Advancement in Science and Technology Research
Vol. 5(1), pp. 1-8, January 2019
ISSN: 2437-1319
Full Length Research Paper

Calibrating the effectiveness of a question paper evaluation tool using automatic keyword extraction

Dimple V. Paul

Department of Computer Science, DM's College and Research Centre, Assagao, Mapusa, Bardez, Goa-403507, India.

Accepted 25 January, 2019

ABSTRACT

Well aligned curricula in every subject of a course place an emphasis not only on the curricular goals but also on the evaluation, teaching and learning activities. Performance evaluation of students in every subject of a course is done by university based examinations with subject specific question paper. Question paper gives proportionate marks to different content areas/topics and include analytical, objective, long and short answer type questions. We address the problem of evaluating question papers at undergraduate level examinations by analysing each question of a question paper on different criteria. University specified syllabus file of a subject along with Bloom's Taxonomy concept is used as a guideline in evaluating difficulty level of an examination question paper. Using "Blooming" of an examination –that is, identifying its constituent parts on Bloom's taxonomy, the difficulty level of an examination paper can be pseudo-objectively assessed. Bloom's Index assist in identifying the proportion of marks allocated to higher order cognitive skills. Text pre-processing and information extraction techniques are used to extract keywords from textual contents in the syllabus file as well as in the question paper. A tool named KEQPBS (Keyword Extraction of Question Paper Based on Syllabus) has been implemented. The tool can be used by the subject expert or question paper setter or question paper moderator or chairperson as a guideline to evaluate or revise the examination paper accordingly.

Keywords: Bloom's taxonomy, educational data, text pre-processing, pruning, keyword extraction.

E-mail: dimplevp@rediffmail.com.

INTRODUCTION

Text pre-processing and information extraction play an important role in assessment and evaluation of learner's performance in academic environments (Delavari et al., 2008). Examination, as the evaluation of teaching and learning methods, has fixed its important position in the practice of education (Liu and Wu, 2009; Garfield et al., 2011; Lemons and Lemons, 2013; Dunham et al., 2015). It is considered as one of the common methods to assess knowledge acceptance of students. Based on the examination result, student's thoughts, behaviour and skills can be studied (Jones and Harland, 2009). In the current system of evaluation at the undergraduate level examinations of Goa University, written examination is one of the major tools in the strategy of evaluation. Whether or not the written examination is able to assess student's ability is very much dependent on the questions asked in the examination question paper (Jones and Harland, 2009). The research work adopts a comparison of Question paper against University Prescribed Syllabus File as well as against Bloom's Classification system in order to verify the effectiveness of a Question Paper Evaluation Tool for theoretical courses such as Software Engineering. The layout of the paper is as follows: discussion of related work, methodology, problem statement and experimental results, and conclusion of the paper.

RELATED WORK

Current studies in educational field has resulted in generating information extraction models to improve quality of decision making process in higher education,

assess student performance, classify students' results, identify students who are likely to drop out, identify students at risk of failure, etc (Deniz and Ersan, 2001; Oladipupo and Oyelade, 2010; Ayesha et al., 2006; Kotsiantis and Pintelas, 2005; Harding et al., 2003; Namdeo et al., 2010). An interesting work in this area has been carried out by evaluating examination question paper based on Bloom's Taxonomy. Bloom's taxonomy is a classification system of educational objectives based on the level of student understanding necessary for achievement or mastery (Jones and Harland, 2009). Educational researcher Benjamin Bloom and colleagues have suggested six different cognitive stages in learning such as Knowledge, Comprehension, Application, Analysis, Synthesis and Evaluation. Details of verbs and question examples that represent intellectual activity at each level of Bloom's taxonomy can be found in Jones and Harland (2009), Matsuo and Ishizuka (2003) and Matsuo and Ishizuka (2003) discusses the mechanisms of keyword extraction by focusing on information retrieval of text mining.

METHODOLOGY

The terminology and notations used in this work are represented in Table 1.

