Tool for Automated Evaluation of Descriptive Answers & Course Outcomes (TADACO)

Dharma Reddy Tetali
Professor, MLR Institute of Technology, Hyderabad, India.

Sowmya Tamma
Research Scholar, MLR Institute of Technology, Hyderabad, India.

Lakshmi Ramana
Research Scholar, MLR Institute of Technology, Hyderabad, India.

Abstract – **Manual evaluation of descriptive answers is an error-prone, tedious and time taking activity. TadaCo is a tool developed for automated evaluation of descriptive answers and assessment of corresponding course outcomes. TadaCo uses a smart and systematic technique to evaluate the answers. This tool is developed using PyQt, Python and its modules: pyuic, xlsxwriter, Platypus and Reportlab. This paper elaborates TadaCo tool, its functionalities and implementation details along with an example evaluation of a subject entitled 'Software Testing Methodologies'. TadaCo works either in a Semi-automated mode or in Complete automated mode. Semi-automated mode has the flexibility to allow the faculty to reevaluate an answer and update the results. It is established that the Semi-automated mode yields comparatively better results than the Complete automated mode.**

Index Terms – **Computer Assisted Assessment (CAA), Multiple Choice Question (MCQ), PyQt, pyuic, xlsxwriter, Platypus and Reportlab.**

1. INTRODUCTION

In the current evaluation system, descriptive answers given by the students are evaluated manually by the faculty. This is an error-prone process, as different professors are likely to award different marks to the same answer. As per reference[1], during last semester Four Hundred and twenty thousands of students in Anna university applied for revaluation of their answer scripts, as they believe that the marks obtained during the first evaluation are incorrect. Nineteen percent of the failed students passed after the revaluation. This clearly indicates that the marks obtained during revaluation differ from the initial evaluation. The same answer scripts are awarded with different marks in the revaluation. This example indicates one of the problems with manual evaluation of answer scripts. Marks awarded during manual evaluation depend not only on the content of the answer script, but also on other factors like hand writing, style of answering and even on the perception of the evaluator.

Manual evaluation is a time taking process, and any delay in this process would ultimately lead to the delay in the announcement of results and discomfort to the students. For instance, Jawaharlal Nehru Technological University conducted the first semester end examinations for the final year engineering students, during 3rd week of November, 2016 and the corresponding results were announced on 9th Feb 2017[2]. There is time gap of 80 days between the conduction of examination and declaration of results.

The current evaluation system is tedious as faculty has to put lot of manual effort to read through and evaluate the scripts of all students. Faculty has to put considerable effort to deal with the following concerns while evaluating the scripts [3]

- Concerns about equity and fairness
- Concerns about comparability of the evaluations
- Concerns about what to weigh in making judgments

The time needed to correct any given answer script depends on its content. On an average, if a teacher spends twenty-five minutes to correct an answer script, then it would take ten hours forty-one minutes to correct 25 scripts [4].

At present, assessment of course outcomes is carried out after the evaluation. The assessment process involves data entry of the marks of all answers for every student in a class. This process again takes a considerable amount of time.

Attempts to use computers in educational assessment started in the early 1990s [5]. Bunderson et al. explored the ways in which hardware and software technologies can be used for effective educational assessment. Their recommendations in this regard can be summarized as follows:

•Enhance the frequency and variety of help services to the disciple through online as-assessments

•Increase the frequency of formative evaluation, and provide incentives to use the evaluation data for ongoing improvement of educational programs

•Make use of alternate methods of assessment that require human judgment and that measure more complex, integrated, and strategic objectives

•Foster new item types and uses of portable answer media in order to utilize the existing testing infrastructure more effectively

•Emphasize the development of localized infrastructure of

Integrated Learning and Assessment Systems, and the coordinated evolution of central sites for development of system, tests, and research

•Stimulate the professional development of faculty and other professionals who are knowledgeable and skilled about both the human judgment and the technical aspects of automated assessment

G. Frosini et al described a tool to build software systems [6]. Their tool is used to replace the role of an examiner during typical exams, and to reduce effort as well as anxiety. The tool uses computerized adaptive testing to increase assessment efficiency. According to their model, the database consists of queries submitted by an author. A Design module handles insertion of the queries, analysis module analyzes the results gathered during the exams and calculates particular indexes to determine the level of difficulty on the submitted questions.

