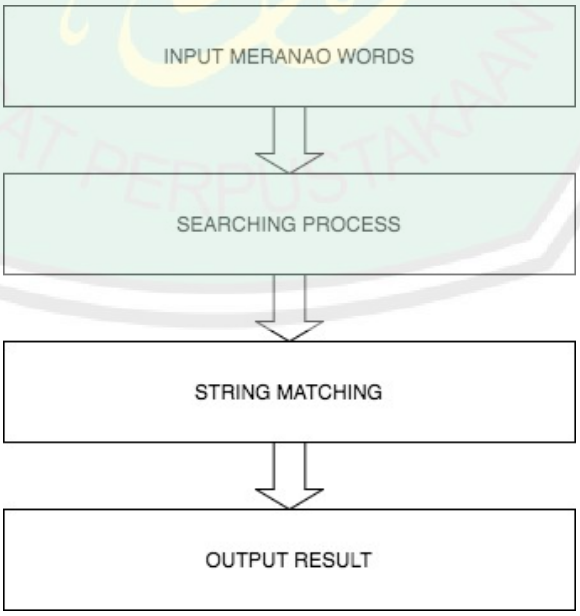

Figure 4.1.1 Experimental Procedure in Block Diagram

Figure 4.1.1, outlined an Input-Process-Output pattern starting with *INPUT MERANAO WORD*, then *SEARCHING PROCESS*, then *STRING MATCHING*, and ends at *OUTPUT RESULT*.

To explain how the procedure is operating along the process, first, we run the application Meranao Translator in Matlab Platform. Second, the user enters Meranao text into the 'input panel' of the window which is shown in Figure 4.1.2.

Figure 4.1.2 A Screenshot of the Meranao Input Panel

Figure 4.1.2 is a partial screenshot taken from the actual Meranao Translator Application interface during the experimental action.

Third, the translation results applied if and only if the given strings match the spelling that is available on the database corpus. Finally, the result of the translation of the text will be executed after String matching is performed throughout the process from the reference corpus then it will display the translation result on the 'output panel' of the window which is shown in Figure 4.1.3.

Figure 4.1.3 A Screenshot of the English Translation Output Panel

Figure 4.1.3 is a partial screenshot taken from the actual Meranao Translator Application interface during the experimental action.

Every word collected as a token proceed to searching for the equivalent string on the Database Corpus and then construct the lists of array into a compound group as a sentence and display the output.

Experimental tests are undertaken in a random Meranao words, starting from single vocabulary, to phrase and sentences. Notice the gradual formation we have conducted to exhibit the increase of the complexity and discover the result as we continue in on ward.

Lists of Meranao text given in Table 4.1, 4.2, 4.3 are provided to test the Meranao Translator accuracy in this section.

Table 4.1 Meranao Vocabulary

No.	Meranao Text
1.	*Seka*
2.	*Ama*
3.	*Ina*
4.	*Pagari*
5.	*Walay*
6.	*Libro*
7.	*Ontoda*
8.	*Layok*
9.	*Ogop*
10.	*Siyaw*
11.	*Opisina*
12.	*Mababa*
13.	*Kaulit*
14.	*Pagti*
15.	*Maputi*
16.	*Maitem*

17.	*Singa*
18.	*Guraok*
19.	*Turog*
20.	Pagnaw

Table 4.2 Meranao Phrase

No.	Meranao Text
1.	*Mapiya kapipita*
2.	*Marata gawii*
3.	*Isa milion*
4.	*Sakto arga*
5.	*Mapasang wata*
6.	*Mamis dolsi*
7.	*Masakit tiyan*
8.	*Seka sabot*
9.	*Tanto malagod*
10.	*Madakel tao*
11.	*Sumiyong sii*
12.	*Tolad ayam*
13.	*Magastos talompa*
14.	*Matenggaw ig*
15.	*Matanog lagam*
16.	*Marata ontong*
17.	*Maloya khakan*
18.	*Makapal libro*
19.	*Tanto diyokawan*
20.	*Maitem kapi*

Table 4.3 Meranao Sentence (1)

No.	Meranao Text
1.	*Saken somiyong sa siti*
2.	*Gyoto na iportante*
3.	*Saken miniga sa kuwarto*
4.	*Dimbantay salida magagawii*
5.	*Saken galebek sii*
6.	*Pirito sda na maregen*
7.	*Permi pkhada so layit*
8.	*Dinggitagita sa liyo*
9.	*Miyakawma sa walay*
10.	*Phamakineg sa radio*
11.	*Inoto masosowa san?*
12.	*Mapita na isnin*
13.	*Muuntod sa park*
14.	*Ilayanga so kalindar*
15.	*Saken madakel piyamasa bangala*
16.	*Sapeng so khakan sa lamisaan*
17.	*Madakel problema na di khasoleb*
18.	*Mbida siran sa pamikiran*
19.	*So turog na kailangan*
20.	*Skaniyan miyakapas sa iksam*

Table 4.4 Meranao Sentence (2)

No.	Meranao Text
1.	*Saken piyamasa madakel bangkala*
2.	*Sapeng so kan sa lamisaan*
3.	*Anda so ipegokit sa walay i Daud*
4.	*siran sumiyong sa padian.*

