Literature Review of Automatic Question Generation Systems

Sheetal Rakangor*, Dr. Y. R. Ghodasara**

*PhD Scholar, School of Science, RK University
**Associate Professor, College of Agricultural Information Technology.

Abstract- NLP is an area of research where many researcher have presented their work and is still an area under research to achieve higher accuracy. Researcher is going to use NLP for automatic question generation system. As in any education system, examination is conducted to judge the caliber of the students. So to conduct the examination, educator needs to generate the questions manually which is very time consuming process, so researcher has decided to develop system through which, automatic questions can be generated, and time and efforts will be reduced. To develop automatic question generation, many researcher has presented their work and many algorithms are proposed to generate the automatic question from given sentence or text. This paper presents review of work to generate questions automatic from inputted text. Paper reviews methodologies with respect to the phases of question generation.

Index Terms- Automatic Question Generation system(AQGS), Multiple Choice Question (MCQ), ,Natural language Processing(NLP), Named Entity Recognizer (NER), Question Generation(QG), Semantic Role Labeler (SRL).

I. INTRODUCTION

This paper represents the study of acceptance and question generation through NLP or Statistical pattern and acceptance for in this area. Now a This paper represents the study of automatic objective review of research work carried out so far in this area. Now a day's MCQ(Multiple Choice Question) type question is widely used for examination when there is huge number of Students, students appearing for examinations like GATE, CAT, NET etc. MCQ is very easy for evaluations and its evaluation is implemented through computerized applications so within a few hours result can be declared and evaluation process will be 100% pure. Researcher has study the examination process and decided if questions can be generated automatically with the help of Computerized application. so it will reduce the task of educator. NLP is an area of research and application that explore how computers can be used to understand and manipulate natural language text or speech to do useful things. NLP holds great promise for making computer interfaces that are easier to use for people, since people will be able to talk to the computer in their own language, rather than to learn a specialized language of computer commands. NLP researchers aim to gather knowledge on how human beings understand and use language so that appropriate tools and techniques can be developed to make computer systems understand and manipulate natural languages to perform the desired tasks. NLP research has evolved from the era of punch cards and batch processing to the era of Google[11]

Automatic question generation is part of NLP i.e. Natural Language processing, is an area of research where many researcher have presented their work and is still an area under research to achieve higher accuracy. Many researcher have worked in the area of automatic question generation through NLP and numerous techniques and models have been developed to generate the different types of question automatically and in many different languages work has been done like English, Punjabi, China, Spain, etc.

There are many changes being made till now in various fields that tend to move from manual systems to automated systems. These automatic systems help us with less cost and time efficient solutions. In the education field, the academicians are majorly dependent on their own for generating questions for various examinations[3].

II. LITERATURE REVIEW

CQG (Cloze question generation)[1] system that generated list of cloze questions given in English article. CQG system is divided into three main module, Sentence selection, key selection and distractor selection. In the first stage, informative and relevant sentences are selected and in the second stage, keywords (or words/phrases to be questioned on) are identified in the selected sentence key selection will not be noun or adjective it would find on the basis of NER. Distractors (or answer alternatives) for the keyword in the question sentence are chosen in the final stage. First two stage are not domain specific. third stage is domain specific, because quality of distractor depends on domain so distractor will be selected on the basis of the key selected and through web, list of distractors will be generated and knowledge based distractor list will generated. And evaluation of the system is done manually through three phases 1). Evaluation of the selected sentence 2). Evaluation of selected keyword and 3). Evaluation of selected distractor.

Automatic question generation on the basis of the discourse connectives[2], question generation system divided into two modules content selection and question formation. Content selection consists of finding the relevant part in text to frame question from while question formation involves sense disambiguation of the discourse connectives, identification of question type and applying syntactic transformations on the content. Researcher concentrates on seven discourse connectives like because, since, although, as a result, for example and for instance on that basis Question type will be decided like if sentence consist since then question type would be *Why*. System

has been evaluated for semantic and syntactic soundness of question by two evaluator.

Automatic Question Generation Using Software Agents for Technical Institutions[3] developed an system in which take an input in form of the text file from user which contains of the text upon which the user desires to fetch questions; the output is produced in form of a text file containing questions based on Bloom's taxonomy. The advantage of generating questions based on Bloom's taxonomy enables to generate the questions that help to assess learning ability of the students. The proposed framework helps in question generation by deploying agents, the agents will perform various operations like document processing, information classification and question generation. Thus system may also be termed as a multi agent system. In Document processing tree tagger tool and stemming process is done to eliminate the human process. Information classification takes an list of keyword generated by Data Processing and finds the Bloom's category of those words, by searching appropriate action verb in the repository which fits with the given keyword. question generation module takes the output of Information classification as input to generate questions. The process is a template based approach, which fits the selected keywords in the question template according to the Bloom's levels.

