

# A Survey of Methods for Automatic Question Paper Generation System

Mr. Shreyas R Chim <sup>#1</sup>, Dr. Geetanjali V Kale <sup>\*2</sup>

*#Department of Computer Engineering  
Pune Institute of Computer Technology,  
Pune, India.*

<sup>1</sup>[shreyas4329ay@gmail.com](mailto:shreyas4329ay@gmail.com)

<sup>2</sup>[gvkale@pict.edu](mailto:gvkale@pict.edu)

**Abstract**— Education is very important in today's era, and to evaluate student's knowledge, examination plays a vital role in every educational organization. Setting a question paper for examination is the tough tasks as well as an insecure one, due to human errors. We have presented an advanced technique that is an "Automatic Question Paper Generation System". This system uses various techniques such as randomization algorithm for question selection, backtracking algorithm, data mining techniques for question categorization, genetic algorithm, shuffling algorithm, Fuzzy logic etc. which is fast and more secure than the traditional one. This system maintains question repository in which questions are in proper unitized manner, they are mapped according to question priority level, Bloom's taxonomy level and question difficulty level etc. This system also contains the tagging of course outcome (CO) and program outcome (PO) with respective its questions to fulfil the paradigms of outcome-based education. This system randomly selects question in a categorized manner from the question repository and generate question paper. This system is useful for the various educational organization such as school, colleges, autonomous institutes, universities etc.

**Index Terms**— Automation, Automatic Question Paper Generation System, Blooms Taxonomy, Randomization, Outcome Based Education (OBE), Artificial Intelligence, Shuffling algorithm, Fuzzy logic, genetic algorithm.

## I. INTRODUCTION

Traditional way of setting question papers was a complex task, it was more time consuming and insure as well. Paper setters need to frame the questions accordingly and also need to prepare with the alternatives of the question sets for the worst condition, if the question paper gets leak.

To overcome this issue automatic question paper generation system (AQPGS) gets introduced. Due to the advancement in technologies, more efficient techniques are developed for the automatic question paper generation.

This paper presents the results of a literature review of various methods use for automatic question paper generation systems. The complexity of different methods and algorithms is discussed, and the paper provides a set of comparison criteria for all methods and a set of recommendations on the best methods to use depending on the type of question paper need to be generate.

Automatic question paper generation system (AQPGS) is the set of processes and techniques that designed to frame accurate question paper, protect confidentiality of question paper and avoid human errors.

There are two types of AQPGS.

- A. Automatic Question Selection and Setting Question Paper.
- B. Automatic Question Generation, Selection and Setting Question Paper.

Based on the above two approaches AQPGS are designed. In the first approach questions are maintain in the question repository, which are added in the database manually and taged with the different parameters which are consider at a time of question selection for the particular question paper set. The seletion is done with the help of different selection algorithm

e.g.: Randomization algorithm, suffling algorithm, Fuzzy logic etc. Where on the other hand second approach for automatic question paper generation is actually means the definition of automatic, questions are generated from the paragraph or text automatically, in this method instead of entering ready made questions in the repository, paragraphs are feeded into the system. Advance systems are capable of generating various types of questions automatically and also tagged question with different tags (e.g.: Blooms Taxonomy level, Course outcome (CO), Program Outcome (PO), Performance Indicator (PI) etc.). These systems are based on artificial intelligence. They use machine learning and data mining techniques question paper generation.

## II. MAJOR STEPS IN QUESTION PAPER GENERATION

In last few years education system has shifted its paradigms to outcome based. To measure the student learning outcome, it is important to maintain the quality of questions asked in the examination. Course Outcome (CO) and the Program Outcome (PO) are two main component of Outcome Based Education System (OBE). In todays era it is mandatory to design question paper which satisfies the learning

outcome. This student-centered system focuses mainly on outcome. Most of the educational organization follows the standards of OBE. Blooms Taxonomy is another aspect of teaching process. It contains the six cognitive level. These six levels classify the student learning objectives at each level hierarchically.

While designing the AQPG it is very much important to consider the above issues.

### III. METHODS FOR QUESTION PAPER GENERATION

Based on the type of AQPGS, below are the several methods that are proposed by the various researchers.

