Comprehensive Study of Textual Processing and Proposed Automatic Essay Evaluation System

[1]Aqsa Zahid Mughal, [2]Asia Samreen

[1] *Dept. of Computer Science, FAST National University of Emerging Sciences, Karachi, Pakistan*
[2]*Dept. Of Computer Science, Bahria University, Karachi, Pakistan*

[1]`aqsa.zahid@nu.edu.pk`
[2]`asiasamreen.bukc@bahria.edu.pk`

Abstract—From last 50 years the work has been conducted on building such systems that can have capabilities by which it can evaluate or check like a human tutor or even better than a human tutor, this is the goal of Automatic Essay Evaluation System. Grading essays is one of the most tedious and time-consuming task, subjectivity of topic, bias nature of human grader are also key points which affect the process of assessing, this becomes initial motivation for advancing the method of assessment resulting human written essays are now assessed by humans and also by computer system Automatic Essay Evaluation System. In this paper a detailed study is conducted on AEE systems and its building approaches such as text mining and text processing for the purpose to bring the exposure to this research field as technology upgrades, it has become more commercialized raising to the most important problem in the development of AEE system, the lack of its exposer. This paper also addresses our approach replicating all possible qualities of existing AEE system for the students and teachers of Pakistan.

Keywords—Text mining, AEES, NLP, Essay Evaluation

I. INTRODUCTION

Computer-based testing and assessment is not actually a myth but a reality since from 1959 and it is influencing our current educational systems more than anything else and institutes from all over world have transferred their current conventional techniques of assessing into CBAS (computer-based assessment systems) replacing paper-based tests and more new systems are also in way of introducing for the betterment of existing systems. Currently, more than 40 commercial systems are available. CBAS has grown quite rapidly past few years as leading universities have adopted and increase in the number of students.

Mostly the systems are based on:
- Short question answers.
- Objective type questions.
- Multiple choices.

But collectively many researchers of education field agreed that essay writing is a more effective way to assess the student capabilities of skills and hold on knowledge as essay writing requires high order thinking. Essay writing is one of effective test which tests the learning outcomes, check the ability to recall, how the ideas are organized and integrated.

A. Essay Writing:

Essays are described as short literary composition on a theme or subject, usually in prose and generally analytic, speculative, or interpretative [1].

B. Initial Problem Statement:

But grading essays is time-consuming as the content is more written as compare to other tests. Around 30 to 40% of subject instructors spend their time in grading the students. Which is great consumption and as humans are not machines, inconsistency, tiredness and loss of interest effect in the final grading. Also, human beings hold bias nature and subjectivity of essay is considered as a problem in grading essays and due to this, there is seen a variation in assessing and awarding grades resulting less interest and disappointment by students.

C. AEES (Automatic Essay Evaluation System):

To encounter the initial problems as we have mentioned above, the automatic evaluation system was introduced which aimed to provide help and assistance to teachers, consistent in assigning grades to the student as compared to human beings. Automated essay evaluation is described as a task as the process of evaluating and scoring the written prose via computer programs. Shermis and Burstein [2] AEE is even considered itself as a field in which incorporates research from many other disciplines like cognitive psychology, computer science, educational measurement, linguistic methods.

But at the same time, this merging technology has faced hard criticism on the concept of replacing a human teacher with the computer system and still some researchers from various fields hold the belief that does not except automatic

evaluation. Due to the invention of internet and development in various fields of computer science like artificial intelligence, NLP, machine learning, neural network and software like word processing make it possible to make AEE systems reality. Not only providing a specific grade but also instance feedback to student; the applications and usage of AEE systems are vast. Nowadays, AEE became support and intellectual technology in education. The AEE systems are used in combination with human graders in different high-stakes assessments such as:

- Graduate Record Examination (GRE).
- Test of English as a Foreign Language (TOEFL).
- Graduate Management Admissions Test (GMAT).
- SAT, American College Testing (ACT).
- Test of English for International Communication
- (TOEIC).
- Analytic Writing Assessment (AWA).
- No Child Left Behind (NCLB).
- Pearson Test of English (PTE).

