iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VII **Month of publication:** July 2021

DOI: https://doi.org/10.22214/ijraset.2021.36449

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

Weighing Students Narrative Answers using NLP-RAKE Algorithm

B. Tharun Kumar[1], K. Lakshmi Prasanna[2], K. Vivek Chaitanya[3], B. Avinash[4], B. Vishnu Vardhan[5]
[1, 2, 3, 4]*UG.Students,*[5]*Assistant Professor, Department of C.S.E, Prasad V. Potluri Siddhartha Institute of Technology, A.P. India*

Abstract: ***The Examination portals that are deployed over several servers are used to conduct online Examination for various purposes among which some may include conducting a test for Entrance Examinations, or Olympiads at National and International level and while some portals are designed to conduct a test for placement purposes. But mostly existing portals are designed to conduct tests that contain multiple choice Questions. So, the aim of this project is to evaluate marks for Descriptive Answers.***

Keywords: Natural language Processing, stop words, text summarization, Candidate and adjoining keywords, keyword scores.

I. INTRODUCTION

Student performance is evaluates using different assessment strategies. Most commonly used method is descriptive examination. In this method, a student expresses his/her opinion in response to the answer in a text pattern. With automatic descriptive answer evaluation system, we can achieve good evaluation system for the education system by checking students answers under various parameters like keywords, length of the answer, grammar of a sentence, spelling mistakes. A student may answer a question by following different grammatical styles like answering in active voice or passive voice or simple present tense, and chooses different words similar to the actual answer. The motivation behind the automated answer script evaluation is to
reduce the time of evaluation i.e., it is very difficult for a human to read each and every line and award marks, prohibiting human evaluator's psychological changes, and very easy to keep record and extraction. It also assures that mood swings or change in perspective of the human assessor will not affect the evaluation process. The automatic answer script evaluation based on Natural Language Processing (NLP) will help us to overcome the difficulties faced in the manual evaluation. Here a student's written answer is fed to the system and the system will automatically score marks after the evaluation. The system considers all possible factors like spelling error, grammatical error, and various similarity measures for scoring marks. The natural language processing technique is used to make the handling of used English language much easier.

II. PROBLEM STATEMENT

Evaluating the Multiple Choice Questions is easy because those questions will have the correct answer in the database. If the student kept the same answer, then the student is awarded mark otherwise not. But, coming to descriptive answers it is not like that. Aim of this project is to create an application, to evaluate student's descriptive answers using natural language processing. So, to evaluate the descriptive answers we have to consider many factors. Keeping mind that, what is the input, features to consider and how to evaluate the answer plays a vital role in this application.

III. EXISTING SYSTEM

Many portals are evaluating the Multiple Choice based questions for Entrance Examinations, or Olympiads at National and International level and while some portals are designed to conduct a test for placement purposes. During the lockdown times, all most all the placement examinations are conducted through online examinations. Those examinations have multiple choice based questions, some coding challenges and many more. In such portals we have a questions and multiple choices for every question. So, the questions along with the correct answers are stored in the database. Only correct answer for each question is stored in the database. So, whenever the student submitted the exam the evaluation will be like for every question the correct answer is fetched from the database and after fetching the fetched answer and the student's answer is checked. If the fetched answer and the student's answer is same then the student is awarded mark for that question. If the fetched answer and student's answer is not same then that question is not awarded mark. Because, the multiple choice questions will only have options and one option from those given options must be correct. So, if the student answer and fetched answer does not match we will award zero mark for that question. And finally, all the marks are added and those marks will be shown. This is the easy method to evaluate the marks of a student in multiple choice based exam.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

IV. PROPOSED SYSTEM

In the existing system, students are evaluated with multiple choice questions only. The drawback with this kind of system is, the answers are displayed and student may guess/randomly choose the answer, with this student may get good marks without attaining the proper knowledge. The Student's Descriptive Answers evaluation will evaluate the Student's descriptive answers and award the marks to the student's answers by checking many factors such as keywords matching, length of an answer, grammar checking and spell checking. So, this system will test the students' knowledge in a subject, eliminates physiological changes and mood swings of an evaluator while evaluating the answers. The below figure. 1 describes the design methodology of this application. The Rapid Automatic Keyword Extraction (RAKE) algorithm is used in this application. What this method does is split the text into a list of words and remove stopwords from that list.

