Crowd Sourcing Model for Generating Question Banks and Question Papers

V. Soniya,

M.E., Assistant Professor, Department of Computer Science & Engineering, St. Peter's Engineering College, Hyderabad

Akhil Munagala,

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

Vamshi Krishna Goud.S.P,

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

Kudidhi Samyukth,

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

Abstract: *The traditional method of setting a question paper by a particular person teaching in the university. This method is limited in terms of number of questions, the difficulty of the question paper, and the experience of the lecturers. Generating large question banks can aid in generating a question paper with a random difficulty level. This can be useful for schools, colleges, autonomous Institutes, and Universities. We propose, designing a portal that allows the lecturers from various institutes to login and submit the questions to the board and another login for the board or the administrators who can check all the questions and validate the questions so that the question paper or a question bank 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 level of difficulty i.e., Easy, average, difficult for each question while submitting the questions 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.*

Keywords: Automatic generation, Natural language processing, artificial intelligence system

1.   INTRODUCTION

Examinations are good way to test skills and knowledge of individuals, and this system can be able to adopt by almost everyone who wish to test the knowledge, skills. Exams are, conducted almost every day around the world. Question papers for examinations, are generally set by lecturers belong to that subject or by the people who have good knowledge in them. Setting question the paper, this way has its own advantages and disadvantages. There are different techniques, which can used to generating question papers automatically by using Discourse Cues [1], Cloze Question Generation (CQG) techniques [2], etc.

For setting question papers, we propose a system, which will generate the question paper with little to no need of human interaction, which will hugely reduce the manual efforts of institutions. This system can be, adopted by almost every institute, which wish to conduct examinations. This system could decrease the burden of gathering question banks and choosing the questions to set question paper. This is because question banks, are generated by gathering different questions from different lecturers all over the country, this would create huge database of questions and from this, and completely random questions can be selected for generating question papers. With the help of natural language processing[3][4][5] techniques, we can make sure that no questions, are repeated more than once in the question paper, and this technique is also helps in similarity checking, which is completely done by using python language [6] and it's frameworks such as pandas [7], numpy [8][9] etc. We also use stop words [10][11][12] to remove unwanted words from questions and creating corpus [13][14][15] for better similarity checking.

2. EXISTING SYSTEM

There are different methods, which propose for generating question papers, but unfortunately, Institutes or Universities rarely use them because of their complexities. So they are forced to stick to the traditional methods of setting the question paper i.e., lecturers setting the paper manually. However, the problem with this traditional method is it is having lot of disadvantages.

2.1. Disadvantages:

1. It is more time consuming as lecturer has to set them manually.
2. Selecting different questions can be tedious task as there is no specific question bank to choose questions.
3. Difficulty of question paper can predicted easily, as well as questions.
4. There are less variety of questions because question paper set by the knowledge of one or two individuals.
5. It also has huge risks, of paper to leak before examination.

These challenges threaten the genuineness of the question paper. If any of these situations arises motive, conducting examination will compromise. This results in examination being 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 does not solve the chances of these challenges occurring.

3. RELATED WORK

So in order to solve this problems, we propose a system that allows us to gather differ questions from various sources and from a question bank so that question paper can be generated by selecting questions form this question bank. This is done diversify the range and type of questions. In addition, this system make sure that question paper is completely unpredictable and at the same time reducing the human efforts.

3.1. Advantages

1. It reduces the cost and time of the institute and requires less human involvement.
2. Chances of question paper leaked reduced as the question paper generated a little while before exam.
3. Question paper generated by the system is unpredictable since it randomly picked by the system.
4. Presence of large question bank to select wide range of questions.
5. Large variety of questions available since knowledge is not depended on single individual.

