A Knowledge Based System for Automated Assessment of Short Structured Questions

Teshna Luchoomun Muhammad Chumroo Vidasha Ramnarain-Seetohul

Department of Information and Communication Technologies, University of Mauritius
Reduit, Mauritius

teshna.luchoomun@umail.uom.ac.mu ; muhammad.chumroo@umail.uom.ac.mu ; v.seetohul@uom.ac.mu

***Abstract*— Assessment is a key tool of the educational system to gauge the level of understanding of students. Assessment is traditionally carried out through various mediums ranging from written evaluations to oral evaluations or even through practical sessions in some cases. Over the years, it has been noticed that the level of education is escalating and that academic competition is growing fiercer. Thus, educational institutes are faced with the need to upgrade their process of assessment. Although assessment is an essential part of the educational process, the correction of structured questions represents a substantial portion of an instructor's worktime. Besides, with the increase in cohort size throughout institutions and the popularity of Massive Open Online Courses (MOOCs), this problem seems to worsen. Teachers are therefore faced with a conundrum; they can either devote their time to the correction of scripts or to teaching activities. According to research, computerized marking give rise to faster correction and it saves staff time in terms of supervising and marking. In this context, a knowledge based automated assessment system has been developed to help correction and grading of short structured questions online. The system will compare the meaning of two sentences by taking into account both semantic and syntactic content and the corresponding grade will be returned.**

Keywords*— *Automated, Assessment, Education, Tutors

I. INTRODUCTION

One of the key focuses of the current educational system is assessment. Indeed, it empowers the faculty to evaluate the performance of students and hence deduce valuable insights for the faculty in terms of pedagogical approach. Traditionally, this evaluation is carried out through various mediums ranging from written evaluations (article, summaries and reports, notes, charts and graphs, maps, projects, structured questions, assignments among others) to oral evaluations (presentation & class participation) or even through practical sessions in some cases. Given the paramount importance of continuous assessment to prepare students for final examinations, they are normally provided with tutorials for them to self-assess their performances along with marked tests. Over the years, it has been noticed that the level of education is escalating and that academic competition is growing fiercer: be it at higher level (University level) or at lower level of schooling [1]. Thus, educational institutes are faced with the need to upgrade their process of assessment. Furthermore, self-assessment of students is becoming more difficult to implement as summative assessments (graded tests) are particularly time-consuming. Beyond being time consuming for tutors, studies [2] have highlighted that the corrections of scripts is inherently subjective based on the perception of the tutor. Moreover, humans are prone to errors due to fatigue or automatism developed due the repetitive action of correction. These slightly iniquitous and random corrections lead to different grades assigned by various tutors for the same quality of work which entail a feeling of unfairness among students. Fortunately with the advent of computers, novel methods have been investigated. Indeed, a large part of research on assessment has been allocated in developing intelligent automated system that would facilitate the task of assessing students [3] since a few decades. Various studies [4] [5] [6] have given rise to many automated systems that helped tutors to automatically correct questions; some of those systems are quizzes including multiple choice questions, fill-in the blanks, yes/no responses, label matching, ordering options and entering text, calculations/randomly generated answers and essays correctors. Overall, these have greatly eased the tasks of tutors. According to [7], automated assessment systems give rise to faster correction; hence it saves time in terms of marking. In this paper, a knowledge based system has been implemented for correction and grading of short structured questions online. Paper is organized as follows: Section II relates to the work already done in the automated assessment field, Section III gives an overview of the proposed system, section IV is based on implementation and testing and finally section V gives the conclusion.

II. RELATED WORK

According to [8], automated text assessment can be classified into three groups namely assessment of content only, style only or both style and content. The Intelligent Essay Assessor (IEA) is an automated assessment system which focuses on content only. It is based on Latent Semantic Analysis which was initially used for document indexing and text retrieval. It uses a two dimensional matrix. However LSA does not check for word order and a lot of computations are required with the matrix. Project Essay Grade (PEG) is a system which is based on only text-style and is used to automatically grade essay. PEG grade an essay based on writing quality rather than content based. Paperless School free-text Marking Engine (PS-ME) is an automated assessment system based on a mixture of content and style assessment. It uses Natural Language processing (NLP) techniques to evaluate student essay so as to report