A question paper generally, comprises of objective answers as well as descriptive answers to evaluate the skills of a disciple in any course. Objective questions require a user to choose or provide a response to a question whose correct answer is predetermined. Such a question might require a disciple to:

Select a solution from a set of choices (MCQ, true/false)

Supply brief numeric or text responses (text input)

An MCQ is composed of two parts: the stem and the answer options or usually referred to as choices. The stem is the main question, and the options include both the correct answer as one option and one or several distractors which are incorrect. Following is an example

Stem:

What country is Chennai from?

Answer:

India

Distractors:

China, Mongolia, Sri Lanka

Selecting distractors is a difficult task when creating an MCQ: the quality of an MCQ relies heavily on the quality of these options [7]

Multiple Choice Questions (MCQs) and questions with true/false answer are easy to evaluate. Answers to such questions can be automatically evaluated by using Computer Assisted As-sessment (CAA) systems [8].

More recently, Jannat et al developed an Intelligent Classroom System for the Analysis of Students Conceptual Understanding [9], but that system is not dealing with the assessment of course outcomes. Also, this system is deviating a lot from our current educational structure, and hence it is quite difficult to adopt it. TadaCo not only deals with the evaluation of the question papers, but it also assesses the course outcomes automatically after the evaluation. Another advantage with TadaCo is that it can be easily implemented in the current educational system.

TadaCo evaluates the answer scripts based on keywords and phrases, and has the flexibility to update the result of an evaluation. The tool automatically generates the evaluation and assessment reports in either .PDF format or .XLS format as chosen by the faculty.

2. RESEARCH METHODOLOGY

TadaCo is applied during the First Internal examination conducted at MLR Institute of Technology [10].The tool is applied to evaluate the answer scripts for the subject 'Software Testing Methodologies'. The answer scripts of three batches of students are taken for evaluation. Each batch comprises of 40 students. The answer scripts of first batch of students are evaluated manually. Second batch scripts are evaluated in 'semi automated mode' of TadaCo, and 'Complete automated mode' is used to assess the scripts of the third batch. A feedback survey is conducted after the evaluation, and the results of this survey are depicted in the 'RESULTS ANALYSIS' section.

3. ROLE OF TADACO TOOL IN EVALUATION OF DESCRIPTIVE ANSWERS

TadaCo evaluates the descriptive answers by matching keywords and Phrases in the answer given by the disciple, with the keywords and phrases of the original answer. The keywords and Phrases of the original answers are stored in the answer base of the system. Answer base contains the entities for keywords and phrases, along with the no. of marks to be awarded for each of their occurrences. The structure of the answer base is discussed in detail, in the later section entitled 'ANSWER BASE OF TadaCo'.

TadaCo accepts disciple answer scripts as text files. The name of the text file, containing an answer script is provided as an input to the system. The system provides an appropriate error message, if that file doesn't exists. Otherwise, it accepts the file and initiates it's processing. The system gets the keywords and the corresponding marks to be awarded, from the answer base. It then checks for the existence of those keywords in the answer file. If a keyword exists in the given answer, then the marks

corresponding to that keyword are added to the marks to be given to that answer. This process is repeated for the phrases as well. The total marks to be awarded are calculated by summing up the marks for keywords and phrases. The system has the flexibility to update these total marks. The system generates a report, in the format (either in .pdf, or in .xls format) chosen by the faculty. This report contains the details of marks awarded to each keyword and to each phrase of the given answer. Faculty can update the marks, if needed, after going through the report.

3.1.ROLE OF TadaCo TOOL IN ASSESSMENT OF COURSE OUTCOMES

The course outcome corresponding to each question is stored in the database. TadaCo assesses the course outcomes after evaluating an answer script. It stores the marks awarded against each question in the database. Using this information, the system generates a report on the attainment of course outcomes.