Automatic Question Generation system called G-Ask[4], which generates specific trigger questions as a form of support for students' learning through writing. They conducted a large-scale case study, including 24 human supervisors and 33 research students, in an Engineering Research Method course and compared questions generated by G-Asks with human generated questions. authors identified the most frequent question types, derived from the human supervisors' questions and discussed how the human supervisors generate such questions from the source text. Compared and Citation Classification performance is done through precision and recall, and Question Quality evaluation is done through Cohen's Kappa coefficient.

Automatic Multiple Choice Question Generation System[5], The system provided selects the informative sentence and the keyword to be asked based on the semantic labels and named entities that exist in the sentence, the distractors are chosen based on a similarity measure between sentences in the data set. In this paper, automatic question generated like multiple choice questions which asks about a word in a given sentence, the word may be an adjective, adverb, vocabulary, etc. For generating question Semantic Role Labeler and NER (Named Entity Recognizer) is used to identify whether its Name, Location or Name of Organization. Once Question sentence is prepared, then measures the similarity between the Question sentence and each sentences from the Question knowledge based. Sort the obtained similarity values from other sentences and Get three keyword from three different sentences as an distractor values. In this research out of nearly 145 parsed sentences, there were 109 considered good according to the keywords that are extracted from them.

Semantic Based Automatic Question Generation[6] system uses both Semantic Role Labeling and NER (Named Entity Recognizer) technique is used to convert the inputted sentence to semantic pattern. and developed and Artificial immune system which that will be able to classify the patterns according to the question type. The question types considered here are set of WH-

questions like who, when, where, why, and how. Immune system utilizes feature extraction, learning, storage memory, and associative retrieval in order to solve recognition and classification tasks. Inputted sentence will first parse using NER and SRL technique, and from NER and SRL identifies whether, sentences contain person name, location, date, on the basis of this identification question pattern will be identified for example if person name then question pattern would be WHO, sentence pattern for who and question patterns is interpreted as two features vector in the training set, one vector for each question type. and For Evaluation sentences where extracted from Wikipedia or TREC 2007. The percentage of truly generated patterns increased 87% which appears to be promising ratio in this problem comparing it to other techniques used in generating questions automatically.

Automatic question generation in multimedia-based learning [7] here author investigate whether questions generated automatically by two Natural Language Processing (NLP) based systems (one developed by the authors, the other a state-of-theart system) can successfully be used to assist multimedia-based learning. Researcher examine whether the questions produced by our system can be successfully used as pre-questions and thus support creators of assessment materials. Two different types of pre-questions are investigated: text-based and with supporting image. experiment also serves to test whether pre-questions have a beneficial effect in combination with audio-visual learning material as opposed to reading material; Researcher analyze the effect pre-questions have on test-takers performance on a comprehension test about a scientific video documentary. In this paper present QG for documentary video and image related to Question will display, automatic Question will generate as well as manually experts generated question by watching video, and both question will compared and select the questions which will appear in both type. For image, screenshot is taken from the video at respective time a source sentence occurs in the video. and developed the interface, and experiment on 29 students, will watch the video once and answered the question generated by the interface and tracked each participant's answers and time spent to answer each question. Supplying a screenshot alongside a prequestion will result in a statistically more significant difference of correctly answered questions when comparing to no prequestions. The ability to supply a screenshot alongside a question is unique to our system. The average time taken to answer a question is not statistically significantly different between the pre-question settings.

Automatic Generation of Question Bank Based on Predefined Templates[8], In this paper All possible questions are generated by parameterized concepts from a set of pre-defined templates. The generated questions cover all selected topics in all level of difficulties the form of a multiple-choice question (MCQ). The proposed system contains three subsystems: Knowledge Descriptor, Questions Generator, and E- learning Executer. The Knowledge Descriptor subsystem allows the instructor to describe the learning contents. The Questions Generator subsystem receives the learning contents and generates the corresponding multiple questions. The E-learning Executer subsystem uses and allows the students to use the generated questions in education process. system generated correct and wrong options question (CWOQ) algorithm is generated and had

sub algorithm like onlyOneSoultion, twoSoultions, allOfTheAbove, and noneOfTheAbove algorithms. A question bank is developed for course General Biology course (Bio110) at Faculty of Science, King Abdulaziz University (KAU): The bank contains 12 chapters with 239 sub topics and total of 46345 questions. The system allows the student to solve the training of each chapter and then display his answer and the correct answer

AUTOMATIC GENERATION OF MULTIPLE CHOICE QUESTIONS FROM DOMAIN ONTOLOGIES[9], approach presented in this paper is based on domain specific ontologies and it is independent of lexicons such as WordNet or other linguistic resources. System creates multiple choice question items using the Semantic Web standard technology OWL (Ontology Web Language). The proposed approach is independent of the domain since questions are generated according to specific ontology-based strategies. class based, property based, terminology based strategies were used to generate the multiple choice question. Property-based strategies may produce a large number of multiple choice questions but are very difficult to manipulate syntactically. Class and terminologybased strategies on the other hand are much easier to handle syntactically but generate fewer questions for ontologies of the same depth and population. The generated questionnaires were evaluated in three dimensions: Pedagogical quality, linguistic/ syntactical correctness and number of questions produced.

