Volume 66, Issue 3, 2022
Journal of Scientific Research
of
The Banaras Hindu University

Context-Based Text Completion System for Amharic Language

Alemebante Mulu[*1], Vishal Goyal [2]
[*1] Department of Computer Science, St. Mary's University, 1211, Ethiopia, alembantemulu184@gmail.com
[2] Department of Computer Science, Punjabi University Patiala, 147002, India, vishal.pup@gmail.com

Abstract: **This paper discusses the practical aspects of easiness in communication using Short Message Service (SMS), E-mailing, correcting misspelt words and checking the grammatical mistakes. There are different data entry mechanisms to insert a text on the computer machine as well as a mobile device, such as a keyboard, soft keys, speech etc. The paper proposed to develop a context-based auto text completion system for the Amharic language specifically to correct misspelling on Short Message Service (SMS), E-mailing and helps to correct the grammar mistakes as well. Data entry technique can be inserted with the support of text completion (predictive) or non-predictive. Therefore, we are using a statistical model, Predictive Partial Match (PPM) and Support Vector Machine (SVM) approaches for implementing the Amharic context-based text completion system. Since the system is developed by using the context-based and statistical model, we adopted the Amharic Part of Speech (POS) tagger system. For training and testing the system, we are using 395,464 unique words with frequency and 750,000 sentences that has been prepared by the Walta Information Centre (WIA) and Ethiopia News Agency (ENA). All those data have been used to build the Amharic dictionary, the corpus of the system and to calculate the frequency occurrences of each word as well. Finally, the results show a 14% improvement from traditional frequency-based Amharic word prediction system.**

Index Terms: **Context-based text completion, Part of Speech, Predictive Partial Match, Support Vector Machine.**

I. INTRODUCTION

A. Background

Amharic language has been in use since 4th century AD and it's originated from Ge'ez. The name Amharic comes from the people of Amhara in northern Ethiopia, which is thought to be the historic center of the language (Gebremichael, 2011). Now a days, Amharic is the official language of Ethiopian which is the second most vastly semantic language in the world. As we know, the language is spoken by different countries such as Ethiopia, Eritrea, Canada, USA and Sweden (Gebremichael, 2011). Majority of the 25 million speakers or so of Amharic are found in Ethiopia (Gebremichael, 2011). Since Amharic language is an official language of Ethiopian government, the political issues, mostly the magazines, newspapers, industries, commercial organizations, and governmental offices are using this script.

The Amharic language has used of 238 Ge'ez scripts with 56 additional labialized alphabets that are represented by two sounds like ጓ for ግዋ (Alemebnate & Vishal, 2015, pp.113-118).The area of coverage of Amharic language is very vast and is second most spoken sematic language next to Arabic. The script used to write for different languages like Ge'ez, Amharic, Tigrinya, Tigre, Awngi, Harari, Silt'e, Argobba, Blin, Chaha, Dizin, Inor and Xamtanga (Gebremichael, 2011). Auto text completion system in its place uses the frequency-based disambiguation method and predicts the most commonly used word above other possible words.

Context-based text detection and completion of misspelling are done by taking into account an adjacent letter which is much more difficult task for computational linguists and software developers than just checking orthography. Text completion system is handling those errors which are violating or misspelling the correct words, for example, missing one or more letters including with incorrect letter order arrangement as well.

B. Motivation

Now a day, the development of technology has increased at a great speed and stretch upward to ever in this world. Therefore, every country and the individuals have taken advantage of technology. However, utilizes of technology has dependent on how much the country is developed. Practically, to make better use of technology, it should be assigned to a particular place. Humankind in a given country should have the technology which helps their day to day activities, socio-economic live, such as for their languages and culture as well. As we know, the majority of computers and mobile applications are used for

* Corresponding Author

specific language such as English and others European language. However, this makes unable to be used the technology those who don't know these languages (Mahamed &Vishal, 2015, pp. 274-279). This is a big problem for the development of the technology, to handle this problem and improve the usability of computer devices to make reachable and let them know the people can be convey their own ideas in their own languages. After observing the gape, this research has required to be done for handle the problems. Although it has a big role and motivation to minimize the gap between all language technologies. Explicitly the relationship of motivation with effect is used to handle the availability and accessibility of language resources as well as the globalization of language technology. As a researcher, we has to needs to devote more attention for globalization and practical effects of motivational mechanisms than to the get associated with them by creating software or application.

