FREQUENT QUESTION MINING USING APRIORI ALGORITHM

[1]Deepanayaki M, [2]Deepa C
[1] Associate Professor, [2]II MCA Student,
[1 2]Department of Computer Applications,
[1 2]Priyadarshini Engineering College, Vaniyambadi, Vellore, Tamilnadu, India

***Abstract*:** Examination process is an important activity for the education system and to analyze students' performance. The nature of the exam questions would determine the quality of student talent by the institutions. With the help of the questions asked in the examination the performance of individual student will be evaluated. By Apriori Algorithm, a seminal algorithm for mining frequent item sets, this paper works on identifying the important and repeated questions from previous question papers.

IndexTerms: Data mining, Apriori Algorithm, Market Basket Analysis, Question Paper Generation.

1. INTRODUCTION

1.1 Examinations in Autonomous Institutes

Examinations predominantly use question papers as a vital constituent to discover the caliber of students. A good exam gives all students an equal opportunity to fully demonstrate their learning. The National Board of Accreditation (NBA), India states that Autonomous institutes should encourage the standard of teaching, self-analysis and responsibility in higher education. Automatic Question paper generator system considers question paper creation as its fundamental activity, and the caliber of questions is the key in enhancing examination standard, which depend on intelligent and random choice of a set of questions.

1.2 Automation

Automation means to replace the manual operations with computer procedures and other machines. Automation is aimed at increasing productivity, manufacturing prowess. It also reduces costs, labor and eliminates human error.

Manual Paper Generation	Automatic Question paper Generation
Human process	Automated process
Patterns or repetitions may occur	Totally random and unbiased process
Low Security as chances of paper leaking are high	Higher Security as chances of paper leaking are zero percent
Slow as human labor involved	Faster due to computer based automation
Less variety of different types questions	Huge variety of different types of questions

Table 1: Showing comparison of Manual process and automated process of question paper generation.

Types of Automation

Automation helps to increase productivity and reduces costs in industries. Automation plays a crucial role in manufacturing industries. Automation can be of different types: -

a) Information technology (IT)
b) Computer-aided manufacturing (CAM)
c) Numerically controlled (NC) equipment
d) Robots
e) Flexible manufacturing systems (FMS)
f) Computer integrated manufacturing (CIM)

1.3 Information Communication Technology (ICT)

ICT refers to technologies that use telecommunication to facilitate transmission of information. Internet and wireless networks can be thought of as medium for ICT. ICT has presented new communication resources over the years.

ICT in Automation

ICT are capable of sustaining powerful automation as they provide an unparalleled infrastructure. ICT plays important role in automation because of following reasons:-

a) ICT allows user to access resources from any location at any time and improves the System Performance.
b) ICT facilitates access to remote resources and allows effective utilization of the resources.
c) Communication channels are efficiently optimized.
d) ICT enhances the quality of education which proves vital.

2. CHALLENGES FOR MAKING AUTOMATIC QUESTION PAPER GENERATOR SYSTEM

2.1. Course Outcome mapping

Course outcomes are statements that describe significant and essential learning that learners have achieved, and can reliably demonstrate at the end of a course or program. mapping of questions with the Course outcomes is one of the important challenges faced while making Question paper. So Question paper validation is to be ensured to get desired standard from outcome attainment perspective as well as learning levels perspective

2.2. No repetition of questions

While making question paper many a times patterns or repetitions may occur. So to avoid this situation the challenge faced is to make use of such algorithms which keep track of questions and do not allow the repetition of questions.

2.3. Chapter wise allocation

As there are different types of exams like unit test and end semester exam, so each exam has different modules of chapters covered on which the question paper has to be generated. So challenge faced here is to make system such that it is able to allocate the question chapter wise or module wise for the respective exams.

2.4. Analysis of database

In order to effectively assess the students, the important step is to plan the question paper which covers all the vital components to test knowledge of student. One such taxonomy was presented by Bloom (1956), who described degree of learning to be classified into six unique domains, namely 1.Knowledge 2.Comprehension 3.Application 4.Analysis 5.Synthesis 6. Evaluation, the benefit of making questions based on Bloom's taxonomy.

