_____

Design for Semi-Automatic Generation of Question Paper from A Semantically Tagged Distributed Question Repository

Gauri Nalawade
Dept. of Computer Technology
Shah & Anchor Kutchhi Engineering College
Chembur, Mumbai, India
suvarna.benke@gmail.com

Prof. Rekha Ramesh
Dept. of Computer Technology,
Shah & Anchor Kutchhi Engineering College
Chembur, Mumbai, India
rekha.sakec@gmail.com

*Abstract:-*In any educational course curriculum, the courses are defined with Learning Objectives. Teacher assess students to check whether the students have achieved the objectives or not for the particular course. So examination becomes an important activity to test the performance of the student. In fact, the nature of the exam questions determines the quality of the students produced by the institutions. Preparing the exam questions is very challenging, tedious and time consuming job for the teachers. Question repository (QR) is one solution which is useful to schools, Institutes, publishers and test paper setters. There might bevariety of questions in a repository but they could be useful only if properly tagged. Researchers recommend different set of tags such as cognitive level, difficulty level, type of question, content /topic, etc. Hence it is important to tag questions in a QR with such a set of tags. Here we are proposing a system design for semi-automatic   generation of question paper from semantically tagged distributed question repository.

*Keywords:-*Keywords: Question Repository, Moodle, tagging, cognitive level, Blooms Taxonomy
_____*****_____

1. INTRODUCTION

In any educational course curriculum, the courses are defined with Learning Objectives. Teacher assess students to check whether the students have achieved the objectives or not for the particular course. The assessment can be in online or offline mode. In today's scenario where use of Internet is widespread, teachers as well as students find it convenient to fulfill the learning objectives of the course using online mode. Use of LMS is thus inevitable for educational institutes.

Examination becomes an important activity to test the performance of the student. In fact, the nature of the exam questions determines the quality of the students produced by the institutions. Different types of examinations get conducted in a year in any academic course. Those vary from term test, practice tests, online quizzes, weekly tests, mock tests etc. Making question papers for these exams is tedious and time consuming job. It is very challenging for the teachers to meet certain criteria such as it should be able to verify the intellectual growth of the students, check understanding of the topic/concept, should cover all aspects of the course objectives and avoid duplication of questions in the subsequent exams. It would be difficult for any teacher to create such questions frequently as they are occupied with various other academic activities other than delivering just lectures. At times, all these factors may deteriorate quality of the exam paper.

Technology may assist in this case if we have software which can store questions and generate question paper frequently with ease. Teachers can keep on adding questions to the database while can use existing question bank for generating variety of question papers as per their need without being concerned about repetition and duplication from the past exam.

Question repository (QR) is one solution which is useful to schools, Institutes, publishers and test paper setters who want to have a huge database of questions and generate test papers frequently with ease. Questions in a repository could be useful only if they are properly tagged. Researchers recommend different set of tags such as cognitive level, difficulty level, type of question, content /topic, etc. Hence it is important to tag questions in a QR with such a set of tags. There exist many popular learning management systems (LMS) like Moodle, Desire2Learn, ATutor, Sakai, Blackboard, Sumtotal, Totara, etc. which facilitate generation of question repository and their management. Choosing right LMS is challenging as different organizations have different needs. Corporates may need LMS for conducting trainings to staff whereas educational institutions need it for conducting courses and assessments. Size of organizations also varies. Universities need LMS who can handle more than 10000 users at a time whereas in case of colleges, it may be in thousands.

Having tagged question repository is not enough. Teachers or question paper setters should be able to generate the question paper from this repository as per their needs within few seconds. The questions could be shortlisted for online quiz, term test, unit test, weekly test etc. by teacher. The aim of this paper is to understand the importance of tagging and commonly used tags while creating questions and propose a system design to generate question paper based on these tags. Section 2 discusses the extensive literature survey done in this regard. Section 3 describes need for question repository, need for tagging and various question tags. Existing LMS features and their limitations are discussed in section 4. Section 5 is all about new system architecture. Section 6 consists of summary and conclusions and scope of future work.