II. RELATED WORK

Ellis Batten Page in 1966 introduces the automatic assessing by mean of computer systems and worked in his idea. In 1968, he presented the outcome of his 2-year work in the form of PEG (project essay grader).

In 1982, a UNIX program called Writer's Workbench could offer punctuation, spelling, and grammar advice. The practical approach begins with the 1990s, e-learning systems were commercially coming into been and accepted. Another reputable system after PEG was IEA (Intelligent essay assessor) used to grade essays of undergraduate courses in 1997, develop by Peter Foltz and Thomas Landauer. It is now owned by Pearson Educational Technologies used for different national tests.

Many state's different departments of education of the US utilizes CRASE which was developed by Howard Mitzel and Sue Lottridge, Pacific Metrics. The system is also in use in US Department of Education-funded Enhanced Assessment Grant.

In the initial phase of development, only limited methods of language processing and algorithms were available which includes spell checks, word counts, grammatical errors; the first generation automated assessing were developed using this and syntax, lexical properties of the essay were evaluated. For scoring regression techniques of machine learning were implemented. With the time, there have been prominent developments in lexical analyses and advances in knowledge representation extending the current AES scoring model to semantic meaning.

More linguistic features can be extracted from essay content. [1, 2, 3]

A. Existing System:

Table 1. Current Existing Systems.

Developing Company	System Name	Technique/ Methodology
e-rater	ETS	NLP
Pacific Metrics	CRASE	NLP
University of Maryland	BETSY	Statistical
AIR	AutoScore	NLP
PKT	IEA	LSA, NLP
OBIE based AEE	University of Oregon	OIE, DL
CTB	Bookette	NLP
Measurement Inc.	PEG	Statistical
MetaMetrics	Lexile	NLP
Curtin University	OzEgrader	NLP
Blue Wren Software Pvt Ltd	Markit	NLP, PMT
University of Ljubljana	SAGE	NLP
Bangladesh University	GLSA based AEE	GLSA
Paperless School	PS-ME	NLP
Robert Gordon University	SEAR	Statistical

III. SCIENTIFIC APPROCH FOR BUILDING AEES

The most prominent and fascinating fields of computer science are AI Artificial Intelligence and Natural Language Processing. If we look at the main objective of AI that it is quite clear that AI is meant to create intelligent systems exhibiting intelligence as human have. The applications created using AI can be categorized into the following ways:

- Building such applications which do such tasks which required compulsory intelligent computers.
- It includes computer robots, real-world simulators, games, weather forecasting etc.
- Building such applications replacing humans or making human work easier.

Automatic information processing like speech recognition, helpdesks, computer vision, and natural language processing.

A. Text Mining

One of the technologies of Artificial Intelligence which is also known as Text Analysis. Text mining is defined as —the non-trivial extraction of hidden, previously unknown, and potentially useful information from (large amount of) textual data" [4]. It involves natural language text from which meaningful information is attempted to extract for a specific situation. It is processing of unstructured textual information for the purpose to discover the new high-quality information.

The unstructured data have the nature of ambiguity making it hard for processing as compared to structured data.

Typical tasks in text mining are text clustering, text categorization, entity recognition, sentiment analysis, and summarization. Text mining systems are the one that analyses the collection of large natural language text for the extraction of useful information and detection of a pattern of lexical or linguistic usage. Text analysis involves information retrieval information extraction, data mining techniques including association and link analysis, visualization and predictive analytics.

The goal is, essentially to turn text (unstructured data) into data (structured format) for analysis, via the use of natural language processing (NLP) methods. [5]

Text mining is also known as Intelligent Text Analysis, Text Data Mining and Text Data Mining or Knowledge-Discovery in Text (KDT) as text mining is similar to data mining.

Variations are present between these two interdisciplinary fields as data should be in the form of structured for applying data mining whereas text mining can handle semi-structured or unstructured text data. In spite of being a difficult field, most of the research and development is conducted in data mining. Text mining hindrance is that natural languages are meant for humans not machines and making machines to comprehend the context and accurate prediction is something which is a long way.