2.5. Formatting of paper

The generated questions from the system are to be presented in proper format or proper exam template. So the challenge faced is how to format the layout of the paper in the required style and in a compatible format.

2.6. Paper pattern

It is very important to know the question paper pattern and Marks allocation Scheme. So while designing the question paper the challenge is that the system should give flexibility in choosing the desired paper pattern and allocating marks to each question in paper.

3. PROPOSED SYSTEM

In this proposed system i am going to explain about how the automatic question paper is going to generate. By the basic Algorithm for finding the frequent itemsets(questions) in Apriori Algorithm. In ten set of question paper the questions are going to select as a transaction questions. The questions are list of questions as aitemset for the five units. The Apriori Algorithm can also used to prune search during data cube computation. After selecting the question paper and questions in ten sets it should generate by algorithm for selecting the frequent questions. Number of question appear in the unit wise for generating the question paper.

4. EXISTING SYSTEM

The Apriori Algorithm for finding frequent questions:

Procedure for frequent question selection:

Step 1: In the first iteration of the algorithm, each item is a member of the set of question 1-itemsets, C1. The algorithm simply scans all of the transactions in order to count the number of occurrences of each item.
Step 2: Suppose that the minimum support count required is 2, that ismin_sup=2. The set of frequent 1-itemsets, L1, can then be determined. It consists of the question 1-itemsets satisfying minimum support. In this example, all the questions in C1 satisfy minimum support.

Step 3: To discover the set of frequent 2-itemsets, L2, the algorithm uses the join L1 ⋈L1 to generate a question set of 2-itemsets, C2. C2 consists of (|L1|/2) 2-iemsets. If no question are not removed from C2 during the prune step the each subset of the question is frequent.

Step 4: Next, the transactions in D (transactions questions) are scanned and support count of each question itemset in C2 is accumulated in the figure:

UNIT 1:

TQP	**List of question _IDs**
QP1	q11, q12, q14
QP2	q11, q13
QP3	q14, q15, q16
QP4	q11, q12, q14, q16
QP5	q12, q13, q14, q16
QP6	q13, q15, q16
QP7	q13, q14
QP8	q12, q16
QP9	q11, q12, q16
QP10	q13, q14

UNIT 2:

QP	**List of question _IDs**
QP1	q21, q23, q25, q26
QP2	q22, q24, q26
QP3	q25, q26
QP4	q24, q25, q26
QP5	q22, q23, q24
QP6	q26, q27
QP7	q21, q23, q24
QP8	q22, q23, q24, q25
QP9	q24, q26, q27
QP10	q25, q2

UNIT 3:

TQP	**List of question _IDs**
QP1	q32, q34
QP2	q31, q32, q35
QP3	q31, q32, q34
QP4	q32, q33
QP5	q31, q32
QP6	q32, q33
QP7	q31, q32
QP8	q31, q32, q33
QP9	q31, q32, q33, q35
QP10	q33, q35

UNIT 4:

TQP	**List of question _IDs**
QP1	q41, q43, q45, q46

QP2	q42, q44, q46
QP3	q42, q43, q44
QP4	q41, q43
QP5	q44, q46
QP6	q41, q43, q44
QP7	q43, q44, q45
QP8	q42, q43
QP9	q41, q42
QP10	q43, q44, q45

UNIT 5:

TQP	**List of question _IDs**
QP1	q51, q52, q53
QP2	q53, q54
QP3	q54, q55
QP4	q52, q54, q55
QP5	q51, q53, q54
QP6	q53, q55
QP7	q54, q55
QP8	q52, q53, q54
QP9	q51, q52
QP10	q52, q53, q54

Step 5: The set of frequent 2-itemsets, L2, is then determined, consisting of those questions 2-itemsets in C2 having minimum support.

