RAJIV GANDHI INSTITUTE OF TECHNOLOGY

KOTTAYAM - 686501

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

MINI PROJECT REPORT

DESCRIPTIVE ANSWER EVALUATION SYSTEM

Submitted By

MIDHILA V K

REGISTER NO:KTE17CSCE12

for the award of the degree of

Master of Technology in Computer Science and Engineering

Project Guide

Prof.Raji.R.Pillai

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

April 2018

RAJIV GANDHI INSTITUTE OF TECHNOLOGY

KOTTAYAM - 686501

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the project report entitled ***DESCRIPTIVE ANSWER EVALUATION SYSTEM*** *submitted by* ***Midhila V K (Register No:KTE17CSCE12)*** *in partial fulfillment of the requirement for the award of Master of Technology Degree in Computer Science and Engineering by the* ***APJ Abdul Kalam Technological University*** *is a record of her bonafide work under our guidance and supervision.*

Project Guide PG Co-ordinator

Head of the Department

Place: Kottayam

Date:

ACKNOWLEDGEMENT

The project on Descriptive Answer Evaluation System wouldn't have been a success if it wasn't for support and guidance from various sources throughout the development. I use this opportunity to show gratitude towards everyone who have had impact on this project, directly or indirectly.

First of all, I thank God almighty for His grace and blessings for without his unforeseen guidance this would have remained in dreams.

I express my sincere gratitude to **Dr. Jalaja M.J** Principal, Rajiv Gandhi Institute of Technology, Kottayam, for providing the ambiance for carrying out the work of this project.

I am deeply indebted to **Dr.Kavitha.N**, Head of the Department, Computer Science and Engineering,for providing permission and availing all required facilities for undertaking the project in a systematic way.

I express my sincere gratitude to **Dr.Sobhana N.V** ,PG Coordinator for her constant support in every phase and for the source of strength in completing this work. As a Project Coordinator,she has provided necessary assistance throughout the project.

I am extremely grateful to **Prof.Raji.R.Pillai**, Project Guide,Department of Computer Science and Engineering, for her constant guidance throughout all the phases of the project and timely support to overcome difficulties in the project

Gratitude is extended to all teaching and non-teaching staffs of Department of Computer Science and Engineering, for the sincere directions imparted and the cooperation in connection with the project.

I am also thankful to my parents for the support given in connection with the project. Gratitude may be extended to all well-wishers and my friends who supported me for the project.

Midhila V K

ABSTRACT

Every year many examinations are conducted like competitive, intuitional, non institutional which students apply for. Competitive and entrance exams typically contain objective or multiple choice questions. These exams are evaluated on machine and therefore their evaluation is easy. However as these exams accommodate only objective or multiple choice questions the ability to answer descriptive question is still not provided and evaluation of the same. It will be very helpful for educational institutions if the process of evaluation of descriptive answers is automated to capably assess student's exam answer sheets.

The proposed Descriptive Answer Evaluation System introduces a novel approach that combines the pattern mining unsupervised technique with the similarity measurement. In this, the sequential word patterns that are more common (than the rest of the patterns)among student answers are identified along with the count for each pattern. Along with this, the students answers are matched against the pre-graded answers or manually crafted key concepts using the concept of similarity measurement. Finally a weighted score function is generated for each answer script and evaluated accordingly.

The proposed system seeks to implement an application which will be able to evaluate the descriptive answer to a question.This system could be of great effectiveness to the educational institutes, as it saves time and trouble of checking bundle of papers. Initially the current system evaluates answers written in English.

Contents

Acknowledgement **III**

Abstract **IV**

List of Figures **V**

Definitions, Acronyms and Abbreviations **VIII**

1 Introduction **1**

1.1 Motivation 2

1.2 Organization of the report 2

2 PROBLEM DEFINITION **3**

2.1 Approach to NLP 3

2.2 The Infrastructure 3

2.2.1 **[A] Data set and Text Preprocessing** 4

2.2.2 **[B] Sentence Feature** 4

3 LITERATURE SURVEY **6**

4 SYSTEM REQUIREMENTS **9**

4.1 Operating Environments 9

4.1.1 Software constraints 9

4.1.2 Hardware constraints 9

4.2 Functional Requirements 9

4.3 Non Functional Requirements 9

4.3.1 Performance 9

4.3.2 Reliability 10

4.3.3 Portability 10

5 SYSTEM MODEL AND DESIGN 11

5.1 Architecture of the System 11

5.2 Detailed system design 12

5.2.1 Preprocessing Module 12

5.2.2 Pattern Mining Module 12

5.2.3 Similarity Comparison Module 13

5.2.4 Weighted Function Generation Module 13

5.2.5 Scoring Module 13

6 IMPLEMENTATION 14

6.1 Graphical User Interface 14

6.2 Coding of Algorithm 16

6.3 Display Output 17

7 PERFOMANCE ANALYSIS 18

8 LIMITATIONS AND EXTENSIONS 20

8.1 Limitations 20

8.2 Extensions 20

9 CONCLUSION 22

BIBLIOGRAPHY 23

List of Figures

2.1 Cosine Similarity Formula 5

5.1 General System Architecture 11
5.2 General Architecture of the Evaluation Module 11
5.3 Cosine Similarity Function 13

