

EnglishBuddy – An approach for structured answer evaluation and feedback for O/L English language examinations in Sri Lanka.

A.M. Gurusighe, amilamadhushanka@gmail.com,

W.K.D.P.Wijenayaka, daraniabasara@gmail.com

C.N.Nawagamuwa, chami485@gmail.com

W.K.D.L. Priyadarshani, lalitha_priyadarshani@ymail.com

M.P.A.W. Gamage, anjalie.g@slit.lk

*Faculty of Computing
Sri Lanka Institute of Information Technology*

Abstract

At present, English has become a universal language. It may not be the most spoken language in the world, but it is the official language in a large number of countries. Proficiency in English both spoken and written has become a basic and a crucial requirement to get a decent white collar job and also to pursue higher studies or career development. Therefore, passing O/L with a good grade for English has become critical. But due to the busy life styles of students, teachers as well as parents, do not see this as a major problem and pay less attention to English compared to the other subjects. Since there is a limited time available for each subject at school, teachers might not be paying their full attention to the students who need teachers' help. Sometimes parents also feel it is difficult to attend to the parents meetings and they might not know the actual grades of their children until the final results are given. This research provides a solution to the above problems by developing an automated system called "English Buddy" which will mark student's structure based answers in English and help the students to learn and evaluate their knowledge alone. This web solution will be useful for teachers to upload material and check progress of the students and for students to learn and practice exercises and get feedback. It'll be helpful for parents to be alert and follow the progress of their children. The systems is mainly build using techniques in Natural Language Processing and checked for accuracy with manual marking.

Keywords: Natural Language Processing, Context Free Grammar, Word Tokenizing

Introduction

English has become a language for communication worldwide and it is the most spoken language in the world, it is also the official language in a large number of countries. People who are fluent in English immediately opens up opportunities for their career development. Moreover, most of the

universities conduct their courses in English so it is very important to at least have a basic knowledge in both written and spoken English.

Currently there is limited number of facilities for students to improve English knowledge. Especially when it comes to writing answers for essay type questions students faces difficulties. Even for the teachers, marking essay type questions is time consuming and in large classes, the opportunity for students to get feedback for essay type questions from their teachers is less. Also they might be having a little time in the school to personally consult teachers.

This report consists with a detail explanation about structure based question analysis of English examination for Ordinary Level (O/L) using python language and NLTK tool kit. It evaluates the answer writing skills of each student by providing marks and highlighting the areas where they need to improve.

The research started with 200 college students around Sri Lanka to complete four sample O/L question papers, which consist of all type of questions, and according to the outcome of the research, it was found that more than 50% of the students are weak in the essay type of questions such as invitation writing, dialogue writing, summarization and grammar questions. These areas were incorporated in the new system through giving proper exercises and instructions on how they have to write the answers and how the marks are being allocated. Therefore, a student can attempt these types of questions and understand how they have to improve writing skills. Furthermore, the system will monitor the performance of students as they progress.

Methodology

The system was implemented using Python and Python Natural Language Tool Kit (NLTK). The system was implemented in levels as shown in figure 1. Each level contains questions where students can improve knowledge progressively.

