EasyChair Preprint

№ 2905

AI Based E-Assessment System

Saloni Kadam, Priyanka Tarachandani, Prajakta Vetal and Charusheela Nehete

March 9, 2020

AI Based E-Assessment System

Saloni Kadam
Dept of Information Technology
Vivekanand Education Society Institute of Technology
Maharashtra, India
2016.saloni.kadam@ves.ac.in

Priyanka Tarachandani
Dept of Information Technology
Vivekanand Education Society Institute of Technology
Maharashtra, India
2016.priyanka.tarachandani@ves.ac.in

Prajakta Vetal
Dept of Information Technology
Vivekanand Education Society Institute of Technology
Maharashtra, India
2016.prajakta.vetal@ves.ac.in

Mrs. Charusheela Nehete
Dept of Information Technology
Vivekanand Education Society Institute of Technology
Maharashtra, India
charusheela.nehete@ves.ac.in

Abstract -- We have seen that a number of students apply for various examinations which may be institutional, non-institutional or even competitive. The competitive exams mostly have objective or multiple choice questions(mcqs). The automation of scoring of subjective or descriptive answers is a need considered nowadays. This paper focuses on designing an efficient algorithm that will automatically evaluate the answers given by students and assign a score based on the AI technologies which are as good as scores given by a human being.

Keywords : E-assessment, Online Subjective Answer Checker, Similarity, Grammar, Descriptive..

I. INTRODUCTION

Examination is a test of a person's knowledge in a particular area which is either subjective or objective or both. Usually, competitive examinations consist of multiple choice questions or mcqs. Automatic evaluation of the objective exams is beneficial as it saves time, provides efficiency and reduces usage of resources. However this automated evaluation technique is only for the objective exams and not for the subjective ones. Subjective answer sheet checking is one of the huge administrative tasks for any education institute. In this examination process, candidates need to write answers, an examiner collects those answer sheets and submits them to authority for further checking process. This process involves 3 levels of paper checking:-

- First Level Paper Checker,
- First Level Moderation,
- Second Level Moderation.

So, the amount of pressure education systems and teachers hold is understandable as the number of answer sheets to evaluate is too large. So there is a necessity for an approach which will reduce the usage of resources which will automatically evaluate the answers given by students and provide results. Such a system is the goal of this paper. We have developed an E-assessment system that checks the answer sheet of the student and provides marks to the same. The system consists of an algorithm that compares the student's answer against the reference answer given by the teacher. It takes certain parameters into consideration while evaluating, such as keywords, grammar, similarity between sentences, synonyms. Both the answers need not be exactly the same or word to word. This approach can be a quick and easy way for the examiners by reducing their workload.

II. LITERATURE SURVEY

[1] Sheeba Praveen, "An Approach to Evaluate Subjective Questions for Online Examination System:, International Journal of Innovative Research in Computer and Communication Engineering. Vol. 2, Issue 11, November 2014. This system will solve the problem of deducing knowledge represented by partially or grammatically incorrect sentences, and will interpret the meaning conveyed by the student in different forms and sentences, propose a normalized strategy for grading the answers, ways to interpret the mathematical formulas and expressions however the system will be limited to non mathematical subjects only.

[2] Amarjeet Kaur, M Sasikumar, Shikha Nema, Sanjay Pawar(2013), "Algorithm for Automatic Evaluation of Single Sentence Descriptive Answer", International Journal of Inventive Engineering and Sciences (IJIES). In this paper, they have considered only text in single sentence descriptive answers which are grammatically correct and with no spelling mistakes. Their approach is to represent learners and standard answers in the form of graphs and then comparing it by applying some of the similarity measures for the allocation of marks.

[3] Aditi Tulaskar1, Aishwarya Thengal2, Kamlesh Koyande3, "Subjective Answer Evaluation System", Department of Information Technology Vidyalankar Institute of Technology, Mumbai, India. The proposed system will allot the marks according to the percentage of accuracy present in the answer. This is a software system in which users will be authenticated by using user login. After the authentication, users will be provided with the questions. The proposed system is designed to evaluate answers for

five users providing five different answers. The standard answer is stored in the database with the description meaning and keywords. Then it will evaluate each answer by matching the keywords or the key concepts as well as its synonyms with the standard answer. It will also check the grammar and spellings of the words. After the evaluation, it will grade the answer depending on the correctness of the answer.

