DESCRIPTIVE ANSWER EVALUATION SYSTEM

MIDHILA V K
ROLL NO :11
S2 MTECH CSE

GUIDED BY **RAJI R PILLAI**
PROF. OF COMPUTER SCIENCE

RAJIV GANDHI INSTITUTE OF TECHNOLOGY
KOTTAYAM

April 9, 2018

CONTENTS

PROBLEM STATEMENT

- Natural language processing (NLP) is the ability of a computer program to understand human language as it is spoken

- My project is to evaluate the descriptive answers using Natural language processing (NLP) and to improve the accuracy and efficiency.

INTRODUCTION

- Automatic assessment is preferred to Manual Assessment .
- The idea of using computers to assist learning process has surprisingly changed the field of learning system.
- Descriptive answers assessment remained a tedious job for teaching instructors.
- A computational technique for automatically grading student answers in natural language is introduced.

OBJECTIVES

- Provide instructor independent evaluation for descriptive answers.
- Generic, not domain specific.
- Excludes human efforts and saves time and resources.
- Calculates the score and provides results instantly.

LITERATURE SURVEY

Paper	Year	Author	Limitations
Text Data Analysis: | Computer Aided Automatic Assessment System	2017	Nisarg Dave Harsh Mistry Jai Prakash Varma	After conducting many experiments the accuracy is only up to 80% of human expectancy.
A Review of an Information Extraction Technique Approach for Automatic Short Answer Grading	2017	Uswatun Hasanah Adhistya Erna Perumanasari Sri Suning Kusumawardani Feddy Setio Pribadi	This paper does not conclude the best method which can be used for general cases. System does not find the semantic meaning of words on students answer.
Automated Question Answering System based on Ontology and Semantic Role	2017	S.Jayalakshmi Ananthi S	Requires huge knowledge base. Document similarity based technique is used.
Automatic Short Answer Grading System	2010	P.Selvi Dr A.K.Banarjee	Results up to only 59 % with respect to the similarity based technique.

EXISTING SYSTEM

- Prior work comprises extensive intervention from instructors and susceptible to human errors..
- Automatic grading systems have been in practice primarily for multiple choice questions (MCQs).
- Evaluation model based on similarity between documents is used.
- Grading done using concept mapping.
- Completely unsupervised technique where commonalities among the students answer used to evaluate.

LIMITATIONS OF THE EXISTING SYSTEM

- Evaluation done by human graders leads to great variation in score awarded to students.
- Requires rich knowledge bases.
- No standardization around how model answers are written.
- Based on the perspective of the instructor model description is given.
- Does not work properly when multiple correct answers are given.
- Performs the worst when all answers are wrong in the same manner.

PROPOSED SYSTEM

- Proposed system use an approach that combines the pattern mining unsupervised technique with the similarity measurement for the evaluation of descriptive answers.
- Based on this finding the proposed technique requires:
  - Preprocessing Module
  - Pattern Mining Module
  - Similarity Comparison Module
  - Weighted Function Generation Module
  - Evaluation Module

PROPOSED SYSTEM CONTD..

PREPROCESSING MODULE

- Complex statements are converted into simple statements.
- The given statements are converted to a sequence of words.
- Stop words removal and stemming technique helps to convert into their respective base forms.

PATTERN MINING MODULE

- Tokens are generated after preprocessing.
- The sequential pattern is generated from the tokens of each sentences.

PROPOSED SYSTEM CONTD..

The count for each obtained pattern is also determined.

Most frequently occurred patterns are obtained based on the count.

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.

PROPOSED SYSTEM CONTD..

- 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.

- The cosine similarity is measured by using the word vectors as in below equation:

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

PROPOSED SYSTEM CONTD..

WEIGHTED FUNCTION GENERATION MODULE

- A novel approach that combines the pattern mining unsupervised technique with the similarity measurement is introduced.
- Pruning on the data set based on the novel strategy determines the empirical formula to be used.

SCORING MODULE

- The given answers are evaluated according to this novel strategy.
- Identifying both commonalities and similarity between the students answers scores are assigned.

