Computer aided environment for drawing (to set) fill in the blank from given paragraph.

Sheetal Rakangor and Dr. Yogesh Ghodasara

Lecturer, Depatment of Computer Science, Saurashtra University, Rajkot 360001
Associate Professor, College of Agricultural Information Technology, Anand 388110

Abstract: *In this paper, we present Development of computer aided environment for drawing (to set) fill in the blanks that can generate for given paragraph. The System finds fill in the blanks, blanking key generates from the selected statement. Syntactic and lexical features are used in this process. NLP parser is used, part of speech taggers are applied on each of these sentences to encode necessary information.*
We present our work in designing and implementing the system which generate the blanks. The System is developed in Java using JDBC which is open source.
Keyword: *Natural Language Processing NLP, Key selection, POS tagging, Sentence selection*

I. Introduction

Now a day, increasing in population affects the number of students in every education stream. If we compare the scenario before 10 to 15 years with the current scenario, the number of students in each stream increases rapidly.

In any educational system, student has to appear for the examination to determine their performance and depending upon their performance they will be promoted for the next year. To evaluate the performance of the student different evaluation methods i.e. examination system are implemented. Fill in the blank questions where one or more words are removed from a sentence/paragraph, are easy to evaluate. Preparing this question manually required lots of time and efforts so with the help of this system teacher can automatically generate the fill in the blanks question from given paragraph.

Some of the earlier attempts in generating fill in the blanks are given below:

One of the first automatic question generation systems proposed for supporting learning activities was AUTOQUEST [2]. Used pattern matching approach to generate question from one of the sentences and grade the student answer. The purpose of these questions is to help novices to learn English.

Automatic gaps fill question generation system by Manish Agrawal and prashant mannem [4] generates fill in the blanks with options generated from text book without external resources except the document from text book. Sentence Selection, Key Selection and distractor are selected from chapter only. POS tagger and typed dependency is used for generating question.

TEDDCLOG [5] System take input as key and finds distractor from distributional thesaurus and find the sentence having inputted key and does not have distractor. Then system presents the sentence with distractor for approval, modification and rejection.

MCTI [6] system is semi automatic system consist following task Key Term Extraction, Source Clause Selection, Transformation to stem and Distractor Selection. In this system first key Term Identification is done than on that basis sentence And then Transformation of stem means convert the sentence into wh-phrase. And last distractors are created.

Computer aided multiple choice test items [8] is created post-editing interface where term extraction, sentence transformation, and for distractor selection Wordnet and corpus external resources were used. In this system post editing is used and user friendly interface is used for modification of question. And provide list of multiple distractor, user can select distractor's from that list for MCQ generated by system.

Most of the system approaches use is to select key first, and then sentence is selected. We have given priority to sentence and then from the important sentence key is extracted.

II. Model And Methodology

The following goals were set for the proposed computerized system for drawing question paper, from given paragraph:

Fill in the blanks questions are generated from it in two stages: sentence selection and key selection *Sentence selection* involves identifying *sentences* in the paragraph which can be used to generate question. These sentences are then processed in the *key selection* stage to identify the *key* on which to ask the question.

Architecture of System

In the Architecture of the system, User will input the paragraph, through interactive user friendly editor.
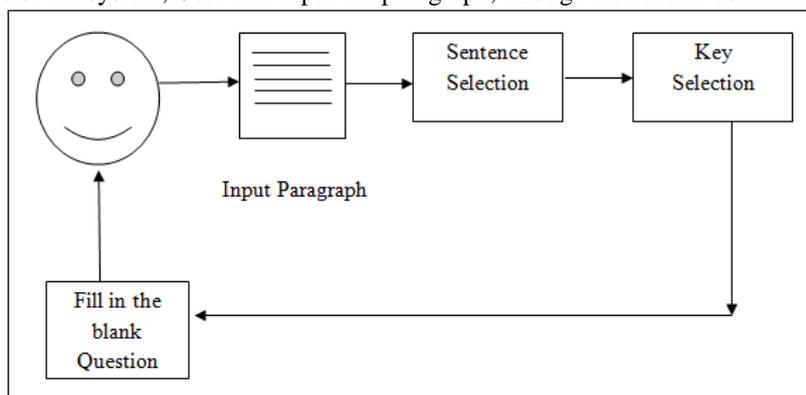
Figure 1 : Architecture of System

In this approach user will enter the paragraph from the system. And extract the sentences and select the important sentence from the given paragraph. Using certain features which are explained below, then key list is generated from the important sentence selection. From the key list, potential key is extracted from system and blank is generated from the given paragraph.