[4] Asmita Dhokrat,Gite Hanumant R.,C.Namrata Mahender(2017), "Automated Answering for Subjective Examination",(IJCA). The proposed system has two login facilities. The user login is the login allocated for the students. As soon as you click the student login button you will be asked to enter login id and password. The system will check for the id and automatically display students name, email id and phone num for verification. The user login will be able to write answers with respect to the question uploaded. The system will show marks scored as soon as you enter the next button.~~The~~ Admin login will let the teachers login. The admin login's each user will have his own password and id through which they can login in. The admin can add subtract questions, check for students marks and so on. Just like the teachers can do manually.

[5] Merien Mathew, Ankit Chavan, Siddharth Baikar, "ONLINE SUBJECTIVE ANSWER CHECKER", International Journal of Scientific & Engineering Research, 2017. The system requires you to store the original answer for the system. This facility is provided to the admin. The admin may insert questions and respective subjective answers in the system. These answers are stored as notepad files. When a user takes the test he is provided with questions and area to type his answers. Once the user enters his/her answers the system then compares this answer to the original answer written in database and allocates marks accordingly. Both the answers need not be exactly the same, word to word. The system consists of in built artificial intelligence sensors that verify answers and allocate marks accordingly as good as a human being.

III. IMPLEMENTATION

We have developed the process of Subjective Answer Evaluation which includes short answers. It goes through preprocessing-case normalization, stopwords removal, tokenization which is done using the seven features discussed below. We have assigned weights to these features based on their accuracy and importance in the evaluation process. Each answer coming from a student is evaluated using all the seven features against the reference answer provided by teacher.

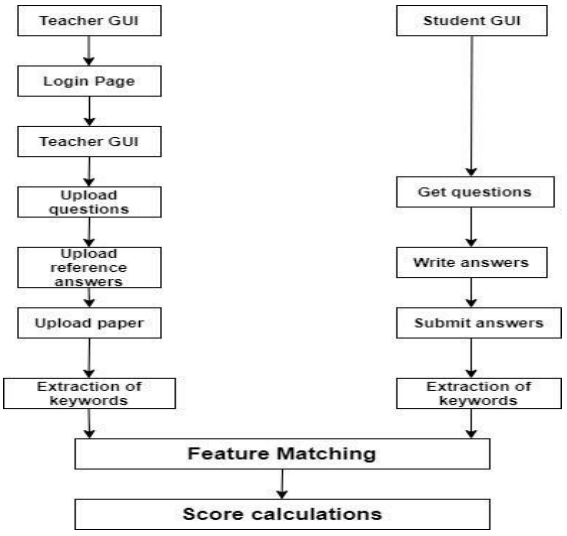

Fig.1 Flow diagram

A. Weighing Features

1. Spacy Similarity
2. Tfidf Vectorizer
3. Difflib Similarity
4. Jaccard Similarity
5. Grammar check
6. Cosine Similarity
7. WMD Distance

1) Spacy Similarity

SpaCy is a number which ranges between 0 to 1 and tells us how close two words are, semantically. This is done by finding similarity between word vectors in the vector space. spaCy is one of the NLP libraries which is widely used today, providing a simple method for this task. spaCy's Model supports two methods to find word similarity: using word vector and context-sensitive tensor.

2) Term Frequency Inverse Document Frequency

TF-IDF stands for "Term Frequency - Inverse Document Frequency".

Term Frequency (tf): It gives us the frequency of the word in each document in the corpus. It is the ratio of the number of times the word appears in a document compared to the total number of words in that document. Term Frequency increases as the number of occurrences of that word within the document increases. Each and every document has its own tf.

Inverse Document Frequency (idf): It used to calculate the weight of infrequent words across all documents in the corpus. The words that do not appear often in the corpus have a high IDF score. It is given by the equation below.

Example Sentence 1 : The cat is sleeping on the table.

Sentence 2: The dog is sleeping on the bed.

In this example, each sentence is considered as a separate document.

The TF-IDF for the above two documents, which represent our corpus is given below.