6.1 Graphical User Interface Window 14
6.2 Selecting File 15
6.3 Running File 15
6.4 Sample Code 16
6.5 Sample Input 17
6.6 Sample Output 17

7.1 Comparison Of Marks Obtained 18

Definitions, Acronyms and Abbreviations

NLP	Natural Language Processing
MCQ	Multiple Choice Questions
AICTE	All India Council of Technical Education
ASAGS	Automatic Short Answer Grading System
GUI	Graphical User Interface
PYQt	Python Qt
NLTK	Natural Language ToolKit

Chapter 1

Introduction

Automatic grading systems have been in practice in the educational domain for many years now, but primarily questions where students have to choose the correct answer from given options such as multiple choice questions. Prior research has shown that such recognition questions are deficient as they do not capture multiple aspects of acquired knowledge such as reasoning and self-explanation. In contrast, recall questions that seek students' constructed answers in natural language have been found to be more effective in assessing their acquired knowledge.

However, automating assessment of such answers is non-trivial owing to linguistic variations (a given answer could be articulated in different ways); subjective nature of assessment (multiple possible correct answers or no correct answer); lack of consistency in human rating (non-binary scoring on an ordinal scale within a range); etc. Consequently, this has remained a repetitive and tedious job for teaching instructors and is often seen as an overhead and non rewarding.

Attempts at developing automatic evaluation systems have been made by researchers at various points using different techniques, which have found some acceptance. Descriptive answer evaluation system seeks to implement an application which will be able to evaluate the descriptive answer to a question which reduces the human effort. The proposed system is not domain specific and so can be applied to any questions related to an any domain.

1.1 Motivation

In India, the quality of education system is severely hampered due to ever-growing population and poor infrastructure. The amount of pressure education systems and teachers hold is imaginable as the number of answer sheets to evaluate is huge. The latest (2016-17) report by the All India Council of Technical Education (AICTE)states that, there are approximately more than 10330 institutions offering education in India to about more than 2 million students. The AICTE reports also states that there are 6432 engineering colleges in India with a yearly student intake capacity of over 3 million and more than 1.6 million truly enroll with only 578245 faculties. Typically every institute has 4 examinations every year; so on calculating 6.4 million answer sheets are generated.

The assessment for descriptive answers have several drawbacks such as it consumes time and human effort.The proposed system could be of great effectiveness to the educational institutes, as it saves time and trouble of checking bundle of papers.

1.2 Organization of the report

Each chapter begin with a brief introduction to its content. Chapter 2 includes Problem Definition. Chapter 3 provides a study and review of relevant literature materials in relation to the topic and chapter 4 provides the specifications related to proposed system. Chapter 5 discuss about system modelling and design. Detailed implementation of the system is described in Chapter 6. Chapter 7 analyses the performance of the system and chapter 8 includes the limitations of the system. The report on the proposed system is concluded in chapter 9.

Chapter 2

PROBLEM DEFINITION

There are a number of commercial assessment tools available on the market today; however these tools support objective type question such as multiple choice Questions or short one-line free text responses. This will assess student's depth of knowledge only at lower level of taxonomy of educational objectives. They fail to assess student's performance at higher level of taxonomy of educational objective.

So, to overcome the problems encountered the proposed system involves parsing of text and find the semantic meaning of student answer and finally assign the final scores using Natural Language Processing

2.1 Approach to NLP

In this study, Natural Language Processing is a vast area. In which, researches are still going on occasionally, improvements in such researches makes contribution of specific features to this system.

In this project the similarity between the model answer and students answer ,and the promotion of pattern matching among the students answer are processed with the NLP technique. Later on the final scores are assigned

2.2 The Infrastructure

The infrastructure created for the evaluation mainly includes :

2.2.1 [A] Data set and Text Preprocessing

Data set for descriptive answer evaluation system usually consists of several questions, instructors answer and students answers. Approximately 20 documents are used to create the answer set ,along with a separate document that includes the model description. These documents are considered as the input to proposed system Each document consists of 100-200 words with an average of 180 words. The goal of this project is to generate an accurate output that depends upon the input text documents.