5.	*Ngkainoto, ba aden a pagangin ngka?*
6.	*Pemisita ko sa ginawai aken*
7.	*Oway, ba ka san song. Na ilay ka bo om*
8.	*Anda so manga tao sa giyai walay?*
9.	*Soled ka, bolos ka.*
10.	*Kolor biro na iyan baya ago kolor gadong na di niyan baya*
11.	*Madakel so problema na di khasoleb igira da pirak*
12.	*Sabagi o siran di sabot ka*
13.	*Skaniyan miyakapas sa iksam ago skaniyan guraok*
14.	*Mbida siran sa pamikiran ago di siran pagayon*
15.	*Miyakawma sa walay ago da aken mailay*
16.	*Sa liyo na madakel a tao imanto pesagad*
17.	*Saken tareg batiya goani saken torog*
18.	*Gyoto iportante karatas na madakel a miyabnasaon*
19.	*Ilayanga so kalindar aden mariga gawii*
20.	*Dimbantay kambola magagawii ago khan madakel khakan*

For the advance experimentation on this application, we have taken a paragraph form of Meranao Language from a website entitled “kataro sa lalag” which means “speech” representing a short classical way of giving speech in Meranao.

Ayakaposan ang kai a lalag ko na panalamatan aken ang kai a mbala a mindiamonga sa dairan kapakandarainon sa kiya paka tokawa iran rekami ago giya darodopa aken sangkai a kalilimod, Na pamangni ako rekanopen sa maaf oba aden a dako katarotopi ko kiyapayag o sariat o lalag ko sabap sa kangodaan akopen a gopen gii sinanad, sa ayamala na sobo so kadnan I daa pawing iyan. Wassalam

Figure 4.1.4 Meranao Text in a Paragraph Form

Figure 4.1, is a collection of text taken from a blog website with the URL https://acmaly88.page.tl/MANGA-KAPAGONGANGUN.htm

1. HAMDALA
2. PAMEKASAN
3. SALAM
4. KAPANABIA
5. KAPANGO-NGODASAN
6. MOAYAN O LALAG (purpose of the speech)
7. PAHALA O LALAG (Importance of speech)
8. KAPOSAN O LALAG

Figure 4.1.5 Meranao Text in Listing Words

Figure 4.1.5 an additional collection of text taken from a blog website with the URL https://acmaly88.page.tl/MANGA-KAPAGONGANGUN.htm for further demonstration of how the Meranao Translator operates.

During the experiment procedure, the scarcity of the corpus in providing the available words on the database, occasionally fail to find a match word which is non existence in the database and so we provide an addition feature for the quick fix on immediate insertion of the word. We added a panel called 'Input New Word' which is shown in Figure 4.1.6.

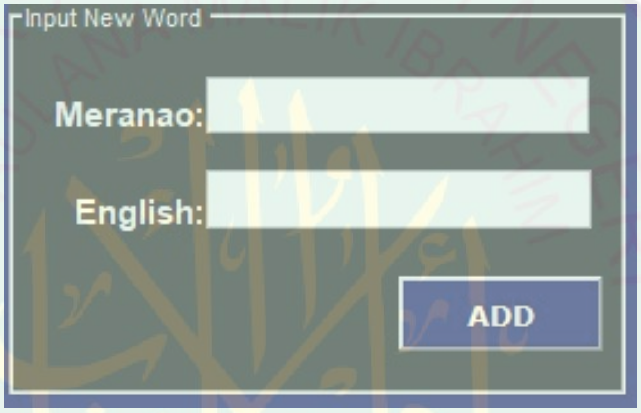

Figure 4.1.6 A Screenshot of the Input New Word Panel

In the Image 4.1.6, adding this feature on the application will enhance the developer to update the number of corpus in the database. Supposing that the database corpus is necessarily be updated and needing a further data entry.

4.1.1 Calculation of the Accuracy Evaluation

Experiment tests are conducted by evaluating the results of the translation based on the output data available in the corpus. Tests carried out on a number of found translation matching the result on a default data source such as the SEAlang

Library Maranaw and verify the value result. The calculation of the accuracy for the tests is in the following formula.

$$Accuracy = \frac{Sum\ Total\ of\ the\ Found\ Predicted\ Match}{Sum\ Total\ of\ the\ Overall\ Testing\ Result}\ x\ 100\%$$

4.2 EXPERIMENTAL RESULT

The Meranao Translator has been tested based on the number of the string operation execution to be translated. The collection of the texts are randomly chose based on common Meranao words used on daily communicational situation.

In every experimental test, each table contain 'Meranao Text' column, 'Translation in English by the Corpus' column, 'English Translation by SEAlang Library Maranaw' column the 'Match Result' column which indicates a 'green' shade if the result matches the word searched from the SEAlang Library Maranaw into the word available on Corpus. Measurement of the accuracy calculation are labeled in the description following each table experiment tests.

Observations and evaluations of the result are based on hands on procedure in achieving a reliable and independent experiment of actions.

Experiment Test I: Demonstration Test

In this test, we have undergone a demonstration procedure that input a single word and expecting a single word result from the corpus as well. In the Table 4.4, we match the result found by the corpus and *SEAlang Library Maranao* available online and see if we get the same translation.