Word	TF (A)	TF (B)	IDF	TF*IDF (A)	TF*IDF (B)
The	1/7	1/7	log(2/2) = 0	0	0
Cat	1/7	0	log(2/1) = 0.3	0.043	0
Dog	0	1/7	log(2/1) = 0.3	0	0.043
Is	1/7	1/7	log(2/2) = 0	0	0
Sleeping	1/7	1/7	log(2/2) = 0	0	0
On	1/7	1/7	log(2/2) = 0	0	0
The	1/7	1/7	log(2/2) = 0	0	0
Table	1/7	0	log(2/1) = 0.3	0.043	0
Bed	0	1/7	log(2/1) = 0.3	0	0.043

Fig.2 Table - TF-IDF corpus

From the above table, we see that TF-IDF of common words is zero, which shows that they are not significant. The TF-IDF of "cat" , "dog", "table", and "bed" are not zero. These words have more significance.

3) Difflib Similarity

The difflib module in python contains tools for computing and working with differences between sequences. It offers a way to compare mult- line strings and entire list of words. get_close_matches(word, possibilities, n, cutoff) function work in Python which returns the best 'good enough' matches. It accepts four parameters in which n, cutoff are optional. Word is a sequence for which close matches are desired and possibilities is a list of sequences against which the word is matched.

4) Jaccard Similarity

Jaccard similarity measures similarity between finite sample sets. It is also called an intersection over union and is defined as the size of intersection divided by size of union of two sets.

Jaccard Similarity = (Intersection of set A and set B) / (Union of set A and set B)

The range is between 0 to 1. If the score is 1, then they are identical and if there is no common word between the first sentence and the last sentence then the score is 0.

5) Grammar check

To check and detect grammatical mistakes and spelling errors Grammar Bot API is used in our system. When the text is sent to Grammar Bot's API , it returns a list of potential grammar and spelling errors.

6) Cosine Similarity

Cosine similarity is a benchmark to measure how similar the documents are irrespective of their size. The similarity between two sentences can be found as a dot product of their vector representation.

Hence, the matching is done based on these methods to calculate the students' score.

7) WMD Distance

The measure of similarity between two blocks of text can be used as a good measure for evaluation of answers. Ideally statically based algorithms like LSA,BLEU etc can capture semantic relation between two documents. So when two documents have no word in common their euclidean distance would be maximum. One way to face this problem is to use the word mover's distance (WMD) that adapts the earth mover's distance to the space of documents. At a high abstraction, the WMD is the minimum distance required to transport the words from one document to another. We assume that we are given a word embedding matrix (word2vec). We use Word Mover Distance (WMD) problem on a matrix of pairwise distances between each state vector of the model and student answers. If a word 'wi' appears 'fi' times in a document, its weight is calculated where 'n' is the number of unique words in the document. The higher its weight, the more important the word is. The dissimilarity between word 'wi' in student answer and word 'wj' in model answer can be computed as where 'xi' and 'xj' are the embeddings of the words 'wi' and 'wj' , respectively.

B. Score Generation

For every feature we have assigned weights based on their accuracy and importance in the evaluation process. Marks for the answer are reliant on percentage of keywords match, grammar, synonyms etc. Hence if a student writes an answer missing any of these, marks will be deducted according to their weightage in evaluation.

Table 2. Score Generation Criteria

Features	Weightage allotted
Grammar check	10%
Jaccard Similarity	29%

TF-IDF vectorization	35%
Spacy	2%
difflib	4%
Cosine similarity	12%
WMD distance	8%

Each question is considered of 10 marks so the accuracy obtained from each feature for a question is scaled out of 10. This is how marks for all questions are calculated and finally added to get the result.

IV. RESULT

The Teacher GUI is as shown below, where the teacher uploads questions and reference answers for the student with unique question paper code. Also, she can click on 'check marks' and check marks of students.

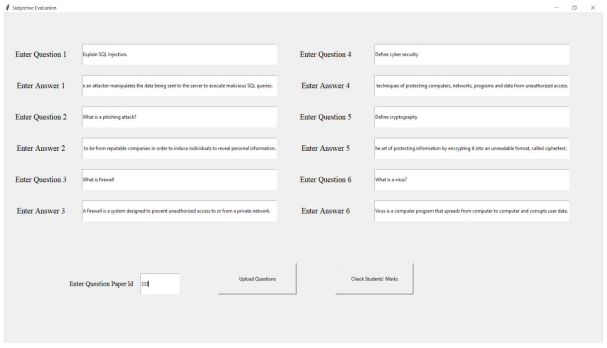

Fig.3 Teacher GUI-1

After clicking on the 'check marks' button, the page below appears.