Currently, input document are of plain text format. There are four main activities performed in this stage: Sentence Segmentation, Tokenization, Removing Stop Word, and Word Stemming. Sentence segmentation is boundary detection and separating source text into sentence. Tokenization is separating the input document into individual words. Next, Removing Stop Words, stop words are the words which appear frequently in document but provide less meaning in identifying the important content of the document such as 'a','an', 'the', etc.. The last step for preprocessing is Word Stemming; Word stemming is the process of removing prefixes and suffixes of each word.

2.2.2 [B] Sentence Feature

The score of each sentence are based on several sentence features and by sorting sentences based on this scores, helps to evaluate the answers.

[1] Sentence to Sentence similarity

This feature is a similarity between sentences. For each sentence S, the similarity between S and each other sentence is computed by the cosine similarity measure with a resulting value between 0 and 1. The term weight wi and wj of term t to n term in sentence Si and Sj are represented as the vectors. The similarity of each sentence pair is calculated based on similarity formula.

The score of this feature for a sentence S is obtained by computing the ratio of the sentence similarity of sentence S in instructors answer with each other sentence over the students answer.

$$S_{F\ (S)} = \frac{\sum_{t=1}^{N} W_{it} * W_{jt}}{\sqrt{\sum_{t=1}^{N} W_{it}^2} * \sqrt{\sum_{t=1}^{N} W_{jt}^2}}$$

Figure 2.1: Cosine Similarity Formula

[2] Mining Sequential Patterns

The task of finding commonalities from student answers was done in a manner similar to the sequential pattern mining problem.

The objective of this step is to extract commonly occurring patterns and quantify the notion of commonalities using support. A student answer (s) is converted to a sequence of words (wi1,wi2, . . . ,w in) by removing stop words and stemming content words to their respective base forms.

A sequential pattern (SP), p of length l, is a sequence of l tokens from si i.e. p = wi1 ,wi2 , . . . ,wil such that i1 <i2 <. . . <il. Support of p is defined as the number of student answers containing p.

Chapter 3

LITERATURE SURVEY

There are many existing implementations of Descriptive Answers Evaluation System. In this survey a few of them are included

[A] Concept Mapping Method

"Text data analysis: Computer aided automated assessment system" ,Nisarg Dave;Harsh Mistry;Jai Prakash Verma Computational Intelligence Communication Technology (CICT), 2017 3rd International Conference

Computer aided evaluation systems are generally considered objective types of questionnaires. Evaluation based on subjective answer is consider a problem under text analytics, where text answer will be compare with available correct text answer. This paper is emphasizing the issues of computer aided automated assessment and proposing a model for handling these issues. Using such mechanism a faculty can avoid the evaluation process manually. Students can be automatically graded using the application and given a summary report.After conducting many experiments the results showed only 80 percentage of expectancy.

[B] Information Extraction Method

"A review of an information extraction technique approach for automatic short answer grading" Uswatun Hasanah ; Adhistya Erna Permanasari ; Sri Suning Kusumawardani ; Feddy Setio Pribadi Information Technology, Information Systems and Electrical Engineering (ICITISEE), International Conference

The requirement for automatic short answer grading (ASAG) system brings researchers to discover more knowledge about this field. Many techniques have been developed to reach the highest accuration. It can be processed by following stages: creating data set, pre-processing, model building, grading, and model evaluation. One of the techniques which commonly used is information extraction technique. Information extraction is a technique that employing finding fact on the student answers as patterns and then matches these to the teacher answer. The accuration is pointed out in computer and human raters agreement. The goal of this paper is to present a review of several ASAG research which using information extraction technique. However, this paper does not conclude the best method which can be used for general cases.

[C] Knowledge Analysis Method

"Automated question answering system using ontology and semantic role"S. Jayalakshmi ; Ananthi Sheshasaayee, Innovative Mechanisms for Industry Applications (ICIMIA), 2017 International Conference

Semantic similarity is an essential part for question answering, it is used various fields such as Artificial Intelligence, Natural Language Processing, information retrieval, Document Retrieval and Automatic evaluations. This paper mainly focuses on similarity measure based on the posted query, and finding the appropriate meaning between the words. Accessing an accurate answer from the web document is challenging task.

The proposed approach is used to analyze and measuring the similarity between the words. It presents the Web And semantic knowledge-Driven automatic question answering system (WAD). It encompasses three phases to enhance the performance of QA system using the web as well as the semantic knowledge. Initially, the WAD approach determines the user query, query expansion technique and entity linking method. The ontology based information is used in WAD to rank the answers and experimental results provide the result with high accuracy than the baseline method.The drawback of this paper is that we need to require rich knowledge base for the evaluation purpose.