2. LITERATURE REVIEW

We did extensive literature survey to understand the need for tagging, types of tags, Various LMS and need for semi-automatically generation of question paper from semantically tagged question paper.

_____

Authors Vijay KP and others [1] proposed an adaptive question bank management system that chooses questions from question bank and generated the question paper based on the inputs or provided by the instructor. The system is implemented using database having huge collection of questions represented using concept map graph. Authors KapilNaik and ShreyasSule with colleagues [2] presented the solution in form of Automatic Question Paper Generator System (QGS) which makes use of shuffling algorithm as a randomization technique.

Author Rekha R.[3] proposes a software system that semi-automatically generates metadata for questions in a question repository. This paper elaborates the need for question tagging and various tags recommended by expert teachers. The most common set of tags used are cognitive level, difficulty level, question type and content/topic. Saha A.K. [4] considered Stack overflow site for tagging and thus categorizing the questions to help ensure that their users are aware of questions related to their areas of expertise or interest.

Sucunuta [5] proposes the architecture for a Question-Answering System for a storing the documents in a standard way. The system consists of three modules: the first doing analysis operation and provides GUI for question entry, second focuses on searching the documents whereas third module extracts the response. Caminero, A.C. [6] performed a comparative study between the most widely used LMS namely .LRN, Sakai, and Moodle. The study was based on hardware parameters such as CPU utilization, memory usage, disk read/writes, network I/O. The second evaluation was from administration point of view. From both system and admin point of view Sakai stands on the top of rest all in system performance. But to support large number of users easily, there is no other option than Moodle. Author Ramesh V M presented rubric to evaluate LMS against various criteria from student as well as teacher's perspective and found that Moodle is the best LMS to be used for any educational institution.

3. QUESTION REPOSITORY MANAGEMENT SYSTEM

3.1 Question Repository
Question repository (QR) is one solution which is useful to schools, Institutes, publishers and test paper setters who want to have a huge database of questions and generate test papers frequently with ease. Question repositories are organized collections of questions that are useful for teachers in many ways. It mainly deals with the gathering, sorting and administration of a large amount of questions about different levels of toughness from scientific as well as non-scientific subjects related to various classes.

3.2 Need for Question Repository
The question repository brings various advantages to user when compared to the traditional manual question paper generation method. Listed below are some of the advantages of the QR:

1. Once we have a pool of questions, teachers can use it for different purposes like generating question bank, asking questions to students while teaching, conducting an oral/ written examination

2. QR can be used by schools, Institutes, publishers and test paper setters who want to have a huge database of questions and generate test papers frequently and instantly with ease, thus saving a lot of time.

3. QR can help examiners to generate the question paper based on the learning outcomes elements. Different sets of test papers could be generated without any limitation.

4. Various algorithms/ tools/ techniques can be used to choose questions from the central question repository to prevent duplication and repetition of questions in the paper.

5. A new question can be added to the repository at any instance by authorized users

6. QR can be shared with anybody. If it is web based QR, multiple persons can append some questions or query it from multiple places

7. Moreover, not only teachers, but students also require questions for self-learning and self-assessment

8. The question paper generation can be accomplished in minutes and even in absence of operators from the question repository.

3.3 Need for Tagging
Though we have question repository with all kinds of questions added into it, having just QR is not enough. Requirement for each exam is different like for conducting oral exam you require short answer questions, for written exam you require questions of higher cognitive level. Questions may be useful in a certain context depending on parameters such as cognitive level, difficulty level, type of question, content /topic, etc. Question repository may have different types of questions but teachers will find it useful only when they are properly tagged. Hence it is very much necessary to tag questions in a QR with such a set of tags. If the repositories contain questions that are tagged with such properties, then the process of selection becomes simply querying the Question Repository with required attributes.

