Analyzing word Error Rate using Semantic Oriented Correction Approach on Bing Search Engine

Theint Zarni Myint
Researcher, UCSY
University of Computer Studies, Yangon
Yangon, Myanmar

Abstract— **Voice input has already required in many application environment. This paper represents to search the data that are users' query using android application. So, this system used the speech recognition that is done via the Internet, connecting to Google's server. The system acquires speech at run time through a microphone and processes the sampled speech to recognize the uttered text. The recognized text can be stored in a file. We are developing this on android platform using eclipse workbench. The speech-to-text system directly acquires and converts speech to text. The users can search the data using the application that can search for famous places in Myanmar. The application is adapted to input messages in English. But this output is not matched with users' data because English phonetics and Myanmar phonetics are different. Therefore, this system proposed a method for obtaining more detail about actual translation errors in the generated output by using the word error rate (WER) based on the semantic oriented approach. We investigate the use of WER for automatic error analysis using a dynamic programming algorithm like Levenshtein distance over POS tagging. This paper gives a better overview of the nature of translation errors as well as ideas of where to put efforts for possible improvements of the translation system. After this process, the resulting text will be matched and they are given as input to Bing Search Engine to be searched.**

Keywords—Speech Recognition; WER; POS tagging; Semantic ErrorCorrection; Bing Search

I. INTRODUCTION

Mobile phones have become an integral part of our Everyday life, causing higher demands for content that can be used on them. Smart phones offer customer enhanced methods to interact with their phones but the most natural way of interaction remains speech. Market for smart mobile phones provides a number of applications with speech recognition implementation. Voice recognition technology is the process of identifying and understanding the voice signals of a user, which is converted into text or commands for a program.

In this work the voice recognition technology is applied into a laboratory information system for identifying each technician's voice. However, due to the low recognition rate of the current speech recognition systems, in order to increase the accuracy of information retrieval, appropriate post-processing is required.

Therefore a variety of automatic evaluation measures have been studied over the last years. The most widely used are Word Error Rate (WER), Position independent word Error Rate (PER), the BLEU score and the NIST score. These measures have shown to be valuable tools for comparing different systems as well as for evaluating improvements within one system. However, these measures do not give any details about the nature of translation errors. Therefore some more detailed analysis of the generated output is needed in order to identify the main problems and to focus the research efforts. A framework for human error analysis has been proposed in, but as every human evaluation, this is also a time consuming task.

This article shows a framework for calculating WER over POS tagging that is for estimating the contribution of each POS tagging to overall word error rate. Although this work focuses on POS classes, the method can be easily extended to other types of linguistic information. The translation system used is the phrase-based statistical machine translation system. This work presents a framework for extraction of linguistic details from standard word error rates WER use for an automatic error analysis. We presented a method standard word error rates WER over basic POS classes. We also carried out a detailed analysis of inflectional errors which has shown that the results obtained by our method correspond to those obtained by a human error analysis. In addition, we proposed a method for analyzing missing word errors.

We suggest a more robust semantic-oriented error correction approach, which is an improvement over but easily integrated into the previous fragile lexical-based approaches. In our approach, in addition to the lexical information, we use high level syntactic and semantic information of the words in the speech transcription. We obtain semantic information from some knowledge base such as general thesauri and a special dictionary which we construct by ourselves to contain some domain specific knowledge to the target application. This work presents framework for extraction of linguistic details from word error rate (WER).

II. BACKGROUND THEORY

A. *Speech Recognition on Android*

Speech recognition is the process of converting voice signal into corresponding text or commands. In this research this kind of voice to text conversion is done at the time of smear test result entry. The technician speaks the result of the smear test of each patient through the microphone and the converted text is inserted in the corresponding result field and stored in the database. The speech recognition application in smart phones is incorporated to implement this speech to text conversion operation. Figure 1 depicts the basic system architecture of the speech recognizer.

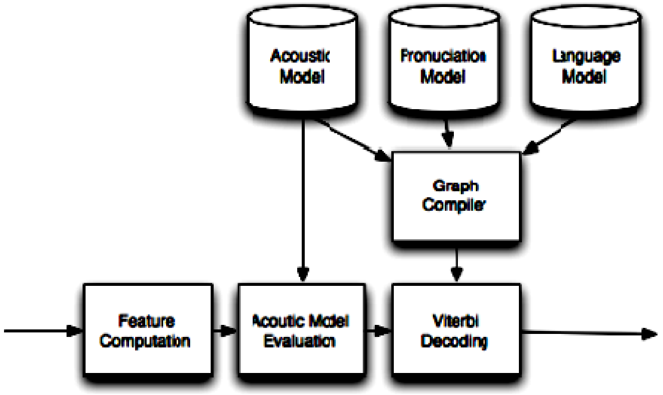


Fig. 1. Basic Block Diagram of Speech Recognizer

Speech recognition system can be divided into several blocks: feature extraction, a coustic models database which is built based on the training data, dictionary, language model and the speech recognition algorithm. Analog speech signal must first be sampled on time and amplitude axes, or digitized. Samples of speech signal are analyzed in even intervals. This period is usually 20 ms because signal in this interval is considered stationary. Speech feature extraction involves the formation of equally spaced discrete vectors of speech characteristics. Feature vectors from training database are used to estimate the parameters of acoustic models. Acoustic model describes properties of the basic elements that can be recognized. The basic element can be a phoneme for continuous speech or word for isolated words recognition.Dictionary is used to connect acoustic models with vocabulary words. Language model reduces the number of acceptable word combinations based on the rules of language and statistical information from different texts. Speech recognition systems, based on hidden Markov models are today most widely applied in modern technologies. They use the word or phoneme as a unit for modeling. The model output is hidden probabilistic functions of state and can't be deterministically specified. State sequence through model is not exactly known. Speech recognition systems generally assume that the speech signal is a realization of some message encoded as a sequence of one or more symbols . To effect the reverse operation of recognizing the underlying symbol sequence given a spoken utterance, the continuous speech waveform is first converted to a sequence of equally spaced discrete parameter vectors.