[D] Document Similarity Based Technique

"Automatic Short –Answer Grading System (ASAGS)" P.Selvi, Dr.A.K.Bnerjee

Automatic assessment needs short answer based evaluation and automated assessment. Various techniques used are Ontology, Semantic similarity matching and Statistical methods. An automatic short answer assessment system is attempted in this paper. Through experiments performed on a data set, we show that the semantic ASAGS outperforms methods based on simple lexical matching; resulting is up to 59 percent with respect to the traditional vector-based similarity metric.

Chapter 4

SYSTEM REQUIREMENTS

4.1 Operating Environments

4.1.1 Software constraints

Operating system	Windows 10
Programming languages	Python
GUI	PyQt
Editing tools	Python IDLE(version 2.7.6), Notepad++ (version 6.7.5)

4.1.2 Hardware constraints

Processor	32 bit Intel Processor
Main memory	1GB

4.2 Functional Requirements

- User Input : Inputs the documents,that is the answers of each student which he/she wants to evaluate.
- User Output : Evaluated result corresponding to each student is obtained.

4.3 Non Functional Requirements

4.3.1 Performance

Performance requirements are concerned with quantifiable attributes of this system such as response time, throughput, availability and accuracy. The system has high performance with very less delay. The performance mainly depends on the size of the input

text document, number of paragraphs in the document, relation between sentences, length of sentence.

4.3.2 Reliability

Reliability species the capability of the software to maintain its performance over time. Documents is not domain specific,but the is specific to answers written in English language. Also the system does not evaluate answers with diagrams and mathematical expressions,otherwise it effects the whole reliability of the system.

4.3.3 Portability

The Document summarizing software can be run in any platform that supports .exe file. Software developed for the evaluation technique is portable if needed, this software can be installed and run on any PC or Laptops.

Chapter 5

SYSTEM MODEL AND DESIGN

5.1 Architecture of the System

The proposed system architecture is shown below:

Figure 5.1: General System Architecture

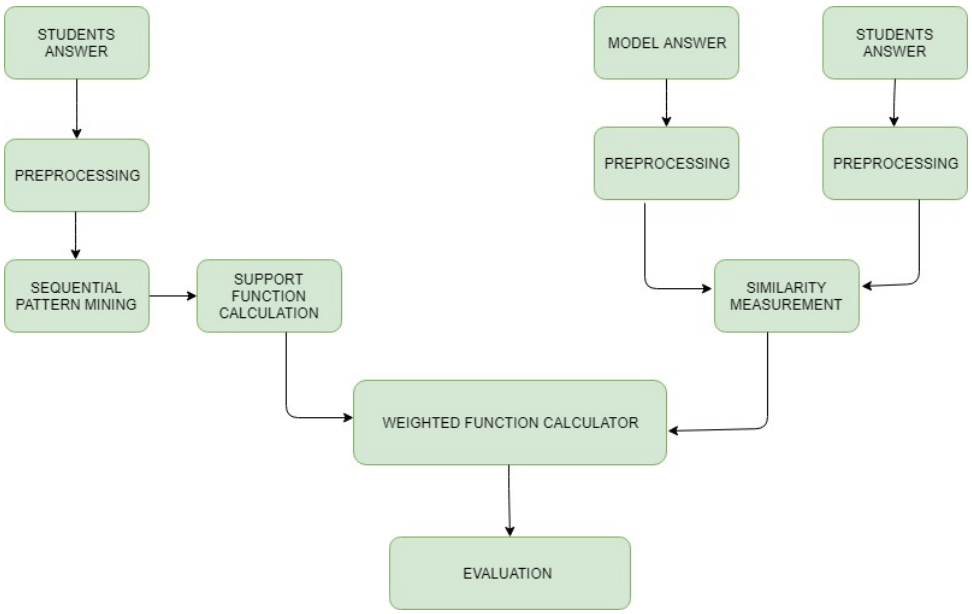

Figure 5.2: General Architecture of the Evaluation Module

5.2 Detailed system design

The system design starts with the user who selects the input text document. Then, the evaluation procedure follows the steps below.Mainly there are five modules

- Preprocessing Module
- Pattern Mining Module
- Similarity Comparison Module
- Weighted Function Calculator Module
- Scoring Module

5.2.1 Preprocessing Module

The input text document is changed into a normalized form using the preprocessing. There are four main activities performed in this stage:

- Segmentation
- Tokenization
- Removing Stop words
- Word Stemming

Sentence segmentation is boundary detection and separating source text into sentence. Tokenization is separating the input document into individual words. Next, Removing Stop Words, stop words are the words which appear frequently in document but provide less meaning in identifying the important content of the document such as ‘a’, ‘an’, ‘the’, etc.. The last step for preprocessing is Word Stemming; Word stemming is the process of removing prefixes and suffixes of each word.

