Comparative Study of Different Algorithms for Automatic Extraction of Objective Test from Given Corpus

Sheetal C. Rakangor[1] and Y. R. Ghodasara[2]

[1]Department of Computer Science, Rajkot-360001, Gujarat, India
[2]College of Agricultural Information Technology, Anand Agricultural University, Anand-388110, Gujarat, India

Abstract—**Automatic Generation of Objective test from given corpus is an area where many of researcher has present their work using different NLP parser or statistical pattern. NLP parser features like Named Entity and POS tagger are used for extracting objective question paper generation. Manual extraction of question and evaluating is very time consuming and tedious process for teachers. So researcher has decided to work in this area. And the presented research paper deals with the comparative study of different techniques which performs Automatic Generation of objective questions.**

Keywords— **Corpus, Key Selection, Multiple Choice Question (MCQ), NLP Parser, Named Entity Feature, POS Tagger, Sentence Selection and WordNet**

Fig. 1: Managing examination preparation process

I. INTRODUCTION

In any educational system, student has to appear for the examination to determine their performance and depending upon their performance, student caliber will be judge. Let's start from ancient era, there was a Gurukul system. Guru will teach to sisya, to judge the sisya, guru will assign task to the sisya, and on that basis of task allotment sisya performance will be judged. Due to less number of sisya, Guru can observe the sisya and assign the task; assignment of task was nothing but one kind of examination only. Due to less number of students, Guru can observe the sisya and assign the task according to their observation. And then when number of students increase examination came into existence. To evaluate the performance of the student different evaluation methods i.e., examination system are implemented.

First it was started with Oral examination, where teacher will ask the question orally and student will reply to that question. As the number of student increases oral examination changed to theory examination approach, students are suppose to write the long and narrative answers to the question, the theory examination process for teachers is divided into three parts: 1) Question paper preparation,   2) Conduct test and, 3) examine the paper and declare the result.

On examination day, teacher distributes the question paper to students; test will be conducted of students, to determine their performance. Teachers check the students answer sheets and according to the student answer, marks will be allocated to the student and result will be declared.
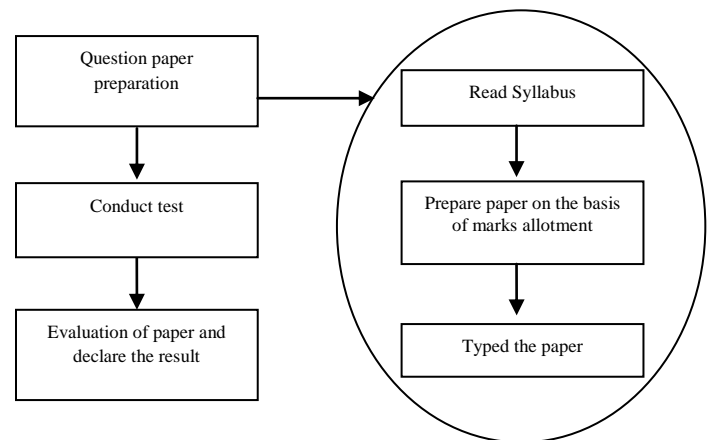
Based on the current situation, some of the problems for the teachers have been identified. Preparing question papers would take a long time because teacher would have to type the question. As number of student increases, paper to evaluate also increases. So paper assessment will be difficult and time consuming process for the teachers.

This process would take a lot of time and burden for teacher, due to overcome this problem examination system has decided to implement the new method which removes the concept of cramming and student can score the marks only if concepts are cleared.

Instead of theory approach objective examination came into existence, the process of the examination system will be the same as describe above, only different is instead of theory or descriptive paper. Teacher has to prepare objective question paper and student has to give answer in short form, instead of long and descriptive answers. Objective examination approach is divided into many different approaches as stated below:

True or False questions

Multiple Choice questions

Fill in the blanks from the given options

Fill in the blanks by writing the answer in the blank space.

In True or False approach, teacher has to prepare the statement, whether that statement is true or false, that answer student has to give. In this approach if concepts are clear then only student can say True or False. Concept of cramming remove and 100% pure evaluation of this question paper and very easy so evaluation process will be faster as compare to theory approach.

Fill in the blanks from the given options approach, teacher has to prepare the incomplete statement with four to five options, students has to just select the correct answer among the options available. Fill in the blanks approach, teacher has to prepare the incomplete statement, students has to complete the statement.