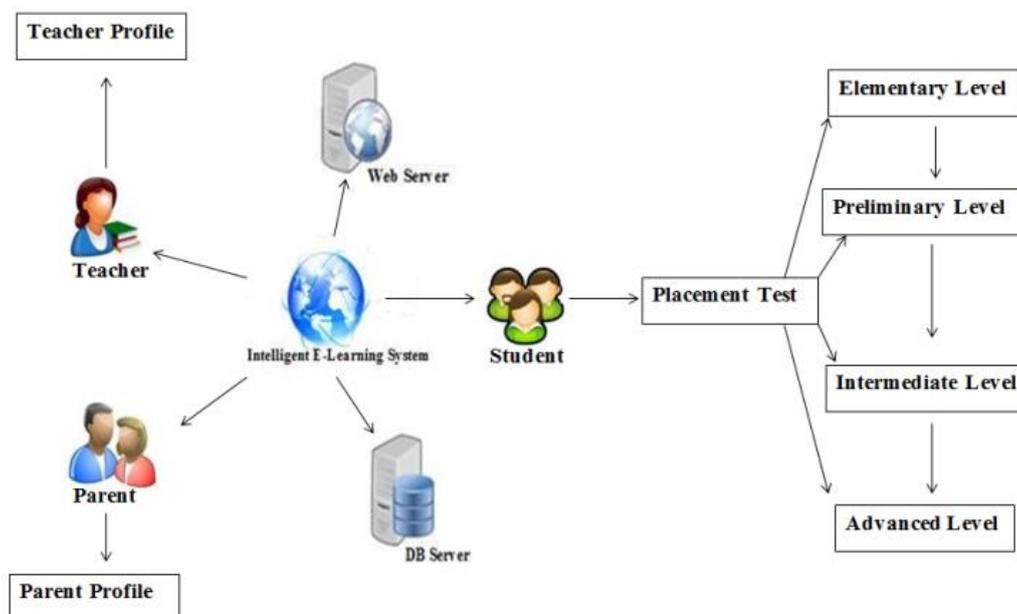


Figure 1- Overall System diagram

A. Dialogue Type Questions

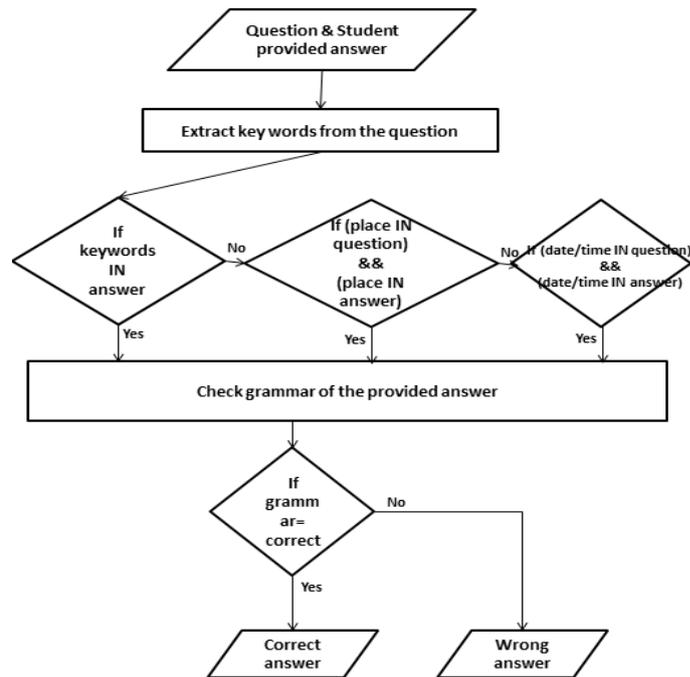


Figure 2- Dialogue evaluation process

Once the student submit answers the system extracts the key words from the previous and next sentences to the provided answer using python en library [1]. Then it checks whether the keywords are in the answer. It also repeats the same process with the next question. If keywords are not found then it checks whether the question is about a place, date/time, color and checks necessary answer provided. It also finds adjectives in the question and checks if it is in the answer. If the above mentioned method gives a positive result then it checks the grammar of the provided answer using Ginger API. Answer should be relevant and grammatically correct to get marks. Otherwise no marks are given. If the answer matches with the selected pattern then it checks the grammar of the sentence. Both pattern and grammar are correct then the provided answer is correct otherwise it is considered wrong.

B. Active-passive sentences

The process begins by reading a plain input text file containing sentences written in English. For this purpose, a parser is used to parse the English sentences to obtain their part-of-speech (POS) tags before further processing. The parser used is CSTs Part Of Speech Tagger. The parsed text is then fed into the system to detect verb in tense and convert it into the required tense. The final result consists of displaying the wrong answers of the natural language input. The final result is displayed with a friendly user interface that can be used as a feedback for students to improve the quality of their writing.

From the linguistic aspect, we usually say that the main “building blocks” of a sentence are Noun Phrases (NP) and Verb Phrases (VP). The noun phrases are usually the topics or objects in the sentence or in simple words this is what the sentence is talking about, while verb phrases describe some action between the objects in the sentence.