Our work involves the use of university specified syllabus file and Bloom's taxonomy used in Jones and Harland (2009) as a guideline in evaluating the difficulty level of an examination question paper. The steps involved are as follows:

A. Extract words from subject question paper as well as syllabus file of the subject.
B. Match the above extracted question words against subject syllabus words. Record all successful matches as keywords. Based on the matched keywords, classify questions as direct and indirect. (questions that match with the syllabus are termed as direct questions and others are considered as indirect).
C. Classify each question using Bloom's taxonomy (Jones and Harland, 2009; Matsuo and Ishizuka, 2003; Anderson et al., 2014) into one among the possible set of taxonomy based cognitive levels such as Knowledge (know), Comprehension (comp), Appli-cation (appl), Analysis (anal), Synthesis (sys) and Evaluation (eval) by matching question words against pre-defined Bloom's classifier.
D. Find the set of direct questions identified at (B) and perform its unit-wise classification. Use unit-wise classified questions and calculate unit-wise marks or unit-wise marks of question paper.
E. Assign difficulty level to the question paper on the basis of the total number direct and indirect questions identified at (B), total number of questions matching with different cognitive levels of Bloom's classifier identified at (C) and the percentage of unit-wise marks allotted in the question paper respectively.

For the verification of results obtained above, we use the following steps:

A. Accept input regarding direct/indirect question type as well as Bloom's class type as per (B) and (C) for each question from the question paper setter of the question paper under study.
B. Find recall and precision values by comparing the results obtained using the tool against the values provided by the paper setter.

In general, the work focuses on information extraction approach with emphasis on text extraction and text classification. Pre-processing has been done on textual contents of Question Paper as well as on the textual contents of Syllabus File in order to remove white spaces and special characters. As a first step towards text extraction from pre-processed question content as well as from pre-processed syllabus content, we perform operations of stop word removal and stemming. Stemming is implemented using Porter Stemmer algorithm (Paice, 1994). Stemmed question content is represented as question words and stemmed syllabus content is represented as syllabus words respectively. In order to complete keyword extraction from question words as well as from syllabus words, we focus on adjacency of location of words. A sequence of words that form a keyword will be co-occurring within the stemmed question words as well as within the stemmed syllabus words. Accordingly, our approach permits extraction of the longest matching keyword of a question which successfully maps to the most appropriate keyword of the unit/topic in the syllabus file. Main modules of the tool KEQPBS are as follows:

A. Course_Syllabus_Registration: This module assists in accepting syllabus details, perform stemming and stop word removal of syllabus content, find syllabus words and generate syllabus_word_list.
B. Question_Paper_Registration: This module assists in accepting question details, perform stemming and stop word removal of question content, find question words and generate question _word_list.
C. Question_Paper_Difficulty_Calculator: This module finds difficulty level of a question paper on the basis of three different criteria such as: question-to-syllabus keyword-match, question-to-blooms keyword-match and question to unit-marks match. It is carried out using the following sub- modules:

1. Question_Syllabus_Keyword_Match: This sub-module matches syllabus_word_list and question_word_ list against each other and store the result of the match in question_syllabus_keyword file. During Question_

Table 1. Terminology and notations used.

Notation	Term	Meaning
SSO	subject_syllabus_outline	Syllabus outline of all subjects
SSUC	subject_syllabus_unit_content	Content of each unit in the syllabus for all subjects
SWL	syllabus_word_list	Unit wise list of syllabus words after removing stop words
SSW	syllabus_stem_words	Unit wise list of stemmed SWL
QPO	question_paper_outline	Outline of question papers of all subjects
QPC	question_paper_content	Individual questions of question papers of all subjects
QWL	question_word_list	Unit wise list of question words after removing stop words
QSW	question_stem_words	Unit wise list of stemmed QWL
QSKW	question_syllabus_keywords	All SSW which is matching with QSW
BMC	bloom_main_class	Outline of six levels of blooms taxonomy
BCV	bloom_class_verbs	Details of verbs belonging to each level of taxonomy
QBKW	question_blooms_keywords	All QSW which is matching with BCV
IIDB	instructor_ind_direct _blooms	Instructor's input for type of question as direct/indirect and blooms class
CIQSKW	compare_instructor_question_syllabus_keywords	Compare syllabus keywords given by instructor with the same extracted by the tool
CIQBKW	compare_instructor_question_blooms_keywords	Compare blooms class given by instructor with the same extracted by the tool

Syllabus_Keyword_Match, there is one more step of Question_Syllabus_Pruning. Pruning step verifies whether a question finds its match with more than one unit of the syllabus, and if found, the question successfully maps to the most appropriate unit to which the longest keyword matches.