5.2.2 Pattern Mining Module

Tokens are generated after preprocessing. The sequential pattern is generated from the tokens of each sentences. The count for each obtained pattern is also determined. Most frequently occurred patterns are obtained based on the count.

5.2.3 Similarity Comparison Module

Given two sentences, the measurement determines how similar the meaning of two sentences is.The higher the score the more similar the meaning of the two sentences.Cosine based similarity approach is used to calculate the similarity between the input answer and the model description.To compute cosine similarity between two sentences s1 and s2, sentences are turned into terms/words, words are transformed in vectors .Each word in texts defines a dimension in Euclidean space and the frequency of each word corresponds to the value in the dimension.

$$\text{Cos}(s1,s2) = s1 \,.\, s2 \,/\, ||s1||\,||s2||$$
$$\text{Where } s1 \,.\, s2 = \sum_{i=1}^{n} s1_i \, s2_j$$

Figure 5.3: Cosine Similarity Function

5.2.4 Weighted Function Generation Module

A novel approach that combines the pattern mining unsupervised technique with the similarity measurement is introduced.The input text document is processed using this approach.Pruning on the data set based on the novel strategy determines the empirical formula to be used.Thus we give equal weight age to both the method.

5.2.5 Scoring Module

The given answers are allotted with a score according to this novel strategy.Identifying both commonalities and similarity between the students answers scores are assigned.

Finally the result of the evaluated answer is obtained and displayed through the Graphical User Interface.

Chapter 6

IMPLEMENTATION

The proposed system is implemented according to the Figure 5.1. It consists of 3 parts namely, the Graphical User Interface, Coding of Algorithm and Display output.

- User will open the System user interface.
- Click on the Select file button.
- Select the answer using the opened window.
- Run the file from the drop down menu File Run.

6.1 Graphical User Interface

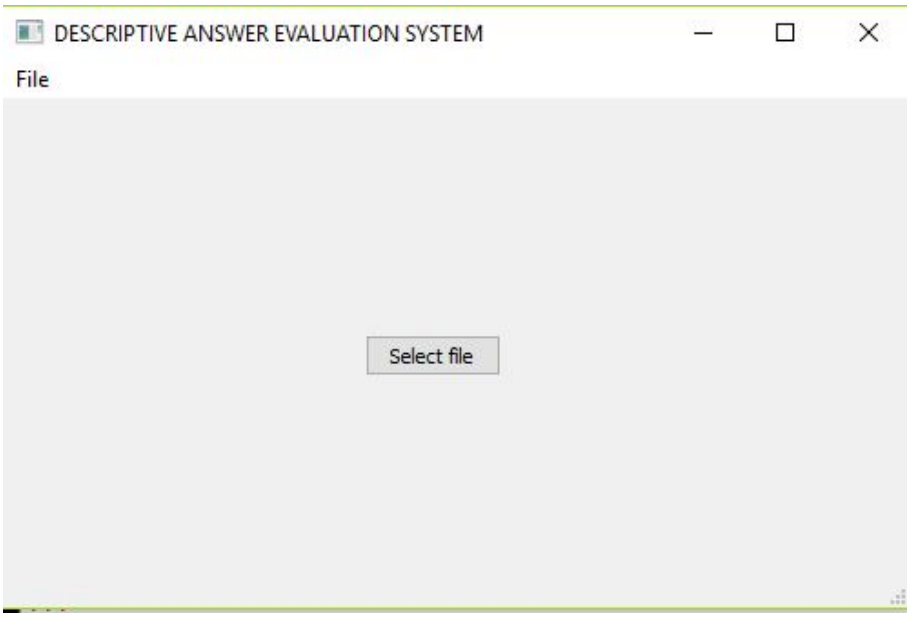

Figure 6.1: Graphical User Interface Window

GUI is a type of interface that allows users to interact with electronic devices through graphical icons and visual indicators. The actions in a GUI are usually performed through direct manipulation of the graphical elements. Here the project uses PYQT for creating the user interface.