B. *Standard Work Error Rate (Overview)*

The word error rate (WER) is based on the Levenshtein distance (Levenshtein, 1966) - the minimum number of substitutions, deletions and insertions that have to be performed to convert the generated text hyp into the reference text ref . A shortcoming of the W ER is the fact that it does not allow reordering of words, whereas the word order of the hypothesis can be different from word order of the reference even though it is correct translation. In order to overcome this problem, the position independent word error rate (PER) compares the words in the two sentences without taking the word order into account. The P ER is always lower than or equal to the W ER. On the other hand, shortcoming of the PER is the fact that the word order can be important in some cases. Therefore the best solution is to calculate both word error rates.

Calculation of WER: The WER of the hypothesis hyp with respect to the reference ref is calculated as

$$WER = \frac{1}{N^{*}_{ref}} \sum \min_{r} d_L(ref_{k,r}, hyp_k)$$

(1)

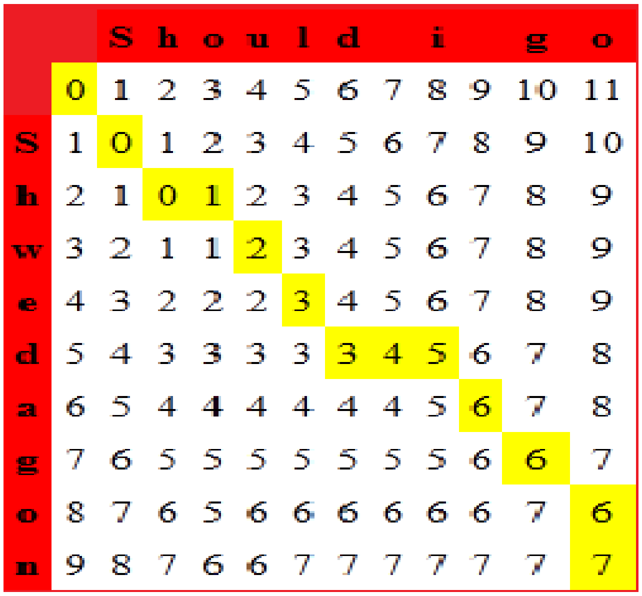
where $d_L(ref_{k,r}, hyp_k)$ is the Levenshtein distance between the reference sentence $ref_{k,r}$ and the hypothesis sentence hyp_k. The calculation of W ER is performed using a dynamic programming algorithm.

C. *Levenshtein distance*

Levenshtein distance (LD) is a measure of the similarity between two strings, the source string (s) and the target string (t). The distance is the number of deletions, insertions, or substitutions required to transform s into t. The greater the Levenshtein distance, the more different the strings are. In our case, the source string is the input, and the target string is one of the entries in the dictionary.

The Levenshtein algorithm (also called Edit-Distance) calculates the least number of edit operations that are necessary to modify one string to obtain another string. The most common way of calculating this is by the dynamic programming approach. A matrix is initialized measuring in the (m,n)-cell the Levenshtein distance between the m-character prefix of one with the n-prefix of the other word. The matrix can be filled from the upper left to the lower right corner. Each jump horizontally or vertically corresponds to an insert or a delete, respectively. The cost is normally set to 1 for each of the operations. The diagonal jump can cost either one, if the two characters in the row and column do not match or 0, if they do. Each cell always minimizes the cost locally. This way the number in the lower right corner is the Levenshtein distance between both words.

An example of finding LD between the user speech "Shwedagon" and the speech recognition output "should I go" is shown in figure 2.



		S	h	o	u	l	d		i		g	o
	0	1	2	3	4	5	6	7	8	9	10	11
S	1	0	1	2	3	4	5	6	7	8	9	10
h	2	1	0	1	2	3	4	5	6	7	8	9
w	3	2	1	1	2	3	4	5	6	7	8	9
e	4	3	2	2	2	3	4	5	6	7	8	9
d	5	4	3	3	3	3	3	4	5	6	7	8
a	6	5	4	4	4	4	4	4	5	6	7	8
g	7	6	5	5	5	5	5	5	5	6	6	7
o	8	7	6	5	6	6	6	6	6	6	7	6
n	9	8	7	6	6	7	7	7	7	7	7	7