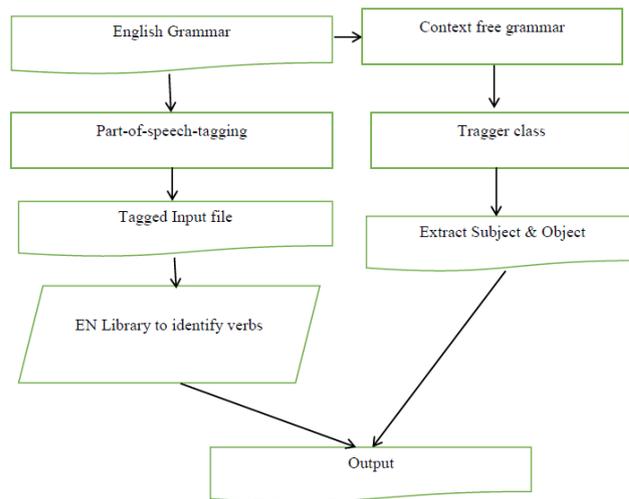


Figure 3- Active passive sentence evaluation flow

Goal was to extract only the noun phrases from the sentence, so to define some simple patterns which describe the structure of a noun phrase. For example:

- NN = content
- JJ+NN = visual content
- NN+NN = content marketing
- ...
- *NN = noun, JJ = adj

First, we defined our own Part of Speech tagger and define some “Semi-CFG”, which holds the patterns of the noun phrases. So that the above mention problem of grammar and verb confusion can be minimized. The code just tags the sentence with my tagger, and then searches for NP patterns in the sentence. Then as previously parsed text is fed into the system to detect verb in tense and convert it into the required tense. Then the active or passive voice format is formatted with “subject + verb + object” or “object + verb + subject”. The final result is displayed in the user interface.[5]

C. Writing “Summary”

The process begins by reading student answer written in English. Sentences are tokenized by using sentence_tokenize in NLTK tool kit. Then to calculate the word count RegExTokenizer is used. To evaluate the grammar mistakes and spelling mistakes Ginger API is used. The system extracts the keywords from a sentence to check whether it is relevant to the given question. The final result is displayed with a friendly user interface can be used as a feedback for students to improve the quality of their writing. So the main function of this process is to query the English sentence patterns through keywords. The second function is to detect grammatical errors in written English. [2]

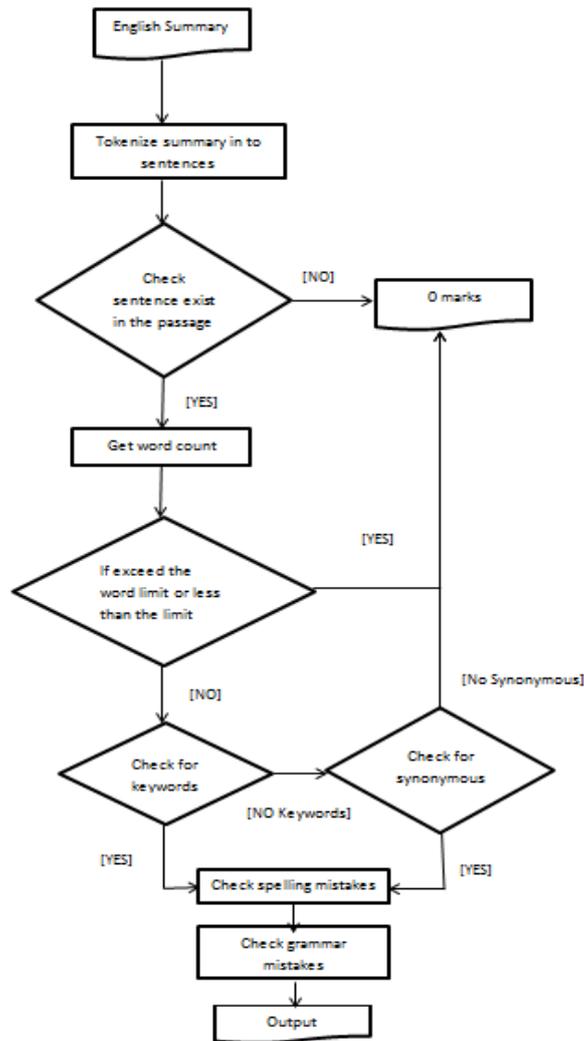


Figure 4- Summary evaluation flow

D. Writing “Notice/Invitation”

When the student gives the answers for the given question type (Notice or Invitation), those answers will be evaluated based on Content and Language. Content consists with answer’s words count and the format. Language consists with grammar, spellings and punctuations. So total marks are allocated out of five. The word count should be calculated by using *RegExTokenizer* [3]. The format is most importantly based on the heading of the question type (like ‘NOTICE’), date, time, venue, designation and other most important keywords. Marks for all those feature are allocated out of two. Sentences are tokenized by using ‘*sent_tokenize*’ and words are tokenized using ‘*word_tokenize*’ in the *NLTK tool kit* [4]. Keywords are extracted by using the ‘*keywords*’ library. Date, time, venue and designation are searched with the use of writing regular expression patterns by ‘*re*’ library [3] and using nouns with the help of ‘*pos_tag*’ in ‘*en*’ library [1]. Other than that, grammar, spellings and punctuations mistakes are evaluated using *Ginger API*. Marks for these features are allocated out of three. Above evaluation has been illustrated in the flow chart shown in Figure 5.