Text completion is one of the most needed research areas which demands the researcher's attention for a particular language. Therefore, the system is important for writing text with minimum time, reduce misspelling and improves writhing the correct messages as well. All the reasons as mentioned earlier are motivated us to research text completion for the Amharic language to utilized by the Ethiopian people and playing a crucial role in the improvement of technology.

ግዕዝ	ካዕብ	ሣልስ	ራብዕ	ኃምስ	ሳድስ	ሳብዕ	ዘመደ ራብዕ
ሀ	ሁ	ሂ	ሃ	ሄ	ህ	ሆ	ኋ
ለ	ሉ	ሊ	ላ	ሌ	ል	ሎ	ኗ
ሐ	ሑ	ሒ	ሓ	ሔ	ሕ	ሖ	ኍ
መ	ሙ	ሚ	ማ	ሜ	ም	ሞ	ሟ
ረ	ሩ	ሪ	ራ	ሬ	ር	ሮ	ሯ
ሰ	ሱ	ሲ	ሳ	ሴ	ስ	ሶ	ኗ
ሠ	ሡ	ሢ	ሣ	ሤ	ሥ	ሦ	ሧ
ሸ	ሹ	ሺ	ሻ	ሼ	ሽ	ሾ	ሿ
ቀ	ቁ	ቂ	ቃ	ቄ	ቅ	ቆ	ቋ
በ	ቡ	ቢ	ባ	ቤ	ብ	ቦ	ቧ
ቨ	ቩ	ቪ	ቫ	ቬ	ቭ	ቮ	ቯ
ተ	ቱ	ቲ	ታ	ቴ	ት	ቶ	ቷ
ቸ	ቹ	ቺ	ቻ	ቼ	ች	ቾ	ቿ
ኀ	ኁ	ኂ	ኃ	ኄ	ኅ	ኆ	ኋ
ነ	ኑ	ኒ	ና	ኔ	ን	ኖ	ኗ
ኘ	ኙ	ኚ	ኛ	ኜ	ኝ	ኞ	ኟ
አ	ኡ	ኢ	ኣ	ኤ	እ	ኦ	ኧ
ከ	ኩ	ኪ	ካ	ኬ	ክ	ኮ	ኳ
ኸ	ኹ	ኺ	ኻ	ኼ	ኽ	ኾ	ዃ
ወ	ዉ	ዊ	ዋ	ዌ	ው	ዎ	
ዐ	ዑ	ዒ	ዓ	ዔ	ዕ	ዖ	
ዘ	ዙ	ዚ	ዛ	ዜ	ዝ	ዞ	ዟ
ዠ	ዡ	ዢ	ዣ	ዤ	ዥ	ዦ	ዧ
የ	ዩ	ዪ	ያ	ዬ	ይ	ዮ	
ደ	ዱ	ዲ	ዳ	ዴ	ድ	ዶ	ዷ
ጀ	ጁ	ጂ	ጃ	ጄ	ጅ	ጆ	ጇ
ገ	ጉ	ጊ	ጋ	ጌ	ግ	ጎ	ጓ
ጠ	ጡ	ጢ	ጣ	ጤ	ጥ	ጦ	ጧ
ጨ	ጩ	ጪ	ጫ	ጬ	ጭ	ጮ	ጯ
ጸ	ጹ	ጺ	ጻ	ጼ	ጽ	ጾ	ጿ
ፀ	ፁ	ፂ	ፃ	ፄ	ፅ	ፆ	ፏ
ፈ	ፉ	ፊ	ፋ	ፌ	ፍ	ፎ	ፗ
ፐ	ፑ	ፒ	ፓ	ፔ	ፕ	ፖ	ፘ

Figure 1.1: Overview of Amharic characters

C. Objective

The objective of this research is to develop a text completion system for Amharic language by using the context-based method.

D. Research Methodology

To achieve the goal of this research work, the following research methodology steps have followed. Those methods are literature review, data collection, data analysis, design and implementation. To build this research work effectively, we investigated and reviewed different materials in a different way. The literature review of this research work has covered different materials and supportive tools, such as books, articles, journals and internet. To address the present problem Amharic text completion system, we are using the implementation tools, such as to Python, SQLite and window operating system. Based on the literature survey, we had been collected the relevant data for building the corpus. More than 750,000 Amharic sentences have been collected from Walta Information Centre (WIA) and Ethiopia News Agency (ENA). After that, 390,464 unique words with their frequency has extracted from all the data collected from Walta Information Centre (WIA) and Ethiopia News Agency (ENA) and stored the unique word with the frequency of the word.