This returns a list of what is known as content words.

Suppose our list of stopwords and phrase delimiters look like these:

stopwords = [is, not, that, there, are, can, you, with, of, those, after, all, one]

delimiters = [., ,]

Then, the algorithm splits the text at phrase delimiters and stopwords to create candidate expressions. So, the candidate keyphrases would be the following:

Keyword extraction is not that difficult after all. There are many libraries that can help you with keyword extraction. Rapid automatic keyword extraction is one of those.

A. Rake Algorithm

RAKE also known as Rapid Automatic Keyword Extraction is a keyword extraction algorithm. RAKE Algorithm is a well-known keyword extraction method which uses a list of stopwords and phrases delimiters to detect the most relevant words or phrases in a piece of text. Using keywords, the system will assign marks to the answer.

```
def split_sentences(text):
    sentence_delimiters = re.compile(u'[\\[\\]\n.!?,;:\t\\-\\"\\(\\)\\\'\u2019\u2013]')
    sentences = sentence_delimiters.split(text)
    return sentences
```

For example if this below input is feed to the system:

"Compatibility of systems of linear constraints over the set of natural numbers. Criteria of compatibility of a system of linear Diophantine equations, strict inequations, and non strict inequations are considered."

After execution of split_sentences(text) function, the result will be the list of sentences separated by delimiters.

['Compatibility of systems of linear constraints over the set of natural numbers' , ' Criteria of compatibility of a system of linear Diophantine equations', ' strict inequations', 'and nonstrict inequations are considered.']

B. Text Extraction

The text summarization is a process of creating a short, accurate summary of the longer text. It is very time wasting task to generate a summary of longer article manually. Hence an NLP-based automatic text summarization technique is used to facilitate and speed up the text processing.

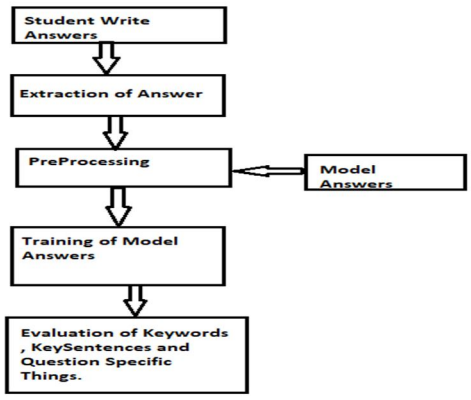


Figure 1

©IJRASET: All Rights are Reserved

Two types of text summarization techniques are used for generating the summary. The extractive text summarization technique is used to select phrases and sentences from the source document, and generates a new summary. The abstractive text summarization technique is the opposite of extractive technique. It generates entirely new phrases and sentences to hold the meaning of the source document. The NLP-based strategies are very well suited for generating summary rather than the manual process. The summarized text will be fed as input to compute various similarity measures.

C. Stop Words

Stop words are the meaningless words in the text. They have no meaning they are just used to form a sentence(E.g, the, in, a …..). The input parameters for the **RAKE Algorithm** comprise a list of stop words also a set of phrase delimiters and word delimiters. It uses **stop words** and **phrase delimiters** to partition the document into candidate keywords, these candidate keywords are mainly the words that help a developer in extracting the exact keyword necessary to get information from the text.

D. Candidate Keywords

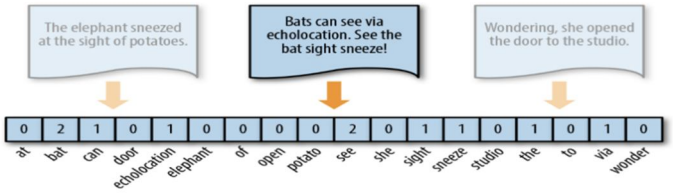
RAKE classifies the main content bearing word as Candidate Keyword by parsing the document with the help of stop words and phrase delimiters. This is done basically by some of the following steps, firstly the document text is split into an **array of words** by the **specific word delimiters**, and secondly, the array is again split into a **sequence of contiguous words** at **phrase delimiters** and **stop word positions**. Finally, the words that lie in the same sequence are assigned the same position in the text and together are considered as a candidate key.

