Improvement of E-Governance and M-Governance in polyglot Countries with Digital Etymology exploitation Indo-Aryan linguistics

Muppala Sunny Chowdhary[1], R. Aishwarya[2], Basam pravallika[3]
[1][3]Student, [2]Assistant Professor
[1][2]Sathyabama Institute of Science and Technology, [3]Lovely Professional University

Abstract: With large improvement of digital property (Wifi, 3G, 4G) and digital devices access to cyberspace has reached within the remotest corners today. Population of square measure resident within the country will simply access the net or applications from laptops etc. This can be a chance to the government to succeed into the state throughout a large vary, get their feedback, associate them in policy call with e-governance whereas not deploying material or resources. But the government of polyglot countries face tons of drawback within the victorious implementation of the presidency to national (G2C) and national to Government (C2G) governance because the rural folks tend and like to move in their native languages. Presenting equal expertise over internet or app to completely different language cluster of speakers could be a real challenge. Throughout this analysis, we have got sorted out the problems featured by Indo Aryan speaking netizens that are normally conjointly applicable to any language family groups or subgroups. Then we have got tried to present probable solutions victimization Etymology. Etymology is one way to correlate the words hardship their ROOT forms. Within the fifth century BC linguistic scientist wrote Ashtadhyayi wherever he pictured sutras or rules but, a word is modified in keeping with the number, gender, person, tense etc. Further this book was followed in Indo-Aryan countries together to obtain their synchronic linguistics of relatively different languages. Later system will be trained for automatic root extraction from the bottom level or morphed type of words from Panian Grammatical rules. The next step is to test our system over 10 thousand Tamil verbs and extracted the muse sort with cardinal accuracy. The unit of measurement presently operative to extend the program to that success lemmatize any words of any language and correlate them by applying those rule sets in Artificial Neural Network.

Keywords- Digital Etymology; Multilingual e-Governance; Root form extraction; Panini, Neural Network; Semantic Search; Android Application

I. INTRODUCTION

There are variety of nations within the world wherever folks speak a unique kind of languages ninetieth of the developing and transit countries square measure multilingual that is there square measure quite one language used for the communication between the voters of identical country, let's take associate degree example Asian nation is such a rustic wherever there are twenty-nine states and have twenty nine totally different languages that have their own official recognition and gets to market, Excluding this Asian nation has 122 major languages (spoken by quite 10,000 people) and 1599 alternative languages, the key population of the country lives in an exceedingly country wherever they impart within the native language. The people who live in the rural area are comfortable with speaking their mother tongue or in their local language if they get a chance to interact in on the net then most of them feel comfortable with their local or native language. As Asian nation India has polyglot states Naturally it becomes the responsibility and intention of the govt to draw in those individuals to be governance initiative. As there is wide variation in the languages of the countries like India it can be taken as an opportunity to gain political strength. The central government has taken up this opportunity and made changes like displaying the web text in the local language. But for the government, the feedback of the public is very important. Nowadays the public gives the feedback in their own native language, It is scheduled and the normal soft keyboard is available. As the posed question is in a particular language and the answer to the same question is in another language then the system fails to retrieve. For example, if a farmer of Tamil Nadu poses a question on rice farming in Tamil and the answer or the discussion is already over in English in the USA between two farmers. Search engines fail to retrieve or revisit the solution within the language wherever the question was posed. As there area unit sorts of languages in Republic of India the system fails to retrieve the search, kind the solution to the actual question. During this context, there's the sole answer is computational linguistics of all the text in alternative languages to English language because the web has the best information within the English language. Even though there is nothing to depend any automatic machine translator that translate any of the different universal languages to English with complete accuracy and also it is the least likely to change or update and it's technically an abnormal tedious job to convert every universal languages into English and store them on-line that is not possible.

In the ancient time our ancestors knew the way to speak that were the sound signals Then conversational of languages came into the image and synchronic linguistics was shaped and literature was written. The first successful answer was proposed by Universal Networking Language (UNL), which is metaphysical and worldwide language for communicating over computer networking. Even though for every 1:20 years ratio
unable to provide a worldwide mathematical framework called “Digital Etymology”. Etymology is a study about the history of words, their origins including the style of their type, which suggests that have been modified over time known as root kind extraction in Indo-Aryan. This paper is organized as follows. In section II we explained about the problem statement. The detailed proposed approach is shown in section III. In section IV methodology is detail explained. The application of visual survey and android application is shown in section V. We concluded this paper with scope of improvement and application is shown in section VI and VII.

II. PROBLEM STATEMENT

As per the analysis statements from the scientists and linguists, there's twenty three variety of language teams within the world. because the population is everywhere the planet completely different civilization area unit born and also the evolution of language happened. all told ancient books referred like “The previous Testament” or “Hindu Legends,” it's thought of that there was just one language at the beginning of the human civilization. The twenty 3 form of language groups unit of measurement Indo-European, Uralic, Basque, Afro-Asiatic, tongue, tongue, Khoisan, Altaic, Korean, Japanese, Chukotko-Kamchatkan, Sino-Tibetan, Daic, tongue, Austronesian, Andamanese, Australian, linguistic communication, Na-Dene, Amerindian, Caucasian, Dravidian, and Burushaski. of those groups area unit derived or obtained from one language.