The next steps of this research work were designing the system and select the preferred approaches. For developing a text completion system, we are selected Support Vector Machine (SVM) and Predictive Partial Match (PPM), designing models. Moreover, we adopted the Amharic Part of Speech (POS) tagging system for statistical prediction method. Finally, we implement the system and evaluate the performance of the system accuracy in a different aspect, such as speed, text completion accuracy, and the system flow has well gone or not at the running time.

E. Previous Work

During this research work, we investigate the previous work related to Natural Language Processing (NLP) specifically for Amharic text prediction and correction systems. The Amharic text prediction systems in place have used the dictionary and frequency-based technique to predict the most commonly used words above other possible words. The previous text prediction system predicts the correct word based on the highest word frequency and the alphabetical order. Most of the time, the frequency-based prediction system is not solved properly. The problem of giving correction for Amharic text. For example, two or more word has occurred by the same number of frequency, the system cannot detect the exact word and predicted correctly. Therefore we initiated to solve such kind of problems by using context-based text completion system method. Finally, the main outcomes of this research are to develop an auto text completion system for Amharic language by using a context-based method.

F. Purpose and Contribution of Paper

Computer machine uses limited language specifically English and views others. This makes a problem for the usability of the technology. Those who don't know the languages, they face a problem which is how to use these devices (Mahamed &Vishal, 2015, pp. 274-279). Technology has its benefit to develop the country. Because of this and related benefits, why Amharic speaker is not using their language? And why Amharic language is a part of text completion language technology? In addition to this, the system can give alternative service for text completion and entry method for E-mailing, SMS, Grammar checking and any other social media like Facebook, Messenger, WhatsApp, etc. (Mahamed &Vishal, 2015; Alemebnate & Vishal, 2015, pp.113-118). To increase the usability of the computer or mobile phone and let people express their ideas by using their languages. Therefore, in this research work, we stand to develop auto text completion system for Amharic language and the developed system is contributing a big role for exchanging accurate message as well as increasing rapid technology growth over the world.

II. Literature Survey

The context Based Word Prediction system has for Short Message Service (SMS) messaging in which context is used to predict the most appropriate word for a given code. As the researcher mentioned, the system to allow informal words (short forms for proper English words). In addition to this, the mapping from informal word to its proper English words is done using Double Metaphone Encoding based on their phonetic similarity (Sachin & Shilpa, 2007, pp. 1-9).

The predictive text entry method for the Somali language has prepared the text prediction model. As the researchers mentioned, the text prediction is a technique which has predicts what the user wants to write on the mobile phone. The user wants to write some characters, and the system will predict the remained ones. Moreover, the text prediction has reduced the redundancy and makes the system more efficient because of time-saving. In this research work, it was prepared text prediction model for the Somali language (Mahamed &Vishal, 2015, pp. 274-279).

To explore and examine the efficacy of mobile keypad text entry is the growth of Short Message Services (SMS) messages on mobile phones, especially in Europe. In this research work, in earlier times, the majority of SMS users were using multi-press as their method of text entry. Always there are some errors when the users were typing on mobile keyboards with multi_ press keys. These errors are mainly coming from hitting keys adjacent to the desired key (Hedy, Eugene, Terry, Howard & Scott, 2001).

III. System Proposed Method

This research work has proposed to implement by using the Support Vector Machine (SVM) model. The methods Support Vector Machine is commonly used and quite mature methods to generate good results. However, the support vector machine does not perform very well, when the data set has large and more noise, i.e. target classes are overlapping. In addition to this, the method has works by putting data point, above and below the classifying hyperplane there is no probabilistic explanation for the classification and underperform. Due to this and the related problem, we are applying the other method to solve drawbacks of the method. Therefore, we are using two main methods in this research work.

When the data set has large and more noise, we apply the Predictive Partial Match (PPM), and when the putting data points have, above and below the classifying hyperplane, we apply the statistical method for handling probabilistic explanation of classification. Therefore, we were implementing this research work by using multiple models such as support vector machine, predictive partial Match and statistical prediction to achieve the research work effectively. The Bi-gram/Tri-gram model is used to predict the most probable word and Part of speech (POS) pair given its code and previous words of part of speech (Sachin & Shilpa, 2007, pp. 1-9). Predictive partial match (PPM) is another method of a machine learning algorithm that is used to analyze words and characters based on the history of the sequence. The second method of this research work is statistical prediction method. This method is operated by counting the frequency of the word and the user is used recently as the probabilities of the words are used. For further achievement of this research work effectively, we shall he using the following supportive research methodology that is listed below.