4. ANSWER BASE OF TadaCo

The answer base of TadaCo consists of three entities: Keyword, Phrase & Question. The Keyword entity is used to store the marks to be awarded to each of the keyword, in the answer of a given query. Similarly, the Phrase entity stores the marks to be given to each substantial phrase in the answer. The Question entity consists of the question ID along with the description of the question and the maximum marks that can be awarded to that question. The following diagram depicts the relationship between the entities of TadaCo.
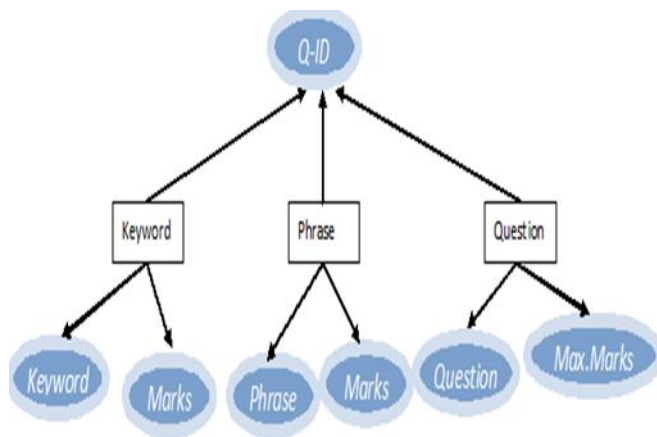

Figure1: Entity Relationship Diagram of TadaCo

5. USER INTERFACES OF TadaCo

Following are user interfaces of the TadaCo System.

☺ TADACO.UI   ☺ QUESTION.UI

☺ PHRASE.UI   ☺ KEYWORD.UI

☺ EVALUATE.UI   ☺ GENREP.UI

☺ ANSWER.UI

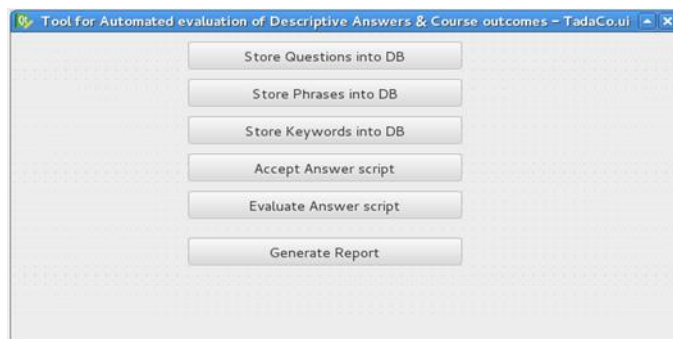TADACO.UI IS THE PRIMARY USER INTERFACE, WHICH IS SHOWN IN THE FOLLOWING FIGURE.

FIGURE2: PRIMARY USER INTERFACE OF TADACO

The functions of the push buttons in the primary interface are as follows:

Store Questions into DB:

This push button instantiates the user interface question.ui, which is used to store the Question ID, Question Description and the maximum marks that are allotted to the question, into the data base.

Store Phrases into DB:

This push button instantiates the user interface phrase.ui, which is used to store the Phrase ID, Phrase Description and the maximum marks that can be awarded to that phrase, into the data base.

Store Keywords into DB:

This push button is almost similar to the above mentioned 'Store Phrases into DB' push button, except that this button is used to store the Keyword details into the data base, instead of phrases. This pushbutton instantiates the user interface keyword.ui

Accept Answer Script:

This push button instantiates the user interface answer.ui, which in turn accepts the name of the answer script file, and verifies if that file is existing or not.

Evaluate Answer Script:

This push button instantiates the user interface evaluate.ui, which in turn accepts the ID of the question to be evaluated.

Generate Report:

This push button instantiates the user interface genrep.ui, which is used to generate the report either in .PDF format, or in .XLS format, as chosen by the user.