Fig. 2. Example of Levenshtein Distance

D. *WER decomposition over POS tagging*

The dynamic programming algorithm for WER enables a simple and straightforward identification of each erroneous word which actually contributes to WER. Let err$_k$ denote the set of erroneous words in sentence k with respect to the best reference and p be a POS class. Then n (p, err$_k$) is the number of errors in err$_k$ produced by words with POS class p. It should be noted that for the substitution errors, the POS class of the involved reference word is taken into account. P OS tags of the reference words are also used for the deletion errors, and for the insertion errors the P OS class of the hypothesis word is taken. The W ER for the word class p can be calculated as:

$$WRE(p)=\frac{1}{N^{*}_{ref}}\sum_{k-1}^{K} n(p,err_k) \quad (2)$$

The sum over all classes is equal to the standard overall WER. An example of a reference sentence and hypothesis sentence along with the corresponding P OS is shown in figure 3. The W ER errors, i.e. actual words participating in WER together with their P OS classes can be seen in table1. The reference words involved in W ER are denoted as reference errors, and hypothesis errors refer to the hypothesis words participating in WER.

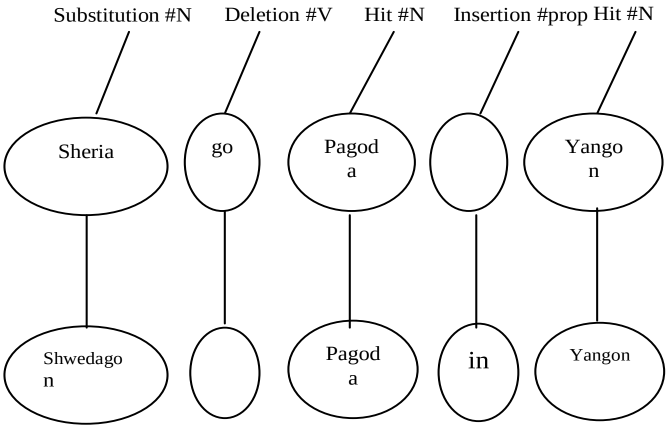


Fig. 3. Example for illustration of actual errors : a POS tagged reference sentence and a corresponding hypothesis sentence

Table I WER Errors: Actual Words Which Are Participating In The Word Error Rate And Their Corresponding POS Tagged

Reference errors	Hypothesis errors	Error type
Sheria #N	Shwedagon #N	substitution
Go #V		deletion
	In #prop	insertion

Standard WER of the whole sentence is equal to 3/5=60% . The contribution of nouns is WER (N)= 1/5=20%, of verbs WER (V) =1/5=20% and of proposition WER (prop) =1/5= 20%.

E. *Web Search Engine*

A web search engine is a software system that is designed to search for information on the World Wide Web. The search results are generally presented in a line of results often referred to as search engine results pages (SERPs). The information may be a mix of web pages, images, and other types of files. Some search engines also mine data available in databases or open directories. Unlike web directories, which are maintained only by human editors, search engines also maintain real-time information by running an algorithm on a web crawler. The architecture of the speech recognition like voice search on World Wide Web (WWW) as follows:

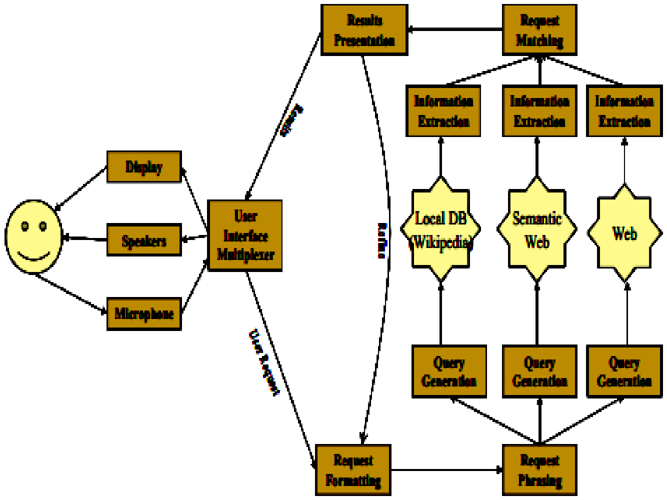


Fig. 4. Architecture of voice search on World Wide Web

F. *Feature of Bing Search Engine.*

1) Interface Feature

- Daily changing of background image. The images are mostly of noteworthy places in the world, though it sometimes displays animals, people, and sports. The background image also contains information about the element(s) shown in the image.
- Video homepage for HTML5 enabled browsers on occasional events, similar to the daily background images.
- Images page shows the main picture from that day and four searches that refers to that image with three preview pictures per search term.
- Left side navigation pane. Includes navigation and, on results pages, related searches and prior searches.
- Right side extended preview which shows a bigger view of the page and gives URLs to links inside of the page.
- Sub links – On certain search results, the search result page also shows section links within the article (this is also done on other search engines, including Google)
- *Enhanced view* where third party site information can be viewed inside Bing.
- On certain sites, search from within the website on the results page
- On certain sites, Bing will display the Customer Service number on the results page.

2) Media Feature

- Ail Preview – where, by hovering over a video thumbnail, the video automatically starts playing.
- Image search with continuous scrolling images results page that has adjustable settings for size, layout, color, style, and people.
- Advanced filters or advanced (query) operators – allow users to refine search results based on properties such as image size, aspect ratio, color or black and white, photo or illustration, and facial features recognition.
- Video search with adjustable setting for length, screen size, resolution, and source.