2. Question_Blooms_Keyword_Match: This sub-module matches verbs in different cognitive levels of Bloom's classifier and question_word_list against each other and store the result of match in question_ blooms_keyword file. During Question_Blooms _Keyword match, there is one more step of Question_Blooms _Pruning. Pruning step verifies whether a question finds its match with more than one cognitive level of the Bloom's classifier, and if found, the question successfully maps to the lower among the levels of Bloom's class to which the question matches.

3. Question_Unit_Marks_Match: This sub-module uses Question_Syllabus_Keyword_Match file, classify matched questions into different units of syllabus file, add up unit-wise question marks and for each unit, find unit_marks allotted in the question paper. Also find out the variation of unit_marks allotted in the question paper as compared to the actual unit-marks allotted in the syllabus file.

4. Display_Results: This module initially tabulates Question_Syllabus_Keyword_Match and find out the number of direct and number of indirect questions in the question paper. Secondly it tabulates questions under each level of Bloom's class and display the cognitive level wise marks allotted in the question paper. Lastly the module calculates unit_marks allotted in the question paper and display the variation of unit_marks against the actual unit-marks allotted in the syllabus file.

D. We calculate precision and recall values for indirect questions represented by precision_indirect and recall_indirect respectively as given below:

$$\text{precision\_indirect (prec\_indir)} = \frac{\text{Number of relevant indirect questions retrieved by tool}}{\text{Total number of indirect questions retrieved by tool}}$$

$$\text{recall\_indirect (recl\_indir)} = \frac{\text{Number of relevant indirect questions retrieved by tool}}{\text{Total number of relevant indirect questions given by paper setter}}$$

Similarly, we calculate the Bloom's class precision and recall values namely precision_bloom and recall_bloom i.e. the match of each question of a question paper with Bloom's classifier as shown below:

$$\text{precsion\_bloom (prec\_blm)} = \frac{\text{Number of relevant Bloom's class questions retrieved by tool}}{\text{Total number of Bloom's class questions retrieved by tool}}$$

$$\text{recall\_boom (recl\_blm)} = \frac{\text{Number of relevant Bloom's class questions retrieved by tool}}{\text{Total number of relevant Bloom's class questions given by paper setter}}$$

Problem statement

Given the Syllabus file of a theoretical subject such as Software Engineering and the question paper of a particular examination for that subject, analyse the question paper with respect to the following criteria:

a. Number of Direct and Indirect questions
b. Number of Question under different cognitive levels of Bloom's Classification System
c. Marks assigned to different units in the question paper.

Also find the accuracy of the result obtained from the tool by finding recall and precision values with the help of corresponding user input values for the different criteria listed above.

Pseudo code of the main module
Begin
1. Call Remove_Syllabus_Stop_Words (val SSUC, var SWL)
2. Call Get_Syllabus_Stemmed_Words (val SWL, var SSW)
3. Call Remove_question_Stop_Words (val QPC, var QWL)
4. Call Get_Question_Stemmed_Words (val QWL, var QSW)
5. Call Gen_question_syllabus_keywords (val SSW, val QSW, var QSKW)
6. Call Gen_question_blooms_keywords (val QSW, val BCV,var QBKW)
7. Call compare_instructor_tool_direct (val IIDB, val QSKW,var CIQSKW)
8. Call compare_instructor_tool_blooms (val IIDB, val QBKW,var CIQBKW)
End

EXPERIMENTAL RESULTS

KEQPBS is implemented using Microsoft Visual Basic.NET as Front End Tool and MySQL as Back End Tool. Testing of the tool is carried out using question papers of Software Engineering (SE) subject in the yearly examinations of 2014, 2013, 2012 and 2011 respectively. SE course is offered in the third year of the three year bachelor's degree program of computer science (B.Sc Comp. Sc) at Goa University.