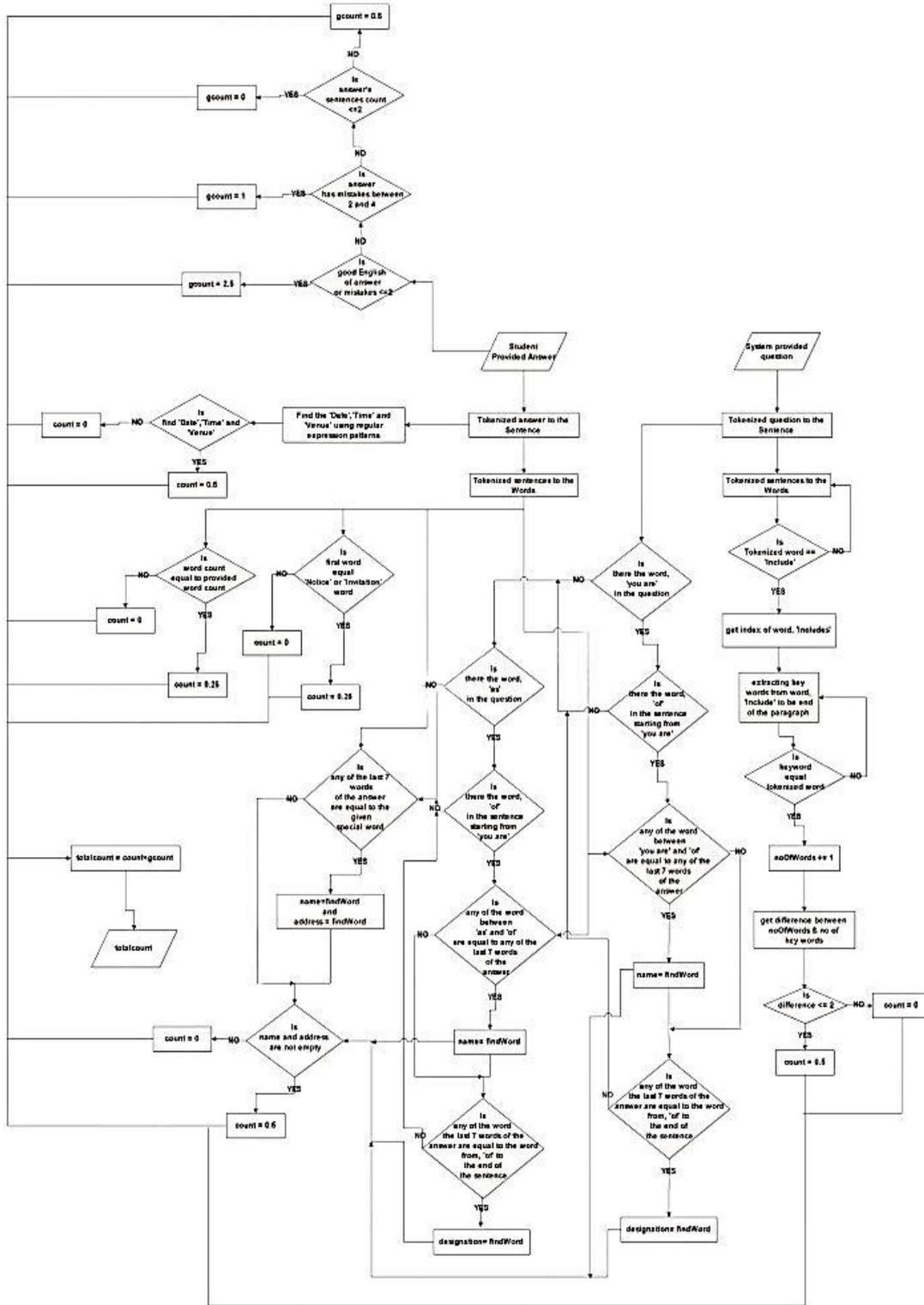


Figure 5- Notice/Invitation evaluation flow

Research findings

The system provides accurate answers for the verb form grammar correction. Therefore, for these types of tenses the accuracy is almost 100% because it identifies the verb of the tense after tokenizing it. Then the verb is checked with the user entered answers. Since the exact verb identify by identifying the location of blank space the next word contains the verb, because of that there is less confusion and mistakes in the grammar parts in evaluating.