Most of the questions created by teachers have insufficient tags. And without adequate tagging, they are difficult to use in practical scenarios. So, it is desirable to have enough tags for all questions in repository. Many online tools provide the features to create questions with tags [3]. However question creators hardly seems to be using this feature extensively. Moreover, large number of questions in a repository has only 'basic-level' tags such as topic, subject and so on [4]. So, it becomes teacher's responsibility to verify the suitability of the question with respect to required attributes such as its Blooms level, type of question, difficulty level, the content or topic of the question, etc. for the desired assessment. This becomes an additional overhead for teachers to tag the questions in a repository. The teachers really need to be expert in the subject domain to tag the questions correctly. It may be sometimes difficult to get subject matter experts considering the workload on the teachers in the required timeframe. Hence it is necessary to have an automatic tagging system.

3.4 Tagging types

Based on the literature survey [3][5] and commonly used set of tags recommended by teachers, we are focusing on four set of tags namely, cognitive level, difficulty level, question type and content/topic. The scope of the questions is limited to the subject of Advanced Algorithms and Complexity (AAC) of post-graduation course in engineering curriculum. The various tags along with their value range is given in table 3.1.

Table 3.1: Various tags with value range

Tags	Values
Cognitive level	Six levels of Blooms taxonomy: Recall, Understand,Apply, Analyze, Evaluate, Create
Question type	Objective:, Multiple-choice, True-false, Answer-in-one-word,Match the following, Fill-in-the-blanks
	Subjective: Short answer, Program-implementation, Differentiate/ Comparison, Short-note, Long-answer
	WH-type: Why, When, What, Who, Whom, How
Content	Topics and subtopics from the syllabus that forms the node names of the ontology.
Difficulty level	Low, Medium, High

3.4.1 Cognitive Level Identification

In our system, Blooms taxonomy is the basis for identification of cognitive level of a question in our system as there are also classification schemes. Bloom's Taxonomy is a classification of learning objectives of education. It is a way of distinguishing the fundamental questions within the education system. The questions are classified on this taxonomy as it is considered to be foundational and essential elements of educational objectives (categorization shown below in Figure 1).
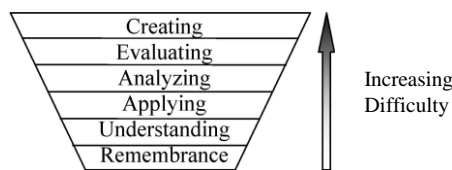

Figure 1. Bloom's Taxonomy.

Here the assumption is made that the QR contain a tagged repository. Tags can be of different types depending upon which repository is being considered.

3.4.2 Question Type identification

Various types of questions are stored in the dictionary. The classification is subjective type, objective type and WH-Type. Subjective questions are classified as short-answer, Program-implementation, differentiate/Comparison, short-note, long-answer, and essay type etc. Objective type is classified as True-False, Fill-in-the-blanks, Match-the following, Multiple-choice, True-false, Answer-in-one-word, etc. WH-Type question includes How, Why, When, What, Who and Whom.

The keyword is extracted from the question and matched with that in dictionary to decide question type.

E.g. Write an algorithm to find single source shortest path on a weighted directed graph.

The keywords extracted will be write an algorithm and refers to Program Implementation. Question is of type "Program Implementation", Question category is "Subjective" and since there is no WH-Type words, it is "Not a WH-Type".

E.g. what are the different steps while developing dynamic programming algorithms?

The keyword What will be extracted from the question and then classified as Wh-Type of question.

3.4.3 Content Type identification

The content of the question is identified by mapping the concepts from a question to the contents of the syllabus. Here the syllabus is represented using Ontology. Ontology defines a subject domain using concepts, instances, attributes, relations and axioms .The Advanced Algorithms and Complexity (AAC) subject from Semester I of Masters in Engineering is considered as an example here. If we represent syllabus ontology for the AAC domain then each node in the ontology is a concept/topic from the syllabus. The root of the ontology tree is the subject name whereas nodes at level 1 are major topic names. As the major topics have subtopics, the level 1 nodes further have subclasses. The ontology tree can be extended to have smallest subtopics as their leaf nodes.