B. NLP (Natural Language Processing):

To understand the meaning and analysing the essence from the human language the computational technique used is NLP (natural language processing). NLP is used to perform a task by organizing and providing structure to knowledge; task includes automatic summarization, sentiment analysis, topic segmentation, relationship extraction etc.

The field of science deals with such models and techniques including algorithms which can be used in analysing and processing the essays, also can interpret. NLP techniques are mainly used to analyse the free text by which a computer machine attempts to grasp the human language. It is used for text mining and automated question answering.

The techniques classified for automatic scoring of free text responses in three main kinds: Statistical, Information Extraction and Full Natural Language Processing. The Statistical approach, when it is only based on keyword analysis, has usually been considered a poor method, given that it is difficult to tackle problems such as synonymy or

polysemy in the student answers, it does not consider the order of the words and it cannot deal with lexical variability. On the other hand, a full-text parsing and semantic analysis are hard to accomplish and very difficult to port across languages [6].

Information Extraction (IE), is in the middle of the Statistical and the full NLP approaches. The only requirement is shallow NLP and no need for in-depth analysis. Information extraction techniques aim to acquire structured information for text by identifying the entities present in the text. Also, dependencies are determined between concepts. In figure 1, Pattern Matching is shown as a technique commonly used for IE. It consists in looking for specific Information in the student's answer to fill in the template that the human experts have previously done. The filled template is compared to the model to calculate the final score. [6]
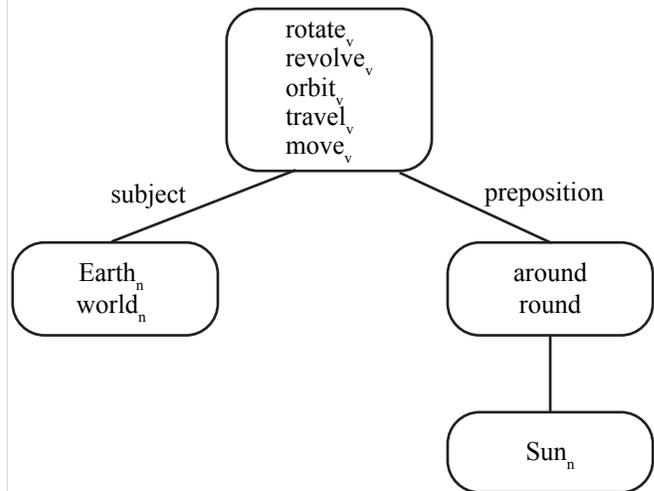

Fig. (1). Example of a scheme used in Automark to score the answer to the question like "What movement relates the Earth and the Sun?" (Source: Mitchell et al., 2003). [6]

a) LSA (Latent Semantic Analysis)

A natural language processing technique LSA (latent semantic analysis) containing the methods using statistical computations of a large set of a corpus for extraction, representation of usage in terms of context and essence (meaning) of words. Also known as latent semantic indexing it is quite much closely relate to human language learning and understanding.

For understanding, deriving and analyzing the essence of essay; Latent semantic analyses (LSA) and similar semantic content analysis and representation models have been successfully integrated into the automated essays scoring solutions. Such integration extended the range of features that can be extracted and used in automated scoring to cover the meaning of words, sentences and paragraphs used in an essay.

These models are constructed using a large structured set of texts and thus provide an external frame of reference for extraction and assessment of semantic features in essay autoscoring. LSA uses hundreds of syntactic, lexical and semantic dimensions to represent the meaning of essays. Such complex statistical modelling has been enabled by rapid developments in computer capacity and power and advances in artificial intelligence theory. Recent developments have also allowed the use of comprehensive methods of machine learning and modelling to establish the connection between lexical and semantic features of essays and their scores. [1]

IV. THE CURRENT PROBLEM

No open source systems of AEE are available, resulting in no standard grading methodology. The commercial organization with heavy investments conduct valuable research which is protected by restricting and concealing the deep insight, technical details; Making AEE systems the most challenging and big ticketed.