Pyuic tool is used to automatically generate the python code of all the above mentioned user interfaces. This automatically generated code is imported into the main programs. These main programs are briefly described in the following section.

6. MODULES/MAIN PROGRAMS OF TadaCo

Following are the main programs of TadaCo.

Question_main.py:

This Python program accepts the Question ID, Question Description and the total marks that can be awarded to the question, and using the line edits of the user interface it inserts those values into the Question entity. Likewise, Phrase_main.py & Keyword_main.py are used to deal with the details of phrases and keywords.

Answer_main.py:

This program uses isfile() function in os.path module [11], to verify whether the answer script file exists or not.

Eval_main.py:

This program obtains the details of the keywords and phrases, of the answer to be evaluated, from the answer base. It then verifies for the existence of the required keywords and phrases in the answer script to be evaluated. If they exist, then it adds the corresponding marks to the 'marks to be awarded'.

Genrep_main.py:

This program is used to generate the details of the awarded marks in the form of either a .PDF report or as a .XLS report. This program imports Python's xlsxwriter [12] module to generate the report in .XLS format. Also, this program imports Python's Reportlab [13] module and Platypus [14] module to generate the report in .PDF format. 'reportlab.platypus.SimpleDocTemplate.multiBuild' is used to generate the report body. 'xlsxwriter.Workbook' is used to create the workbook and 'workbook.add_worksheet' is used to add the worksheets to the workbook.

7. RESULTS ANALYSIS

Following, is part of the .pdf report generated after evaluating the answer for a question with question id 001.By default, TadaCo has assigned 0.5 marks to the phrases and 0.25 marks to the keyword. TadaCo allows these default values to be changed and each keyword or phrase can have separate marks.

TadaCo is tested on a sample of 120 students in MLR Institute of Technology. The students were divided into three batches, each batch comprising of 40 students. The answer scripts of these students for the subject 'Software Testing Methodologies', were evaluated in the following three modes of evaluation:

(i)   Complete Manual evaluation

(ii) Semi-Automated evaluation using TadaCo
(iii) Completely automated evaluation using TadaCo

A feedback survey is conducted after the above three evaluations, and the following graphs depicts the results of this survey.
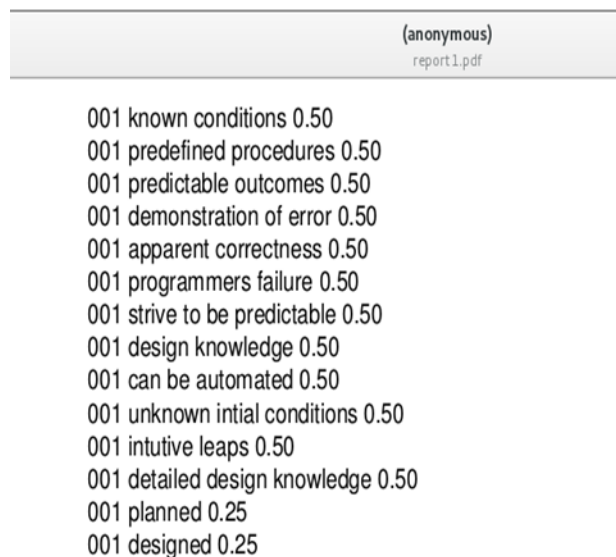

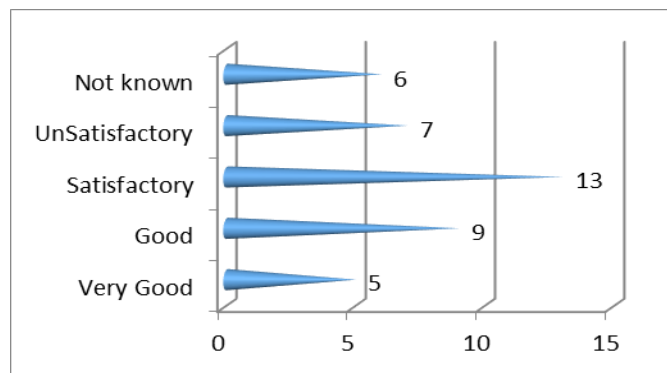Figure3: Part of the .pdf report generated by TadaCo