Eg. Write an algorithm to implement Longest Common subsequence (LCS) for given set of strings.

Here the concept LCS will exactly match with the subclass names in the ontology. So these become the content tags associated with the question.

Sometimes the question may not contain the exact node names represented in ontology. In such cases every node need to be annotated with synonyms. Synonyms are the alternative names that teachers may use instead of actual node names in the ontology tree while preparing questions in the QP.

Eg. How would you apply divide and conquer algorithm to sort into non-increasing order?

In this case, the word Divide and Conquer is not directly mentioned in the topic and subtopic list of syllabus. Hence does not directly map with the node names in the ontology. Then the annotations will be searched to get a suitable match as Quick Sort.

3.4.4 Difficulty level identification

It is observed that difficulty level of a question is decided jointly by cognitive level, concept involved, concept's difficulty and question type. Higher cognitive level questions are more difficult than lower cognitive level questions. First, the cognitive levels are assigned numerical values to calculate the difficulty level. Secondly questions having multiple concepts are assumed to be more difficult than questions with less number of concepts [3]. Moreover all concepts are not of same difficulty. So every concept is assigned the difficulty level ranging from 1 to 4 (smallest to highest). Eg. Sorting, Searching, dynamic programming, greedy algorithms are considered to be in the increasing order of difficulty level. For

the question with more than one concept, the highest value of difficulty level of concepts is considered. So, the difficulty level of a question is actually the addition of the cognitive level value, number of concepts and difficulty level of concept.

3.5 Question Repository Management System

Once we have all questions available in a repository then we can add features over that QR. It may include search feature, extract feature, etc. which would help to put a query with the specifications and generate a question paper.

➢ *Search Feature*: This feature will enable the user to do keyord search or content search. For example, if user wants to search a particular question with a keyword "dynamic programming" or "Matrix chain", the system should be able to respond with proper output.

➢ *Extract Feature*: Given the specifications for kind of questions, number of questions, topics, types, etc, the system is expected to extract set of questions for required quiz or test. For example, if the requirement is to conduct a quiz of 10 marks with short answer questions, if should be fulfilled by the QRMS quickly and easily.

➢ *Auto generation of ready to print question paper*: The system may generate the softcopy of question paper automatically with all required details on front page including instructions, total marks, subject, date etc. The user interface provided is required to have the provision for supplying all this information [2].

➢ *Learning outcome based paper*: Different sets of test papers could be generated without any limitation. It may be for assignment, daily practice tests, weekly tests, test series, online tests, and term test or semester exam conducted by university.

This facility which we are providing on the top of question repository is called as managing QR

4. EXISTING LMS FEATURES AND LIMITATIONS

There are many LMS providers available of which some are free and open source. Open Source LMS include Moodle, Sakai, ATutor, whereas Blackboard is commercial.Considering the cost factor of Blackboard, we will exclude it from our study though it is being used by many educational institutions around the world. We will concentrate on Moodle and Sakai who are similar counterparts in open source LMS world.Moodle and Sakai are supported by large and active community of users and can run off the cloud.Bothsystems provide question bank feature which can be tagged. Moodle support 10 different types of questions whereas Sakai supports 5 to 10 different types of questions. In Moodle, the question can be decided at runtime based on question type. Sakai allows quiz generation only from question bank. Moodle supports all kinds of questions and their automatic evaluation but in Sakai support for automatic evaluation is limited. Moodle proves to be superior to Sakai in case of assignments in a way that it does not only allow uploading of multiple file formats but also check for plagiarism where necessary. Moodle has one best feature which helps to keep track of activity status of when and where from student logged into the system.