978-1-5386-9506-7/19/$31.00 ©2019 IEEE 9–11 April, 2019 - American University in Dubai, Dubai, UAE
2019 IEEE Global Engineering Education Conference (EDUCON)

their capabilities. This system can give students formative feedback regarding their performance in different areas within a given subject. In this paper, [5] the authors focus on development of an Automated Short Answer Scoring (ASAS) systems. It is based on a program that can compute the similarity index between the correct answer and the student answer. It uses the Word Overlap technique to measure the similarity between short texts. The Word Overlap method makes use of the following coefficients: Jaccard, Dice and Cosine. Authors state that the similarity between sentences cannot rely only on word overlap. The answer can have a different structure and diverse words can be used to construct a sentence with same meaning. Test performed on different answers show that the Cosine Coefficient algorithm is more precise than the Dice and Jaccard Coefficient and it always inhibits a higher correlation value. However, these techniques can be used to compute short text that has a restricted number of words only. The authors in [4] proposed a technique which extracts text similarity from semantic and syntactic information in texts which are being compared. The proposed technique combines words by utilizing all the different words in the set of sentences. For every sentence, a semantic vector and a word order vector are derived respectively with the help of a lexical database. The meaning of a word is then measured by data extracted from a corpus. By integrating the semantic vector and the data from the corpus, another semantic vector is formed for each of the two sentences. The two vectors similarity is calculated according to the semantic. Then an order similarity is computed from the two order vectors. Consequently the similarity of the sentence is obtained by merging semantic and order similarity. However, the proposed method does not actually carry out word sense disambiguation for words having multiple meaning. Additionally, several techniques and algorithms have been leveraged by researchers, to implement their systems. These approaches can be clustered into different categories: Natural Language Processing techniques, Information Extraction (IE), Machine learning techniques and Ontologies [9].

III. Proposed System

A web knowledge based automated system was implemented for the correction and grading of students' scripts. The system comprises of a dictionary which was created to store synonyms of keyword of different answers. These synonyms were searched from Rhymzone API that uses Wordnet. The different combinations and variations of words (keywords) are stored in the dictionary and gradually the system will become more resourceful. In addition to that the knowledge based system contains the test questions, sample answers, keywords list, total marks and test answers. The system also has two web interfaces; admin and students. The admin part was designed for tutors. Tutors could create, edit and add sample answers to the system. The Admin part also keeps track of performances of each student for every test taken. In addition to that the admin part comprises of a setting section which has the following options:

- Grammar Deduction Point –this allows the tutors to set the number of marks that should be deducted for grammatical mistakes,
- Type of Correction for the questions-this option allow the tutor to select the type of correction i.e, based of semantic check only, based on both syntactic and semantic or perform an average of both
- Activate Syntactic Coefficient -This option was implemented to test three different algorithm(dice, jaccard or overlap coefficient.) to make comparison and check of keywords matching

The steps for building the knowledge base is very important as it make the system learn different type of answers possible with different keywords. The steps are elaborated as follows:

1. The system takes the model answer of the tutor and extracts meaningful keywords from it by using Dandelion API. A high level tokenization is performed and these keywords are stored in the dictionary and knowledge base
2. The system also extracts keywords from test answers by using Dandelion API. The synonyms of these keywords are found using Rhymzone and stored in the created dictionary.
3. The test answer is subjected to grammatical check using TextGear API that will check for possible grammatical mistakes and according to the setting selected by the admin, the system will deduct marks if necessary.
4. From the sets of keywords obtained above, the system will carry out the Lexical Analysis part whereby the meaningful keyword from the answer will be checked against model answer keywords that is stored in the dictionary.
5. Lastly the Semantic Analysis is carried out. The "test" answer is check against the model answer inserted by the tutor. This check consists of comparing the meaning of the two answers and tries to predict if it has the same meaning. This is carried out by the Dandelion API.
6. The score is then computed depending on the selection of the admin (based on semantic only, based on all aspects or an average of both)

In the students' part, students can login the system and take the test set by the tutors. The system will auto correct the test questions based on sample answers fed into the system by the tutors. In addition to that the students will receive a grade which will be displayed on the website after the correction. The steps involved for the correction of short structured questions include the following:

1. Students enter and submit answer for a specific question
2. The answer is first checked for grammatical error using TextGear API
3. The answer is then stored in knowledge base (student part)
4. The student answer is subjected to a high level tokenization. Keywords are extracted automatically from the student answer using Dandelion API.