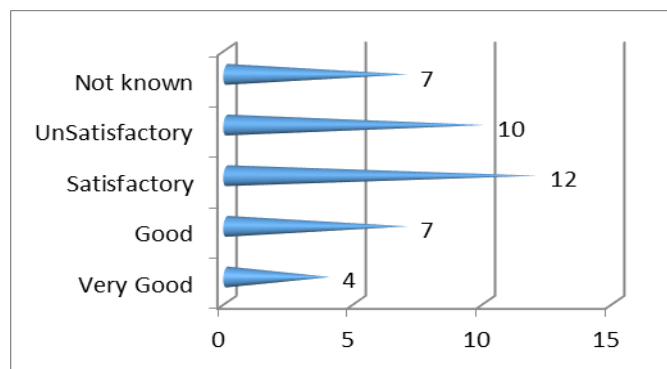Figure 4: Survey results of Complete Manual evaluation

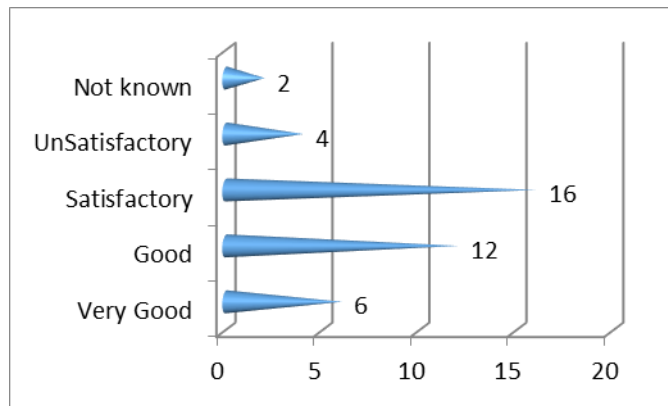Figure 5: Survey results of completely automated evaluation

Figure 6: Survey results of Semi-automated evaluation

8. CONCLUSION

Analysis of the results clearly indicates that the semi-automated evaluation method clearly dominated the remaining two evaluation methods. This is obvious since the semi-automated evaluation involves a high level manual evaluation after the automated evaluation. Even though semi-automated evaluation is not as fast as the completely automated evaluation, it can still be adopted since it is efficient than the complete manual evaluation.

REFERENCES

[1]   http://timesofindia.indiatimes.com/city/chennai/Engineering-grades-rise-after-revaluation-of-papers/articleshow/19718892.cms
[2]   http://jntuh.ac.in//bulletin_board/II_III_IV_Year_BTech_BPharm_2016_17.pdf
[3]   https://www.theatlantic.com/national/archive/2013/01/why-teachers-secretly-hate-grading-papers/266931/
[4]   http://indianexpress.com/article/cities/mumbai/evaluation-season-school-teachers-sweat-it-out-for-a-pittance/
[5]   Computers in Educational Assessment: An Opportunity to Restructure Education practice by C.V. Bunderson, J.B. Olsen & A. Greenberg Published by Institute for Computer Uses in Education, 1990
[6]   Performing automatic exams by G. Frosini, B. Lazzerini & F. Marcelloni, Computers & Education, vol. 31, pp. 282, 1998
[7]   Rodriguez, M.C.: Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. Educational Measurement: Issues and Practice, 24 (2), 3{13 (2005)
[8]   http://caacentre.lboro.ac.uk/
[9]   Intelligent Classroom System for Qualitative Analysis of Students' Conceptual Understanding by Jannat Talwar, Shree Ranjani & Anwaya Aras Published by IEEE,6th International Conference on Emerging Trends in Engineering and Technology,2013.
[10] http://www.mlrinstitutions.ac.in/
[11] https://docs.python.org/2/library/os.path.html
[12] https://pypi.python.org/pypi/XlsxWriter
[13] https://pypi.python.org/pypi/reportlab
[14] https://pypi.python.org/pypi/Platypus