Sentence Selection

In Sentence Selection module goes through all sentences from given paragraph and used NLP Stanford parser which parser the sentences and divided into small fragments called *token*. And from that token POS tagger is used, which provides a representation of grammatical relations between words in a sentence. And extract a set of features like,

Count number of Sentences: Paragraph entered count number of sentences from that paragraph.

Count number of words: Count number of words in the sentence. Short sentence generate unanswerable question because short content and very long content might have enough content to make the question generated.

Count number of nouns: Noun gives an idea about the sentences, if maximum number of noun in sentence means potential key can be generated from that sentence and that sentence having good content which can generate the key for fill in the blanks.

Superlative: Superlative degree defines exaggerated mode of expression or height of quality. Superlative are typically formed with suffix –est (healthiest) or the word most, good, best are used. Sentence which contains superlative degree can generate good fill in the blanks.

On the basis of these features important sentence will be extracted from the given paragraph.

Algorithm for sentence selection

1. Enter the paragraph P
2. Read the statements from the paragraph S.
3. Calculate number of sentences CtS.
4. Calculate number of words from each sentence CtW
5. For each CtNoun and CtSuper from S do
   Select the sentence which contains superlative degree and then calculate maximum number of nouns which contain superlative degree.
6. IF CtSuper and CtNoun from S then

            SetenceSelected SS
        ElseIF Max (CtNoun) from S then
            SetenceSelected SS
        If there is no superlative degree then select that sentence which having maximum number of noun
        Else
            Without Noun and Superlative degree, blanks will not generated
        EndIF
        EndFor
    7.   Display SentenceSelected SS

Key Selection

Key selection is most important stage, to identify the key from important sentence to ask the question on. Previous work in this area, [5] takes key as input and [6] select key on basis of regular expression on noun. Or first search the key and then basic sentence is selected. Our key selection approach is divided into two stages. Generate *Potential keys* from the statement and select *Best Key* form that key list [4].

First Stage: Potential keys are generated from selected sentences, POS tagger is used to identify the words and there type. Suppose we need to generate key for noun then select total numbers of noun available in sentence and list them, and if any noun word is repeated in that list, it would be removed from list [4].

Second Stage: Best keys are generated from the key-list, select the word from key-list and search that word in paragraph. Count how many times that word has been used in paragraph. Select best key which has found in selected sentence and noun repeated in paragraph maximum number of times.

In our system we have used Stanford parser through which POS tagger will use for inputted paragraph.

Algorithm for key selection

1. For Each Word from SS do
    From selected sentence extract the potential key now suppose we need to generate key of noun then extract the nouns from the sentence. And add them into the keylist.
    KeyList =Select Noun from SS
    BestKey = No of Occurrence of that key in SS and Height of that key in the syntactic tree Structure.
    End For
2. Remove that BestKey from sentence SS and generate Fill in the blanks.

