
Common Sourcing Model for Generating Large Question Banks & Automatic Question Paper Generation

Dr. T. Sreenivasulu,

M.Tech, Ph.D., Professor

Department of Computer Science & Engineering, St. Peter's Engineering College, Kompally, Hyderabad, Telangana

Veeral M. Thakker,

UG Scholar, Department of Computer Science & Engineering, St. Peter's Engineering College, Kompally, Hyderabad, Telangana,

V. Sudhamsh Reddy,

UG Scholar, Department of Computer Science & Engineering, St. Peter's Engineering College, Kompally, Hyderabad, Telangana

P. Megha Raj,

UG Scholar, Department of Computer Science & Engineering, St. Peter's Engineering College, Kompally, Hyderabad, Telangana

***Abstract:** At present, we follow the conventional method of setting a question paper in which, a lecturer sets the paper based on his/her experience and understanding of the subject. Using this method limits the adversity of questions and their difficulty. Generating large question banks can aid in generating a question paper with a random difficulty level and the questions that may be asked cannot be predicted. This can be useful for schools, colleges, autonomous Institutes, and Universities. We propose designing a portal that provides two kinds of logins, one for the lecturers who can submit the questions to the board and the other for the administrators or the board member who check all the questions and validate the questions so that the question paper can be generated automatically by the system. At the time of registration, the lecturer has to submit an Identity Proof (for example, Institute/University ID Card) to prove his/her authenticity in order to avoid a non-authentic user to submit a question paper (such as a student, non-teaching staff, etc.).*

Lecturers can set the different levels of difficulty namesly, Easy, average, difficult for each question while submitting the paper and board members can instruct the automatic question paper generator system to set the difficulty of question paper by allotting the number of questions to be picked from each difficulty type. The AI system will detect the questions that have the same semantics but a different representation to avoid having ambiguity while generating the question paper. Also, we show the frequency of the topics from which the questions are asked frequently. This is done in order to determine the most important topics or the hottest topic in the question banks.

1. INTRODUCTION

Examinations are conducted almost every day across the world. For every examination, a question paper has to be set. This is usually done by the lectures, if the examination is a curriculum based examination, and by the subject experts, if it is an entrance examination for admissions in Universities, etc. Question papers almost always follow a pattern if they are set by a lecturer. This has its own benefits and drawbacks which are discussed later in the paper. There are several techniques which have been implemented in generating question papers automatically such as using Discourse Cues ^[1], Cloze Question Generation (CQG) technique ^[2], etc.

We propose a system in which the question paper is set by the system without any interference of humans. Question paper generation via system or software could reduce the workload of the Institutions and the Universities. This system can be used in primary and high schools, colleges, Institutions, Universities and also those organizations which wish to conduct an examination. Using this system would even decrease the burden of gathering question banks and then choosing the questions to set the question paper. This is because we allow lecturers and subject experts from all over the country to submit a question bank. This would create a database consisting of huge number of questions. Also, there would be a variety of questions while setting a

question paper as the question banks are provided by numerous lecturers. Then the questions are randomly picked from this pool of questions. The whole process is discussed later in the paper.

2. EXISTING SYSTEM

Although there have been many systems that have been proposed for generating question papers, they are rarely used by Institutes and Universities because of the complexities in them and their unavailability or inaccessibility. So, the most widely followed method of setting a question paper is the traditional method in which the lecturer from a particular Institute submits a question paper or by a subject expert in a University. This method is being followed since decades. Albeit it has some disadvantages, the lack of other options has forced people to stick with this method.

2.1. Disadvantages:

1. It is more time consuming as the lecturer has to first prepare a question bank and then carefully select the questions to be set in the question paper.
2. It also possesses several security or privacy issues since the question paper can be leaked before the examination through various ways and for various reasons.
3. The number of questions available is very less.
4. There is less variety or diversity in the questions as their creativity is limited to the knowledge of one individual.
5. The difficulty of the question paper can be predicted and so can the questions.

These challenges threaten the genuineness of the question paper. If any of these challenges occur then the motive of the question paper is compromised. This can result in the whole examination's results being biased or predictable. In other terms, there would be no point in conducting the examination. These issues cannot be overlooked. So some sorts of measures are taken by the Universities and Institutes to maintain the secrecy and confidentiality of the question paper. However, this doesn't nullify the chances of these challenges occurring.

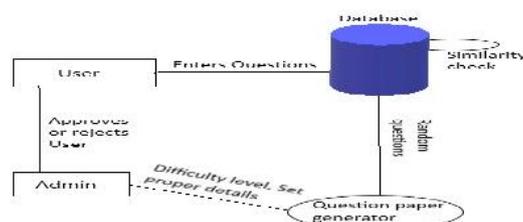
3. PROPOSED SYSTEM

We propose a system that allows us to gather questions from various sources and then generate a question paper with questions randomly^[4] picked from the large question bank generated. The idea behind gathering questions from various sources is to diversify the range and type of questions. And the idea behind generating a question paper by randomly picking the questions from this large database of questions is to make the question paper unpredictable and to reduce the human effort required.