The implemented method provides correct answers for the Past tense / Past continuous tense, Future tense / Future continuous tense and Present tense / Present continuous tense

However, for the active passive questions the sentences should rewrite again in the required format and Context Free Grammar (CFG) is used to identify the subject object and verb of the sentence [5]. The tagger class is used to identify the patterns of the sentence and it train the brown corpus. The only problem here is that it works properly with the currently tested data but accuracy is 60-70% because there might be some new patterns in the tense. Also for some new tenses there may be not any CFG defined which might prompt an error at that situation. Therefore, for this part testing is necessary considering almost all the types available.

There is also some confusion with the NLTK because it identifies some verbs as nouns, and sometimes adjectives as noun. If that type of confusion appears there is a conflict, which cannot be avoided by programmers. However, the program was developed considering the possibility of minimum problems.

For an example, consider the word “refuse”. “Refuse” appear both as a present tense verb (VBP) and a noun (NN). E.g., refUSE is a verb meaning “deny,” while REFuse is a noun meaning “trash” (i.e., they are not homophones). Thus, we need to know which word is being used in order to pronounce the text correctly. (For this reason, text-to-speech systems usually perform POS tagging.)

Most existing approaches focus on the manual assignment of keywords by professional curators who may use a fixed taxonomy, or rely on the authors’ judgment to provide a representative list.[6] To eliminate this problem we used Nodebox English Linguistics library to extract keywords from the questions and to check those keywords in student answers. The keywords library filters the most important words from a text.

Keywords uses a list of connectives that filters out word like ‘the’ and ‘because’. Tags are stripped as well. Optionally, it queries WordNet to filter out any words that are not nouns. The remaining words are counted, for example if the word typography occurs five times in the text and the word workshop occurs three times, and no word occurs more than that, the text will likely be dealing about typography workshops.

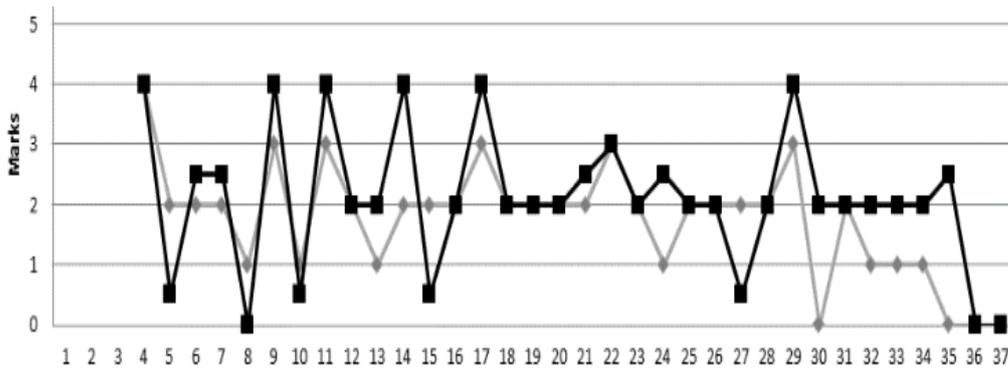


Figure 6 - "Summary" question evaluation accuracy

In summary questions marks are depended on the number of words in the answer. If a student exceed the word limit he will not get the marks even though the answer is correct. To count the words in the answer, word_tokenize in NLTK tool kit is used. After using it, it was noticed that it counts special characters and punctuations also as words. To overcome that problem RegEx Tokenizer [7] in NLTK tool kit was used which helps to split string to substrings using a regular expression.