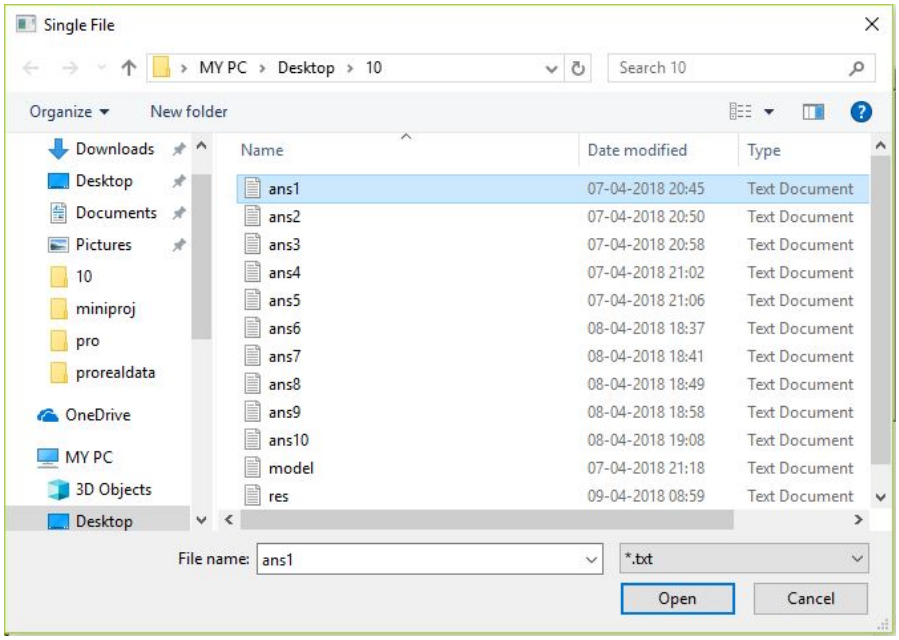

Figure 6.2: Selecting File

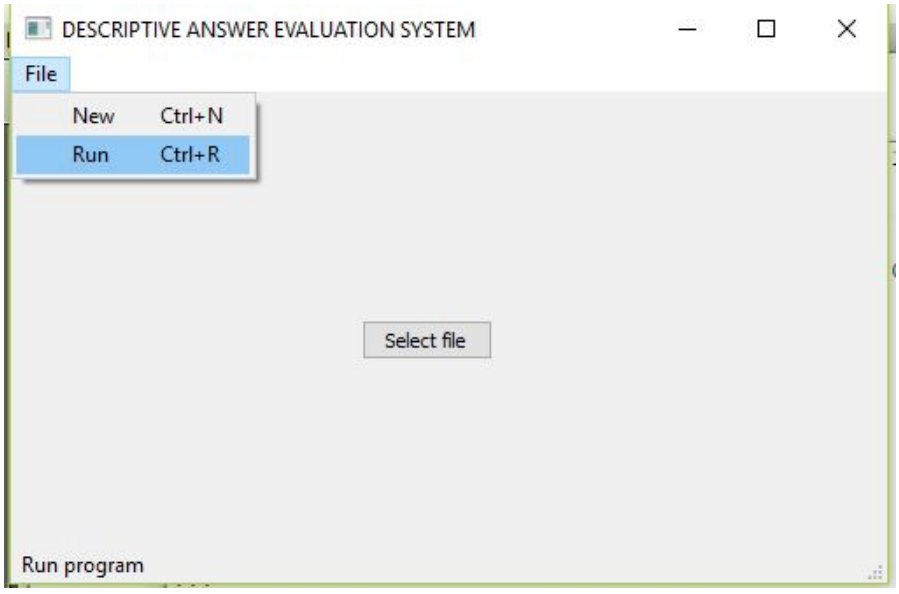

Figure 6.3: Running File

6.2 Coding of Algorithm

The algorithm is implemented using Python. Algorithm is purely based on Python's Natural Language Tool Kit (NLTK).

Why NLTK, is by considering several advantages over the older natural language processing tools is basically the efficiency. It was created to support education. Most advanced techniques like Tokenization, POS tagging etc are included in NLTK. Another advantage of NLTK as a programmer is the efficient scripting employed in NLTK

Over the past several years, however, mainstream open source software libraries like the Natural Language Toolkit for Python (NLTK) have emerged to offer a collection of high-quality reusable NLP functionality. NLTK enables researchers and developers to spend more time focusing on the application logic of the task at hand, and less on debugging an abandoned method for sentence segmentation.