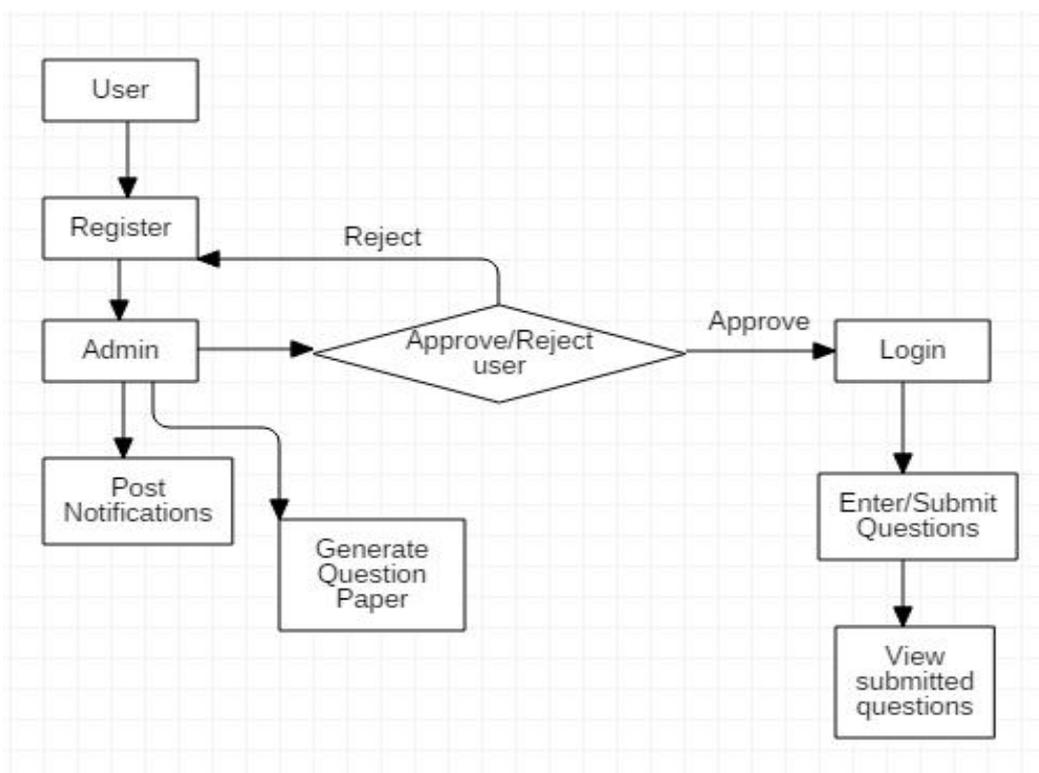
3.1. Advantages:

1. It requires less time and cost in generating the question paper. Also, the human involvement is minimal.
2. The question paper so generated would be unpredictable since they are picked randomly from the question bank.
3. The question paper cannot be leaked as it will be set dynamically by the system.
4. Large pool of questions to choose from.
5. Large variety of questions as the knowledge is not limited to a single individual but to all the people who have knowledge in a particular subject or field.
6. The question paper would be of random difficulty, too, as the questions are chosen randomly.

3.2. Architecture:



3.3. Flowchart:



The system provides two kinds of login, one for the user (the lecturer) and one for the administrator. Before being able to login, the users have to register themselves into the system. At the time of registration, the users have to specify their full name and the organisation that they are working in and a “Proof of Identification” for proving their authenticity. This proof could be the “Letter of Authentication” signed by the Head of the Organisation or any equivalent document that proves the authenticity of the user. The admin would verify this “Proof of Identification” by contacting the organisation and then confirm the registration of a user by sending an e-mail to the e-mail Id. provided by the user at the time of registration. The users even need to specify their field of expertise or teaching and the experience (in years) they have in that particular field. After successful registration, the user is provided with a default password which can be changed after logging in based on user’s discretion. After logging in the user can perform actions such as updating his/her profile, submitting the question bank they have, view the submitted question bank, etc. While submitting the question bank, the user has to specify the difficulty and the key term or the keyword which the question is related to. There are three types of difficulties namely easy, medium and hard. Here, even if a lecturer submits a question bank in which all the questions have “Easy” difficulty level it would not ensure that the question paper generated will be of “Easy” difficulty as there are various other questions to choose from. Thus, this neutralizes the biasness of the question paper and preserves the genuineness of the question paper. The purpose of asking the user to enter the keyword related to the question is to generate a question paper which can cover the questions from the entire syllabus or the curriculum. The system takes the questions for different weightage of marks such as 2, 3, and 5. The user needs to provide a specified minimum number of questions for each category of marks. The user, after providing the specified number of questions along with filling all the necessary fields, can submit the question paper.