Step 6: The generation of the set of question 3-itemsets, C3, is detailed in figure. From the join step, we first get C3=L2 ⋈L2. Based on the Apriori property that all subsets of a frequent itemset must also be frequent, we can determine that the four latter questions cannot possibly be frequent. We remove them from C3, by saving the effort of unnecessarily obtaining their counts during the subsequent scan of D to determine L3. When the give question k-itemset, we only need to check if its (k-1)-subsets are frequent since the Apriori algorithm uses a level–wise search strategy. The resulting pruned version of C3 is shown in figure.

UNIT 1:

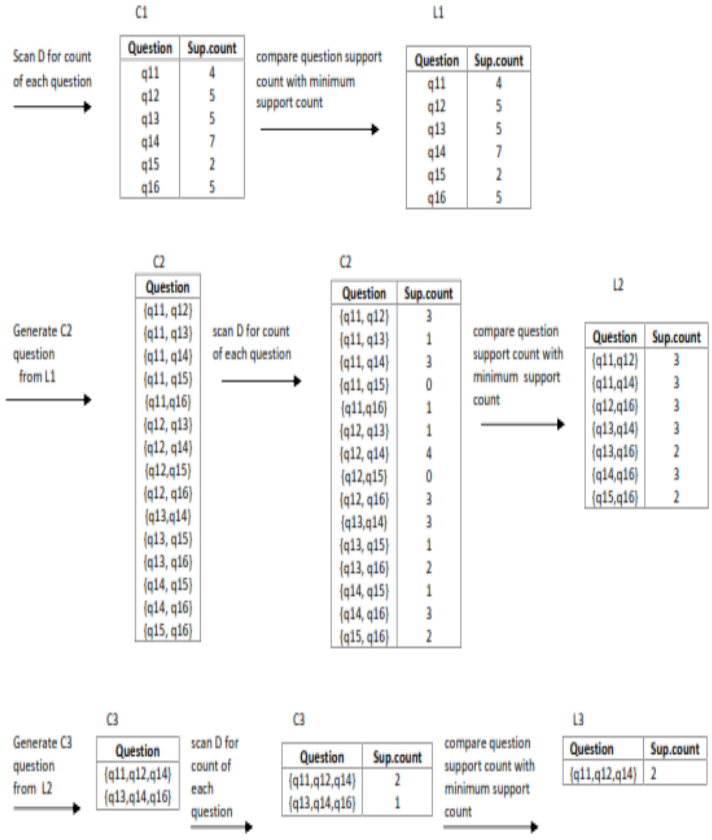

Figure 1: Generation of question itemsets and frequent itemsets, where the minimum support count is 2.

UNIT 2:

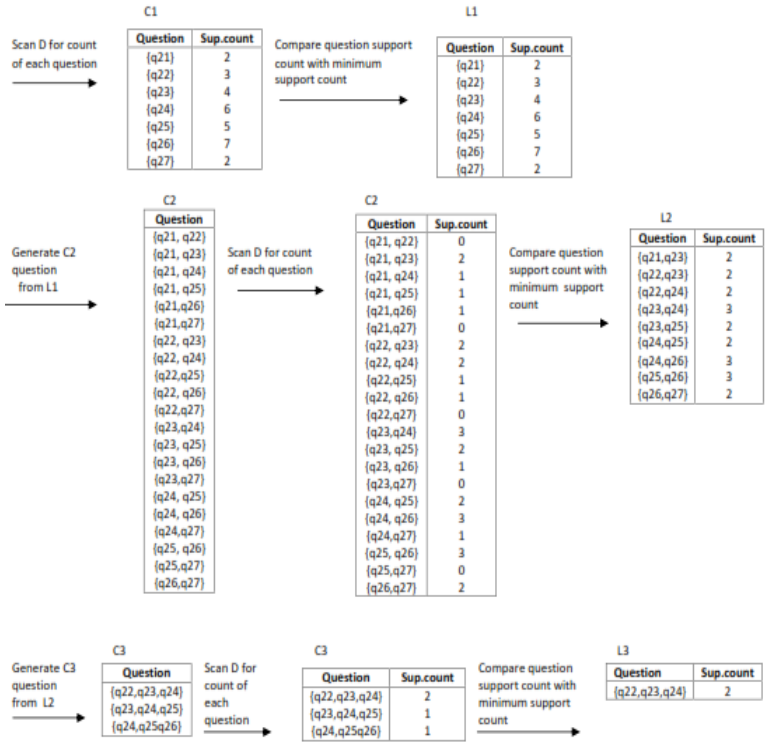

Figure 2: Generation of question itemsets and frequent itemsets, where the minimum support count is 2.