Table 4.5 Meranao Translation on Vocabulary

No.	Meranao Text	Translation in English By the Corpus	English Translation by SEAlang Library Maranao
1.	*Seka*	You	You; thee; thou
2.	*Ama*	Father	father
3.	*Ina*	Mother	mother
4.	*Pagari*	Sister	Sister; Brother
5.	*Walay*	House	House; building
6.	*Libro*	Book	book
7.	*Ontoda*	Chair	Seat; chair; bench
8.	*Layok*	Friend	Friend; chum; pal
9.	*Ogop*	help	Help; service
10.	*Siyaw*	nine	Niine
11.	*Opisina*	Office	Office
12.	*Mababa*	Short	Short; low
13.	*Kaulit*	Fat	Fat(kaolit)
14.	*Pagti*	Thin	Thin(to be) lean; slender
15.	*Maputi*	White	White(mapoti)
16.	*Maitem*	Black	Black
17.	*Singa*	Smile	Laugh
18.	*Guraok*	Cry	Cry(goraok)
19.	*Turog*	Sleep	Sleep(torog)
20.	*Pagnaw*	Awake	Awake

Table 4.5 is a table demonstration on how the application delivers the string to string effectiveness if the case is 1 to 1 string searching within a procedure. This sample is not going to be included in the evaluation result as it is only to showcase the non complexity side of the application system.

Experiment Test II: Phrases

In this test, we have undergone a Phrases through a test for translating the Meranao Text based on the corpus and see if the translation has an equivalent match at the *SEAlang Library Maranaw*. In table 4.6 the operation for finding if the translation has the equivalent result considering the number of words in one phrase. We declare the matching rate of the result as 'not found' if one of the word did not match or non existing in the source provided.

Table 4.6 Meranao Translation on Phrases

No.	Meranao Text	Translation in English By the Corpus	English Translation by SEAlang Library Maranao	Match Result
1.	*Mapiya kapipita*	Good morning	Good; morning	Found
2.	*Marata gawii*	Bad day	bad; day	Found
3.	*Isa milion*	One million	One; million	Found
4.	*Sakto arga*	Exact cost	exact; cost	Found
5.	*Mapasang wata*	Smart child	Expert; child	Not Found
6.	*Mamis dolsi*	Sweet candy	Desert; candy	Not

				Found
7.	*Masakit tiyan*	Painful stomach	Painful; stomach	Found
8.	*Seka sabot*	You understand	You; understand	Found
9.	*Tanto malagod*	Very athletic	Very; athletic	Found
10.	*Madakel tao*	Many people	Many; people(pagtao)	Found
11.	*Sumiyong sii*	Went here	Go(song); here	Found
12.	*Tolad ayam*	Draw animal	Draw; animal	Found
13.	*Magastos talompa*	Expensive shoes	; shoe	Not Found
14.	*Matenggaw ig*	Cold water	Cold; water	Found
15.	*Matanog lagam*	Loud voice	loud; voice	Found
16.	*Marata ontong*	Bad luck	Bad; luck	Found
17.	*Maloya khakan*	Spicy food	Peppery; food(kan)	Not Found
18.	*Makapal libro*	Thick book	Thick; book	Found
19.	*Tanto diyokawan*	Very tired	Very; tired(dokaw)	Found
20.	*Maitem kapi*	Black coffee	black; coffee	Found

Table 4.6 Measurement calculation:

$$Accuracy = \frac{Sum\ Total\ of\ the\ Found\ Predicted\ Match}{Sum\ Total\ of\ the\ Overall\ Testing\ Result}\ x\ 100\%$$

$= \frac{16}{20} x100\%$

= 80%

Experiment Test III: Sentence (1)

In the sentence experiment test the translation from the corpus occur once in a process which is shown in the column '*Translation in English by the Corpus'* in the Table 4.7 while the bases for the matching result if found which is the '*SEAlang Library Maranao*' column has taken more than one translation process to see if they match the same translation in the corpus.

Table 4.7 Meranao Translation on Sentences and its Matching Result (1)

No.	Meranao Text	Translation in English By the Corpus	English Translation by SEAlang Library Maranao	Match Result
1.	*Saken sumiyong kagai*	I went yesterday	I; go(song); yesterday;	Found
2.	*Gyoto iportante karatas*	Those important paper	Those; important; paper	Found
3.	*Saken miniga sa kuwarto*	I slept in room	I; lie down;	Not Found
4.	*Dimbantay salida magagawii*	Watching drama late night	; show; night	Not Found
5.	*Saken galebek sii*	I work here	I; work; here	Found
6.	*Pirito sda na maregen*	Frying fish is difficult	Fried; ; difficult	Not Found
7.	*Permi pkhada money*	Always lose money	Always(pirmi); loss(kada); money	Found
8.	*Dinggitagita sa liyo*	Playing in out	Game(gitagita); ;	Not