In Multiple choice questions approach, teacher has to prepare the statement; four to five answer of that question is given. Students have to just select the correct answer among the options available.

Evaluation process of Objective question paper approach is much faster that then the theory approach, but for the teacher examination process will be the same, Examination preparation process, the objective examination process is also divided into three parts 1) Question paper Preparation. 2) Conduct test and 3) examine the paper and declare the result. Preparing question papers would take a long time because teacher would have to typed the question, conduct test and examine the paper and declare the result. This process would take a lot of time and it is burden for teachers, but in this approach partiality to the student decreases, evaluation process will be 100% pure, cramming of the student decreases.

After invention of the computer and software programs, has change the objective question evolution from manual to computerization in which an objective test, which consist question with multiple choice, is promoted to the student, student has to select the appropriate answer and computer software stores, evaluate and give the appropriate marks to the students. In this approach it is possible to handle only the multiple-choice questions. But in manually system one can evaluate multiple choices, True or False, Fill in the blanks, one word answer, and one sentence answer.

The researcher has study manual as well as computerized evaluation of online test and identified that is it not possible to develop a model, which automatic extract the question from the given corpus. And researcher has determined to develop such a model and prototype that will deals with computerized automatic generation of questions i.e. Fill in the blanks, Multiple Choice Question, and True or false objective question will generate from the given text.

The system we propose provides teachers with a method that reduces time and expenditure for setting manual paper setting. The user chooses phenomena s/he wants to study, and what types of questions s/he wants to create i.e. MCQ, True or false and Fill in the blanks. The system generates, automatically, the question types that the user has requested.

II. STUDY OF ALGORITHMS

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 to learn English. In this AUTOQUEST doesn't require human frame presentation, system takes ordinary text and present it to the students on computer. Students first need to read the paragraph which appears and on the basis of the paragraph question will be generated on the sentences of the text. Answer is evaluated, if incorrect answer again paragraph will be shown to the students.

Automatic gaps fill question generation system [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. Based on the certain features sentence selection, key selection and distractor are selected.

TEDDCLOG [5] is Testing English with Data-Driven Cloze Generation. System take input as key and finds distractor from distributional thesaurus based on the key and find the sentence having inputted key and which does not have distractor , and find the best sentence which have best collocation, sentence which is sort, Begin with capital letter and ends with full stop, having minimum two commas. All other sentence will get rejected, sentence with this features get selected and then system presents the sentence with distractor for approval, modification and rejection.

MCTIG [6] system is semi automatic system consist following task Parsing, Key Term Extraction, Source Clause Selection, Transformation to stem and Distractor Selection. In this system first Parsing of the sentence is done using Charniak's NLP parser, 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. Post-editing phase is generated so that user can manually enter the distractor or select the distractor from the generated list.

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 distractors from that list for MCQ generated by system. After the Post editor has finished revising a given item need to click on submit button. And changes on item will stored in database. Statistic Calculation for item selection, distractor selection, has been done.

QG system [11] input the text is sentences from that question are generated. For this they do steps Data Processing, Elementary Sentence Construction, Sentence Classification, and Sentence Generation are used to generate questions from the given text. In data processing all verb will be extracted using POS tagger and Named Entity features(for identifying Person, location, organization, facility, date, money etc) using Oak System. Complex sentence will be converted to elementary sentence with the help of Charniak parser; elementary sentences will be input of this system, sentence classification is done for generating question. They have process each sentence from top – down manner to get it

classified. For eg if preposition in used in sentence then they will add "when" question type of its classification. With the help of TREC-2007 incorrect grammatically generated question will be rejected and evaluate the performance of the system using recall and precision.

Automatic Generation System of Multiple-Choice Cloze Questions and its Evaluation [12] student insert the text into the system and text is decomposed into the sentences and multiple choice cloze questions generated for each sentence from system. To generate the questions from texts automatically, the system (1) extracts appropriate sentences for questions from texts (2) estimates a blank part and (3) generates distracters of existing questions. Firstly system extracts some sentences which are appropriate for generating the multiple choice cloze question from text using Preference Learning; eliminate the blank part using Conditional Random Field. CRF is framework for building discriminative probabilistic model to segment and label sequence data. And thirdly system generates distractor on the basis of the Statistical pattern. Experimental results shows method is workable for selecting appropriate sentences and blank

part. Moreover, method is appropriate to generate the available distracters, especially for the sentence that does not contain the proper noun.