IV. Context-Based Auto Text Completion Models for Amharic Language

A. Predictive Partial Match (PPM)

Predictive partial match is a popular and well-known algorithm that is predicted by using the history of the sequence. In addition to that, the history of word sequences should be available both in word encoder and decoder. In this algorithm, we only need to store those contexts that have occurred in the sequence which is being encoded. At the beginning of encoding, we will need to code the letters that have not occurred previously in the context. Finally, when the signal that the letter to be encoded has not been seen in the context, we will use <ESC> symbol. The BASIC algorithm is one of the implementations of the predictive partial match method. The steps are combined with each symbol which has not occurred in the context. Thereafter, escape symbol is encoded and each symbol is encountered by the count corresponding to that symbol is updated.

Assume the highest order context is 2, then the order is bounded from (-1, 0, 1, 2) context. Therefore, the context is categorized in -1-Order context, 0-Order context, 1-Order context, and 2-Order context. Example for count array for second order is discussed below in the given phrase "**This ~~b~~ is** ".

Table 4.1: Total cumulative of 2nd order PPM.

Context	Letter	Count	Cumulative count
Th	i	1	1
	<ESC>	1	2
	Total cumulative		2
hi	s	1	1
	<ESC>	1	2
	Total cumulative		2
is	~~b~~	1	1
	<ESC>	1	2
	Total cumulative		2
~~s~~**~~b~~**	i	1	1
	<ESC>	1	2
	Total cumulative		2
~~b~~i	s	1	1
	<ESC>	1	2
	Total cumulative		2

The above table shows how to calculate the lower and upper limit of the phrase "**This ~~b~~ is"** by using the count and cumulative count column on the 2nd order context.

$$L^n = (L)^{n-1} + \left[\frac{(U^{n-1} - L^{n-1} + 1) * \text{ cumulative count}(xn-1)}{\text{Total cumulative}}\right]$$

$$U^n = (L)^{n-1} + \left[\frac{(U^{n-1} - L^{n-1} + 1) * \text{ cumulative count}(xn)}{\text{Total cumulative}}\right] - 1$$

B. Support Vector Machine (SVM)

Support vector machine algorithm is used to identify the most appropriate boundary between the possible outputs by using a technique called the kernel trick which can be used for classification problems. To tell the support vector machine story, it is one of a supervised machine learning algorithms and it's focused on the idea of separating data with a large "gap" (Sachin & Shilpa, 2007, pp. 1-9). Since we develop context-based auto text completion, we are adopt the Amharic part of speech tagging system and the tagged data has been prepared manually by the language experts Hence, to use support vector machine the given word sequence is predicted by the part of speech tagging sequence. In addition to this, we combine the Predictive Partial Match (PPM) and Support Vector Machine (SVM) algorithm for getting better system accuracy. The predictive partial match and support vector machine algorithm is used to implement as a specialization of the support vector machine structure package for sequence tagging in the given problem and identify the correct word to be predicted. On the concept of using support vector machine for developing a system, it would require as many numbers of words with their class in the dictionary. The Amharic dictionary has constructed the data that has provided by the Walta Information Centre (WIA) and Ethiopia News Agency (ENA) roughly around 395,464 words and 750,000 sentences; SVM would need to learn classification for these many classes.

To learn good support vector machine classifier for classes, sufficiently large number of examples are required for all the classes i.e. a large training dataset which cover words from all these classes but the training time for support vector machine grows exponentially with the number of training examples.

However, for the given problem of predicting the correct word for a given code, one classifier per code is really what we need to learn the features used for support vector machine as similar to parameters used in the above graphical models i.e. the part of speech tag of the previous word and the given code (Apoorva, 2015, pp. 347-350).