			outside	Found
9.	*Miyakawma sa walay*	Came in house	Come; ; house	Found
10.	*Phamakineg sa radio*	Listening in radio	Listen; ;radio(radiyo)	Found
11.	*Inoto masososowa san?*	Why happening there	Why; condition; there	Not Found
12.	*Mapita na isnin*	Tomorrow is Monday	Tomorrow; ; Monday	Found
13.	*Muuntod sa towak*	Sitting in stairs	Sit(ontod); ; stairs/ladder	Found
14.	*Ilayanga so kalindar*	Look the calendar	Look(ilay); ;	Not Found
15.	*Saken piyamasa bangkala*	I bought shirt	I; buy (pamasa) ;shirt	Found
16.	*Saken tareg batiya*	I stop reading	i; stop; reading	Found
17.	*Madakel problema na di khasoleb*	Many problem is no solution	Many; problem(problima); no;	Not Found
18.	*Mbida siran sa pamikiran*	Different them in thought	; them; thought	Not Found
19.	*So turog na kailangan*	The sleep is needed	; sleep(torog); need	Found
20.	*Skaniyan miyakapas sa iksam*	He passed in exam	He / she (sekaniyan); pass an exam (miyakaapas) ;	Found

Table 4.7 Measurement calculation:

$$Accuracy = \frac{Sum\ Total\ of\ the\ Found\ Predicted\ Match}{Sum\ Total\ of\ the\ Overall\ Testing\ Result}\ x\ 100\%$$

$= \frac{12}{20} x100\%$

=60%

Experiment Test IV: Sentence (2)

In the sentence (2) test, we have translated the words disregarding any Grammatical standard and see if the search has found all the words matching from the database corpus available for matching procedure. Every word is separated by a 'semi colon' indicating an equivalent word from the preceding cell which is the words resulting from the corpus.

Table 4.8 Meranao Translation on Sentences and its Matching Result (2)

No.	Meranao Text	Translation in English By the Corpus	English Translation by SEAlang Library Maranao	Match Result
1.	Saken piyamasa madakel bangkala	I bought many shirt	I; bought(buy); many; shirt	Found
2.	*Skaniyan miyakapas sa iksam ago skaniyan guraok*	He passed in exam and he cry	He / she (sekaniyan); pass an exam (miyakaapas) ; and; he; cry(goraok)	Found
3.	Anda so ipegokit sa walay i Daud	where the way to in house i daud	Where; ; ;walai; daud	Not Found
4.	siran sumiyong sa padian.	them went in market	They/them; went(song); in ; market	Found
5.	Ngkainoto, ba aden a pagangin ngka?	why is it is there is a want you	;interrogation; there is; a; ; you	Not Found

6.	Pemisita ko sa ginawai aken	visiting mine in friend I	Visit(bisita); mine; ;friend; i	Not Found
7.	Oway, ba ka san song. Na ilay ka bo om	oway is you there go. Is see you bo om.	Yes; ; you; there; go, see; you; ; ;	Not Found
8.	Anda so manga tao sa giyai walay?	where the pluralizer people in in this house	Where; ; pluralizer; people(pagtao); this; house	Not Found
9.	Soled ka, bolos ka.	enter you continue you	Inside/Come in; you; continue; you	Found
10.	*Kolor biro na iyan baya ago kolor gadong na di niyan baya*	Color blue is her desire and color green is not his desire	Color; blue; is; her ; desire; and; color; green; is; no/not; his; desire	Found
11.	*Madakel so problema na di khasoleb igira da pirak*	Many the problem is no solution when no money	Many; ; problem(problima); ; igira; no; money	Not Found
12.	*Sabagi o siran di sabot ka*	Some of them don't understand you	Half; if/or; they/them; no/not; understand; you	Not Found
13.	Sapeng so kan sa lamisaan	Cover the food in table	Cover; food; table	Found
14.	*Mbida siran sa pamikiran ago di siran pagayon*	Different them in thought and don't them agree	; them; thought; and; no/not; they/them; agree	Not Found
15.	*Miyakawma sa walay ago da aken mailay*	Came in house and not I saw	Come; ; house; and; no/not; I; ;	Not Found
16.	*Sa liyo na madakel a tao imanto pesagad*	In out is many a people now passing	In; outside; ;many; people(pagtao); ;	Not Found

17.	*Saken tareg batiya goani saken torog*	I stop reading before I sleep	i; stop; reading; before; I sleep	Found
18.	*Gyoto iportante karatas na madakel a miyabnasaon*	Those important paper is many a destroyed	Those; important; paper; is; many; ;	Not Found
19.	*Ilayanga so kalindar aden mariga gawii*	Look the calendar there is red date	Look(ilay); ;there; red; today/date	Not Found
20.	*Dimbantay kambola magagawii ago khan madakel khakan*	Watching basketball late night and eat many food	; ball game; night; and; eat; many; kan(food)	Not Found

Table 4.8 Measurement calculation:

$$Accuracy = \frac{Sum\ Total\ of\ the\ Found\ Predicted\ Match}{Sum\ Total\ of\ the\ Overall\ Testing\ Result}\ x\ 100\%$$

$$= \frac{7}{20}\ x100\%$$

=35%

In the following experiment we have conducted a full paragraph translation in order to provide a demonstration for the case of the grammatical approach regarding translation on the Meranao Translator application. A reliable website source for an article using a Meranao Language is copied and pasted to the application Meranao Translator and resulted to a vague sentence structure which does not have any systematic understanding based on the human level of perception as a reader.

In Figure 4.2.1 Meranao Paragraph to be translated into English translation based on the database system corpus.