E. Adjoining Keywords

As we know that Rake splits candidate keywords by stop words, so the extracted keywords do not contain **interior stop words**, therefore an interest wan expressed in identifying keywords that contain interior stop words as **the axis of evil**. To find keywords that adjoin one another at least twice in the same document and the same order. For this purpose, a new candidate keyword is created as a combination of those keywords and the interior stop words. In this part, we should understand that very few **linked words** are only extracted which add **significance**.

F. Keyword Scores

After identifying all the candidate keywords from the text data, a **graph of word co-occurrence** is generated which calculates the score for each candidate keyword and defined as the **member word score**. With the help of this graph, we evaluate **several metrics** for calculating word scores, based on the **degree** and **frequency** of the vertices in the graph.



G. Extracted Keywords

After the candidate keyword score is calculated, the top **T candidate keywords** are selected from the document. The **T value** is **one-third** the number of words in the graph.

H. Number Of Words

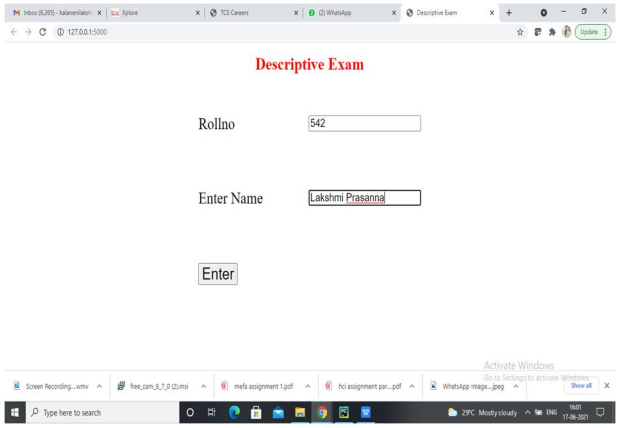
The given text is splitted using the delimeter space and the extracted result is the list of words. Calculate the number of words. Also using the number of words we will assign marks to the text.

I. Language-Tool

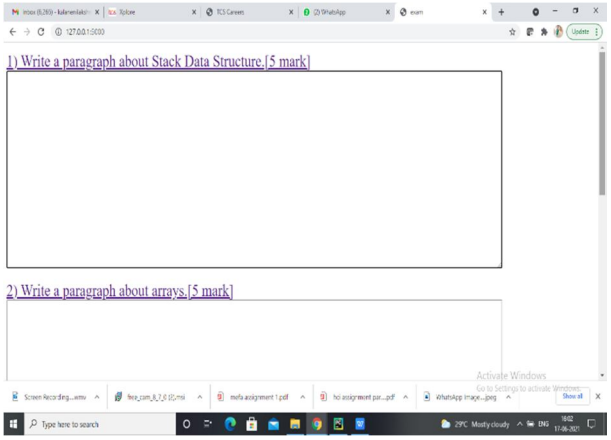
LanguageTool is an open-source grammar tool, also known as the spellchecker for OpenOffice. This library allows you to detect grammar errors and spelling mistakes through a Python script or through a command-line interface. We will work with the language_tool_pyton python package which can be installed with the pip install language-tool-python command. By default, language_tool_python will download a LanguageTool server .jar and run that in the background to detect grammar errors locally.