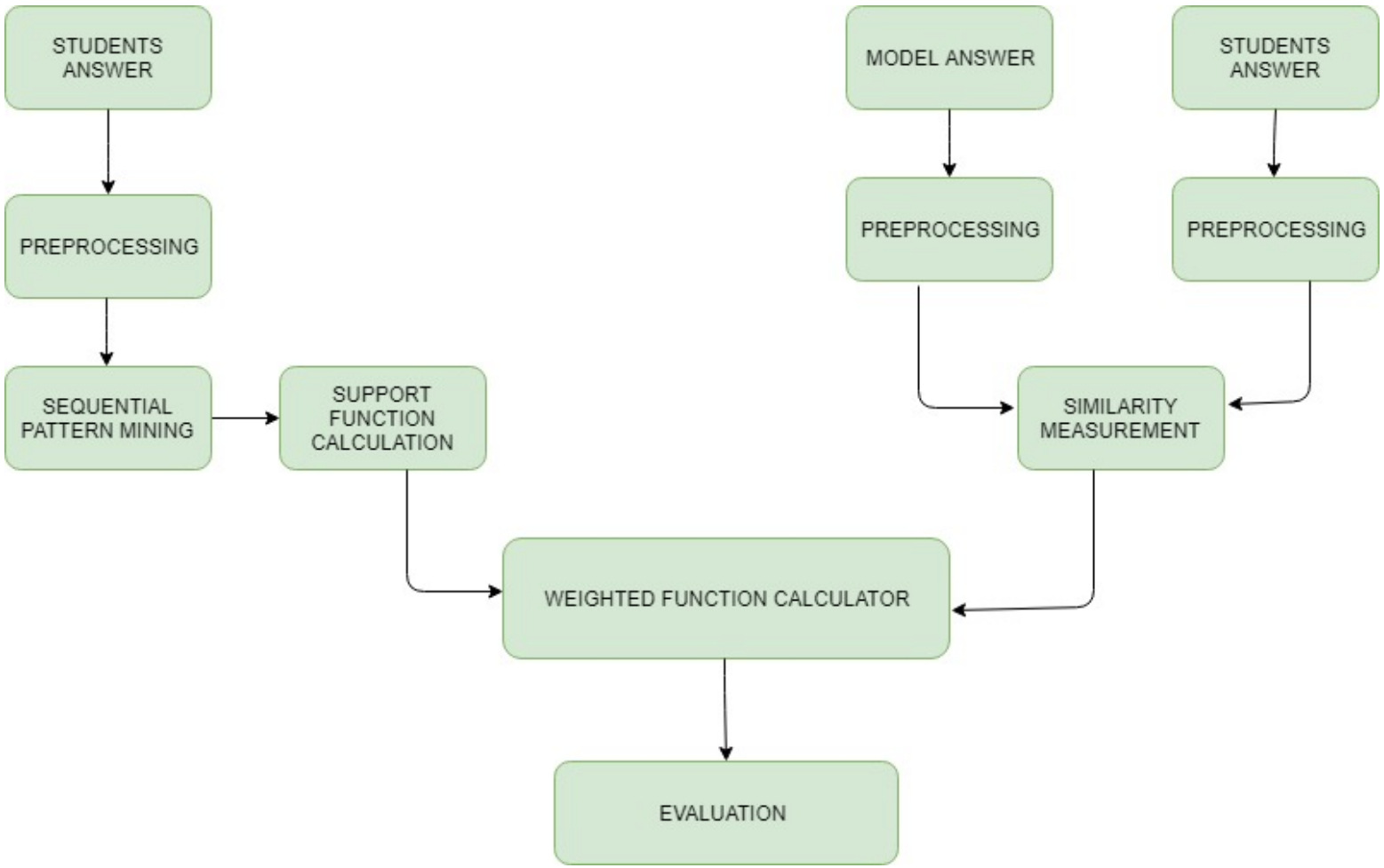
General System Architecture

Figure: Basic Architecture

- Select an input text document
- Evaluation procedure is done.
- Output is obtained on GUI.

DESIGN AND IMPLEMENTATION CONTD...

Figure: Detailed View of the Evaluation Module



DESIGN AND IMPLEMENTATION CONTD...

DEVELOPMENT TOOLS

- Platform and Editing tools : Windows 10 , Python IDLE(version 2.7.6) , Notepad++ (version 6.7.5)
- Test Data : 15-20 UG and PG students answers.
- Programming languages : Python
- GUI : PyQt