Ayakaposan ang kai a lalag ko na panalamatan aken ang kai a mbala a mindiamonga sa dairan kapakandarainon sa kiya paka tokawa iran rekami ago giya darodopa aken sangkai a kalilimod, Na pamangni ako rekanopen sa maaf oba aden a dako katarotopi ko kiyapayag o sariat o lalag ko sabap sa kangodaan akopen a gopen gii sinanad, sa ayamala na sobo so kadnan I daa pawing iyan. Wassalam

Figure 4.2.1 Meranao Text in a Paragraph Form

In Figure 4.2.2 is the output result of the translation from the database provided by the corpus. We are recognizing the large number of errors shown in this demonstration to provide a clearance proof of what it is we are assuming to in the next step of this study.

the end the here a walk mine is grateful I the here a pair a presence in not let attention in in this make knowing their to us ago this behalf I in this a gathering is ask I to you in apologize may there is a not being complete mine agree if condition if walk mine cause in young still a first is learning in firstly is only the God i chant pawing him wassalam

Figure 4.2.2 Translation Result from the Meranao Translator

By the difficulty of finding the exact spelling and words from SEAlang Library Maranaw, the matching procedure fail to establish because of the singularity searching function of the process in SEAlang Library Maranaw online resource.

Figure 4.2.3 here is a Meranao text in a listing order to be translated into English translation based on the database system corpus.

1. HAMDALA
2. PAMEKASAN
3. SALAM
4. KAPANABIA
5. KAPANGO-NGODASAN
6. MOAYAN O LALAG (purpose of the speech)
7. PAHALA O LALAG (Importance of speech)
8. KAPOSAN O LALAG

Figure 4.2.3 Meranao Text in Listing Words

In Figure 4.2.3 is the output result of the translation from the database provided by the corpus. Notice that the input above shown in Figure 4.2.3 is in order list horizontally with characters e.g. '.' dot, '(' and ')' open and close parenthesis which are omitted by the system in the phase of tokenization. A proof that, study ahead has been initially commenced.

1 thanking 2 introduction 3 greet 4 permission 5 acknowledgement 6 essence if walk purpose of the speech 7 reward if walk importance of speech 8 end if walk

Figure 4.2.4 Translation Result from the Meranao Translator

The translation based on the experiment procedure above has been done on Figure 4.2.4 and it resulted as shown on the mentioned figure.

4.3 DISCUSSION

Assuming all the experiment tests are being evaluated for the accuracy measurement of the application Meranao Translator. We regard all the relevant data for the affirmation of the authenticity of the testing reliability to achieve the valid result according to a reliable conduct. We have shown some demonstration within the experimental test but not put into the evaluation for the complication of establishment of the simplicity of the test procedure.

4.3.1 Analysis and Evaluation

The result of the above experiments based on accuracy rate evaluation shows that in the Table 4.5 has the accuracy result of 100%, and the Table 4.6 has 80% accuracy result and 60% accuracy result for the Table 4.7 and a 35% accuracy rate for the table Table 4.8.

In the review of the Table 4.5, we have tested Lists of Meranao word in a single containing word which resulted to have high accurate quality on finding the matching word regarding the translation process because of its one-to-one look up in comparison with the SEAlang Library Maranaw assuming that the spelling and writing style are consider the same type. Thus we verify it as 'found' and calculated with the accuracy formula and resulted to 100% rate.

In the Table 4.6, the listing of Meranao text to be translated consist of two word each phrase and so not all are found and matched the same translation of the words in one phrase provided by the corpus into SEAlang Library Maranaw. The accuracy rate went down to 80%, such reason is that of a diversity and possible changes into update and the old static database.

In the Table 4.7, as we demonstrate the table into sentence form of Meranao text, we have resulted an occurring low accurate on matching the Meranao Translation by the Corpus into SEAlang Library Maranaw because of the variability of text containing the sentence being evaluated the accuracy deflated to 60% rate. In comparison with the following table which is Table 4.8 that is also compose of sentences but added more words, it resulted to a lower percentage of accuracy rate.

The idea of matching the corresponding meaning of a language to another language is very significant in the field of any communication for understanding processes. It is the only way social settings are taken into more productive and informative because of the language a person or an interpreter is possessed. In this cases the language carrier has to have the appropriate and qualified tool for the interpretation closely similar to the other party to convey the specified language as the medium of communication. The interaction on both party should have a clear translation on understanding the other counterpart. But in order to be an effective communicator, it is necessary to have the right knowledge about the language being used.

The understanding of one language to another language indicate the same principle in this verse in the Quran in surah Yassin verse 36:

"سُبْحَانَ الَّذِي خَلَقَ الْأَزْوَاجَ كُلَّهَا مِمَّا تُنبِتُ الْأَرْضُ وَمِنْ أَنفُسِهِمْ وَمِمَّا لَا يَعْلَمُونَ"

which means:

"Exalted is He who created all pairs - from what the earth grows and from themselves and from that which they do not know."

For further elaboration of the essence of matching in Islamic concept, Allah mentioned in the Quran surah Al-Dhāriyāt ayat 59:

"نا زَوجَينِ لَعَلَّكُم تَذَكَّرونَوَمِن كُلِّ شَيءٍ خَلَقَ

which means: "And of all things We created two mates; perhaps you will remember." An indication of correspondence of two things to meet the necessary fulfilment as it is stated in the Quran and scientifically experiment cannot be done without the proof of collected theory based on facts.