Moodle Assessment module provides many strategies that can be used for formative and summative assessment. However Moodle supports only one type of tagging i.e. question type. It allows creating subjective/ objective type of questions, but does not facilitate to define difficulty level or cognitive level of a question at a time. Even the content tags cannot be defined in Moodle. Similar is the case for Sakai; it also lacks all types of question tag provision at one place. Hence it is necessary to overcome these limitations and design a new system which will enable to define cognitive level, difficulty level, content type and question type for each question simultaneously and generate a question paper with equal distribution of questions with all types of tags.

5. PROPOSED SYSTEM

As we understand, in any educational institute different types of exams are conducted every now and then. Teachers have to be fully prepared to set question paper as per exam needs. Exhaustively tagged question repository is best solution in this case. Some researchers have already worked towards it[3]. The authors have suggestedextra tags such as content type, difficulty level and cognitive level which are usually not provided by QR associated with popular LMSs such as Moodle, Sakai, and Blackboard etc.

We are taking this existing tagged question bank as an input to our new system assuming that the questions are well tagged. Here we are focusing on any engineering college examination. However the system could be used for any of the educational institute for that matter.In this paper, we are proposing a distributed question paper generation system which will allow authorized users to generate a question paper using tag specifications.The architecture of the proposed system can be shown in figure 4.1
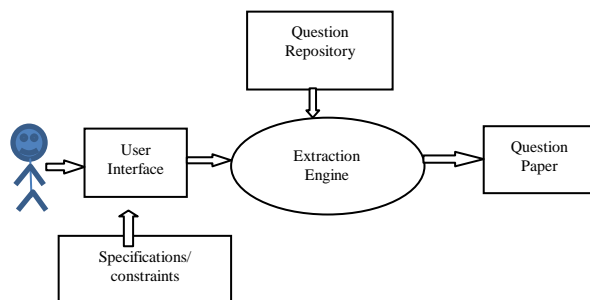

Figure 2. Architecture of proposed system

At the first level user will be asked to enter the question paper specifications such as university, course, year, semester, subject, total marks, and date of examination as shown in figure 3. The details will be validated and stored in a database to print them on question paper at a later stage.

Specifications Form

University
Course
Year
Semester
Subject
Total Marks
Date

Note

OK

Figure 3. Question Paper specification form

The next interface will ask user to select the tag specifications and specify the range in the form of percentage. User may opt for easy, medium or high level of difficulty of overall question paper. Here user will be provided the cognitive level tags; various question types in tabular format with input boxes to specify minimum/ maximum percentage. The high level syllabus topics will also be listed which again has facility to provide upper and lower bounds(Fig 4.3). Once user has entered all the specifications, the search engine will go to the database and extract the questions which fits in all criteria specified.

For example, while generating the term test question paper for "AAC" for first four chapters, we may have the requirement as follows:

• 50% chapter 1, 10% chapter 2, 20% chapter 3, 20% chapter 4.

• Objective questions-20, short-5 and long-5.

• 50% easy, 25% medium and 25% hard.

• 40% remembrance, 30% understanding, 20%application and 10% analyzing.

Now the objective is to generate a question paper with best distribution of questions among cognitive level, question type, difficulty level and topics i.e. chapters.

Specification Form

difficulty level Medium

Submit

Cognitive Level	Min %	Max %
Remember	10	30
Apply	20	40
Create		
Evaluate		
Understand		
Analyze		

Topic	Min%	Max%
#InsertionSort	20	30
#LinkedList	10	40
#InFixExpression		

Type of questions	Min %	Max %
[A] Subjective		
Essay Type		
Short Answer	10	20
Very Short	30	40
[B] Objective		
Fill in the Blanks		
Match the Following		
True-False		
Multiple Choice		

Figure 4 Tag Specification form

The extraction engine generates the question paper and displays it in ready to print format. The question paper can further be saved in .doc format for later use. The system can be implemented as a web based interface in such a way that multiple authorized users can access the system online from anywhere.