C. Statistical Prediction

To accomplish the task of this research work effectively as well to increase the accuracy of the system, we are using statistical prediction as a supportive model for machine learning model such as predictive partial match and support vector machine. The statistical prediction model is supported when the possible outcome of the predictive word is excluded from the context-based model. Specifically, the statistical prediction model needs statistical data which are organized by the occurrence of words with their frequency, means constructing corpus based on the frequent occurrences of the word. Since the Amharic language has not a well-organized corpus, we prepare the corpus from 750,000 Amharic sentences that are collected from various sources that include governmental and private organizations. Moreover, 395,464 unique words are prepared in the form of the fixed lexicon. Fixed lexicon is the easiest method for statistical prediction, means the method is working by using the frequency of a word to predict throughout the general language (Mahamed &Vishal, 2015, pp. 274-279). This method uses a technique to make the prediction that, all words are arranged by their frequencies, and some words are at the top of the list. When the user wants to use this technique, first it is checked whether the desired word is among those words which are at the top (Mahamed &Vishal, 2015, pp. 274-279). If it is so, the system can predict the word and correct the sentence.

The corpus that we collected is referred to as "**Alemebante's Auto text completion corpus**" that is used in this work, is composed of 395,464 unique words. Table 4.2 shows the list of top 10 words frequently occurred from the whole distinct words, their frequencies, and the percentage of occurrence of that word compared to the total size of the corpus.

Table 4.2: Top 10 frequently occurring words.

Words	Word length	Occurrence of word frequency	Word frequency in %
ነው	2	58606	1.90
ላይ	2	42780	1.39
ውስጥ	3	23596	0.77
አቶ	2	22643	0.74
ወደ	2	20252	0.66
ጋር	2	20120	0.66
ጊዜ	2	16331	0.53
ማን	2	13134	0.43
የኢትዮጲያ	6	10201	0.33
እና	2	9900	0.32

V. System Architecture

The Architecture of the system has four basic auto text completion engine components; those are Start-Up Engine, Word Selector, Word Categorizations and Word Ranker. Those components are playing a big role to gain a highly accurate text prediction system for Amharic language (Apoorva, 2015, pp. 347-350).

Start-Up Engine component is used to start up the system for the first time to get one recognized characters from the character recognition engine. After receiving one character from the word repository, the component waits for the time frame set up, which is 1500 milliseconds, to initiate the Word Selector component. To start predicting the intended word, the prediction engine needs at least one character.

Once the **Word Selector** is initiated, it will start searching for words in the dictionary (Lexicon) of which the first characters match the recognized characters and initiate the Word categorization component. The phase uses to check whether the selected word is listed under the word repository or not, then it decides the system direction flow which is redirected to add as a new word or passing to the word categorization component.

On **word categorization** phase**,** the basic and most important task for this research work is running efficiently. Hence, to use word categorization, it compares the most related word from the word selector and initiate word categorization engine to identify whether the selected words have on their part of speech tagging or not. The selected word is available on the corpus it needs to identify the word category (Part Of Speech tagging) and assign to Bigram/Trigram components. The word ranker component will initiate while the selected word hasn't the word categorization.

Word Ranker, by considering the frequency of each word, it provides the rank of each word in the list of the corpus.

This phases for this research work is the supportive task for identifying that text is not predictive by using the context-based method. The method is working, the Words with the highest frequency will get the highest rank and those with the least frequency will get the least rank. In the case of two or more words having the same frequency; the word ranker will be listed of those words by considering the alphabetical order. The architecture of the system is as shown below in Figure 5.1.