In the discourse on the difficulty in finding a higher accuracy expectation in any related scientific problem or simply an idea of solving a subject matter, an entire surah in the Quran As-Sharh specify clearly and imply in its verses as follows:

"أَلَمْ نَشْرَحْ لَكَ صَدْرَكَ(1) وَوَضَعْنَا عَنكَ وِزْرَكَ(2) الَّذِي أَنقَضَ ظَهْرَكَ (3)
وَرَفَعْنَا لَكَ ذِكْرَكَ (4) فَإِنَّ مَعَ الْعُسْرِ يُسْرًا (5) إِنَّ مَعَ الْعُسْرِ يُسْرًا (6)
فَإِذَا فَرَغْتَ فَانصَبْ (7) وَإِلَى رَبِّكَ فَارْغَبْ (8)"

Meaning, "(1) Have We not opened forth for thee thy breast? (2) And We have taken off from thee thy burthen, (3) Which weighed down thy back. (4) And We have upraised for thee thy renown. (5) Then verily along with every hardship is ease. (6) Verily along with every hardship is ease. (7) Then when thou becometh relieved, toil; (8) And Unto thine Lord, attend." A Yusuf Ali translation, consequently conveys the practical idea on how a subject matter dealing with the reality occurrences.

The result of the analysis in correlation to the ayahs mentioned above has so much relevance in attaining the outcome of this research study in dealing with the difficulty which we can also refer to as complexity of the subject matter. A problem or a situation similar to this study as we have resulted the percentage of the Experimental Test II, III, and IV on the accuracy of 80%, 60%, decelerated down to 35% rate is in the greater means as this study is point out the larger problem as we go along. The larger number of the subject matter, the more it is difficult to get the absolute accuracy rate anticipated and so, hoping and dedicating an effort for any situation including statistical matter are aim to declare an ease as the researchers submerge to it and focus on breaking down problems into gradual stages and no matter how large the difficulty is, it will eventually dissipate for those who believe and immerse what the *ayah*s of the Quran are implying for it is only in the favor of those who are the people of knowledge.

CHAPTER V

CONCLUSION AND RECOMMENDATIONS

5.1 CONCLUSION

From the result of this research Meranao Translator Application Implementing a String Matching Algorithm using Matlab-based platform has performed a text translation from Meranao text in its English translation or meaning. The translation result occurrence has been done and come to conclude that the Meranao Translation Implementing String Matching Algorithm has shown differences based on the number of the parameter of the text being experimented. As the parameter increase in number, the rate of the accuracy decreases. With the phrase text test with only two containing words has resulted 80% rate, and the sentence text test containing three words in it resulted at a range of 60% going down to 35% depending on the number of the text element within a sentence. This indicates that the more parameter of text being translated, the more it is to encounter errors in the scope translation result anticipated.

5.2 RECOMMENDATIONS

Further research for this study is a must required for the advancement of the application including the following suggestions:

1. Regular updates on the corpus database for the Meranao lexicon completion.

2. Additional implementation on simple grammatical rules on Meranao Affixation for automatic correction if text is in great number for translation.
3. Further additional research focusing on the Meranao Morphological to be implemented for the advancement of the application function.

REFERENCES

Al-Quran *Color-coded Arabic English Translation* (2008) by Marmaduke Pickthall, Syed Abul Aala Maududi and Muhsin Khan – Taquid Din AlHilali Published Online by Hafiz Khan at www.quranpda.com, 2008

Alonto A. G., Adam A. B., & Zorc R. D., *Maranao Dialogs and Drills*, Defense Language Institute Foreign Language Center, Dunwoody Press, U.S.A., 2009

Aswani S. and Sindhu L. (2014) Morphological Analyzer for Malayalam: A Literature Survery. *International Journal of Computer Applications (0975 – 8887) Volume 107 – No 14, 2014*, 24 – 27.

Bagul, N. L. (2006). Mera Parser. Computer Studies Department, Mindanao State University. Undergraduate Thesis.

Fortes, F. C., & Roxas, R.E., *A Constraint-based Morphological Analyzer for Concatenative and Non-concatenative Morphology,* College of Computer Studies, De La Salle University-Philippines, 2004

IBM: *Language Processing (The Art of Tokenization) CraigTrim*, Retrieved April 29, 2016 12:01:26 PM from https://www.ibm.com/developerworks/community/blogs/nlp/entry/tokenization?lang=en

Gong-shen L, & Sheng-jun Z (2009). *The Factors Analysis and Algorithm Implementation of Single-pattern Matching*, Shanghai Jiaotong University, China, (Sci.), 2009, 14(3): 331-337

Introduction: *EXACT STRING MATCHING ALGORITHM Animation in Java,* Retrieved last October 18, 2016 at 3:10:22 PM from http://www-igm.univ-mlv.fr/~lecroq/string/node2.html

Khumbar D., Braja G. P., Dipankar D., Sivaji B. (2012). Morphological Analyzer for Kokborok. *Proceedings of the 3rd Workshop on South and Southeast Asian Natural Language Processing (SANLP),* 41–52.