Some sample outputs of KEQPBS tool are shown:

A. Analysis of the difficulty level using Direct/Indirect Question Tabulator

Figure 1 shows the experimental results obtained after classifying the questions of SE question paper of 2014 under Direct/Indirect Question Type.

Table 2 named as DIRECT_INDIRECT_DIFFICULTY_CLASSIFIER shows the difficulty level, diff_lvl of SE question papers calculated on the basis of the total marks allotted for indirect questions (mks_of_indirect) in the question paper.

Figure 2 displays the graphical representation of output of Direct-Indirect-Difficulty-Classifier generated from the values of Table 2.

Analysis of the result of Direct/Indirect Question Tabulator:

From the analysis of question papers of Software engineering (SE) of 2014, it is identified that SE question papers are moderately difficult as it includes 20 to 25 percentage of indirect questions and 75 to 80 percentage of direct questions.

B. Analysis of difficulty level using Blooms classifier:

Figure 3 represents the experimental results obtained after classifying questions of SE question papers of 2014, 2013, 2012 and 2011 respectively under different cognitive levels of Bloom's classifier.

Table 3 named as BLOOMS_DIFFICULTY_CLASSIFIER displays the number of questions extracted by the tool under different cognitive levels of Bloom's classifier.

Figure 4 displays the graphical representation of output of Bloom's Difficulty Classifier of Table 3.

Analysis of the result of Bloom's classifier

Results displayed by the tool brings out the

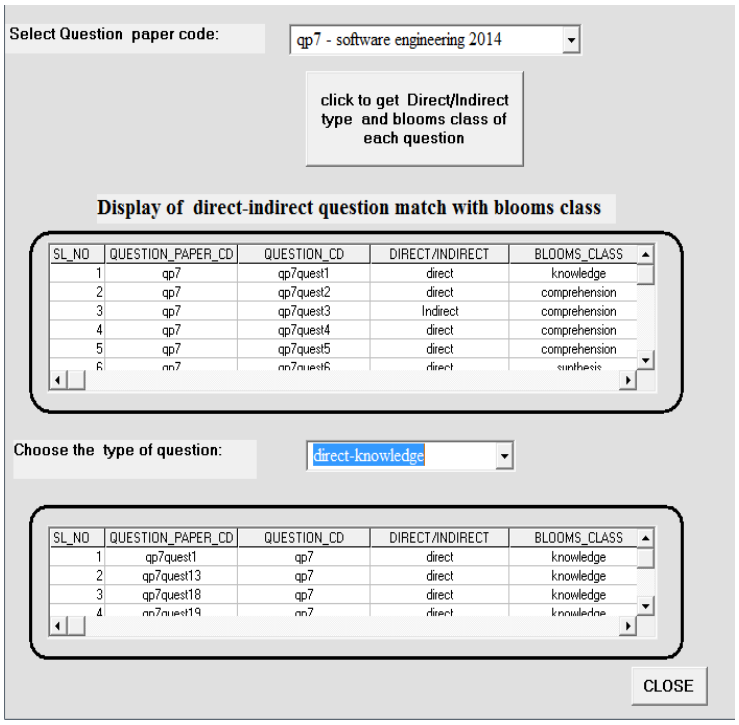

Figure 1. Direct/Indirect Question Tabulator with a display of questions of SE question papers classified under direct or indirect question type.

Table 2. Direct-indirect-difficulty-classifier.

question_paper_code	tot_quest_pap_mks	mks_of_direct	mks_of_indirect	diff_lvl
SE 2014	100	79	21	21%
SE 2013	100	76	24	24%
SE 2012	100	80	20	20%
SE 2011	100	75	25	25%