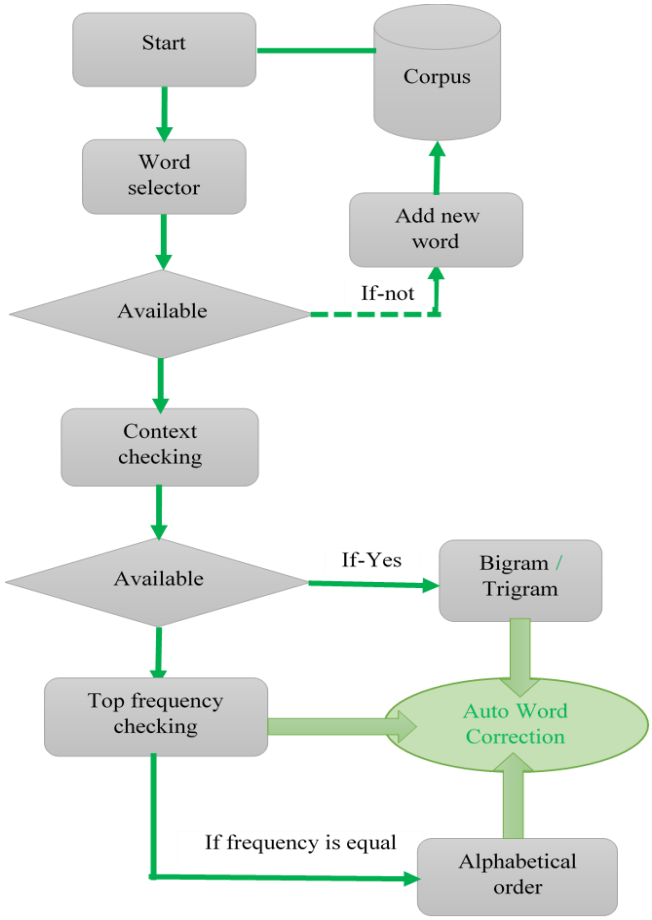

Figure 5.1: Text completion system architecture

VI. Experiment

A. *Implementation*

In order to achieve the objective of this research work we demonstrate the validity and the usability of the system.

Therefore we have selected the designing algorithm and different implemented tools for auto text completion of Amharic language.

The development of environment makes the necessary experiment on the system, so we have used different tools. The tools we are using are Python and SQLite, excel and Microsoft word under the windows operating system.

B. *Result of the Experiment*

The result of the experiment is evaluated by 3000 Amharic test set of words that are collected randomly from Emails, governmental news, official letters and

Books. The result in Table 5.1 depicts the total number of test set data that are categorized by the word length and the data

were correctly predicted or not predicted. The column, both correctly predicted and incorrect words are directly fed in the system and evaluated manually by using precision, recall and F-measure (Apoorva, 2015, pp. 347-350). Based on the test set data the average system accuracy has reached at 97.8%.

Precision: - The result shows that the percent of a selected item is correct on the system and how to predict all words that are out of error.

$$\textbf{Precision} = \frac{\textbf{correct}}{\textbf{output length}}$$

Recall: -The result is expressed as the output percentage are the percent of a correct item is selected.

$$\textbf{Recall} = \frac{\textbf{correct}}{\textbf{reference length}}$$

F-measure: - It is a combination of measurements of precision and recall.

$$\textbf{F} - \textbf{measure} = \frac{\textbf{2} * (\textbf{Precision} * \textbf{Recall})}{(\textbf{Precision} + \textbf{Recall})}$$

Tables 6.1 and Table 6.2 show the list of test data that is categorized by the word length into 9 groups. The test data is collected from different sources and it uses to know the efficiency and accuracy of the system that is categorized by the correctly predicted and incorrect word out of the total 3000 test set data. Finally, this data shows the precision, recall, and F-measure of the system.

Table 6.1: Correct and incorrect testing word result

Word Length	Testing words	Incorrect word	Correctly predicted
2-Letter	573	23	550
3-Letter	529	12	517
4-Letter	696	31	665
5-Letter	504	19	485
6-Letter	301	5	296
7-Letter	209	2	207
8-Letter	88	1	87
9-Letter	65	1	64
10-Letter	35	0	35
Total	3000	94	2906

Table 6.2: Average correct word result

Word Length	Average accuracy
2-Letter	95.98%
3-Letter	97.73%
4-Letter	95.54%
5-Letter	96.23%
6-Letter	98.33%
7-Letter	99.04%
8-Letter	98.86%
9-Letter	98.46%
10-Letter	100%
Total	97.8%

CONCLUSION

The use of Amharic text completion system is to handle the misspelled texts during communicating with each other. The communication gap may happen when the user faces the computer machine and mobile device. Therefore developing the text completion system is important to the group of people who have to use the Amharic language on the computer machine. Text completion is a process of correcting words of the user interests to write after the first character. While Amharic text completion has been developed by the context and statistical-based, we are especially concerned on preparing part of speech tagging for Amharic word and lexicon of the statistical information obtained from the corpus. Since the text writing using computer and mobile needed recognition, follows a complex process that also requires a training set and testing set data for machine learning (Shimeles, 2005; Negussie, 2006). When text completion is used for handwriting recognition, the possible errors of word and character recognition are also minimized. To accomplish the task of context-based text completion for Amharic language, it needs a good collection of sentences that have been tagged by the categorization of their part of speech and organized words which will be used as a corpus is a must. However, there is no such kind of well-organized data for Amharic language; it needs to collect a huge amount of data from a different source. In the data analyses phase, we need to do the following main tasks like POS tagging, the frequency of a unique word, average word-length of Amharic language, the most frequently used Amharic word-length and the experiment conducted, our word prediction system has shown a prediction of accuracy of 97.8%. In this work, the searching efficiency of the system has not been taken into consideration because, searching efficiency is the limitation of the research work.