So this way system will generate the fill in the blanks from given paragraph, below figure shows the how system works when user entered user enters the paragraph with example.
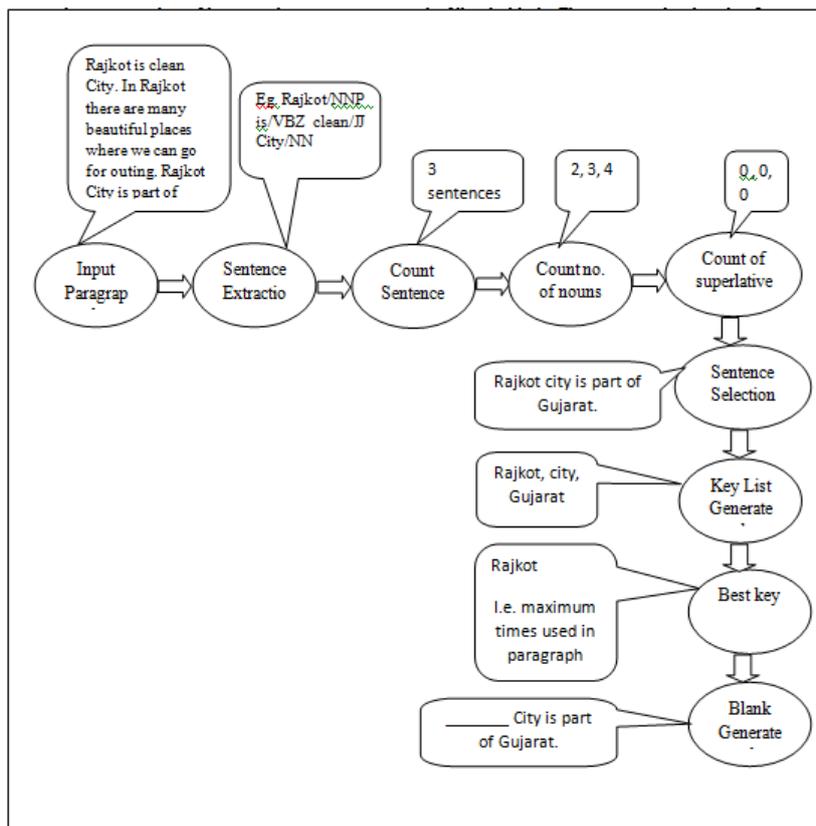
Figure 2: Process flow diagram of System

III. Evaluation And Result

The System has tested with number of small and simple paragraph. The table shows some of the implementation of paragraphs, first sentence is selected on the basis of the features describes earlier and best key is selected from the selected sentence and then fill in the blanks is generated from the given paragraph. And Figure 3 shows screen short of the system how system generates the fill in the blanks. The system is developed in Java using JDBC for generating fill in the blanks.

No.	Entered Paragraph	Selected Sentence	Key selected	Fill in the blank generated.
1	Rajkot is clean City . In Rajkot there are many beautiful places where we can go for outing . Rajkot City is part of Gujarat .	Rajkot City is part of Gujarat	Rajkot	_____City is part of Gujarat
2	Shrusti is the most sincere student in the class . while soniya is the most mischievous student. In fact Shrusti and soniya are most regular student. While kruti is the most beautiful student. Overall my class is the best of all.	In fact Shrusti and soniya are most regular student.	Shrusti	In fact _____ and soniya are most regular student
3	Sachin Ramesh Tendulkar born 24 April 1973. Sachin is a former Indian cricketer widely acknowledged as the greatest batsman of the modern. Sachin took up cricket at the age of eleven, made his Test debut against Pakistan at the age of sixteen, and went on to represent Mumbai domestically and India internationally for close to twenty-four years.	Sachin took up cricket at the age of eleven, made his Test debut against Pakistan at the age of sixteen, and went on to represent Mumbai domestically and India internationally for close to twenty-four years.	Sachin	_____took up cricket at the age of eleven , made his Test debut against Pakistan at the age of sixteen , and went on to represent Mumbai domestically and India internationally for close to twenty-four years
4	Butterflies are one of the most beautiful and interesting creatures on earth. Butterflies attractive, brilliant colors catch interest immediately. They have wings of varying shapes and sizes, and some even appear to	Butterflies are one of the most beautiful and interesting creatures on earth	Butterflies	_____ are one of the most beautiful and interesting creatures on earth

	have eyes on them.			
5	The peacock is the national bird of India. They have colourful feathers, two legs and a small beak. They are famous for their dance. When a peacock dances it spreads its feathers like a fan. It has a long shiny dark blue neck. Peacocks are mostly found in the fields they are very beautiful birds. The females are known as 'Peahen. Their feathers are used for making jackets, purses etc. We can see them in a zoo.	The peacock is the national bird of India	Peacock	The _____is the national bird of India