#### A. Generating question paper by selecting questions from question bank.

In this type of systems, the questions are manually feeded in the database which is used as a question bank, the question paper template is discussed, the questions are set according to its marks and the difficulty level. The paper setter chooses the template and the generate the question paper. All this task is done manually except question selection. The questions are selected randomly form the database [1]. Peijiang Chen in the year 2011 proposed this type of system. In this sytem the question is selected randomly according to requirement and condition specified. On the other hand, Syahaneim Marzukhi provide the advancement in this system by proposing manipulation of question paper, this system generates the question set by selecting random questions from database from which setter need to select exam questions from that set, if user satisfied with the question the he/she can proceed further else can select another set for exam question selection [2]. In 2014 K. Naik, S. Sule, S. Jadhav and S. Pandey proposed the sytem in which question database is mapped with cognitive level [3][4].

#### B. Extraction of Keywords.

Discourse connective based automatic AQPGS is proposed by A Narendra with M Agrawal and R Shah [5]. This system is capable of generating questions from natural language text. It contains syllabus module and semantic based AQPGS.

A new model is proposed by Bo Liu and Fang Yaun for information extraction from PDF. In this approach the main task of information extraction algorithm is to find out the right-left side symbols of each attribute. This process contains three modules, first moduls is for construct PDF parser, second constructs the tag injector and third one process the user defines tags [6].

#### C. System based on Intelligent algorithm.

These papers emphases on improving the question paper quality through various intelligent techniques.

Random selection and backtracking are the most popular techniques in this domin. Selection without following any deterministic pattern is called randomization. Backtracking is the recur-

sive techniques to build incremental solution in one piece at a time. Algorithm are used for selection of question by removing the factors that fails to satisfy the constraints of AQPGS [13].

To implement AQPGS reseachers use various Data Mining and machine learning techniques such as neural network and rule systems. Among neural network recurrent neural network (RNN) is popular for NLP. For the generation of "Wh" question L. Reznik uses the long short-term memory (LSTM) network, which is more advance model of RNN [21].

S. Bindra in 2017 proposed the Outcome Based Predictive Analysis of Automatic Question Paper using Data Mining [22]. In this approach author design the table structure to store required question along with the CO and PO attainment data. Questions are entered in the designed framework. The proper procedure is developed to extract keywords, levels from questions in entered in database which will further use for extraction of CO and PO. Set up repository for CO and PO attainment. The classification is done by using various data mining algorithm such as, Navie Bayes, IBk, J48, ZeroR, oneR etc.

#### D. Genetic Algorithm based system

Genetic algorithm (GA) is a one kind of randomization, it is a random search method which is biologically inspired. It considered the population. Kui Zhang proposed the improved version of GA to solve the AQPGS [11]. In this method, segmented decimal coding is used, this coding is generated on the basis of number of question type and number of each question type. Crossover operation will take place in each coding. In the form of multipoint crossover chromozomes will be express, this results in making segments as per question type.

## IV. OBSERVATIONS AND RECOMMENDATIONS

The extent of research papers found on automatic question paper generation shows that these techniques are pervasive and growing research area for advance education system tools. The main question stands now is, which of these techniques are more effective for question paper generation system? Unfortunately, this not yet established.

#### A. Obesrvation Related to Question Database/Repositories

Framing the appropriate question paper set, it is mandatory that the input data feed in database should be proper, incorrect input will directly affect the quality of question paper, this may despoil the superiority of question paper. The questions in the repository should contains the proper tagging of Blooms taxonomy level, difficulty level, marks occupaied by the particular question, priority etc. These are the basic factor required for the setting question set. Coming to the Outcome Based Education System Paradigms, it is necessary that each question should mapped with CO and PO.

#### B. Factors related to AQPGS Performance

One of the most important factors related to the performance of AQPG is the input data, as we have discussed in the above section, wrong input, incorrect tagging will directly result in the generation of incorrect question sets. To avoid this, there will be the need of taking special care at time of feeding data.

Second Factor related to the AQPGS performance, is the type of algorithm is going to be employed and design of the overall application. The literature survey revealed that every method used for AQPGS are with some drawback. However, study shows that, each of these methods having its own advantages and disadvantages depending upon the type of questions need to be generated e.g.: MCQ's, Short Question, Application Level question, forming equations etc. Integrated use of more than one method and several modules can result in a much better system.