IMPLEMENTATION

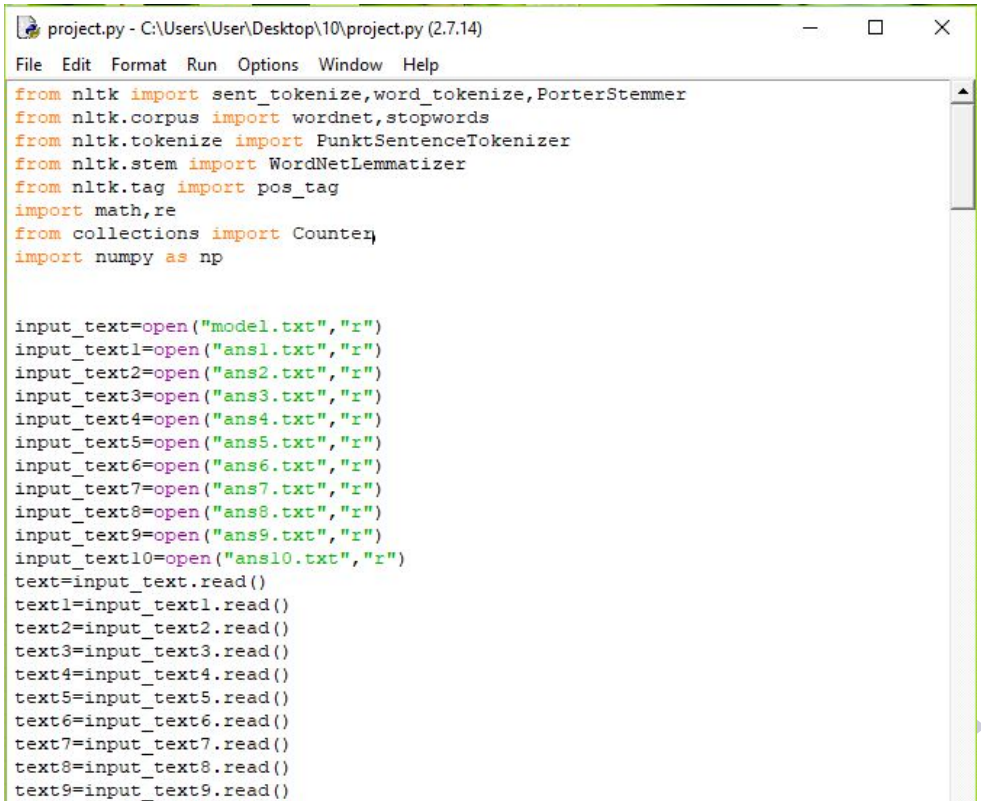
- Coding of Algorithm , the GUI, and Display output the main parts of the proposed system.

DESIGN AND IMPLEMENTATION CONTD...

Coding of Algorithm

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

Figure: Sample Code

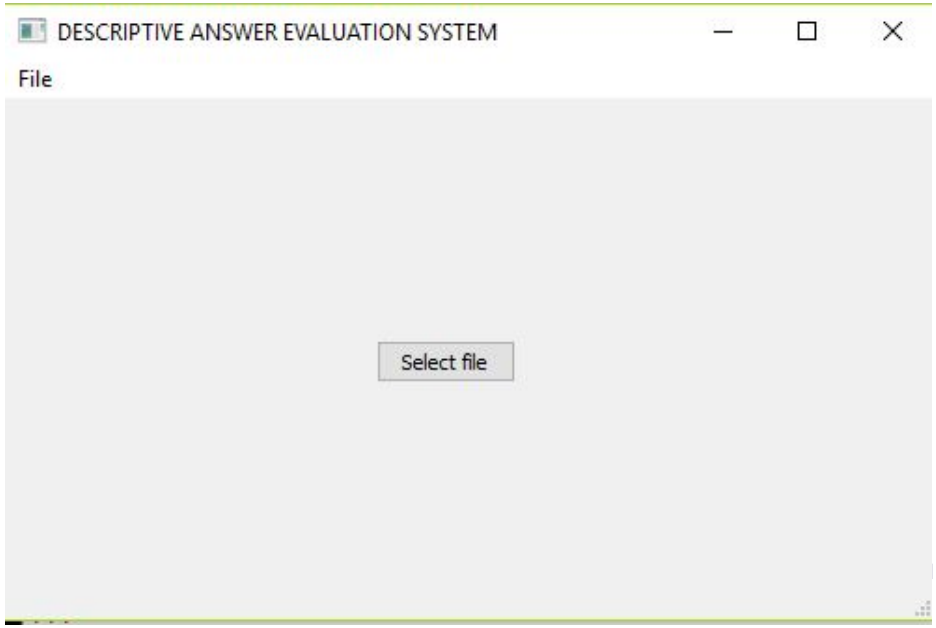


DESIGN AND IMPLEMENTATION CONTD...