UNIT 3:

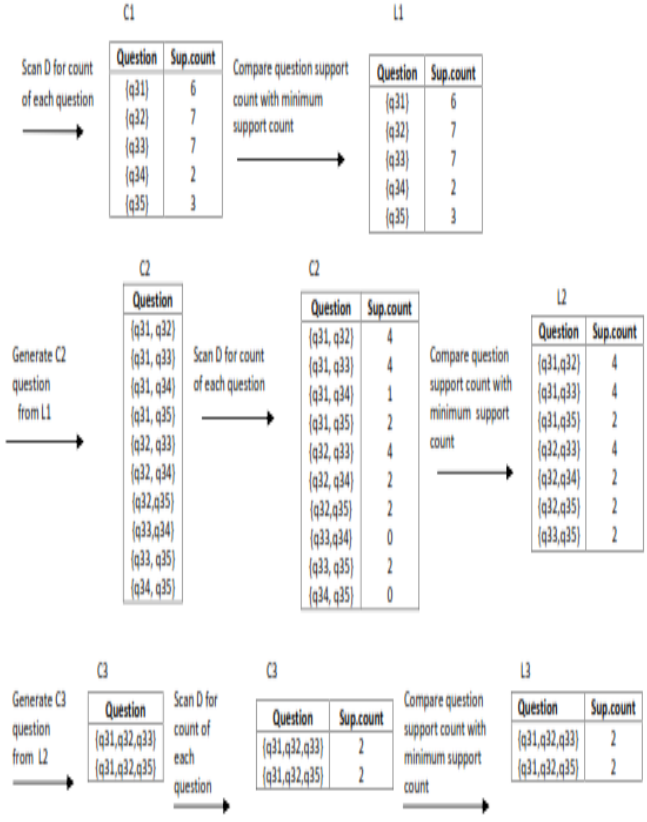

Figure 3: Generation of question itemsets and frequent itemsets, where the minimum support count is 2.

UNIT 4:

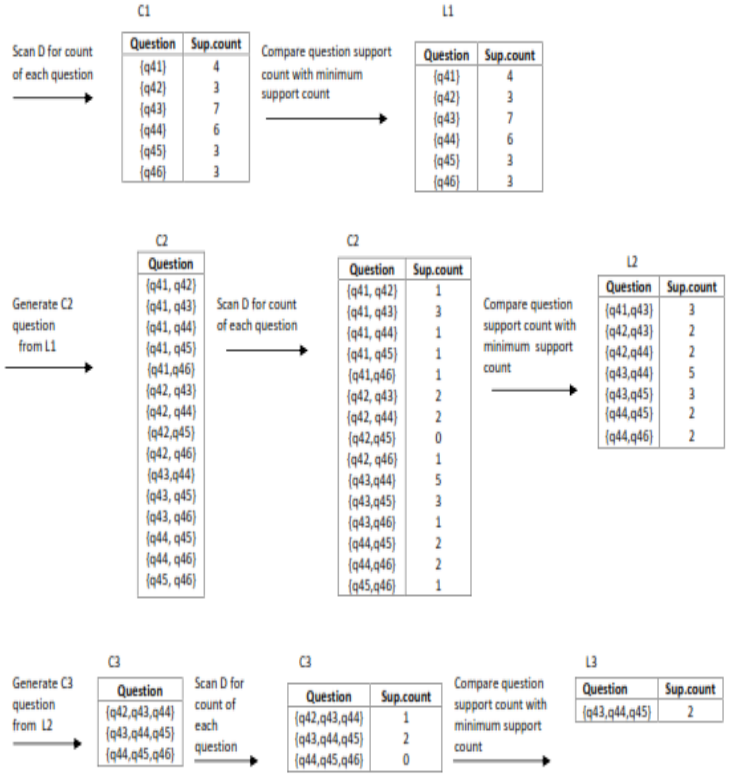

Figure 4: Generation of question itemsets and frequent itemsets, where the minimum support count is 2.