G. *Bing Spelling Suggestion*

Bing Dictionary enables users to quickly search for definitions of English words. Bing Dictionary results are based on Microsoft Encarta World English Dictionary. In addition, Bing Dictionary also provides an audio player for users to hear the pronunciation of the dictionary words.

Bing's spelling suggestion technology can suggest alternative corrections for the often made typos, misspellings, and keyboarding errors. At the core, Bing has a colossal database of billions of online web pages containing trillions of term collections and n-gram words that can be used as groundwork for all kinds of linguistic applications such as machine translation, speech recognition, spell checking, as well as other types of text processing problems. Fundamentally, Bing's spelling suggestion algorithm is based on the probabilistic n-gram model originally proposed by Markov for predicting the next word in a particular sequence of words. In brief, an n -gram is simply a collocation of words that is n words long.

For instance, "The boy" is a 2-gram phrase also referred to as bigram, "The boy scout" is a 3-gram phrase also referred to as trigram, "The boy is driving his car" is a 6-gram phrase, and so forth. The Bing's algorithm automatically examines every single word in the search query for any possible misspellings. It tries first to match the query, basically composed of ordered association of words, with any occurrence alike in Bing's database of indexed web pages; if the number of occurrence is high, then the query is considered correct and no correction is to take place.

However, if the query was not found, Bing uses its n-gram statistics to deduce the next possible correct word in the query. Sooner or later, an entire suggestion for the whole misspelled query is generated and displayed to the user in the form of "Including results for spelling-suggestion". For example, searching for the word "conputer" drives Bing to suggest Including results for Computer". Likewise, searching for "Thehord disk sturage" drives Bing to suggest "Including results for the hard-disk storage". Searching for the proper name "jahn cenedy " drives Bing to suggest "Including results for John Kennedy".

This system is constructed by integrating above methods and in addition the semantic oriented error correction approach discuss in section 3. On the other hand, our methods take far less training corpus, and it is possible to implement the method easily and in a short time to obtain the same or better error correction rate because it utilizes the semantic information of the application. This system used the Bing Spelling suggestion to be accurate our results if it will be needed.

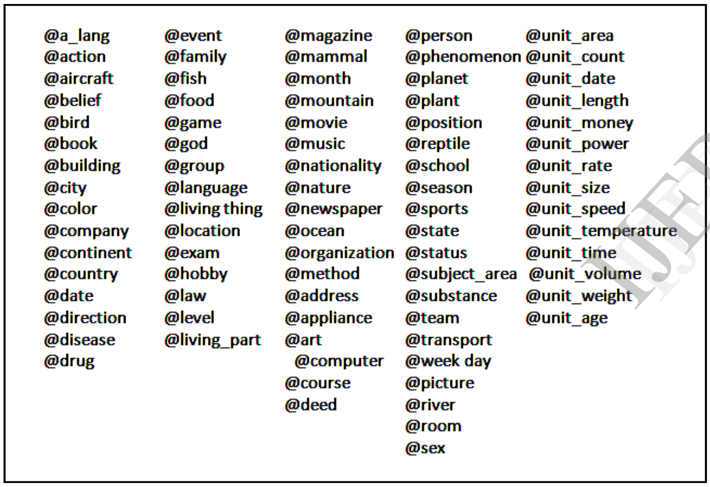
III SEMANTIC ORIENTED ERROR CORRECTION APPROACH

This section presents a semantic-oriented approach to correct erroneous outputs of speech recognizer using knowledge. This system focuses on real-world error detection and correction over our application because this approach is based on the domain knowledge for performing semantic level error detection and correction.

A. *Lexico-Semantic Pattern*

A lexico-semantic pattern (LSP) is a structure where linguistic entries and semantic types are used in combination to abstract certain sequences of the words in a text. It has been used in the area of natural language interface for database.

In an LSP, linguistic entries consist of words, phrases and part-of-speech (POS) tags, such as 'YMCA,' 'Young Men's Christian Association,' and 'NNP'. Semantic types consist of common semantic classes and domain specific (or user-defined) semantic classes. The common semantic tags again include attribute-values in databases and predefine semantic category values such as '@location' for location name like "Yangon" and '@pagoda' for pagoda name like "Shwedagon".



@a_lang
@action
@aircraft
@belief
@bird
@book
@building
@city
@color
@company
@continent
@country
@date
@direction
@disease
@drug

@event
@family
@fish
@food
@game
@god
@group
@language
@living thing
@location
@exam
@hobby
@law
@level
@living_part

@magazine
@mammal
@month
@mountain
@movie
@music
@nationality
@nature
@newspaper
@ocean
@organization
@method
@address
@appliance
@art
@computer
@course
@deed

@person
@phenomenon
@planet
@plant
@position
@reptile
@school
@season
@sports
@state
@status
@subject_area
@substance
@team
@transport
@week day
@picture
@river
@room
@sex

@unit_area
@unit_count
@unit_date
@unit_length
@unit_money
@unit_power
@unit_rate
@unit_size
@unit_speed
@unit_temperature
@unit_time
@unit_volume
@unit_weight
@unit_age

Fig 5 **Common semantic category values**

In domain-specific application, well defined semantic concepts are required, and the domain-specific semantic classes represent these requirements. The domain specific semantic classes include special attribute names in databases such as '%action' for 'active' and 'inactive,'. And also includes semantic category names, such as '%hobby' for 'reading' and 'recreation,' for which the user wants specific meaning in the application domain.