Graphical User Interface

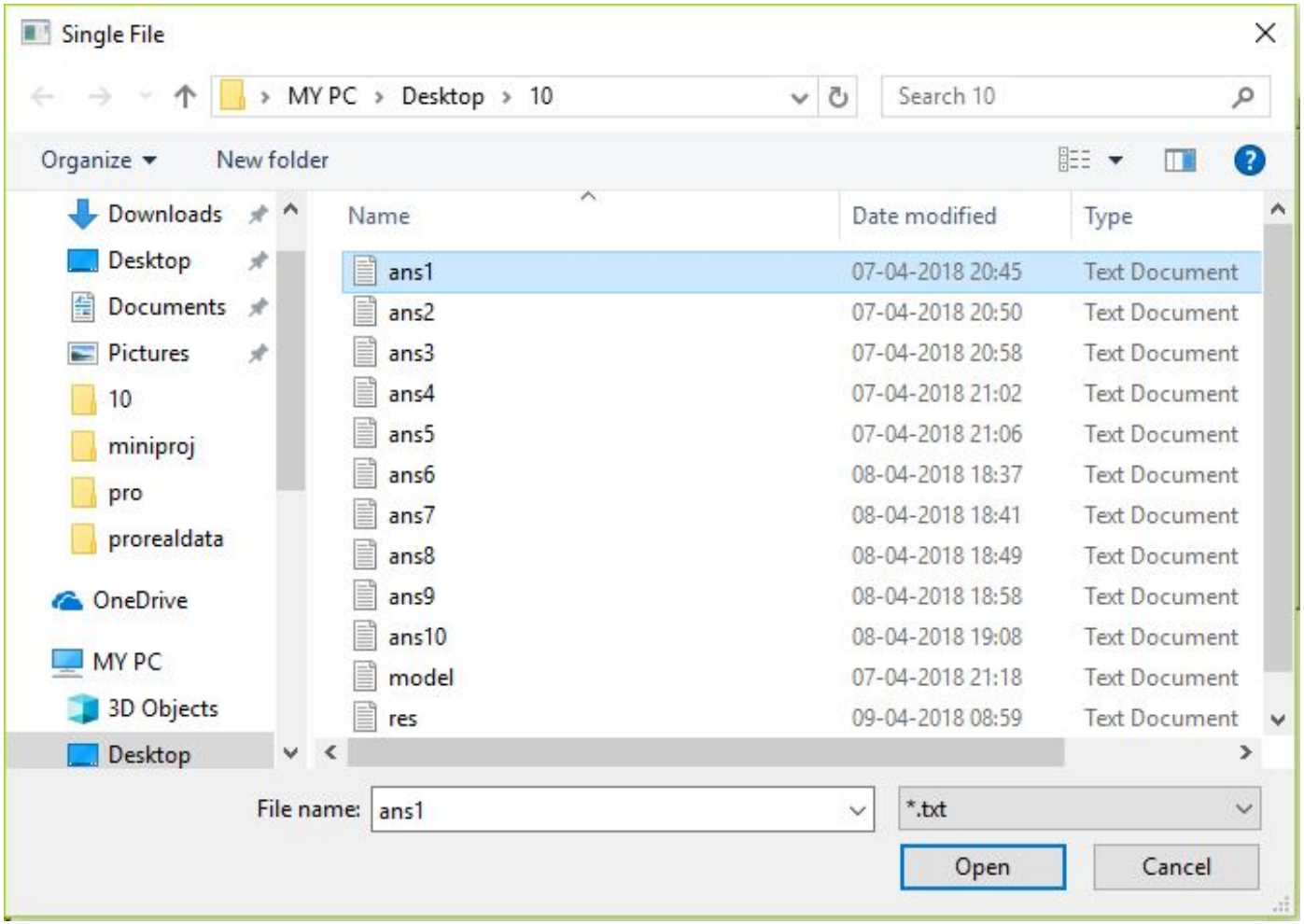
- The actions in a GUI are usually performed through direct manipulation of the graphical elements.
- Here the project uses PYQT for creating the user interface.

Figure: GUI Window



DESIGN AND IMPLEMENTATION CONTD...