The relation between the derived sub-group languages of an different cluster finally ends up within the giant forms of bottom level words. As among the case of the amount of internet communication medium English is that the only real language introduced then Spanish, Portuguese, French, German, Hindi, Mandarin distended as a medium. but as recently, internet is reaching the agricultural area of developing and nation's with multiple languages unit of measurement victimization native or linguistic communication to talk, a great deal of it will unfold and a great deal of it will attract the untouched native speakers. area unit often} an opportunity for the government if the roadblock will be overcome and embarrassing if the native speakers get distracted whereas not getting support for his or her first language over internet access. In Asian nation for all basic and religion languages support for keyboard, codification is gift i.e. support for all insertion and retrieval ahead end in native languages done.

Now the overall era of the problem statement is,
- How to translate the same statements to other languages express the same meaning?
- Even though they express the same expression then how to retrieve the answer? (where both the question and answer are in different language.)
- How to make sure that the opinions are expressing the same meaning in both the language? (how to consider the Telugu opinions were the major opinions are in Hindi/English)
- How to recruit the government officials for the fieldwork? (there are some village where there is no minimum knowledge of the smartphone)
- How to design the app based on the search engine translator/ machine translation?

However, the above-posed questions are extremely relevant in background processing beneficial to the central/state government.

III. PROPOSED APPROACH

First convert all the text into a universal language to represent the same data as extracted from the native/local language as formal or declaration language which is specifically designed. Where it is used as the way to approach for the machine translation system or it is used for the knowledge representation for retrieving the answer to the native/local language. UNL(Universal Networking Language) was the first to make such an approach to building a global or universal language. We have represented such proposed universal language that may be easily approached or could overcome the difficulty faced by the UNL(Universal Networking Language). First collect all the probable information of some language which helps us in the proposed universal language. Germany and Sanskrit are best as per framing and skeletal work for our proposed language. The main reasons behind this are:
- Firstly, both languages have a strong grammatical foundation where compared to other languages.
- Secondary, both languages are very rich in terms of vocabulary.
- Final and the most important thing the meaning of the sentences doesn't change when there is a relocation of the words in the sentence of a particular language.

As with the analysis Sanskrit is the only language which has system architecture. Word Forms are classified into tabular form of 8X3 size where they are classified related to the changes made according to the number, gender, vibhakti (to retrieve the changes to be made in the preposition part) and changes made in the root verb (dhatu roop) it's a 3X3 tabular form where it shows the changes that to be made according to the tense, person and number. In any language (excluding some rare expression) each sentence is of the shape either SVO (Subject Verb Object) or SOV (Subject Object Verb). A special comment field is proposed whereas the encoding part of the noun, verb, and preposition in universal language take part and store them. Therefore, if the expression of sentence or words given in one language does not have a unique code. As of the sentence from one language and the retrieved sentence from another language is expressed in the language in which even though there is no common term between the two different languages the system unable correlate and translate and them.

Therefore by the theory which was proposed through the theoretical approach, we can enlarge our domain of reach for
any intelligent system to the whole world without any local/native language barrier, which gives the power of search, sort, retrieve, decision making same as human keyword matching to the system or translating machine. As per the research, the root forms are unique and the linguist has performed many research works on the etymology.

The technical approach will perfectly generate backend process for analysis which can even be extended to the paragraphs. By the help of the algorithm, even encode the subject and object of the sentence by making a composite code for any language. So it is proved that once the statements are encoded by the making use of the set of rules mentioned above then it is easy to retrieve the given sentence to any other language. But now the problems faced are:

- How can we generate the codes in the universal language from all the native/local languages?
- What is the mathematical framework to generate the code in the universal language?
- How can we create an Android application by inserting the code which is generated by mathematical expressions?

We have worked on a possible way to solve the above-mentioned problems and formulated a possible route to achieve the goals.

IV. METHODOLOGY

The technique is explained in step by step process where used to solve the problem statements shows in below flowchart and points

- First, we need to identify that in which parts of speech the sentences are written and then determine each word in the given sentence using POS(Parts of Speech) tagger.
- Second, we need to identify and Mark the noun, pronoun, verb and some other reference from pronoun to a noun.
- Simultaneously we want to use the particular methodology to each and every linguistic communication to lemmatize them that is supervised learning. (e.g. running to run).
- Passing the sentences to the neural networks where we determine the Sanskrit root words.
- This is the point where we generate the codes using the mathematical framework.
- This is the critical point where we create an Android application for the translation, here we insert the algorithm and the mathematical framework expressions to generate the codes.
- At last, we add a comment field where this handles the phrasal verbs and idioms.