Automatic Generation of Cloze Items for Preposition [13] This system is used for improving the vocabulary level its take following steps to produce cloze item source corpus. (1) Determine the key from source corpus (2) Select the sentence from the corpus (3) Generate the distractor incorrect choices for the key. Mostly preposition are familiar to the user, making it more difficult to select good distractor. This paper mostly focused on the distractor selection which should be similar enough to key as well as it must be result in an incorrect sentence. For distractor baseline and collocation and non native English corpus is used to generate the distractor

Computer Aided Generation of Multiple-choice test [14] The System generates test question and distractors offering the user the post-edit the test items. The system is built on separate components, which perform the following tasks: (1) term extraction, (2) selection of distractors and (3) question generation. To retrieve the term noun or noun phrase, verb phrase, adjective and adverb phrase is identified using the

Table 1: Comparative analysis of various algorithms which generates automatic question

Algorithm	Methodology	MCQ Generation	Fill in the blanks	True or False	Question Generation
AUTOQUEST [2]	Sentence Selected from paragraph	No	No	No	Yes
Automatic gaps fill question generation system [4]	1.Sentence selected 2. Key selected 3.Distractor selected	Yes	No	No	No
TEDDCLOG (MCQ for English learning) [5]	1. Input Key 2.Distractor Selected 3.Question Selection	Yes	No	No	No
MCTIG [6]	1.Key Term Extraction 2.Source Clause Selection 3.Transformation to stem 4.Distractor Selection	Yes	No	No	No
Computer aided multiple choice test items [8]	1.Term extraction, 2.Sentence transformation, 3.Distractor selection	Yes	No	No	No
QG system [11]	1.Data Processing, 2.Elementary Sentence Construction, 3.Sentence Classification, 4.Sentence Generation	No	No	No	Yes
Automatic Generation System of Multiple-Choice Cloze Questions and its Evaluation [12]	1.Extracts sentences 2.Estimates a blank part 3.Distractor Seletion	Yes	No	No	No
Automatic Generation of Cloze Items for Preposition [13]	1. Key Selection. 2. Sentence Selection. 3.Distractor Selection	Yes	No	No	No
Computer Aided Generation of Multiple-choice test [14]	1.Term Extraction 2.Select the Distractor 3.Question Generation	Yes	No	No	No

FDG shallow parsing. Selection of distractors, WordNet retrieves hyponyms, hyponyms, and coordinates of the term, if applicable. If WordNet returns too many concepts, those appearing in the corpus are given preference. Generation of test question a number of simple question generation rules have been implemented i.e., (SVO rules). The system makes use of agreement rules which ensure the grammaticality of the question generated. These rules also check for agreement between concepts mentioned in the question and the distractors.

III. COMPARISON OF AUTOMATIC OBJECTIVE QUESTION GENERATION ALGORITHMS

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 are extracted.

All Algorithms mostly concentrated on MCQ objective question generation. We are going to develop system through which MCQ, True or False and Fill in the Blanks can be generated from the Paragraph and Important Sentences is extracted from the paragraph and keys are extracted from the important sentences and then distractor are selected.

IV. CONCLUSION

By studying and analyzing the literature, it was determined to develop a tool which gives an excellent IDE. With the help of this IDE teachers can use a computer for drawing question paper, which consist true or false, fill in the blanks, and MCQ question generation from given paragraph.

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 Simon Smith, P.V.S Avinesh and Adam Kilgarriff. 2010.Gap-fill Tests for Language Learners: Corpus-Driven Item Generation .

[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

[10]. Vasile Rus, Brendan Wyse, Paul Piwek, Mihai Lintean, Svetlana Stoyanchev and Cristian Moldovan:The First Question Generation Shared Task Evaluation Challenge (2010).

[11]. Husam Ali Yllias Chali Sadid A. Hasan: Automatic Question Generation from Sentences (2010).

[12]. Takuya Goto, Tomoko Kojiri, Toyohide Watanabe, Tomoharu Iwata, Takeshi Yamada: Automatic Generation System of Multiple-Choice Cloze Questions and its Evaluation.

[13]. John Lee, Stephanie Seneff: Automatic Generation of Cloze Items for Prepositions (2007).

[14]. Ruslan Mitkov, Le An Ha: Computer-Aided Generation of Multiple-Choice Test.