At the time of submission, the questions that are entered or submitted by the user are cross-checked or compared with the questions that already exist in the database. This is done to avoid have too many similar questions in the database. The chance of redundancy here is high as there are many users who can submit a question bank. So, it is necessary to reduce the redundancy as much as possible. For this, we use the following formula,

$$si \quad it \quad (S) = \frac{t_i \quad n_i \quad o \quad si \quad w \quad i \quad the \quad q}{t_i \quad n_i \quad o \quad u \quad w \quad i \quad the \quad q} \quad \dots (1)$$

We compare each question entered by the user with all the questions in the database and calculate the SI between those two questions. The similarity index is compared by converting the sentence into tokens ^[3] and then comparing them. We compare the SI with a threshold value. Here, if the SI value is greater than the threshold value then the question is rejected and not stored in the database and also, it is not compared with any further questions. If the SI value is lesser than the threshold value, then the question is compared with the next question in the database. This continues till either the entered question is compared with all questions in the database or till the SI value is greater than the threshold value. To avoid making the calculation of the SI value difficult, we perform the “stemming” operation on the questions. This is done so that the uniqueness or similarity of the question is calculated on the base or the stem words fairly and easily. After the questions are submitted and stored successfully into the database, the user can only view the questions that he/she has submitted. In other words, each user can submit only one question bank. In the question bank that the user enters, he/she can enter as many questions as he/she wants. This is done to make sure that the questions in the database are not from a single user but from multiple users. The purpose behind being able to view the submitted question bank is to make the user aware of the questions that have been accepted or rejected. If many questions are rejected, then the admin can allow the user to submit another question bank and also preserving the previously accepted questions.

The administrator has a major role in the functioning of the system. The admin has to verify the “Proof of Identification”. Based on the verification, the admin can approve a user by sending him/her the password to the registered e-mail ID, or reject the user’s registration request. The admin also has to notify the users about the various updates related to the system and also about the important notifications such as commencement of the submission of the question bank, last date for submission of the question bank, dates of the examinations, etc. The admin is also responsible for generating the question paper. The question paper generated is stored in a separate file and then sent to the respective Head of the Organisation. This ensures that the question paper doesn’t fall into the wrong hands and is kept confidential. The question paper can be generated a few hours before the examination to avoid any misuse of the question paper and preserve the motive of conducting an examination.

The question paper generated would cover different topics of the syllabus and won’t solely concentrate only on the important questions/topics. This is done by choosing only one question from every keyword from the question bank under each weightage category of marks. In simple terms, we instruct the system to select only a single question from every distinct keyword in the database. This is done for every weightage category of marks. To make sure that all the questions in all the keyword categories are selected. Also, the difficulty of the question paper can be selected by the admin (if required) in order to make the question paper fair to all the students or candidates appearing for the examination. This can be done by assigning a value for each difficulty type. For example, “Easy” can have value as 2 or 3, “Medium” can have value as 4 or 5 and “Hard” can have value as 6 or 7. Then, we sum the values of all the questions generated to obtain an average difficulty value of the questions. This can be shown by the following equation,

$$A \quad D = \frac{S_i \quad o \quad d \quad v \quad o \quad a \quad q}{T \quad a \quad n \quad o \quad q} \quad \dots (2)$$

While generating the question paper, the required range of the difficulty can be set by the admin and the question paper will be generated by keeping the range in mind. This is, again, done by calculating the average difficulty value.

This system can also be used in “Online-based Examinations” such as entrance tests, or National level competitions, competitive examinations, etc. In such examinations, the question paper is generated randomly for every candidate. This would ensure that the question paper is unique for almost every candidate.

4. CONCLUSION & FUTURE WORK

The system developed aids in decreasing the time, cost and effort required to gather questions and generate a question paper. It also makes sure that the source of question banks is authentic and the generate question

paper is safe from any kinds of security threats. The questions are randomly chosen and with random difficulty. In future, we look forward to simplifying the similarity index equation to decrease the overhead caused in comparison. Also, we plan on implementing the same system for Online based examinations.

ACKNOWLEDGEMENTS

The authors express their gratitude towards the Institution for providing us this opportunity. We also are grateful to Dr. Raja M, Head of Department, Department of Computer Science & Engineering, St.Peter's Engineering College, Hyderabad, for supporting us throughout the research.

5. REFERENCES

- [1] Automatic Question Generation using Discourse Cues by Manish Agarwal_, Rakshit Shah_ and Prashanth Mannem.
- [2] Automatic Cloze-Questions Generation by Annamaneni Narendra, Manish Agarwal and Rakshit shah LTRC, IIIT-Hyderabad, India.
- [3] Tokenizing: <https://docs.python.org/3/library/tokenize.html>
- [4] Random: <https://docs.python.org/2/library/random.html>