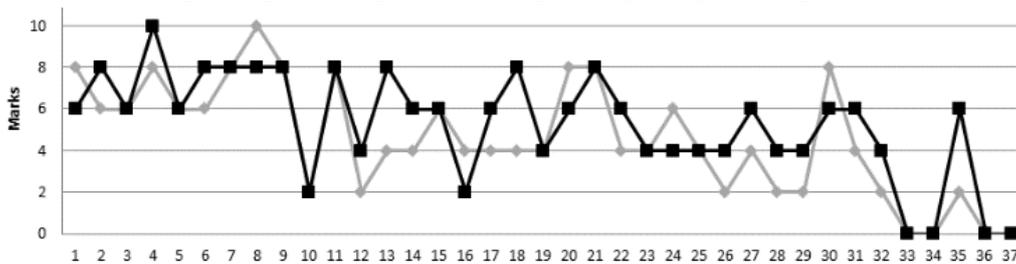


Figure 7- Dialog question evaluation accuracy

Gray color line in the graph represents the marks of manually evaluated answers and black color line represents the marks of the system evaluated answers. In this graph both series are going very closer, that means the accuracy of the implemented system is closer to the manual marking.

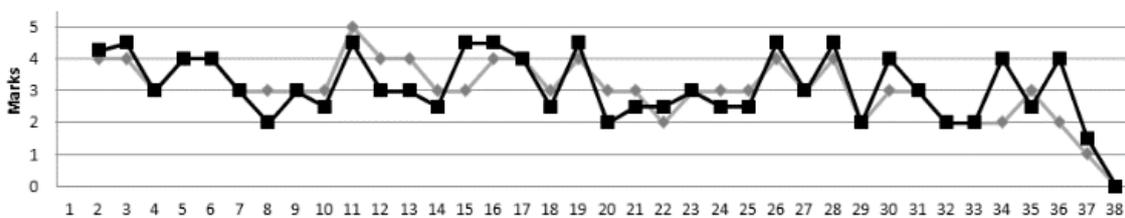


Figure 8- Notice question evaluation accuracy

Conclusions & future work

As discussed so far, the goal of this project English Buddy was to improve the English knowledge of the G.C.E. Ordinary Level students and increase the passing rate of the G.C.E. Ordinary Level exam.[8] By this system the development team hopes to provide guidance to the students how they

get marks for their answers, so they can improve the areas that they get less marks. Also students can practice model papers which consists some key areas of the G.C.E O/L exam paper. This was implemented as a web based system to give the access to students all over the country.

Since system is not intelligent as humans, there may be incompatibilities that are related to marking the answers automatically. Therefore students might not be getting the exact answer they expect from the teachers always. To avoid this problem system has a special feature, if student is not satisfied with the system result he / she can mail the marked answer to the subscribed teacher to get more clarifications.

Since the system only focus on the structure of the answer of invitation, notes and notices it is difficult to measure the creativity of different students using Natural Language Processing (NLP).

References

- [1] Nodebox English Linguistics library [Online] Available: <https://www.nodebox.net/code/index.php/Linguistics> [Accessed ; 05 June 2015]
- [2] Nazlia Omar, Nur Asma Mohd. Razali, and Saadiyah Darus “AUTOMATED ESSAY MARKING TOOL FOR ESL WRITING BASED ON HEURISTICS “ in Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600, Bangi, Selangor, Malaysia
- [3] Usage of regular expressions in NLP [Online] Available: <http://www.ijret.org> [Accessed ; Jan 2014]
Author : Gaganpreet Kaur
- [4] Natural Language Toolkit [Online] Available: <http://www.nltk.org/> [Accessed ; 02 March 2015]
- [5] Nazlia Omar*, Nur Asma Mohd. Razali, and Saadiyah Darus “Automated Grammar Checking of Tenses for ESL Writing” in Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600, Bangi, Selangor, Malaysia
- [6] Stuart Rose, Dave Engel, Nick Cramer and Wendy Cowley “Automatic keyword extraction from individual documents” Text mining: Applications and Theory, 2010 John Wiley & Sons, Ltd.
- [7] RegEx-Tokenizer –NLTK Online Available: <http://www.nltk.org/modules/nltk/tokenize/regexp.html>
- [8] Research and Development Branch, National Evaluation and Testing Service, Department of Examinations., —National Symposium on Reviewing of the Performance of School Candidates G.C.E.(O.L.) Examination - 2011 23 April 2011
Online Available: <http://www.doenets.lk/exam/docs/ebooks/Symposium-2011-new.pdf>
- [9] Steven Bird, Ewan Klein and Edward Loper , Natural Language Processing with Python , O’Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472, June 2009 [E-book] Available: <http://victoria.lviv.ua/html/fl5/NaturalLanguageProcessingWithPython.pdf>
- [10]Hsien-Chin Liou “DEVELOPMENT OF AN ENGLISH GRAMMAR CHECKER A PROGRESS REPORT” CALICO Journal, Volume 9 Number 1