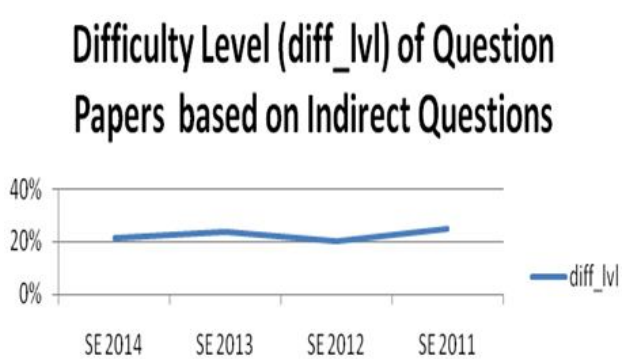

Figure 2. Direct-Indirect difficulty _classifier based on the values listed in Table 3.

generalization about SE question papers of 2014, 2013, 2012 and 2011 examinations that they have allotted 80 to 100% marks in the question paper for lower cognitive levels of Bloom's taxonomy and have only the rest of the marks in the question paper for higher levels of the taxonomy.

C. Analysis of unit-wise marks allotted in the question paper

Table 4 named as UNITWISE-MARKS-_TABULATOR displays the comparison of unit wise marks allotted in the Syllabus File as well as the corresponding marks specified in the Question paper.

Figures 5 and 6 display the graphical representation of output of Unitwise-Marks-Tabulator generated from the values of Table 4.

Analysis of the result of Unitwise-Marks-Tabulator

From the analysis of question papers of SE 2014 and SE

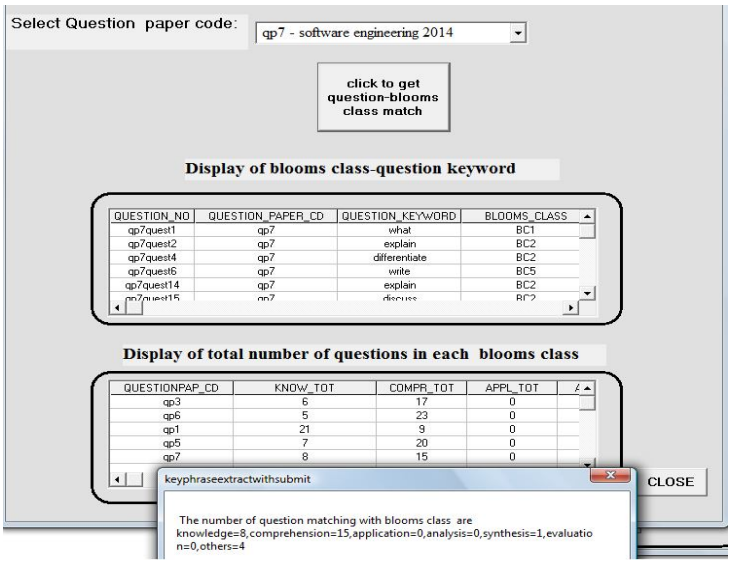

Figure 3. Bloom's Difficulty Classifier with a display of questions of SE question papers classified under different cognitive levels.

Table 3. Blooms-difficulty-classifier.

question_paper_code	no_of_quest	know_tot	compr_tot	appl_tot	anal_tot	synt_tot	eval_tot	other_tot
SE 2014	28	8	15	0	0	1	0	4
SE 2013	28	6	17	0	1	3	0	1
SE 2012	28	5	23	0	0	0	0	0
SE 2011	28	7	20	0	0	0	0	1

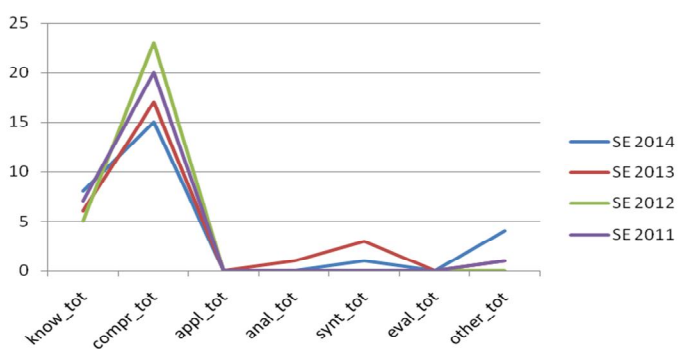

Figure 4. Bloom's Difficulty Classifier based on the values listed in Table 2.