The words in a query sentence are converted into the LSP through two steps. First, a morphological analysis is performed, which segments a sentence of words into morphemes, and adds POS tags into morphemes. Next, each morpheme of the sentence is converted into a suitable semantic symbol by searching several types in the semantic dictionaries. For semantic-oriented error correction, this system constructed a domain knowledge, which consists of domain dictionary, thesaurus, and template queries. Query sentences are semantically abstracted by LSP's, and are collected for the template database. The following table shows example of template abstracted by LSP converting in a predefined domain of famous places.

Table II Example of the template abstracted by LSP

Phrases	**LSP**
Shwedagon Pagoda	%name@pagoda
Sulay Pagoda	
Mandalay Mountain	%name@mountain
Zwekavin Mountain	

B. *Building of domain knowledge*

The domain dictionary is a subset of the general semantic category dictionary, and focuses only on the narrow extent of the knowledge its concerns, since it is impossible to cover all the knowledge of the world in implementing an application. On the other hand, the ontology dictionary reflects the pure general knowledge of the world; hence it performs a supplementary role in extracting semantic information. The domain dictionary provides the specific vocabulary which is used at semantic representation task of a user query and the template database. Assuming that some speech statements for a specific target domain is predefined, a record of the template database is made up of a fixed number of LSP elements, such as POS tags, semantic tags, and domain-specific semantic classes.

For semantic-oriented error correction, we constructed a domain knowledge, which consists of domain dictionary, thesaurus, and template queries. Query sentences are semantically abstracted by LSP's, and are collected for the template database.

C. *Semantic Error Correction*

In this section will show the working mechanism of post error correction of a speech recognition result using the domain knowledge of template database and domain specific dictionary. The overall process is divided into two stages: semantic recovery and lexical recovery stage. In semantic recovery stage, recognized query is converted into the corresponding LSP. The converted LSP may be ill formed depending on the recognized query. Semantic recovery is performed by replacing these syntactic or semantic errors using a semantic confusion table.

We used pre-collected template database to recover the semantic level errors, and the technique for searching data. The semantic confusion table provides the matching cost to the dynamic programming search process. The 'minimum edit distance' using Levenshtein distance between two words is originally defined as the minimum number of deletions, insertions, and substitutions required for transforming one word into the other. Our process is demonstrated on the following figure.

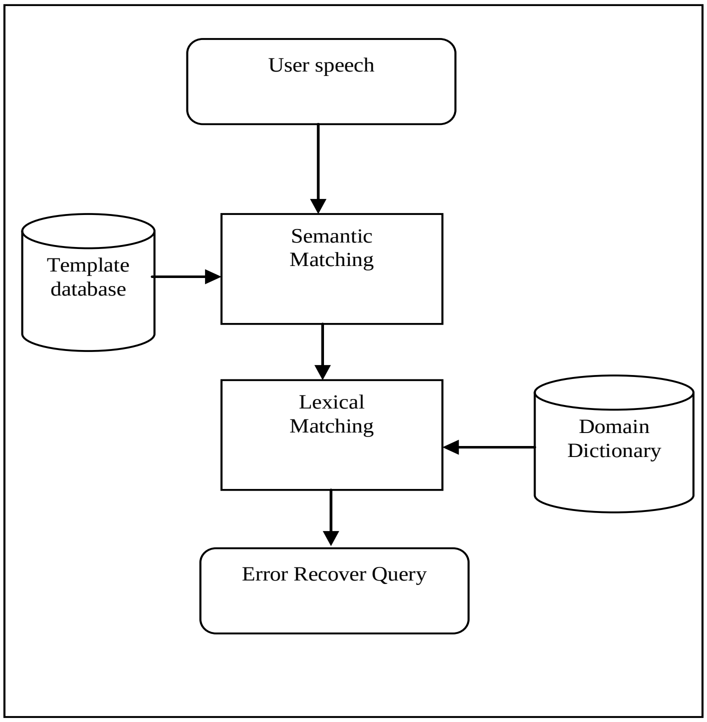


Fig. 6. Semantic Error Correction Process

After this procedure, lexical recovery is performed in the next stage. Recovered semantic tags and the erroneous query produced by speech recognizer are the clue of lexical recovery. Erroneous query and recovered LSP template are aligned by dynamic programming again, and then some candidates are discovered as the most similar words to the original input words in the domain dictionary or ontology dictionary. As many lexical candidates under the same semantic class may exist, we select the most similar one as the final correction word using the minimum edit distance on the lexical level. In minimum edit distance, cost function is important to the sensitive matching, so we use a confusion table of lexical variation at this stage. an example of semantic error correction process using the same data is shown as follow:

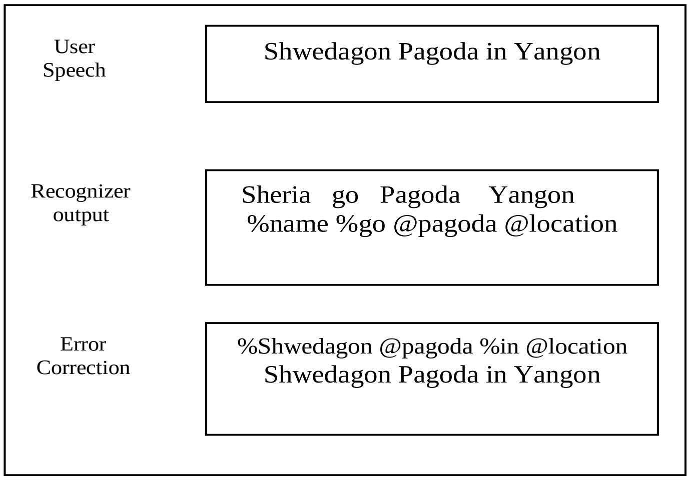

User Speech
Shwedagon Pagoda in Yangon
Recognizer output
Sheria go Pagoda Yangon
%name %go @pagoda @location
Error Correction
%Shwedagon @pagoda %in @location
Shwedagon Pagoda in Yangon

Fig. 7. Example of semantic-oriented error correction

After all of the above processes have already finished, the output will be sent to the Bing Search to retrieve user query.

IV RESULTS AND EXPERIMENTS OVER APPLICATION

A. Application functionality Principle

This application integrates direct speech input enabling user to record spoken information as text message, and send it to Bing Search Engine. After application has been started display on mobile phone shows button which initiate voice recognition process. When speech has been detected application opens connection with Google's server and starts to communicate with it by sending blocks of speech signal. Simultaneously the figure of waveform is generated on the screen. Speech recognition of the received signal is preformed on server. Google has accumulated a very large database of words derived from the daily entries in the Google search engines well as the digitalization of more than 10 million books in Google Book Search project.

The database contains more than 230 billion words. If we use this kind of speech recognizer it is very likely that our voice is stored on Google's servers. This fact provides continuous increase of data used for training, thus improving accuracy of the system. When process of recognition is over, user can see the list of possible statements. Process can be repeated clicking on the button Image Button. Pressing the most accurate option, selected result is entered into interface for searching the data.

B. Results over application

While speech is both convenient and effective as an input method, especially as an alternative to typing on tiny mobile keyboards, spoken output is very limited given its sequential nature. While most mobile speech apps require the user to press a button to initiate recording, only some require the user to manually stop the recording after speaking by pressing it again, or pressing another button.

Firstly, the voice recognition application that handles the intent processes the voice input, then passes the recognized string back to our application. The start activity of our application has a button to speak ,a edit text editor for accepting the output which is error correction from speak recognizer using our proposed system and a web view which is applying the Bing Search Engine for searching this query. It is shown in the following figure:

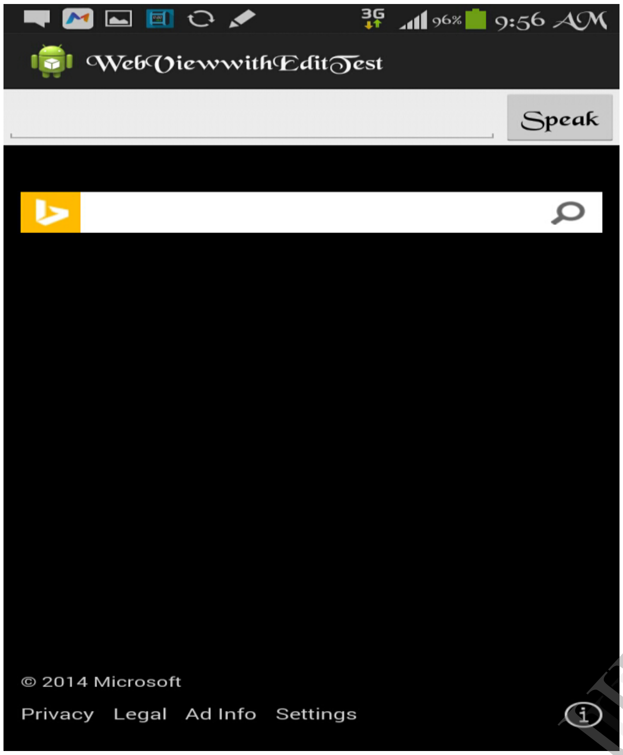
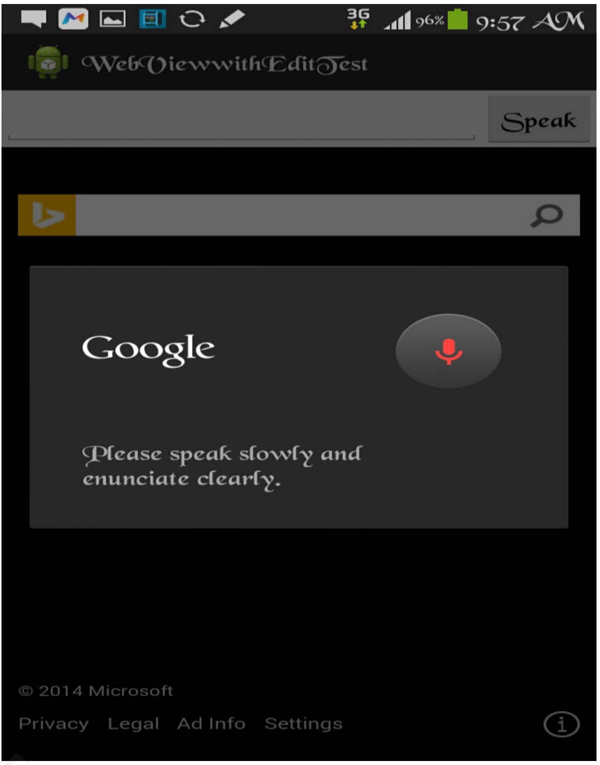


Fig. 8. Start activity of Application

Secondly, after the user press the speak button, it is displayed the screen to talk about the query that is shown as in figure 9. This application only can search the famous places in Myanmar like pagoda, lake, city etc.