Question Generation System From Punjabi Text[10], In this paper author present the review on question generation from documents written in Punjabi language. A NER (Named Entity Recognition) Tool is also need to be created which recognizes the names from a given sentence and generate the appropriate

questions from it. The system generate the question starts with the words "kithe (Where)", "kisne (who)", kinne (how) etc. researcher uses the rule based approach, in which suppose sentence consist the name of the person, then it will generate question who and if location name is found in the sentence then question with "kithe(where)" word can be generated.

Mind the Gap: Learning to Choose Gaps for Question Generation[12], In this paper approaches, the problem of generating good questions into two parts: first, the selection of sentences to ask about, and second, the identification of which part of the resulting sentences the question should address. To achieve the goal of selecting better gap-fill questions, author has broken down the task into stage 1) sentence selection, 2) question construction, and 3) classification/scoring. For generating question, features are used i.e. Token count, lexical, syntactic, semantic and NER feature is used to generate the Gaps fill question. and generated question were analyzed manually and rated the question.

Linguistic Considerations in Automatic Question Generation[13], The system consists of a straightforward pipeline. First, the source text is divided into sentences which are processed by SENNA software, SENNA provides the tokenizing, pos tagging, syntactic constituency parsing and semantic role labeling used in the system. For each predicate and its associated semantic arguments, a matcher function is called which will return a list of patterns that match the source sentence's predicate-argument structure. and paper focuses on evaluating generated questions primarily in terms of their linguistic quality. Evaluation conducted by one file, on chemistry which has 59 questions and 142 questions generated through system.

Table-1. Comparative analysis of various algorithms which generates automatic question generation systems

Sr no.	Algorithm	Methodology	Type of Question	Language	Evaluation and Result
1.	Cloze question generation	Sentence selection, key selection and distractor selection is domain specific and NER feature is used for key selection	Cloze	English	Manually Evaluation is done 1). Evaluation of the selected sentence 2). Evaluation of selected keyword and 3). Evaluation of selected distractor.
2.	Automatic question generation on the basis of the discourse connectives	Content selection and Question formation	Question generation like Why, when, where, In which	English	Manually evaluated for semantic and syntactic soundness of question by two evaluator
3	Automatic Question Generation Using Software Agents for Technical Institutions	Document Processing, Information Classification and Question Generation.	Define, Describes, Give example, long descriptiv e questions	English	-

5.	Automatic Multiple Choice Question Generation System	Citation Extraction, Citation Classification. and Generation Extract sentence from Data Set, Prepare Question sentence, Measure the similarity between the question sentence and all sentences in the knowledge base, Return the three sentences that have the highest similarity values, three keywords of three sentences as	Long descriptiv e questions like Why, when, Does any	English English	Compared questions generated by the system to those produced by humans. and Citation Classification performance is done through precision and recall, and Question Quality evaluation is done through Cohen's Kappa coefficient. In this research out of nearly 145 parsed sentences, there were 109 considered good according to the keywords that are extracted from them.
6.	Semantic Based Automatic Question Generation	distractor selection Input sentence, Feature Extraction through SRL, NER, Choose MCS, Test Sentence pattern and Test the Question type pattern	WH- questions like who, when, where, why, and how.	English	170 sentences are extracted and mapped into 250 patterns using SRL and NER. The 250 patterns are used in training and testing, and Precision, Recall and F-measurement is used for classification of question type. The percentage of truly generated patterns increased 87% which appears to be promising ratio in this problem comparing it to other techniques used in generating questions automatically.
7.	Automatic question generation in multimedia- based learning	A Question generation for documentary video, Definitions, selection of system based generated post-question, Generation of human generated questions, selection of pre-question and selection of images, participants and interface, and procedure.	who, whom, where, whose, when , what	English language with image of question related.	A large-scale experiment investigating the productivity of generating questions (time taken to post-edit questions vs. time taken to generate questions from scratch) is planned.
8.	Automatic Generation of Question Bank Based on Predefined Templates	Knowledge Descriptor, Questions Generator, and E- learning Executer.	MCQ with onlyOneS oultion,tw oSoultions , allOfThe Above, and noneOfTh eAbove	English	A question bank is developed for course General Biology course (Bio110) at Faculty of Science, King Abdulaziz University (KAU): The bank contains 12 chapters with 239 sub topics and total of 46345 questions.
9	AUTOMATI C GENERATI ON OF MULTIPLE CHOICE QUESTIONS FROM DOMAIN ONTOLOGI ES	ontology-based strategies like class based, property based, terminology based strategies	MCQ (Choose the correct sentence)	English	The generated questionnaires were evaluated in three dimensions: Pedagogical quality, linguistic/syntactical correctness and number of questions produced.