Figure: Select File

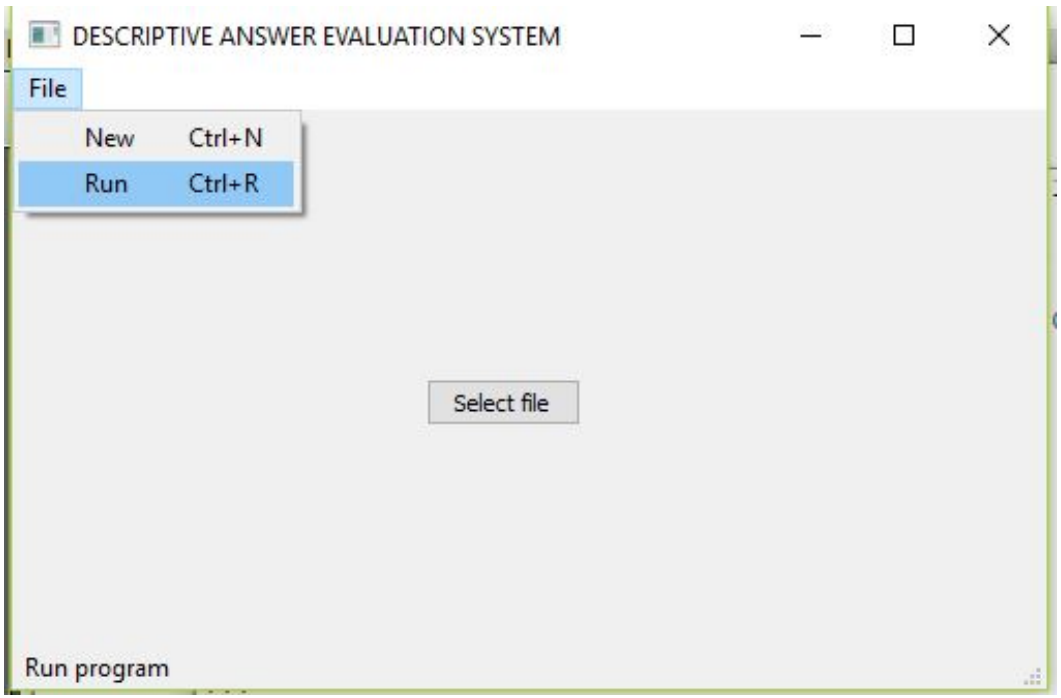


DESIGN AND IMPLEMENTATION CONTD...

Display Output

- The marks obtained for the particular answer is displayed, with the help of GUI.