Klink S. & Kieninger T., *Rule-Based Document Structure Understanding With A Fuzzy Combination Of Layout And Textual Features*, German Research Center for Artificial Intelligence, IJDAR, Germany, 2001

Lee, D. L., Chuang, H., & Seamons, K. (1997). Document ranking and the vector-space model. *IEEE Software*, *14*(2), 67–75.

Lobel, Jason, *Maranao Dialogs and Drills*, McNeil Technologies, Inc. Dunwoody Press, U.S.A., 2009

Majdi S., Eric A., and Mohammad A. M. B. (2013) SALMA: Standard Arabic Language Morphological Analysis, *IEEE,* 2013

Maranao Language: *Maranao Language*, Retrieved last Saturday, March 11, 2017 10:19:31 AM from http://iloko.tripod.com/Maranao.htm

Mckaughan, H.P., Macaraya, B.A., *A Maranao Dictionary*. University of Hawaii Press, Hanolulu, 1967

Nsira, N & Elloumi M., *On-line String Matching in Highly Similar DNA Sequences*, Springer International Publishing, University of Rouen, Normandie University, Rouen, France, 2017

Novak Djokovic Foundation: *The Importance of Children Learning their Native Languages,* Retrieved last Saturday, March 04, 2017 3:01:28 PM from https://novakdjokovicfoundation.org/importance-children-learning-native-languages/

Pangandaman, M. (2008) MELT: *Meranao to English Language Translation.* Computer Studies Department, Mindanao State University. Undergraduate Thesis.

Phonemes: *Learn the Maranao Language*, Retrieved last March 29, 2008 at 4:18:59 PM, from http://www.maranao.info/phonemes.html

Rajashekharaiah, K.M.M, MadhuBabu Ch, and Dr. Raju Viswanadha(2012) *Parallel String Matching Algorithm Using Grid*, International Journal of Distributed and Parallel Systems (IJDPS), Research Scholar Rayalamaseema University, Kurnool, Vol.3, No.3

Robert R. R. and Gersam T. M. (2008) A Morphological Analyzer for Filipino Verbs', *University of the Philippines Visayas*, 467 – 473

Sembok, T.M.T., Abu Ata, B. M. , Abu Bakar, Z., *A Rule-Based Arabic Stemming Algorithm*, Proceedings of the European Computing Conference, Page 392 - 397

Soni K.K, Vyas R., & Sinhal A., *Importance of String Matching in Real World Problems*, International Journal Of Engineering And Computer Science, TIT College Bhopal, India, 2014

Tamames J., *Text Detective: A Rule-Based System For Gene Annotation In Biomedical Texts,* BMC Bioinformatics, Spain, 2005

Vidya S., Prof. Akhtar R., and Dr. Nilay K (2012) *String Matching and its Applications in Diversified Fields*. IJCSI International Journal of Computer Sciences Issues, Vol. 9, Issue 1, No 1, 201

Wang, X. (2012). Text Clustering Based on the Improved TFIDF by the Iterative Algorithm, 140–143.

Wikipedia: *Morphology (linguistic),* Retrieved Friday, March 10, 2017 8:17:12 PM from https://en.wikipedia.org/wiki/Morphology_(linguistics)

William B. Frakes & Ricardo Baeza-Yates, *Information Retrieval: Data Structures & Algorithms*, Udi Manber-Arizona, 2007

Xuri TANG (2006) English Morphological Analysis with Machine-learned Ruled, Dept. Foreign Languages, Wuhan University of Science and Engineering 35-41, China, 2006

Yue H., QingShi G., Li G. & PeiFeng W., Giant complete automaton for uncertain multiple string matching and its high speed construction algorithm, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, 2011

APPENDICES

SAMPLE LIST OF MERANAO WORDS	ENGLISH TRANSLATION
ipegokit	way to
diritso	straight
angkanan	that way
ketoon	find
ka / seka	you
bo	will
aden	there
saya	here
giyai	in this
somiyong	went
ngkainoto	why is it
ba	is
pagangen	want
ngka	you
pemisita	visiting
ginawai	friend
aken	mine
soled	come
bolos	welcome
naino	fine
komasoy	return
baden	anyway
san	there

myailay	saw
mayakataban	winner
lalayon	forever
layok	friend
bedong	cat
tuturogen	sleeping
naba	not
kuminowa	took
karatas	paper
pekhababaya	happy
pamikiran	thought
miyamemesa	amaze
kasigi	ambition
ambolansiya	ambulance
malilipat	amnesia
kadakel	amount
gani	anger
salakao	another
pila	ant
barokan	arm
irek	armpit
atoran	arrange
oma	arrive
arti	art
lagid	as
pakaiza	ask
makatotorog	asleep

domarpa	attend
olaola	attitude
pakabimban	attractive
agosto	August
babo	aunt
kapaar	authority
totolan	authobiography
tomatik	authomatic
abokado	avocado
pananggilaan	avoid
ikaririga	baby
likod	back
timbang	balance
opaw	bald
sapi	beef
pangni	beg
poon	begin
maori	behind
paratiyaya	believe
atag	below
kapiyaan	benefit
pirmiro	best
titipowa	betray
baik	bicycle
bilyon	billion
papanok	bird
kinimbawataan	birthday