In general, it is concluded that conclude the context-based word prediction system performs better than the traditional frequency-based method. On the other hand, in order to increase the accuracy of the system, it is better to use a combination

of context and frequency-based method. Therefore frequency-based method is the best supportive model of a context-based model.

Different graphical models were analyzed to judge in which model there is best causal relationship between PPM and SVM. The model is the combination of PPM and SVM models that are used for sequence tagging and was found an appropriate solution for the given problem due to a large number of classes. For unseen words, since Emailing text is normally short sentences, we will apply a better smoothing method like the Bi-gram/ Tri-gram model. Phonetic encoding scheme with more precision in the given domain would help improve the system performance as well to get better results.

Finally, it is believed that the system can give alternative service to improve the exchange of accurate text messages with each other especially, text entry method for E-mailing, SMS, Grammar checking and any other social Media specifically for Messenger and WhatsApp application.

Appendix A1. List of tag sets with their description

No	Tag set	Description
1	<N>	Noun
2	<PRON>	Pronoun
3	<ADJ>	Adjective
4	<V>	Verb
5	<ADV>	Adverb
6	<CONJ>	Conjunction
7	<NUM>	Number
8	<PREP>	Preposition
9	<PUNC>	Punctuation
10	<AUX >	Auxiliary Verbs
11	<NP>	Noun Phrase
12	<VP>	Verb Phrase
13	<ADJP>	Adjective Phrase
14	<PRONP>	Pronoun Phrase
15	<NPC>	Noun Phrase with Conjunction
16	<VPC>	Verb Phrase with Conjunction

ACKNOWLEDGMENT

First of all, I would like to thank the Almighty God for being completed this research work successfully. All the time, God made me strong enough, exerting great force, and the valuable gift from him and it's really further side of my expectation. An excitation, big thanks to all who supported me up until the present to come.

Next, I have to convey my thanks to the Ethiopia embassy in New Delhi, India next to the Ministry Education of Ethiopian government who have helped me by covering all educational expenses.

Also, I would like to thank Addis Ababa University they have make available library space, internet service and other faculty (linguistic) for extending enriched guidance and advice in accomplishing this gigantic task. Convey my sincere and thankful gratitude to my friends they have been more than a friend and it helps starting from the title selection till to the final stage of this work, they have been very cooperative and supportive. Besides that, my thankfulness is for all the time I feel grateful to stay in Punjabi University Patiala and succeed in attaining throughout this work.

Finally, I would like to thank Harrelson next to YouTube who has helped me in proceeding with this research work with our ideas to their actual implementation joint. They were the ones who have contributed to this thesis beyond the expectation and they supported to build confidence to this research work.

REFERENCES

Alemebnate, M., & Vishal, G. (2015). Amharic Text Predict System for Mobile Phone. International Journal of Computer Science Trends and Technology (IJCST), 3 (4), 113-118.

Apoorva, S. (2015). Context based word prediction for Mobile devices with T9 keypads. International Journal of Scientific & Engineering Research, 6(12), 347-350.

Hedy, K., Eugene S., Terry J., Howard G., & Scott M. (2001). Linguistically Optimized Text Entry on a Mobile Phone. *WordWise Eatoni Ergonomics, Inc: 1-8.* Retrieved Jun 16, 2021, from https://en.wikipedia.org/wiki/LetterWise.

Gebremichael, M. (2011). Federalism and conflict management in Ethiopia: case study of Benishangul-Gumuz Regional State (PhD). University of Bradford, United Kingdom.

Mahamed, D., & Vishal, G. (2015). Predictive Text Entry Method for Somali Language on Mobile Phone. International Journal of Computer Science Trends and Technology (IJCST), 3(3), 274-279.

Negussie, D. (2006). Writer Independent Online Handwriting Recognition for Ethiopic Characters. Unpublished master's thesis, Department of Computer Science Addis Ababa University.

Sachin, A., & Shilpa, A. (2007). Context Based Word Prediction for Texting Language. Conference RIAO2007, *Pittsburgh PA (pp. 1-9), U.S.A. - Copyright C.I.D. Paris*.

Shimeles, A. (2005). Online Handwriting Recognition for Ethiopic Characters. Unpublished master's thesis, Department of Computer Science Addis Ababa University.

***