These keywords are then checked with the keywords in the knowledge based (from tutor model answer and test answer). The keywords can also be compared to words from the created dictionary

5. Step 5 is the Lexical Analysis part (dice, jaccard or overlap coefficient is applied here) depending on what was selected by the admin.
6. The student answer is then compared to the model answer – the semantic analysis; whereby Dandelion API was used to check and compare meaning of the two answers (both syntactic and semantic check is carried out and a percentage of correctness is returned)
7. The final score of the answer is then computed depending on the selection of the admin. The score computed will be either of semantic match only, all aspects (semantic, lexical and grammar) or an average of all (to get a more accurate score)

A. *Architecture of the proposed System*

The architecture of the system is illustrated in Fig 1.0. The webserver hosts the website and the engine. All the questions, sample answers, keywords and marks are stored in the database. The engine will do the similarity check, keyword matching and grammar check.

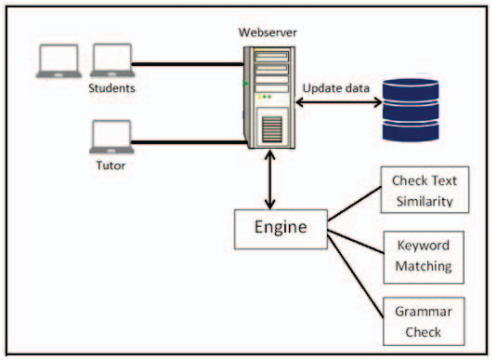

Fig 1.0 Architecture of system

The programs that have been used in the implementation phase are as follows:

- Dandelion has been chosen to check the meaning (syntactic and semantic check) of two short sentences. It works well even if the 2 sentences are not written in exactly the same way but with the similar meaning.
- MySQL database has been used in this system as it is scalable and easy to use.
- WordsAPI has been chosen to find synonym of keywords and compare it to stored data (answers).
- Xampp has been used as local server to host the website. It is a multi-cross platform and support MySQL. Xampp is also open-source and free to use.
- Cortical.io parses through a sentence and extracts keywords from it automatically.
- Curl has been used to transfer data and send request to API
- Scribans API has been used as grammar checker for the student's answers

B. *Design Issues of the proposed System*

Security- Security is an important design issue in this system. Each entity of the system has been assigned a login to authenticate themselves. In addition to that, students may try to communicate during a test and copy the answers of their classmates. To prevent this problem, the test questions have been randomly shuffled. Hence, the students will not have the questions in the same order.

Availability -It is important that the system is available 24/7. Unavailability of the system will hinder the whole process

IV. Testing and Evaluation

The system was tested with ten short structured questions: Two of these questions have been illustrated.
Question 1: Describe what is meant by encapsulation and the marks allocate by the tutor was 3 marks. The sample answer was Encapsulation consists of separating the external aspect of an object which is accessible to other objects, from the internal implementation details which are hidden from other objects.

Question 2: Explain the concept of Inheritance in object oriented and the marks allocate by the tutor was 2 marks. The sample answer was Inheritance define a relationship among classes, where one class share the structure or behavior of one or more classes.
Table 1.0 illustrates the answers given by different students for question 1

Student Answer	Expert Marking	System Marking
Encapsulation separates the external aspects of an object which are accessible to other object from the implementation details which are hidden from the object **(Normal case)**	**3**	**3**
The concept of encapsulation means hiding attributes of an object from other objects **(Normal case)**	**1.5**	**2**
Encapsulation is used to hide the internal representation or state of an object from the outside **(Normal case)**	**2**	**2**
Encapsulation allows hiding of data **(Normal Case)**	**1**	**1**
Encapsulation is the characteristics of being able to assign a different meaning or usage to something in different context **(abnormal case)**	**0**	**0**

Encapsulation consists of separating the external aspect of an object which are accessible to other objects, from the internal implementation details which are hidden from other objects **(Extreme case)**	**3**	**3**

Table 2.0 illustrates the answers given by different students for question 2

Student Answer	**Expert Marking**	**System Marking**
Inheritance is the concept where a class can inherit the properties and methods of other classes becoming a subclass. **(Normal case)**	**2**	**2**
Inheritance is when an object or class is based on another object or class **(Normal case)**	**1.5**	**1**
Inheritance is when one class inherits from another class **(Normal case)**	**0.5**	**1**
which classes acquire properties from other class**(Normal Case)**	**1**	**1.5**
It is the ability to redefine methods for derived class **(abnormal case)**	**0**	**0**
Inheritance define a relationship among classes, where one class share the structure or behavior of one or more classes (**Extreme case)**	**2**	**2**

The graph in fig 2.0 demonstrates the performance of the system compared to corrections of Expert in the specific field. Two questions were taken and various answers that student inserted were corrected by both the system and expert (manually). The graph thus shows the correlation between the two corrections. Data plotted are from table 1.0 and table 2.0