3.2. Architecture:
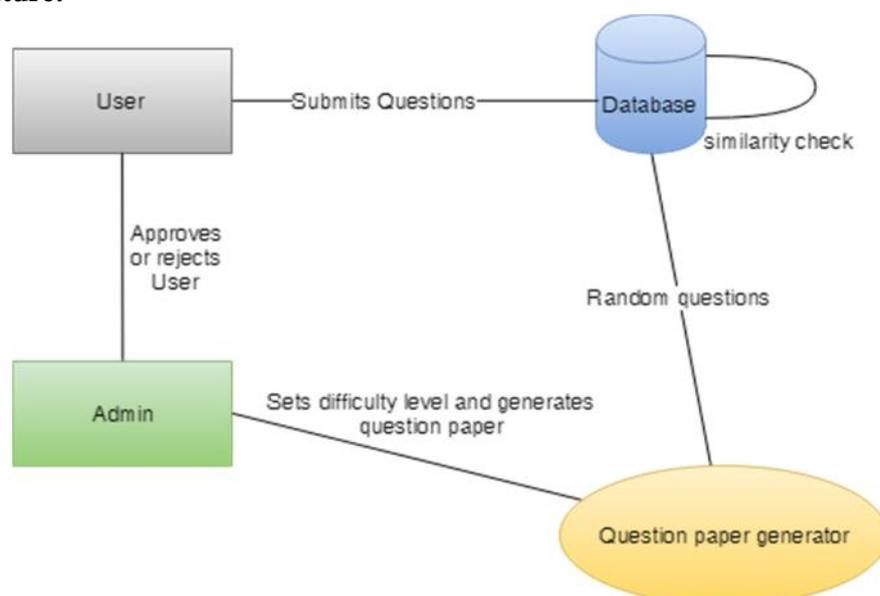
3.3. Flowchart:
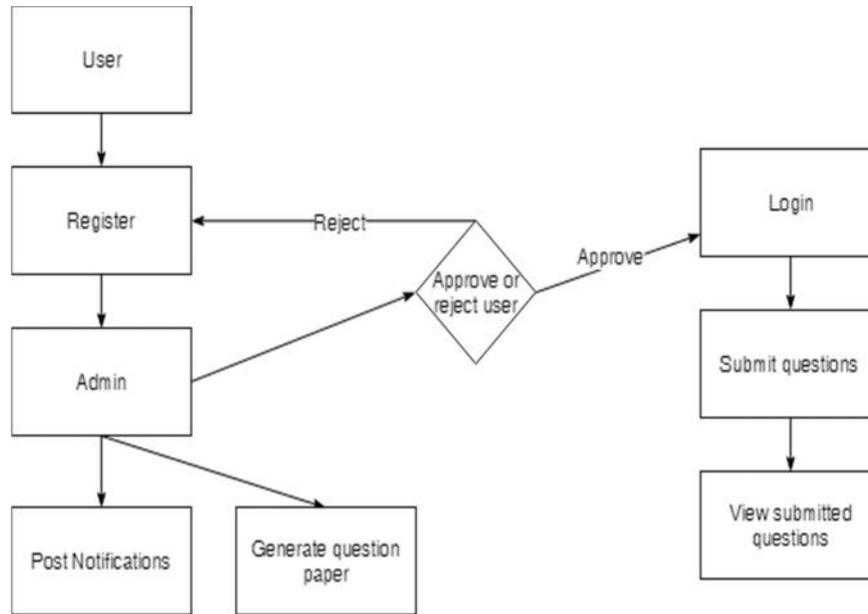
3.4. Design Method:

The system provides two kinds of login, one for the user (i.e. the lecturer) and one for the administrator. Before being able to login, the users have to register themselves. 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. Users can login and perform actions such as updating his/her profile, submitting the questions they have, view their submitted questions, etc. While submitting the questions to the question bank, the user has to specify the difficulty level for each question and the key term, which the question related to. There are three types of difficulties namely easy, moderate and high. 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 i.e. 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,

$$si \qquad i n \qquad (S) = \frac{t \quad n1 \quad o \quad si \quad w \quad i1 \; the \; q}{t \quad n1 \quad o \quad u \quad w \quad i1 \; the \; q} \qquad \dots (1)$$

We compare each question entered by the user with all the questions in the database and calculate the SI between those two questions. We compare the SI with a threshold value. Here, if the SI value is greater than the threshold

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 5
May 2018

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 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 0, "Moderate" can have value as 1 and "High" can have value as 0. 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,

$$A \quad D \quad = \frac{Si \quad o \quad d \quad v \quad o \quad a \quad q \quad es}{T \quad n_1 \quad o \quad q_1} \qquad \dots (2)$$

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.

2.   CONCLUSION & FUTURE WORK

This system aids in decreasing cost, time and effort required to gather questions and generate question paper. In addition, it ensures that source of questions in question bank is authentic and generate question paper avoiding security problems. In addition, we plan to implement the same system for online-based exams as well.

3. 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] Natural language processing by T. Patten Analytic Sci. Corp and P. Jacobs

[4] Using NLTK for educational and scientific purposes by Mykhailo Lobur CAD Department, Lviv Polytechnic National University, 12, S. Bandery Str., 79013, Ukraine and Andriy Romanyuk CAD Department, Lviv Polytechnic National University, 12, S. Bandery Str., 79013, Ukrain.

[5] For natural language processing using python tutorial https://www.analyticsvidhya.com/blog/2017/01/ultimate-guide-to-understand-implement-natural-language-processing-codes-in-python/

[6] For python language basics and syntaxes https://docs.python.org/3/

[7] Pandas framework python http://pandas.pydata.org/pandas-docs/stable

[8] Numpy python framework http://www.numpy.org/

[9] Numpy python framework and tutorial http://cs231n.github.io/python-numpy-tutorial/

[10] Stop words tutorial https://www.geeksforgeeks.org/removing-stop-words-nltk-python/

[11] For removing stop words from questions https://stackoverflow.com/questions/19130512/stopword-removal-with-nltk

[12] https://pythonprogramming.net/stop-words-nltk-tutorial/

[13] Creating a corpus for automatic punctuation prediction in Persian texts by Seyyed Mohammad Saleh Hosseini Department of Computer Engineering, Sharif University of Technology, Tehran, Iran and Hossein Sameti Hosseini Department of Computer Engineering, Sharif University of Technology, Tehran, Iran.

[14] https://www.packtpub.com/books/content/python-text-processing-nltk-20-creating-custom-corpora

[15] https://stackoverflow.com/questions/4951751/creating-a-new-corpus-with-nltk