V. APPLICATION OF THE ANDROID APP AND VISUAL SURVEY

As we have inserted the codes which are generated by the mathematical framework expressions. We can know the problems of the public where the question or problem statement which is posted in the Native / Local language is retrieved by the algorithm into the English language and then convey to the government officials where they discuss the solution for the area or regional problem and convey it to the local or regional government official so that any actions are taken if needed. As in the urban areas everyone is aware of the usage of smartphone technology there is no problem to get the regional or local problems, whereas in the rural areas everyone may or may not be aware of the smartphone and Android technology here the problem arises where there must be a government official in the field to collect the local or regional problems. In this case the government should take a strict action so that any malpractice in the data collection may not take place and here the young India have an opportunity to get the work with government as a course of internship as each and every government official cannot take part in the field/visual survey, as compared it is beneficial to the government and to the young India where every can do the volunteering part with the minimum stipend. Here in the visual survey our team have went to a nearby village with the help of the local person who can have the awareness of the smartphone technology to check the working of the algorithm so the local person have typed the local or religion problem in the local/native language the algorithm and the mathematical framework expressions worked properly where we have retrieved the answer in to the English language and we have given the solution to the problem statement in English and it has retrieved the statements in to the local/native language without any grammatical or vocabulary mistake.

Work Done

Our teams have made a local visual survey where we reached level 3 by taking the Tamil language as a case study. Our system has generated results with 98% accuracy and with no grammatical or vocabulary mistake which was tested over 10,000 verbs and phrases. The algorithm for the above process is mentioned below.

Algorithm

Input: Tamil phrases to Shallow parser
Output: inflected verbs are collected in a file Input.doc
The input to our system: Input.doc
The output of our system: Multiple files classified according to tense and person consisting of root form of verbs
Begin
The statement is taken as input to the shallow parser
All the inflected verbs are collected in a file named Input.doc
Input.doc is taken as our system input
If sentence matched then If person matches then
Verbs are passed to the respective method with the verb and word for processing
Verbs are processed one by one where the mathematical framework rules are applied to those verbs to extract root verb
Root verb is stored in the Output.doc file as well as a separate file according to tense
Else
The verb collected from shallow parser in Input.doc is not a verb
Else
The verb collected from shallow parser in Input.doc is not a verb
End
The next step is to increase the complexity of the neural networks to get the Sanskrit root from the Tamil root forms. This can be solved by approaching the number topological sorting of the lexicon words as signal neurons. Even now we are working with upgrades in the framework and with the algorithm to make the better translation to announce the success of the Android application and we will do publish in the next paper. We have made an approach to the parser to determine and detect the relation between the words that have to be incorporated in the comment field or comment section. The method used in the improvement of the system and to generalize for all languages is semi-supervised by a method where the load is reduced to the programmer from getting the ideas for ongoing research.

VI. APPLICATIONS
If roaring universal language will produce a revolution in transmission addressing the need of computing in natural languages. All the byproduct developed and being developed in each step are terribly essential in language method analysis. If already registered, the machine-driven root verb extractor in ACL. As there is associate large number of the population in trilingual countries analysis on this subject will serve the larger no. of people directly and straightaway as against progressive analysis problems with that outcome comes once a prolonged delay and may profit really very little portion of the society. it'll save the vulnerable languages by attracting their native speakers to digital media. By using common languages at intervals of the backend will impel the North ranked nation of one step forward in making of an same international village with one language and single nation where there is no barriers between country, state, race between the color and community. This can improve the global unity and values of feeling towards united citizenry. This type of education collectively enhances the heritage value of cultural exchange and insulate also it reduces narrowness.

VII. SCOPE OF IMPROVEMENT
As this could be very mental object field of research and specifically we wish facilitate from linguists, Indic descriptive linguistics specialists, and anthropologists, therefore, making a cooperative framework will ease the work. The government is directly involved for demand specification for this universal language a minimum of 1 language for communication among the backend throughout this country with a population of 1 hundred twenty crores of that idiom changes in every eighty-kilometer. Nowadays Govt. is creating an endeavor to approach voters to involve them in governance, this may alone deliver the goods success if rural people be a part of which can they'll be part of providing they get scope to specific their scan within the natural language that scan to boot got to be au fait of upper process thus finance throughout this associate degree lysis will have a recursive positive impact on Digital Asian, Public Administration, Trust worthiness and E-governance.

REFERENCES
[1]. Panini, “Astadhyaayi,” 6th to 5th century BCE.
[3]. Zhenghua Li ; Min Zhang ; Wanxiang Che ; Ting Liu ; Wenliang Chen. “Joint Optimization for Chinese POS Tagging and Dependency Parsing” optimization and language change. Ieee/acm transactions on audio, speech, and language processing, vol. 22, no. 1, January 2014 pp. 274 - 286.
[5]. L.C. Butler ; D. Taylor ; T.I. Pritchard “Effects of response quantisation on the accuracy of transient response test results” TO increase the accuracy of the result pp. 334 - 338
[6]. Ekbal, R. Haque and S. Bandypadhyay, “Maximum entropy based Bengali part of speech tagging,” Advances in natural language processing and applications research in computing science, pp.67-78