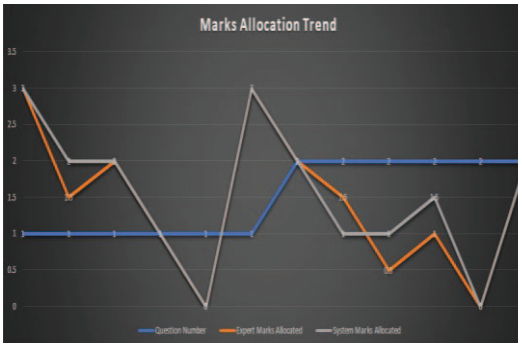

Fig 2.0

Fig 3.0 shows the result for automated and manual correction for two sets of 10 short structured questions answered by 5 students. System gives an accuracy of 79.7% for the two sets of 10 short structured questions answered by 5 students.

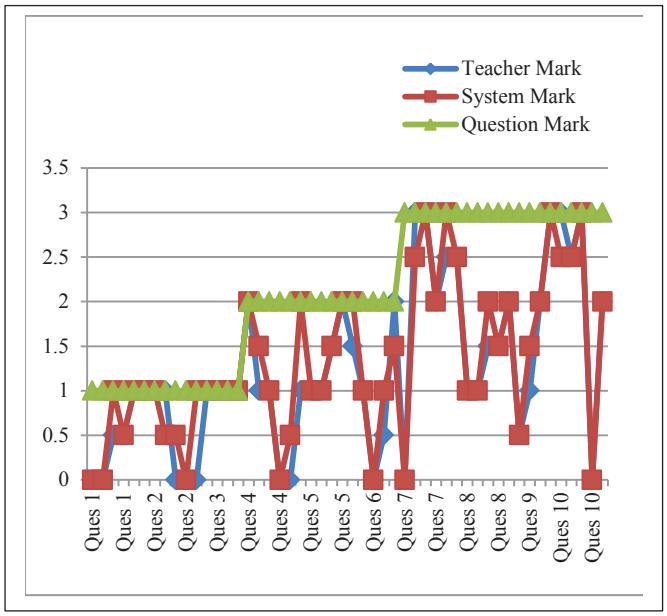

Fig 3.0 Automated and Manual marking of ten short structured questions

V. Conclusion

A knowledge base system which can automatically grade short structured questions was implemented. The knowledge base was fed with different variations of answers to make it learn and thus grade correctly the short-structured answers. A dictionary was also created for comparison of keywords from answers and this was also used in the knowledge base system. The implemented system was able to partly alleviate the burden of manually correcting and grading tests for tutors. The website also caters for the record (tracking) of all students taking the tests. These records consist of the details of the students, the tests taken and the obtained scores in each test. This could be used to monitor progress of students. Currently the system holds data of students for individual tests only. The system could be enhanced so that it cumulates marks for several assessments of the same module for a student and then use the aggregated marks as the final marks for the student.

References

[1] F. Kristensen, O. Troeng, M. Safavi and P. Narayanan, "Competition in higher education – good or bad?," Lund University, 2015.

[2] J. E. Rockoff and C. Speroni, "Subjective and Objective Evaluation of Teacher Effectiveness:Evidence from New York City," February, 2011.

[3] S. Simin and A. Heidari, "Computer-based assessment: pros and cons.," Elixir Edu. Tech 55 (2013) 12732-12734.

[4] L. Yuhua, M. David, B. Zuhair, O. James D. and C. Keeley, "Sentence Similarity Based on Semantic Nets and Corpus Statistics," *IEEE Transactions on Knowledge and Data Engineering,* vol. 18, no. 8, pp. 1138-1150, August 2006.

[5] P. Feddy Setio, A. Teguh Bharata, P. Adhistya Erna, M. Anggraini and U. Aryo Baskoro, "Automatic short answer scoring using words overlapping methods," in *AIP Conference Proceedings 1818*, 2017.

[6] R. &. C. J. Siddiqi, "On the Automated Assessment of Short Free-Text. [Online]," 2008. [Online]. Available: Available at: http://citeseerx.ist.psu.edu/viewdoc/download?. [Accessed 2018 09 21].

[7] O. Mason and I. Grove-Stephensen, "Automated free text marking with Paperless School," in *Proceedings of the 6th CAA Conference, Loughborough University*, (2002).

[8] S. Alber and L. Debiasi, "Automated Assessment in Massive Open," University of Salzburg, 2013.

[9] M. Abdulrahman A. and A. Shaha T., "Hybrid approach for automatic short answer marking," *Southwest decision sciences 43rd annual meeting, 2010.*