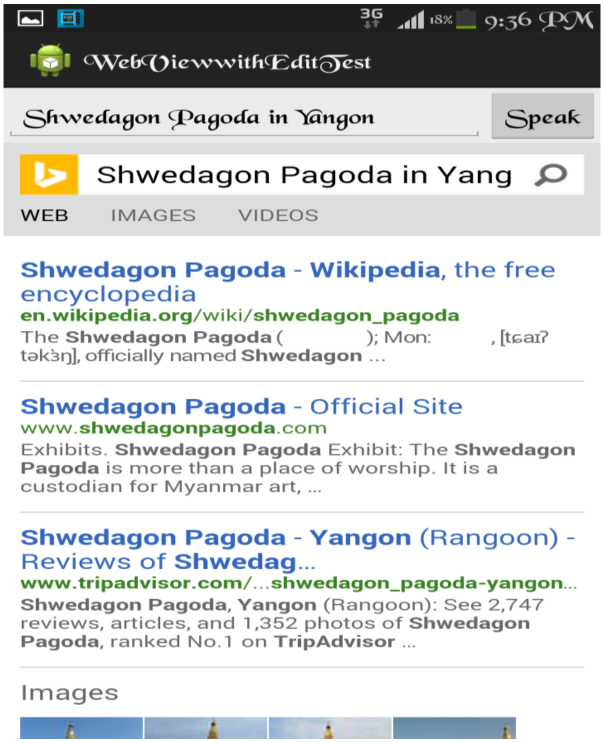


Please speak slowly and enunciate clearly.

Fig . 9. Enable process after clicking speak button

This application is its adaption only for English language, and the need for permanent Internet connection.

If it has no connection , it will be present on the screen like the dialog box "Connection Error" including "Cancle" and Try again". Lastly, we consider the screens in Figure 8 which show the results displayed for the same "Shwedagon Pagoda in Yangon" query using Bing Search feature on Android.



Shwedagon Pagoda in Yangon

Shwedagon Pagoda in Yang

Shwedagon Pagoda - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/**shwedagon_pagoda**
The **Shwedagon Pagoda** ( ); Mon: , [tɕaɪʔ təkɜŋ], officially named **Shwedagon** ...

Shwedagon Pagoda - Official Site
www.**shwedagonpagoda**.com
Exhibits. **Shwedagon Pagoda** Exhibit: The **Shwedagon Pagoda** is more than a place of worship. It is a custodian for Myanmar art, ...

Shwedagon Pagoda - Yangon (Rangoon) - Reviews of **Shwedag**...
www.tripadvisor.com/...**shwedagon_pagoda-yangon**...
Shwedagon Pagoda, Yangon (Rangoon): See 2,747 reviews, articles, and 1,352 photos of **Shwedagon Pagoda**, ranked No.1 on **TripAdvisor** ...

Images

Fig. 10 . Process and give output

Making sure the recognizer at least starts with a clean and complete recording of what the user actually said is key for any speech application.As we've seen, this is far from automatic and many strategies are currently in use. It may be that some are equally effective and address different user preferences. However, we are also likely to discover that some are simply more effective.

C. Experimental Results in the application

` In this section, we show experimental results of the proposed methods. The machine translation system used in this work is based on the semantic error correction approach. The translation is performed in direction over English phoneme and Myanmar phoneme.The error analysis is done on English output. We used WER and POS tagging to bee this output is matched with our traditional names like Shwedagon pagoda.

So, our system analyzed the WER is done over the ten main POS classes: nouns (N), verbs (V), adjectives(A),adverbs (ADV), pronouns (PRON), determiners (DET), prepositoin (PREP), conjunctions (CON), numerals (NUM) and punctuation marks (PUN). In this way, all references and hypotheses are provided with POS tags and based form. Apart from that , the distribution of of missing words ten POS classes is estimated using WER.Figure 9 presents the decomposition of WEr over basics POS classes for English and Myanmar language using the test corpus about 100 sentenses and 280000 running words. The largest part of word error rate come from the two most important word classes, namely nouns and verbs, and that the keast critical classes are punctuations, conjunctions and numbers

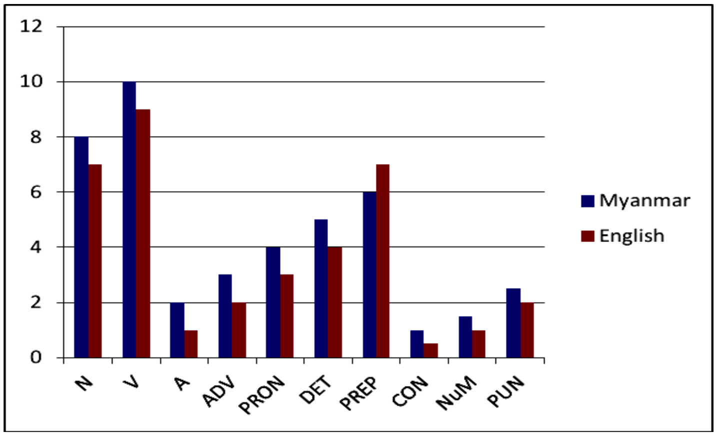


Fig. 11. Decomposition of WER over ten basic POS classes for English Phonatic and Myanmar Phonatic output

In addition, the decomposed error rates give an idea of where to put efforts for possible improvements of the system. For example, working on improvements of verb translations could reduce up to about 10%, working on nouns up to 8 %.