2013, it is found that both SE 2014 and SE 2013 question papers are not exactly matching with the university prescribed syllabus marks of units. At the same time, there is not too much variation except in few cases.

D. Analysis of precision and recall values

Table 5 named as PRECISION_AND_RECALL_VALUES displays the precision and recall values calculated separately for indirect question ,blooms class match, Combination of indirect questions and blooms class match for SE question papers of the year 2014, 2013, 2012 and 2011, respectively.

Analysis of the results of precision and recall

From the analysis of question papers of SE, it is observed that precision on the basis of Bloom's Classifier is more than the recall on the basis of Bloom's Classifier. On the other hand, precision on the basis of Direct-Indirect Question Tabulator is less than recall on the basis of the same. The tabulated precision and recall measures indicate that KEQPBS successfully computes the difficulty level of an examination question paper. Keyword Extraction approach outperforms the approach known as linking the learning outcomes with assessment (Jones and Harland, 2009).

CONCLUSION AND PERSPECTIVES

KEQPBS is efficient in extracting the keywords from

Table 4. Unit wise-marks-tabulator.

question_paper_code	subject_unit_code	syllabus_marks	question_paper_marks	Variation
SE2014	sub1unit1	20	18	2
SE2014	sub1unit2	10	2	8
SE2014	sub1unit3	20	25	-5
SE2014	sub1unit4	5	4	1
SE2014	sub1unit5	5	2	3
SE2014	sub1unit6	5	0	5
SE2014	sub1unit7	10	15	-5
SE2014	sub1unit8	5	5	0
SE2014	sub1unit9	5	5	0
SE2014	sub1unit10	5	10	-5
SE2014	sub1unit11	10	14	-4
SE2013	sub4unit1	20	21	-1
SE2013	sub4unit2	10	16	-6
SE2013	sub4unit3	20	23	-3
SE2013	sub4unit4	5	5	0
SE2013	sub4unit5	5	5	0
SE2013	sub4unit6	5	5	0
SE2013	sub4unit7	10	5	5
SE2013	sub1unit8	5	5	0
SE2013	sub1unit9	5	5	0
SE2013	sub1unit10	5	6	-1
SE2013	sub1unit11	10	4	6

Figure 5. Unit-wise marks Tabulator displaying comparison of marks of SE 2014 question paper with syllabus marks listed in Table 4.

syllabus file as well as from question paper. Also KEQPBS is able to find the difficulty level based on the keyword match of question paper against e Bloom's taxonomy as well as question paper against Syllabus File. KEQPBS can assist paper setters of different academic institution to verify whether the minimum standards are being maintained by teachers while framing questions papers of theoretical papers such as Software Engineering (SE). It can act as a supporting module for university based examination system to select the best question paper among two or more question papers submitted by different question paper setters. It can also act as a guideline in making conclusions on whether sufficient marks is allotted by paper setters for each unit