©IJRASET: All Rights are Reserved

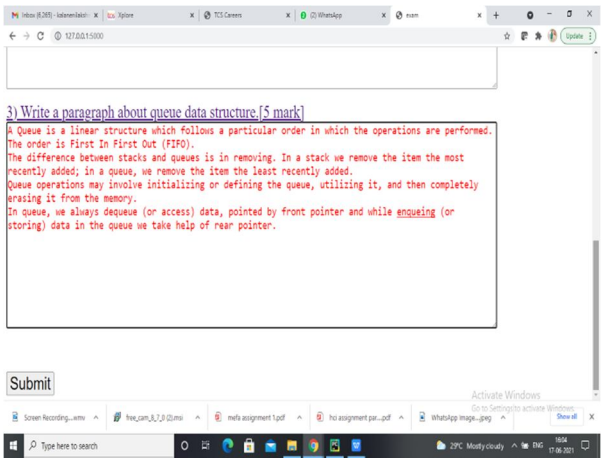
V. RESULTS



Student Registration



Students Answer

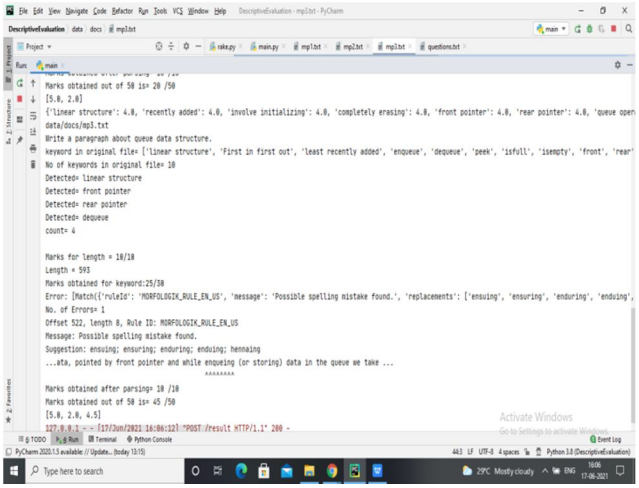


Questionaries'

©IJRASET: All Rights are Reserved



Student Gained marks are Displayed



Extracting keywords and assigning marks for Keywords, no. Of words, grammar for Q3

VI. CONCLUSION

This project is implemented with the help of Natural Language Processing for automatic answer script evaluation and marks scoring. Our study focussed of the following steps (1) text summarization using keyword-based technique, (2) text pre-processing for futher analysis, (3) calculate number of text pre-processing for words (4) Checking for grammar and spelling mistakes (5) marks scoring.

So, this project is all about calculating marks for descriptive answers automatically by the system without intervention of humans. So, this with this kind of system we can save a lot of time and also there is no chance of pointing a evaluator.

VII. SCOPE OF THE SYSTEM

The Descriptive answers Evaluation is used for evaluating the descriptive answers of the student using Natural language processing. This project is used to evaluate only the text based answers. From the text, the keywords are extracted and based on the keywords marks are assigned. Next based on the number of words marks are assigned. And the third criteria is checking the grammar and spelling mistakes of the text.

This project evaluates the text based answers only.