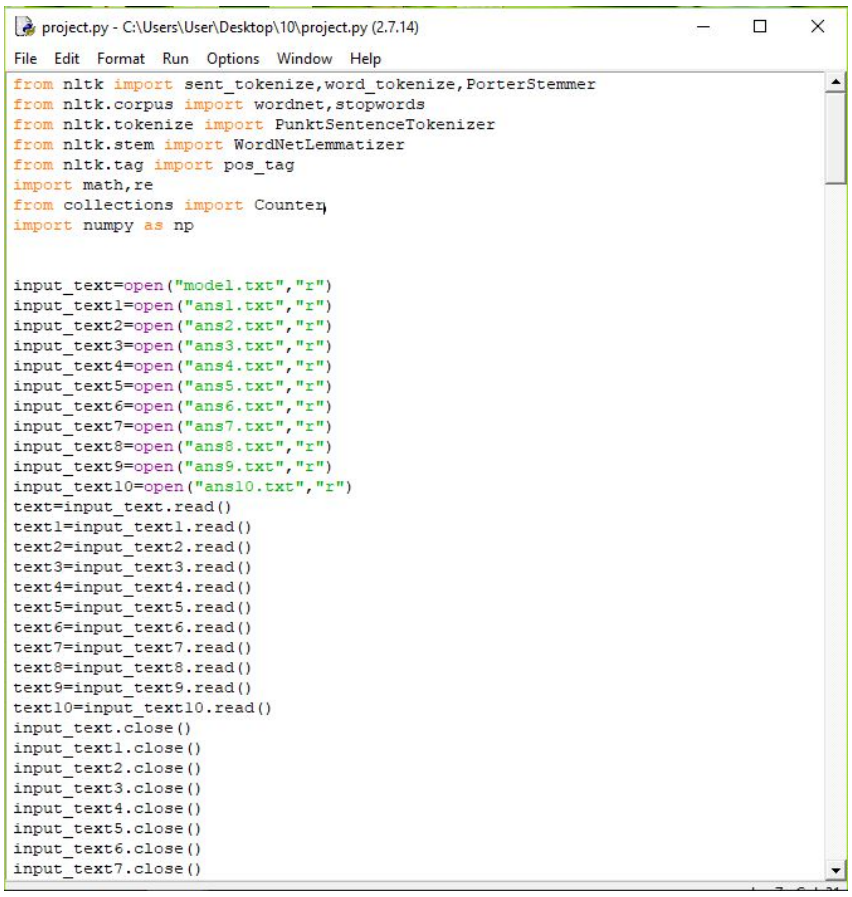

Figure 6.4: Sample Code

6.3 Display Output

A question "Software Process-Way of Producing Software-A Framework.Importance of Software Process".which carries 3 marks.10 students were evaluated by the proposed system.The sample input and sample output are shown below.

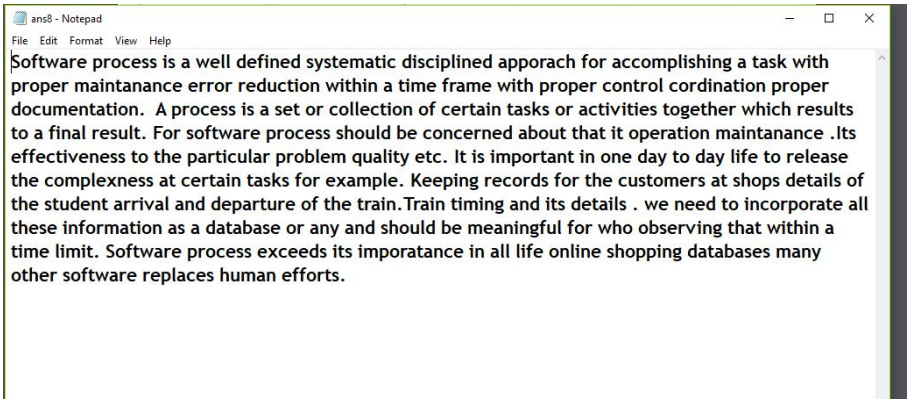

Figure 6.5: Sample Input

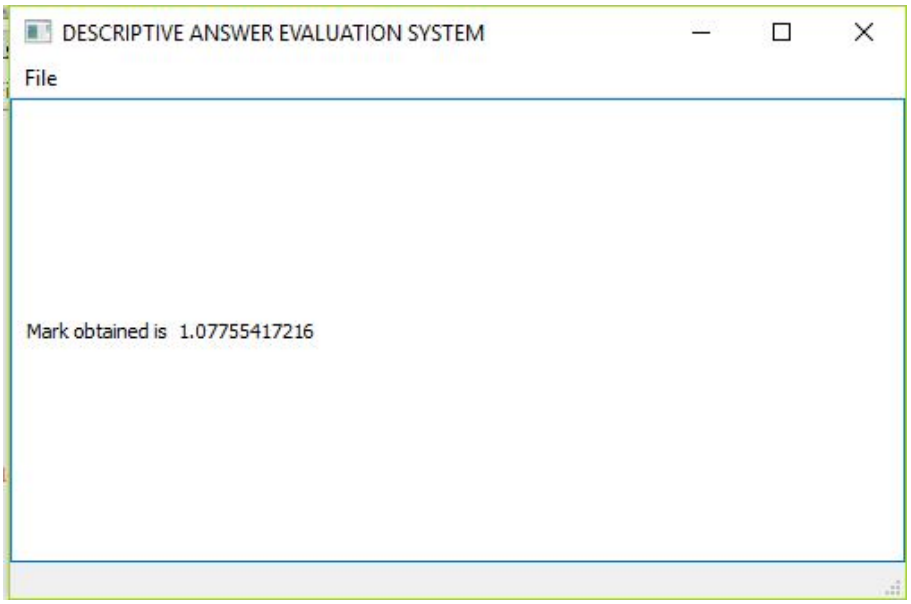

Figure 6.6: Sample Output

Chapter 7

PERFOMANCE ANALYSIS

Performance Analysis deals with the measurement of response time and throughput of the proposed software and comparing it with existing evaluation techniques.A comparison of marks obtained between the most relavant mode of evaluation methodologies are depicted below