UNIT 5:

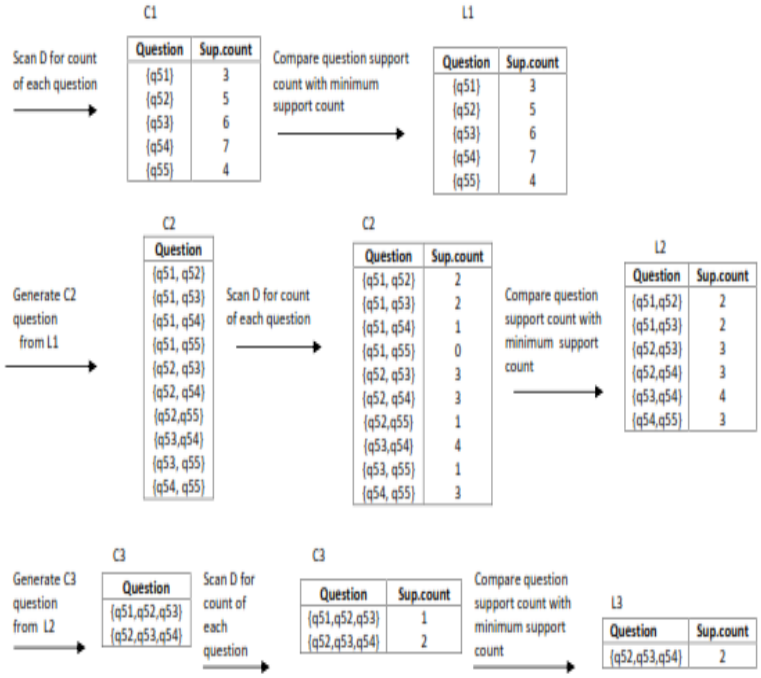

Figure 5: Generation of question itemsets and frequent itemsets, where the minimum support count is 2.

Step 7: The transactions in D are scanned in order to determine L3, consisting of those question 3-itemsets in C3 having minimum support.

Step 8: The algorithm uses L3 ⋈L3 to generate a candidate set of 4-itemsets C4. Although the join results in {{q11, q12, q14, q16}}, this itemset is pruned because its subset {{q12, q14, q16}} is not frequent. Thus, C4=ϕ, and the algorithm terminates, having found all of the frequent itemsets.

5. CONCLUSION

Thus the frequent questions are selected by the Apriori algorithm for five units from ten sets of question paper. The frequent questions from five units are:

Unit 1 {q11, q12, q14}
Unit 2 {q22, q23, q24}
Unit 3 {q31, q32, q33}
Unit 4 {q43, q44, q45}
Unit 5 {q52, q53, q54}

Based on this results, important and repeated questions are generated for all the five units and further the question paper can be set automatically with the above list.

References

[1] RasikaDhondibhauDhavale and M.Z. Shaikh, BharatiVidyapeeth College of Engineering, eSecure Question Paper Generator with Color Pass Technique and Randomization Algorithm.
[2] PritaPatil and KavitaShirsat, Vidyalankar Institute of Technology, An Integrated Automated Paperless Academic Module for Education Institutes.
[3] FenilKiranGangar, HitalGopalGori, and AshwiniDalvi, K. J. Somaiya College of Engineering, Automatic Question Paper Generator System.
[4] RasikaDhondibhauDhavale and M.Z. Shaikh, BharatiVidyapeeth College of Engineering, Automatic Test Paper Generator with Shuffling Algorithm.
[5] RohanBhirangi, and SmitaBhoir, RamraoAdik Institute of Technology, Automated Question Paper Generation System.