Fig.4 Teacher GUI-2

In Students GUI, when the student enters Roll No, Question paper ID and clicks on 'Get question', the question and the answer space will be displayed. The student then submits the answers and checks marks.

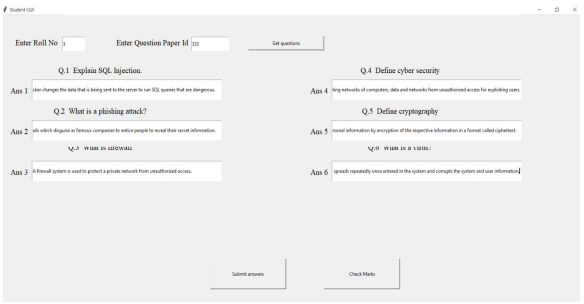

Fig.5 Student GUI

When the student clicks on 'Check marks', a result page is displayed wherein they can see the marks of each question and also the Total marks scored.

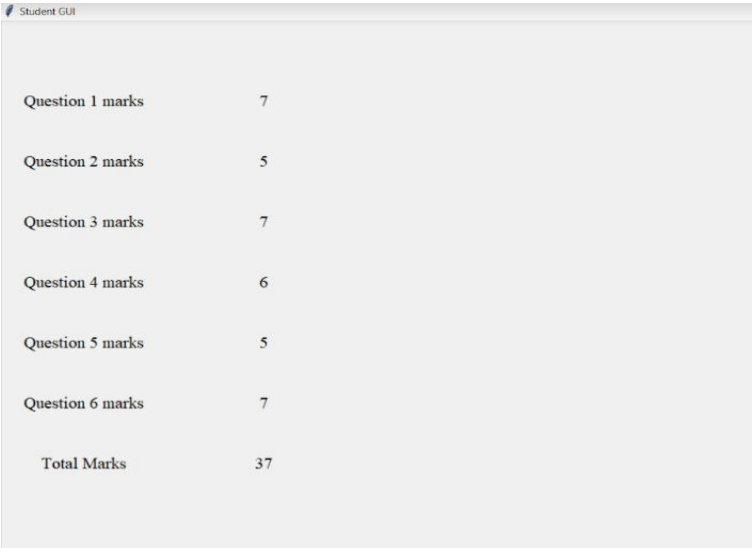

Fig.6 Result Page

V. CONCLUSION

E-Assessment System would be beneficial for the universities, schools and colleges for academic purpose by providing ease to faculties and the examination evaluation cell. Many educational institutes conduct their examinations online, but these exams only contain multiple choice questions which test the student's aptitude, and fail to test the conceptual knowledge a student or learner must possess. Therefore descriptive answers must be included in online examinations. Our proposed system evaluates the answer based on the keywords. By judging against the reference answer, marks are allocated to the student. Highest marks are gained if the student writes grammatically correct answers with all the keywords mentioned in the reference answer. Hence the proposed system could be of great utility to the educators whenever they need to take a quick test for revision purposes, as it saves time and the trouble of evaluating the bundle of papers.

VI. ACKNOWLEDGMENT

We are thankful and would like to express our sincere gratitude towards Prof Charusheela Nehete for her continuous guidance, support and encouragement in making

this project possible. Our sincere thanks to all the Lecturers who helped us in our endeavor, gave valuable advice and thus made our journey possible. Last but not the least we would also like to thank our family and friends for their constant support, motivation and belief in us.

VII. REFERENCES

[1] Sheeba Praveen, "An Approach to Evaluate Subjective Questions for Online Examination System", International Journal of Innovative Research in Computer and Communication Engineering. Vol. 2, Issue 11, November 2014

[2] Amarjeet Kaur, M Sasikumar, Shikha Nema, Sanjay Pawar(2013), "Algorithm for Automatic Evaluation of Single Sentence Descriptive Answer", International Journal of Inventive Engineering and Sciences (IJIES).

[3] Aditi Tulaskar1, Aishwarya Thengal2, Kamlesh Koyande3, "Subjective Answer Evaluation System", Department of Information Technology Vidyalankar Institute of Technology, Mumbai, India.

[4] Asmita Dhokrat,Gite Hanumant R.,C.Namrata Mahender(2017), "Automated Answering for Subjective Examination",(IJCA).

[5] Merien Mathew, Ankit Chavan, Siddharth Baikar, " ONLINE SUBJECTIVE ANSWER CHECKER", International Journal of Scientific & Engineering Research, 2017.