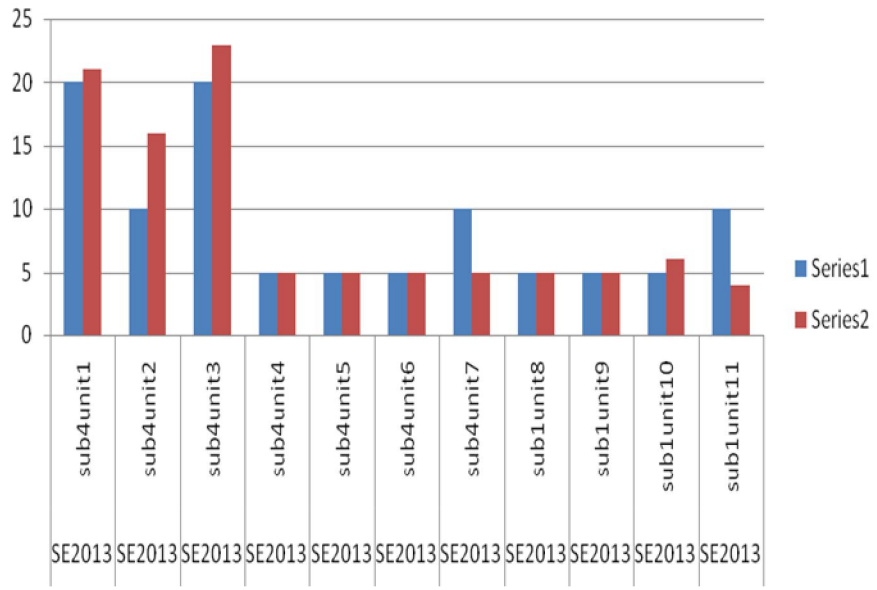

Figure 6. Unit-wise marks Tabulator displaying comparison of marks of SE 2013 question paper with syllabus marks listed in Table 4.

Table 5. Precision_and_recall_values.

question_paper_cd	prec_blm	recl_blm	prec_indir	recl_indir	prec_comb	recl_comb
SE 2014	0.73	0.69	0.50	0.57	0.65	0.63
SE 2013	0.80	0.75	0.60	0.71	0.70	0.73
SE 2012	0.86	0.81	0.57	0.68	0.71	0.67
SE 2011	0.73	0.67	0.50	0.62	0.61	0.65

in the question paper as prescribed in the syllabus file. There is also a provision to accept synonyms of each word while accepting the syllabus file. We are in the process of finding the difficulty level of each question paper by also considering the marks obtained by students in that examination as input. The results will be represented soon in a subsequent paper.

REFERENCES

Anderson L. W., Krathwohl D. R., Airasian P. W., Cruikshank K. A., Mayer R. E., Pintrich P. R., Raths J., Wittrock M. C. (**2014**). A taxonomy for learning, teaching and assessing: A revision of Bloom's. Pearson Education.

Ayesha S., Mustafa T., Sattar A. R., Khan M. I. (**2006**). Data mining model for a better higher educational system. Informat Technol J, 5(3): 560-564.

Delavari N., Phon-Amnuaisuk S., Beikzadeh M. R. (**2008**). Data mining application in higher learning institutions. Informat Educat, 7(1): 31-54.

Deniz D. Z., **Ersan** I. (**2001**). Using an academic DSS for student, course and program assessment. International Conference on Engineering Education, August.

Dunham B., Yapa G., Yu E. (**2015**). Calibrating the difficulty of an assessment tool: The Blooming of a statistics examination. Journal of Statistics Education, 23(3).

Garfield J., Zieffler A., Kaplan D., Cobb G. W., Chance B.L., Holcomb J. P. (**2011**). Rethinking assessment of student learning in statistics courses. The American Statistician, 65(1): 1-10.

Harding T. S., Passow H. J., Carpenter D. D., Finelli C. J. (**2003**). An Examination of the relationship between academic dishonesty and professional behaviour. 33rd ASEE/IEEE Frontiers in education Conference ,November.

Jones K.O., **Harland** J. (**2009**). Relationship between Examination Questions and Bloom's Taxonomy. 39th IEEE Frontiers in Education Conference, December.

Kotsiantis S.B., **Pintelas** P.E. (**2005**). Predicting students' marks in Hellenic Open University. Fifth IEEE International Conference on Advanced Learning Technologies, ICALT.

Lemons P. P., **Lemons** J. D. (**2013**). Questions for assessing higher-order cognitive skills: It's not just Bloom's. CBE Life Sci Educ, 12: 47-58.

Liu J., **Wu** D. (**2009**). The Use of Data Mining in Contents of the Examination and the Relevance of Teaching Research. First international workshop on Education Technology and Computer Science.

Matsuo Y., **Ishizuka** M. (**2003**). Keyword extraction from a single document using word co-occurrence statistical information. Int J Artificial Intellig Tools, 13(1).

Namdeo V., Singh A., Singh D., Jain R.C. (**2010**). Result analysis using classification techniques. Int J Comput Applicat (0975 - 8887), 1(22).

Oladipupo O.O., **Oyelade** O.J. (**2010**). Knowledge discovery from students' result repository: Association rule mining approach. Int J Comput Sci Security, 4(2): 199-207.

Paice C.D. (**1994**). An evaluation method for stemming algorithms, proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 42-50.