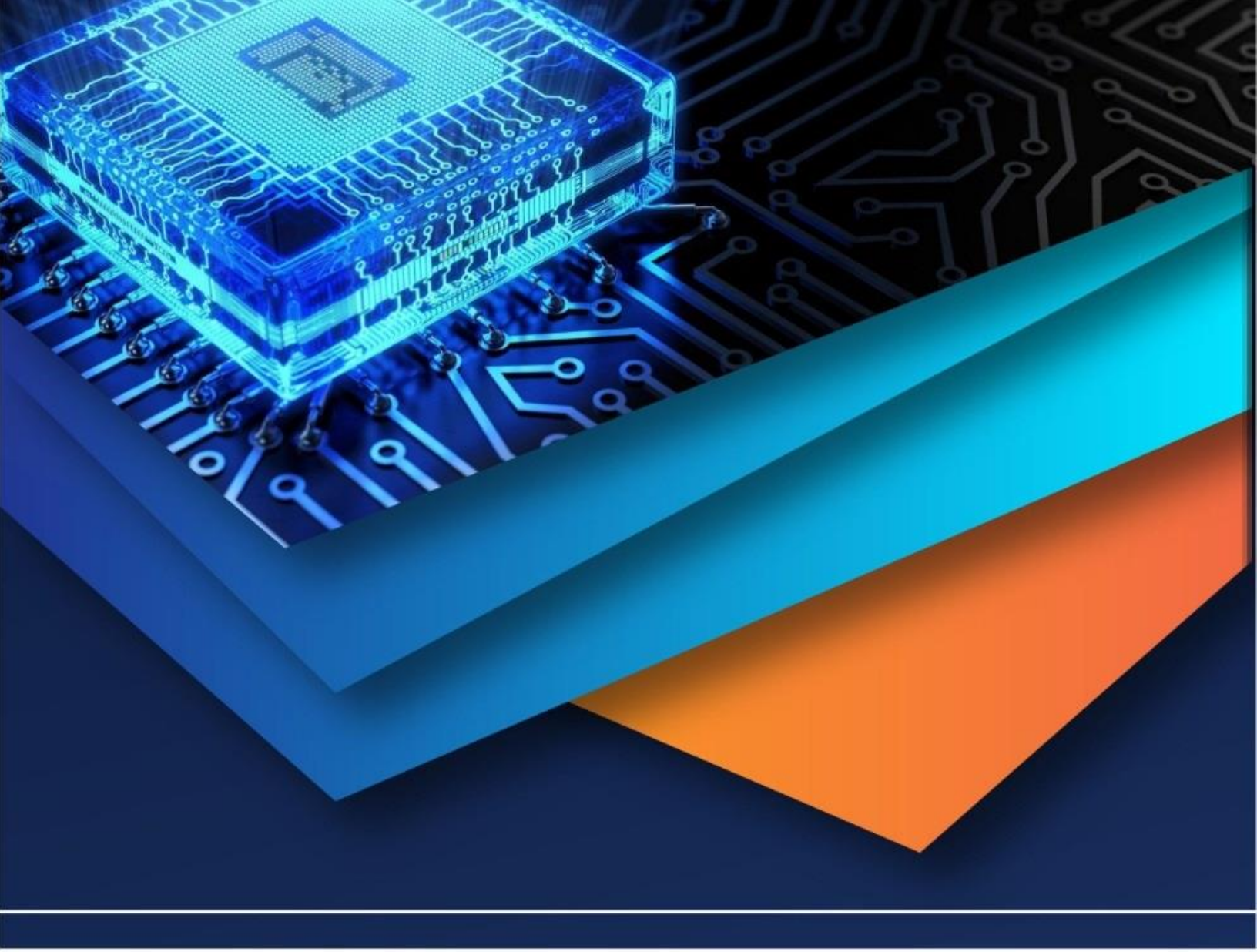
So, the future scope of this project is to build the system to evaluate the graphs, pictures and diagrams etc

©IJRASET: All Rights are Reserved

REFERENCES

[1] Nandita Bharambe, Pooja Barhate, Prachi Dhannawat “Automatic Answer Evaluation Using Machine Learning” (IJIT) – Volume 7 Issue 2, Mar - Apr 2021
[2] B.Vani and M. Shyni Beaulah, "High accuracy Optical Character Recognition algorithms using learning array of ANN", in International Conference on Circuit, Power and Computing Technologies [ICCPCT] 2014, pp. 1474 – 1479
[3] Sk Asif Akram, Mousumi Saha and Tamasree Biswas, "Evaluation of descriptive answer sheet using machine learning", in International Journal of Engineering Sciences \& Research Technology (IJESRT), April 2019, pp. 184-186.
[4] V. Lakshmi and Dr V. Ramesh, "Evaluating students descriptive answer using natural language processing and artificial neural networks", in International Journal of Creative Research Thoughts (IJCRT), Volume 5, Issue 4, December 2017, pp. 3168-3173.
[5] Abhishek Girkar, Mohit khambayat, Ajay Waghmare, Supriya Chaudhary "Subjective Answer Evaluation using Natural Language Processing and Machine Learning" (IRJET) Volume: 08 Issue: 04 | Apr 2021
[6] Vishal Bhonsle, Priya Sapkal, Dipesh Mukadam, Prof. Vinit Raut, “An Adaptive Approach for Subjective Answer Evaluation” VIVA-Tech International Journal for Research and Innovation Volume 1, Issue 2 (2019)
[7] Rosy Salomi Victoria D, Viola Grace Vinitha P, Sathya R,” Intelligent Short Answer Assessment using Machine Learning” International Journal of Engineering and Advanced Technology (IJEAT) , Volume-9 Issue-4, April 2020.
[8] Prince Sinha, Sharad Bharadia, Ayush Kaul, Dr. Sheetal Rathi ,“Answer Evaluation Using Machine Learning” Conference-McGraw-Hill Publications March 2018
[9] Aditi Tulaskar, Aishwarya Thengal, Kamlesh Koyande, “Subjective Answer Evaluation System”, International Journal of Engineering Science and Computing, April 2017 Volume 7 Issue No.4

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)