10.	Review of	Extract person name, Generate	where,	Punjabi	-
	Question	who, Extract location generates	who and		
	Generation	Where and Extract date, generate	when		
	System From	when			
	Punjabi Text				
11.	Mind the	1) sentence selection, 2) question	the Gaps	English	manually analyze the generated
	Gap:	construction, and 3)	fill		questions and rate the question
	Learning to	classification/scoring.	question		
	Choose Gaps	_	_		
	for Question				
	Generation				
12.	Linguistic	The source text is divided into	why	English	This evaluation was conducted with
	Consideration	sentences, tokenizing, pos tagging,	what one		one file (Chemistry: Bonds) which
	s in	syntactic constituency parsing and	line		had 59 sentences, from which the
	Automatic	semantic role labeling is used in	questions		system generated 142 questions. The
	Question	system, matcher function is called	-		average linguistics score per pattern
	Generation	which will return a list of patterns			in this evaluation was 5.0 to 4.18.
		that match the source sentence's			
		predicate-argument structure.			

III. CONCLUSION

In this paper we present the review to generate questions automatically from a given text. As discussed many algorithms created and different methodology is used to generate the automatic question generation system. NLP is used to process the text and NER and SRL is used to identify the semantic relation. Most of work is done in English language and generating MCQ type questions. automatic question generation system is an open area where still there is a scope of a research is there for proposing methodologies by identifying complexities and type of question like one word answer and True or False need to generate.

REFERENCES

- [1] Automatic Cloze-Questions Generation by Annamaneni Narendra, Manish Agarwal and Rakshit shah LTRC, IIIT-Hyderabad, India.
- [2] Automatic Question Generation using Discourse Cues by Manish Agarwal_, Rakshit Shah_ and Prashanth Mannem
- [3] Automatic Question Generation Using Software Agents for Technical Institutions by Shivank Pandey1, K.C. Rajeswari International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970) Volume-3 Number-4 Issue-13 December-2013
- [4] G-Asks: An Intelligent Automatic Question Generation System for Academic Writing Support by Ming Liu and Rafael A. Calvo
- [5] Automatic Multiple Choice Question Generation System for Semantic Attributes Using String Similarity Measures by Ibrahim Eldesoky Fattoh in journal Computer Engineering and Intelligent Systems www.iiste.org ISSN 2222-1719 (Paper) ISSN 2222-2863 (Online) Vol.5, No.8, 2014.
- [6] Semantic Based Automatic Question Generation using Artificial Immune System by Ibrahim Eldesoky Fattoh in Computer Engineering and Intelligent Systems www.iiste.org ISSN 2222-1719 (Paper) ISSN 2222-2863 (Online) Vol.5, No.8, 2014

- [7] Automatic question generation in multimedia-based learning by Yvonne SKALBAN, Le An HA, Lucia SPECIA, Ruslan MITKOV.
- [8] Automatic Generation of Question Bank Based on Pre-defined Templates by Ahmed Ezz Awad and Mohamed Yehia Dahab in International Journal of Innovations & Advancement in Computer Science IJIACS ISSN 2347 – 8616 Volume 3, Issue 1 April 2014.
- [9] AUTOMATIC GENERATION OF MULTIPLE CHOICE QUESTIONS FROM DOMAIN ONTOLOGIES by Andreas Papasalouros, Konstantinos Kanaris, Konstantinos Kotis
- [10] A Review on Question Generation System From Punjabi Text by Payal Garg, Er.Charandeep Singh Bedi in International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) Volume 3, Issue 4, July-August 2014.
- [11] An Overview of Automatic Question Generation Systems by Ms. Deepshree S. Vibhandik, Prof. Mrs. Rucha C. Samant in International Journal of Science, Engineering and Technology Research (IJSETR), Volume 3, Issue 10, October 2014.
- [12] Mind the Gap: Learning to Choose Gaps for Question Generation by Lee Becker, sumit basu and lucy vanderwende.
- [13] Linguistic Considerations in Automatic Question Generation by Karen Mazidi and Rodney D nielsen

AUTHORS

First Author – Sheetal Rakangor, PhD Scholar, RK University and Email: rakangorsheetal@gmail.com

Second Author – Dr. Y. R. Ghodasara, College of Agricultural Information Technology, Associate Professor, Anand Agricultural University, Anand 388 110

Email: yrghodasara77@yahoo.co.uk

Correspondence Author – Sheetal Rakangor, Email:rakangorsheetal@gmail.com, 09428698260