### C. Comparison Criteria

There are several criteria by which the methods for AQPGS could be compared:

- Accuracy
- Time for training a model in case of Machine Learning and Data Mining.
- Speed

If we compare the accuracy of several methods for AQPGS, it is very necessary that the database should be the same. Accuracy is directly dependent upon the input data it contains, the parameters that are defined in particular methods. Same in case of advanced AQPGS which uses Data mining and machine learning techniques.

If one were to compare the accuracy of several intelligent methods, those methods should be trained on exactly the same training dataset and tested on exactly the same testing dataset.

Time for generating and categorizing new questions from the input in case of NLP techniques is another factor for comparison.

### D. Data Mining and Machine Learning Recommendation for AQPGS

The extent of research papers found on automatic question paper generation shows that these techniques are pervasive and growing research area for advanced education system tools. The main question stands now is, which of these techniques are more effective for question paper generation system? Unfortunately, this is not yet established.

### E. Observation Related to Question Database/Repositories

Framing the appropriate question paper set, it is mandatory that the input data feed in database should be proper, incorrect input will directly affect the quality of question paper, this may spoil the superiority of question paper. The questions in the repository should contain the proper tagging of Blooms taxonomy level, difficulty level, marks occupied by the particular question, priority etc. These are the basic factors required for the setting of a question set. Coming to the Outcome Based Education System Paradigms, it is necessary that each question should be mapped with CO and PO.

### F. Factors related to AQPGS Performance

One of the most important factors related to the performance of AQPG is the input data, as we have discussed in the above section, wrong input, incorrect tagging will directly result in the generation of incorrect question sets. To avoid this, there will be the need of taking special care at time of feeding data.

Second Factor related to the AQPGS performance, is the type of algorithm is going to be employed and design of the overall application. The literature survey revealed that every method used for AQPGS are with some drawback. However, study shows that, each of these methods having its own advantages and disadvantages depending upon the type of questions need to be generated e.g.: MCQ's, Short Question, Application Level question, forming equations etc.

The matrices should be predefined for the CO and PO performance e.g.: Kappa statistics, processing time, mean absolute error, root mean square error, correctly and incorrectly classified instances etc. Integrated use of more than one method and several modules can result in a much better system.

### G. Data Mining and Machine Learning Recommendation for AQPGS

Study shows that J48 provides the best classification accuracy [21] which is up to 90% and the highest classification accuracy recorded. Where on the same dataset Naive Bayes is with second highest accuracy of 85%, OneR and ZeroR with 70% accuracy, and IBk with 68.33% accuracy.

## V. CONCLUSION

In this paper, literature review of various methods used for automatic question paper generation is presented. Special emphasis was placed on finding example papers that describe the use of different intelligent techniques in this domain, for selection of questions, generation of questions and categorization of questions.

Unfortunately, the methods that are the most effective for automatic question paper generation system have not been established; and given the richness and complexity of the methods, it is impossible to make one recommendation for each method, based on the type of question paper the system is supposed to generate.

When determining the effectiveness of the methods, there is not one criterion but several criteria that need to be considered. They include accuracy, complexity, time for classifying an unknown instance with a trained model, and understandability of the final solution (classification) of each method. Depending on the particular type of question paper, some might be more important than others.

Another crucial aspect of automatic question paper generation are question selection parameters and the question paper template generation. Advanced techniques like machine learning and deep learning cannot work without proper dataset, therefore for training the learning model, there is a need to maintain proper dataset. If proper question-answer repository data are available, these data must be used for automatic student answer sheet evaluation system in features.

### ACKNOWLEDGMENT

We express our deepest sense of gratitude towards **Dr. Rajendra S. Talware**, Manager, IT-Cell and Coordinator, Exam Automation, Savitribai Phule Pune University, Pune, for giving us an opportunity to work on this project. We thank him for his constant support and providing us the necessary resources to complete this work.

I am grateful to **Dr. P. T. Kulkarni**, Principal, Pune Institute of Computer Technology for his constant encouragement and for providing valuable resources when required.