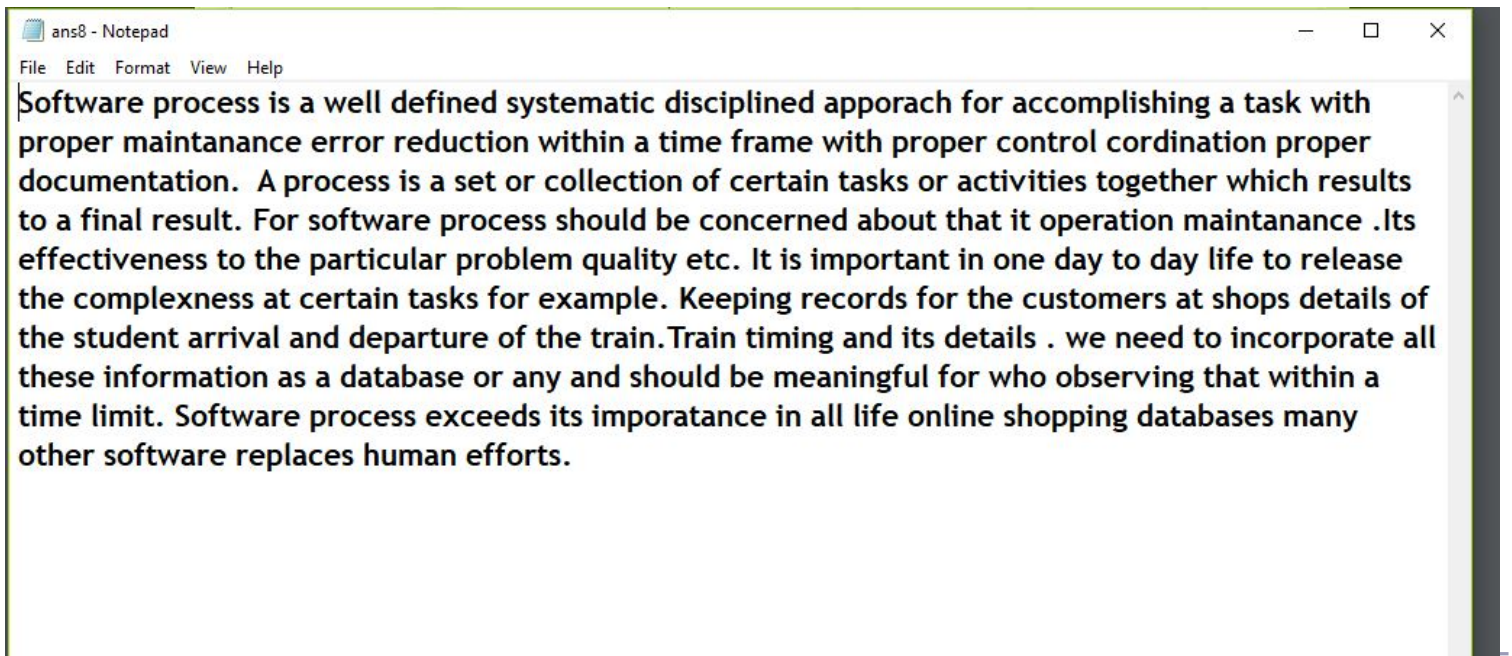
Figure: Run File



RESULTS OBTAINED

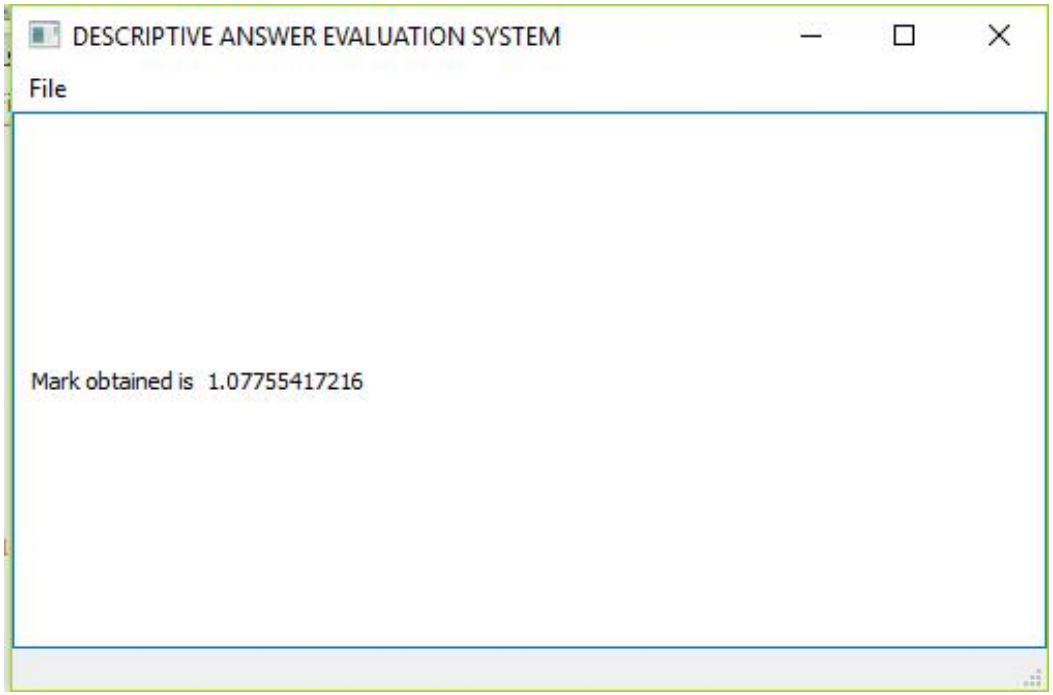
- A question Software Process-Way of Producing Software-A Framework.Importance of Software Process.which carries 3 marks.
- 10 students were evaluated by the proposed system.
- The sample input and sample output are shown below

Figure: Sample Input



RESULTS OBTAINED

Figure: Sample Output



TESTING AND VALIDATION

Figure: Comparison between different modes of evaluation

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

TESTING AND VALIDATION CONTD...

- The result of assessment between system and human can be considered to make design for final grading.
- Much pruning on the data set is required for get an accurate and perfect result.

CHALLENGES

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

CONCLUSION

- Evaluate students a higher level considering assessment of descriptive type questions consisting of multiple sentences.
- The system excludes human efforts and saves time and resources along with the trouble of checking bundles of papers.
- 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.
  - Evaluate answers with diagrams and mathematical expression an checking the polarity .

REFERENCES

[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

REFERENCES CONTD..

[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.


Thank You


shutterstock
IMAGE ID: 1026695770
www.shutterstock.com