teteb	bite
senditan	blame
tanggob	blanket
pirogo	bleed
rogo	blood
oloan	boss
katiya	bote
ipag	brother-in-law
paroparo	butterfly
pamasa	buy
mamasa	buy
pemasa	buyer
taksi	cab
diikapian	cafeteria
arga	cost
tiyan	stomach
ayam	animal
bangkala	shirt
igira	when
baya	desire
sabagi	some
mailay	saw
pesagad	passing
miyabinasa	destroyed
towak	stairs
kambola	basketball
pagnaw	awake

bala	calamity
ikarang	calculate
taloan	call
intad	calm
kodak	camera
kampo	camp
kepakay	can
itareg	cancel
ransok	candle
canoe	awang
kantina	canteen
gora	cap
osayan	caption
karabao	carabao
karamilo	caramel
tagikor	care
karpintiro	carpenter
kabankaban	carton
banggala	cassava
saloon	catch
poonan	cause
silda	cell
look	center
cental	lembak
tangked	certain
titolo	certificate
katatangked	certainly

rar	challenge
sambi	change
gorigao	chaos
ayat	chapter
oring	charcoal
charity	limo
makipagistoria	chat
barato	cheap
akal	cheat
bias	cheek
kiso	cheese
satoran	chess
chief	oloan
matiboron	circle
sigopan	cigarette
palakpak	clap
sosonan	class
lapot	clay
klima	climate
tabid	climb
dinis	cloth
sarosong	clumsy
darpo	cobra
kodigo	code
kapan	coffin
miatalabok	coincidence
magayonayon	collaborate

timo	collect
ikakamotan	cologne
kasosorot	comfort
bantogan	commendation
timooon	compile
sagina	complain
magayon	compromise
itapok	conceal
konsirto	concert
pamotosan	conclusion
isapa	condemn
katibaw	condolence
ipayag	confess
mayon	conform
isompat	connect
kisosompat	connection
kakokomplito	consists
osaren	consume
tagoai	container
madadalem	content
orad	contest
indayon	continue
kapasadan	contract
sopak	contrary
sangkaan	control
loto	cook
petindaan	cooker

sopa	couch
kambilang	count
kawarao	courage
tengged	cousin
talaw	coward
masodi	critic
sagda	criticize
korona	crown
marangit	cruel
pipino	cucumber
paparangayan	cultural
bolongan	cure
masima	curious
kotina	curtain
binasa	damage
sobo	dawn
kepitapita	daybreak
patay	dead
salingat	deaf
kapatay	death
kapawala	debate
miado	decay
bilangatao	decent
korangan	decrease
ipelilalilang	didecation
lebat	deduct
madalem	deep

paawing	defect
abang	defend
osayin	define
ponasen	delete
witen	deliver
dapiya	despite
sima	detect
borawis	diarrhea
manonogo	dictator
tapsir	dictionary
kambidaan	differ
kalot	dig
boringen	dirty
kabinasaan	disadvantage
miyada	disappear
sainsang	disarrange
morala	disaster
lebatan	discount
katokawan	discover
pakilay	display
dimapakay	disqualify
rigaro	disturb
bagi	divide
dokominto	document
galebek	doing
dalar	dollar
sadeka	donation

dobli	double
totol	drama
taginepen	dream
ikadowa	duplicate
inako	during
miyasokat	earn
tanoren	edit
pangendaon	educate
panamar	effort
gogodan	elaborate
sayamot	emotion
datagoian	empty
kaposan	end
ridoay	enemy
makina	engine
tembangen	entertain
mananamar	enthusiastic
sobri	envelope
datar	equal
error	ribat
malagoy	escape
adat	ethic
gagawii	evening
every	tanan
dosa	evil
matangka	exact
ibarat	example

tabia	excuse
pakaoladen	expand
gonanao	explain
sobra	extra
tito	fact
pabrika	factory
paltik	fake
bokag	false
babantogen	famous
robas	fatigue
kalek	fear
pibroari	February
Pilipino	Filipino
pasaden	finish
ondas	flow
tondogen	follow
beneng	forehead
saatan	formula
maradika	freedom
ioman	further
bolawan	gold
salamen	greet
taawi	grief
lopa	ground
meto	grow
parangay	habit
hammer	bakbak

poso	heart
makaogopa	helpful
sambot	immeidately
inggolalan	implement
piyapiyaan	improve
iitong	include
sakatao	individual
pakatokaon	inform
dawat	ink
mangamad	inquire
panononongka	insult
saripitian	interpret
pamekasan	introduction
imbalido	invalid
kapital	invest
magatel	itchy
iniro	January
salak	once
overstay	magtay
pair	rimbang
salad	peso
sabaad	phase
pekaawidan	portable
patatas	potatoe
antoken	predict
igma	procedure
idayon	proceed

okitokit	process
piki	pseudo
pobliko	public
talimbagak	remain
toos	remember
sagipoon	root
kaploba	searching
gedamen	sickness
kalebod	simplicity
masasakop	subject
suddenly	pankao
ditatap	temporary
antangan	theory
bandingan	topic
masa	trend
lainan	underline
alsa	vacation
ilayin	view
makaiiklas	wholehearted
pasosorat	writer
logar	zone