Therefore, we applied the method improved our system using the semantic error correction approach. This system divided the 365 queries into the training sets and the test set and evaluated results of validation for our model. For our processor, we constructed 250 semantic templates for evaluating our systems and 4528 entries of domain dictionary. The experiments results of word accuracies is 79~ 87 % on the utterances in our application.

V CONCLUSION

With the development of software and hardware capabilities of mobile devices, there is an increased need for device-specific content, what resulted in market changes. Speech recognition technology is of particular interest due to the direct support of communications between human and computers. The objective for the future is development of models and databases for multiple languages which could create a foundation for everyday use of this technology worldwide. The main goal of application is to allow sending of text messages based on uttered voice messages. Testing has shown simplicity of use and high accuracy of data processing. The LSP correction has high possibility of generating semantically correct correction due to the massive use of semantic contexts. Hence, it shows the high performance in our application.

In this paper demonstrate the results of recognition give user option to select the most accurate result using semantic error correction approach. This work is planned to implement the model of speech recognition for matching user speech and the output results. The limitation of this system is only applied on the Internet. If the connection is lost, this application will not be used.

VI ACKNOLEGEMENT

This work was partly supported by Dr. Myo Kay Khaing who is our supervisor and my respectfully teachers in University of Computer studies, Yangon. We thank Mr. Shin Dong Chul, Yangon Center Manager, Myanmar ICT Park, Hlaing Campus, 11052, Yangon, Myanmar who guideline to implement our method.

REFERENCES

[1] A. Fujii, K. Itou, and T. Ishikawa, "A method for open-vocabulary speech-driven text retrieval," in proceeding of the 2002 conference on Empirical Methods in Natural Language Processing, pp.188-195, 2002.

[2] H.Jung, G. Geunbae Lee, W.Choi, K.K. Min, and J. Seo, "Multi-lingual question answering with high portability on relational databases," IEICE transactions on information and systems, vol E-86D, No2, pp.306-315, 2003..

[3] R. Malouf., "A comparison of algorithms for max-imum entropy parameter estimation", In proceedings of CoNLL-2002, 49-55, Taipei,Taiwan, 2002.

[4] A.Juhui, L.Seungwoo, and G. G. Lee, "Automatic acquisition of Named Entity tagged corpus from World Wide Web," in proceedings of the 41st annual meeting of the ACL (poster presentation), 2003.

[5] G. G. Lee, C.Jeongwon, and L.Jong-Hyeok, "Syllable pattern-based unknown morpheme segmentation and estimation for hybrid part-of-speech tagging of Korean,"

[6] D. Kim, H. Lee, C. Seon, H. Kim, and J. Seo, "Speakers' Intention Prediction Using Statistics of Multi-level Features in a Schedule Management Domain," in Proc. ofAssociation for Computational Linguistics on Human Language Technologies, pp. 229-232, 2008.

[7] S. Kang, H. Kim, J. Seo, "A Reliable Multidomain Model for Speech Act Classification," Pattern Recognition Letters, vol. 31, no 1, pp. 71-74, 2010.

[8] M. Bacchiani, F. Beaufays, J. Schalkwyk, M. Schuster, and B. strope. Deploying GOOG-411: Early lessons in data, measurement, and testing. In Proceedings of ICASSP, pages 5260{5263, 2008}.

[9] C. Van Heerden, J. Schalkwyk, and B. Strope. Language modeling for what-with-where on GOOG-411. 2009.

[10] S. J. Young et al., "The HTK Book", Cambridge University Engineering Department, 2006.

[11] R. Sproat, C. Shih, W. Gale, and N. Chang. A stochastic _nite-state word-segmentationalgorithm for Chinese. Computational linguistics,22(3):377{404, 1996.

[12] D. Povey, D. Kanevsky, B. Kingsbury, B. Ramabhadran, G. Saon, and K. Visweswariah. Boosted MMI for model and feature-space discriminative training. In Proc. of the Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), 2008.

[13] [Online]. Available: http://dx.doi.org/10.1016/ j.patrec.2010.09.029
C. Seon, H. Kim, J. Seo, "Efficient Appointment Information Extraction from Messages in MobileDevices with Limited Hardware Resources," Pattern Recognition Letters, vol. 32, no 2, pp. 127-133, 2011.

[14] [Online]. Available: http://dx.doi.org/ 10.5626/JCSE.2011.5.2.141
M. Lee, D. Han, "Ubiscript: A Script Language for Ubiquitous Environment," Journal of Computing Science and Engineering, vol. 5, no 2, pp. 141–149, 2011.