### REFERENCES

- [1] Peijiang Chen," Construction of Test Database for Electrical and Electronic Technology", International Conference on Electronic & Mechanical Engineering and Information Technology, IEEE, August 2011, p 5.
- [2] Syahaneim Marzukhi, Muhammad Firdaus Sulaiman. "Automatic Generation Question System to Enhance Preparation of Learning Assessment", p 6.
- [3] Kapil Naik, Shreyas Sule, Shruti Jadhav and Surya Pandey, "Automatic Question paper Generation System using randomization algorithm" IJETR, Vol.2, Issue 12, pp.1-3, Dec 2014.
- [4] Surbhi Choudhary, Abdul Rais Abdul Waheed. Shrutika Gawandi and Kavita Joshi, "Question Paper Generator System," International Journal of Computer Science Trends and Technology, vol. 3, issue 5, Sept –Oct 2015.
- [5] Automatic Cloze-Questions Generation by Annamaneni Narendra, Manish Agarwal and Rakshit shah LTRC, IIIT-Hyderabad, India
- [6] Fang Yuan, Bo Liu. "A New Method of Information Extraction from PDF Files", The Fourth International Conference on Machine learning and Cybernetics, IEEE, August 2005, p 4.
- [7] Husam Ali, Yllias Chali," Automatic Question Generation from Sentences, University of Lethbridge, TALN 2010, p. 19–23.
- [8] Burcu Yildiz, Katharina Kaiser, Silvia Miksch. PDF2table: "A method Extract table information from PDF files. IEEE, 2014, p 6.
- [9] Vasile Rus, Brendan Wyse," Question Generation Shared Task and Evaluation Challenge, Proceedings of the 13th European Workshop on Natural Language
- [10] S. Chen, S. Diggavi," Efficient String Matching Algorithms for Combinatorial Universal Denoising". Generation (ENLG), Association for Computational Linguistics, September 2011, pages 318–320
- [11] Kui Zhang, Lingchen Zhu. "Application of Improved Genetic Algorithm in Automatic Test Paper Generation", IEEE, 2015 p 495-499.
- [12] S. S. Sheik, † Sumit K. Aggarwal," A FAST Pattern Matching Algorithm", Bioinformatics Centre and Supercomputer Education and Research Centre, Indian Institute of Science, J. Chem. Inf. Comput. Sci, 2004, 44, P.1251-1256.
- [13] Jing Mei Li, Jing Li, Jian Pei Zhang, Nan Ding. "Research on intelligent test paper generation based on multi-variable asymptotic optimization". IEEE ,2010, p 6.
- [14] Automatic Question Generation using Discourse Cues by Manish Agarwal\_, Rakshit Shah\_ and Prashanth Mannem.
- [15] Ritvars Bregzis, Calvin Gotlieb, Carole Moore, "The Beginning of Automation in the University of Toronto Library, 1963–1972", in IEEE Annals of the History of Computing, April–June 2002.
- [16] Prof. Godswill Obioma Prof. Ismail Junaidu, Dr. Grace Ajagun, "The Automation of Educational Assessment in Nigeria: Challenged and Implications for Pre-Service Teacher Educaion", Annual Conference of the International Association for Educational Assessment (IAEA) held at the Dan Panorama Hotel, Tel-Aviv, IsrealOctober 20th –25th, 2013
- [17] M. Jou, J.K. Shiauand H.W. Zhang," Application of Web Technologies in Automationon Technology Education", International Journal of Computers and Applications, Vol.31, No. 4, 2009
- [18] Xiang Fu, Boris Peltsverger, Kai Qian, Lixin Tao, Jigang Liu," APOGEE – Automated Project Grading and Instant Feedback System for Web Based Computing", Computer Science and Information Technology, 2nd IEEE International Conference 2009.
- [19] Gerald Weber," Defining the Paperless Workplace with the Paper Metaphor -Not a Contradiction in Terms", Conference: Proceedings of the Fourth Australasian Workshop on Health Informatics and Knowledge Management - Volume 120.
- [20] Vaibhav M. Kale, Arvind W Kiwelekar. "An algorithm for question paper template generation in question paper generation system". IEEE, 2013, p 8.
- [21] Simranjeet Kour Bingra, Akshay Girdhar, Inderjeet Singh Bamrah. "Outcome Based Predictive Analysis of Automatic Question Paper using Data Mining". ICCES IEEE,2017
- [22] Akhil Killawala, Igor Khokhlov, Leon Reznik. "Computational Intelligence Framework for Automatic Quiz Question Generation". IEEE International Conference on Fuzzy Systems (FUZZ), 2018.