Table :1 Example of Paragraphs tested in system.
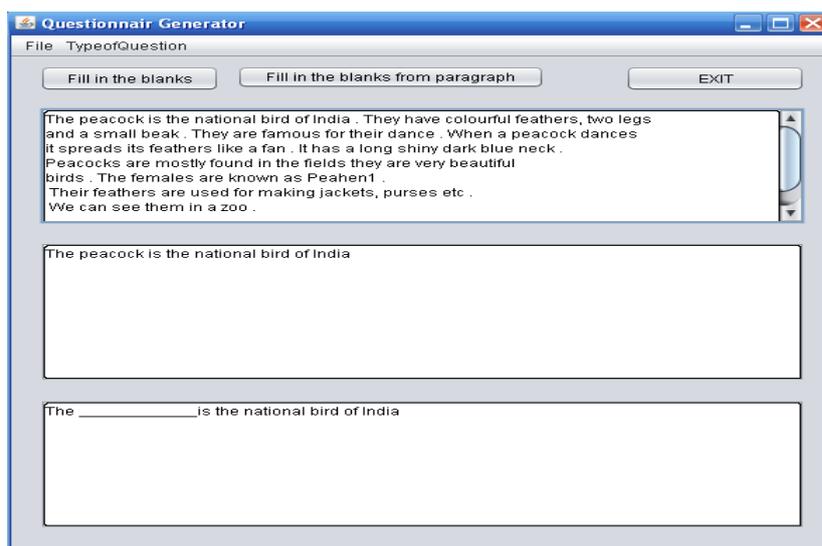
Figure 3: A Sample Screen short of the system.

IV.   Conclusion And Future Work

The system select important sentence from the given paragraph, and generate fill in the blanks question on them. Syntactic features helped in quality of fill in the blanks generated. We look forward to generate large number of blanks from the given paragraph as well as generate MCQS and True or False from given paragraph. Evaluation of Single blank generated is done in Java with the help of Stanford parser.

References

[1]   Pollock, M.J., Whittington, C.D., Doughty, G.F.: Evaluating the Costs and Benefits of Changing to CAA. Proceedings of the Fourth International Computer Assisted Conference CAA, http://www.caaconference.com/. (2000).
[2]   Wolfe, J.H.: Automatic question generation from text - an aid to independent study. SIGCUE Outlook 10(SI) (1976)
[3]   Kunichika, H., Katayama, T., Hirashima, T., Takeuchi, A.: Automated question generation methods for intelligent english learning systems and its evaluation, Proc. of ICCE01 (2001).
[4]   Manish Agarwal and Prashanth Mannem : Automatic Gap-fill Question generation from text books
[5]   Simon Smith, P.V.S Avinesh and Adam Kilgarriff. 2010.*Gap-fill Tests for Language Learners: Corpus-Driven Item Generation* .
[6]   Nikiforos Karamanis, Le An Ha and Ruslan Mitkov: Generating Multiple-Choice Test Items from Medical Text:A Pilot Study
[7]   Naveed Afzal and Viktor Pekar: Unsupervised Relation Extraction for Automatic Generation of Multiple-Choice Questions.
[8]   RUSLAN MITKOV, LE AN HA and NIKIFOROS KARAMANIS: A computer-aided environment for generating multiple-choice test items(2005)
[9]   Le An Ha :Multiple-choice test item generation: A demo Vasile Rus, Brendan Wyse, Paul Piwek, Mihai Lintean, Svetlana Stoyanchev and Cristian Moldovan:The First Question Generation Shared Task Evaluation Challenge (2010)
[10]   Husam Ali Yllias Chali Sadid A. Hasan: Automatic Question Generation from Sentences(2010)
[11]   Takuya Goto, Tomoko Kojiri, Toyohide Watanabe, Tomoharu Iwata, Takeshi Yamada : Automatic Generation System of Multiple-Choice Cloze Questions and its Evaluation
[12]   John Lee*, Stephanie* Seneff*: Automatic Generation of Cloze Items for Prepositions (2007)
[13]   Amisha Shingala, Rinku Chavda and Paresh Virparia : NATURAL LANGUAGE INTERFACE FOR STUDENT INFORMATION SYSTEM (NLSIS). Prajna(journal) Vol.19 41 - 44 (2011)