Here the user of the system could be anyone from school teacher, Institutes lecturer, publishers or test paper setter from college to university.

6. SUMMARY AND CONCLUSIONS

This paper elaborates the need for question repository and performs comparative study of various features of Learning Management Systems. After Literature survey we identified four attributes to tag assessment questions such as Cognitive level, Difficulty level, Question type and content / topic. Moodle, Sakai, Blackboard, SumTotal, Totara are few of the LMS and Question Repository Management Systems which facilitates custom tagging for the questions.

Moodle and Sakai are most commonly used LMS with few variations in features. However theysupport only one type of tagging i.e. question type. Theyallow creating subjective/ objective type of questions, but do not facilitate to define difficulty level or cognitive level of a question. Even the content tags cannot be defined in Moodle. Hence we need a system which will enable to define cognitive level, difficulty level, content type and question type for each question simultaneously. Authors [3] have introduced a system which allows all types of question tagging to QR. In this paper we are proposing a system which will use this tagged question bank as an input and generate a question paper by extracting exact questions meeting the tag specifications. Our future objective is toperform a feasibility study if we could build a system which will generate question paper with more accuracy satisfying all specification criteria. In order to do that, we would have to do rigorous and extensive unit and system testing. The future scope includes making the system more generic be useful to other domains of educational system and improve the accuracy produced by system.

7. REFERENCES

[1] Vijay KrishanPurohit', Abhijeet Kumar', AsmaJabeen', SaurabhSrivastava', R H Goudar' ,Shivanagowda, "Design of Adaptive Question Bank Development and Management System", 2nd IEEE International Conference on Parallel, Distributed and Grid Computing, 2012

[2] KapilNaik, ShreyasSule, ShrutiJadhav, Surya Pandey, " Automatic Question Paper Generation System Using Randomization Algorithm", (IJETR) ISSN: 2321-0869, Volume-2, Issue-12, December 2014

[3] Ramesh, R. ; Mishra, S. ; Sasikumar, M. ; Iyer, S., "Semi-Automatic Generation of Metadata for Items in a Question Repository", Technology for Education (T4E), 2014 IEEE Sixth International Conference, Dec 2014

[4] Saha, A.K. ; Univ. of Saskatchewan, Saskatoon, SK, Canada ; Saha, R.K. ; Schneider, K.A., " A discriminative model approach for suggesting tags automatically for Stack Overflow questions", Mining Software Repositories (MSR), 2013 10th IEEE Working Conference , May 2013

[5] Sucunuta, M.E. ; UPSI-Investig. Tecnol., UTPL-Univ. Tec. Particular de Loja, Loja, Ecuador ; Riofrio, G.E, "Architecture of a Question-Answering System for a specific repository of documents", Software Technology

_____

and Engineering (ICSTE), 2010 2nd International Conference on (Volume:2), Oct 2010

[6] Caminero, A.C. ; Dept. de Sist. deComun. y Control, Univ. Nac. de Educ. a Distancia, Madrid, Spain ; Hernandez, R. ; Ros, S. ; Robles-Gomez, A. , "Choosing the right LMS: A performance evaluation of three open-source LMS", Global Engineering Education Conference (EDUCON), 2013 IEEE, March 2013

[7] Vyshnavi M R , "A rubric to Evaluate Learning Management Systems", Teaching, Assessment and Learning for Engineering (TALE), 2013 IEEE International Conference, 2013

[8] BlackBoard: https://enus.help.blackboard.com/Learn/SaaS/Instructor/040_Activity_Stream, August 2015

[9] Pishva, D. ; Nishantha, G.G.D. ; Dang, H.A., "A survey on how Blackboard is assisting educational institutions around the world and the future trends", Advanced Communication Technology (ICACT), 2010 The 12th International Conference, 2013

_____