Students Answer	Instructor provided mark	Similarity based evaluation	Common pattern based evaluation	Proposed evaluation system(not equally distributed)	Proposed evaluation system(equally distributed)
Answer 1	0.5	1.17	0.79	1.07	0.98
Answer 2	1	1.02	0.99	0.93	0.84
Answer 3	1.5	0.87	1.22	1.26	1.65
Answer 4	2.5	2.71	0.67	2.49	2.26
Answer 5	2	1.06	1.51	1.17	1.29
Answer 6	2.5	1.22	0.12	1.15	1.08
Answer7	1.5	1.20	1.17	1.29	1.381
Answer 8	1	0.79	0.52	0.90	1.01
Answer 9	2	1.09	0.87	1.55	2.01
Answer10	1.5	1.12	1.14	1.32	1.51

Figure 7.1: Comparison Of Marks Obtained

Here , the marks obtained for 10 students after the different evaluation strategies are shown.From above given table we can find that,when evaluated by the proposed model has greater accuracy generated than other method .Thus the proposed system could be of great effectiveness to the educational institutes, as it saves time and the trouble of checking bundles of papers.

From the previous methods we can conclude that ,the proposed system must be simpler technique for the evaluation of students answer,and the results shown are satisfactory.Much pruning on the data set is required for get an accurate and perfect result.Researches related to this techniques are still going on and the new features are taken into account for improving the deepness in relating sentence and obtaining well formed, user convenient, more accurate result.

Chapter 8

LIMITATIONS AND EXTENSIONS

8.1 Limitations

The main limitations of the proposed system are:

- Specific to English language only.
- Does not evaluate answers with diagrams and mathematical expressions.
- Polarity of the sentences is not considered.
- May be incorporated in online examinations.

8.2 Extensions

The proposed system can be further extended as:

- Implementing an option for inputting any type of documents(pdf,docx, etc).
- Implementing additional features that helps to evaluate answers written in other languages also.
- Implementing an option that helps to evaluate answers with daigrams and mathematical expression

- Extending the GUI by adding option in drop down menu for user friendly environment.
- Extending the feature that consider the polarity of the sentences.

Chapter 9

CONCLUSION

In school , colleges and various educational institutes examinations play a very vital role.The automatic assessment of students nowadays only include multiple choice questions,which are efficient in checking the students aptitude skills,in contrast they fail to determine the theoretical knowledge a student possesses.Research work done in this paper was in highlighting these issues and proposing a system for handling them.The proposed system attempts to evaluate descriptive answers.

The purpose of the system is to evaluate students a higher level considering assessment of descriptive type questions consisting of multiple sentences.The system excludes human efforts and errors and also saves time and resources. This system could be of great effectiveness to the educational institutes, as it saves from the trouble of checking bundles of papers.

BIBLIOGRAPHY

[1] *"Wisdom of Students: A Consistent Automatic Short Answer Grading Technique".* Shourya Roy Sandipan Dandapat Ajay Nagesh Narahari Y. 2016

[2] *"A Perspective on Computer Assisted assessment Techniques for Short Free-Text Answers".*In Proceedings of the International Conference on Computer Assisted Assessment (CAA), Shourya Roy, Y. Narahari, and Om D. Deshmukh.2015.

[3] *"Identifying Patterns for Short Answer Scoring using Graph-based Lexico-semantic Text Matching."*Lakshmi Ramachandran, Jian Cheng, and Peter Foltz. 2015.

[4] *"Automated Scoring of a Summary Writing Task Designed to Measure Reading Comprehension"* .Nitin Madnani, Jill Burstein, John Sabatini, and Tenaha OReilly. 2013.

[5] *"Moving Towards a Fully Automatic Knowledge Assessment Tool".*Christian Gutl. 2012. 3(1):111.

[6]] *" Automated assessment of short free-text responses in computer science using latent semantic analysis".*Richard Klein, Angelo Kyrilov, and Mayya Tokman. 2011.

[7] *"Natural Language Processing with Python"*, by Steven Bird, Ewan Klein, Edward Loper.2011

[8] *"Python for Software Design: How to Think Like a Computer Scientist"*, Allen B. Downey, Je Elkner and Chris Meyers, Green Tea Press ACM Trans. Asian Lang. Inf. Process., vol. 2, no. 3, pp. 193–218, Sep. 2011.