Due to its challenging nature, many great minds have attempted to create a more accurate or similar performance to human grader systems, some had introduced the modification under specific circumstances. As there is no standard approach defined there exists multifold approached systems. With advancements, comes criticism and insecurity. The fear of Artificial Intelligence taking over all human jobs is present and questionable. AEE systems also had faced a severe backlash presenting the image of replacing teachers.

V. PROPOSED FRAMEWORK

The automated essay evaluation system is considered as a facilitator aiding teachers by making their assessing task much easier and saving time, for students by providing a platform for practice and self-improvement.

Under such vision we created the system, following below design and structure.The AEE system is mainly composed to output the score and valuable feedback to the user. In this section, we present our design and approach in figure 2.
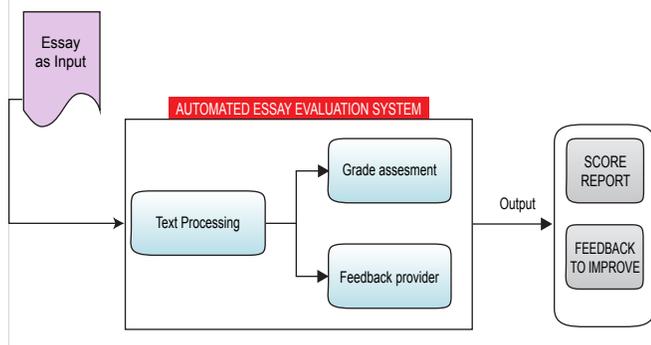

Fig. (2). Block diagram of proposed AEE system.

In general terms, an essay is an input to the system in which it is processed by various algorithms and the outputs are input to next; grader and feedback provider. AEE system is capable to detect specific language errors such as spell, punctuation, grammatical which comes under the category of syntactic errors. The feature of generating automatic feedback on each error to correction is also provided.

AEES is supposed to release the pressure and load from teachers and promoting the essay writing activities by also providing general feedback on the content of the essay. The systems whose work also includes the suggestions and feedback are aid and helping tool, not a replacement technology.

The limitation of feedback till now is reported that feedback is given within bounded with content completion and correctness in terms of syntactic aspect. For semantic evaluation and feedback generation based on semantic approaches are under working and several attempts are made in the progress.

Essay processing is further divided into feature Extraction, text Pre-processing and Advance Text Processing.

A. Basic Feature Extraction:

Extracting or finding feature of the text is an initial step which mandatory, basic part as the result features are used further processing. The important features which are extracted are below:

- Number of words.
- Number of characters.
- Average word length.
- Number of stop words.
- Number of uppercase words.

B. Basic Text Pre-Processing of Text Data:

Basic text processing is needed to be done before heading towards advance processing. Following are some required preprocessing steps.

- Lower casing.
- Punctuation removal.
- Stop words removal.
- Frequent words removal.
- Tokenization.
- Stemming.
- Lemmatization.

C. Advance Text Processing:

After pre-processing of text. Advance algorithms of processing are implemented.

- Information extraction Entity recognition.
- Word checker (spell and vocab).
- Topic relevancy, keyword Extraction.
- Summary Extraction.
- Spell Correction.
- Term Frequency-Inverse Document Frequency (TF-IDF).
- Bag of Words.

D. Information Extraction (Entity Recognition):

Entity identification is a subtask of Information Extraction. It is responsible to find, classify and sort the text into some defined categories; named entities. Named entities can be the name of person, place, year, location (nouns). Named Entity Recognition is designed to take text and find an annotated text which represents entity elements in a text.

Figure 3 is presenting the process of entity recognition. The document is processed by taking all the raw text from it which is split into sentences by sentence segmentation. Each sentence is processed by creating each word a token and next is tagging with parts of speech tags. Tag sentences are passed to entity Detection, finds the potential entities.
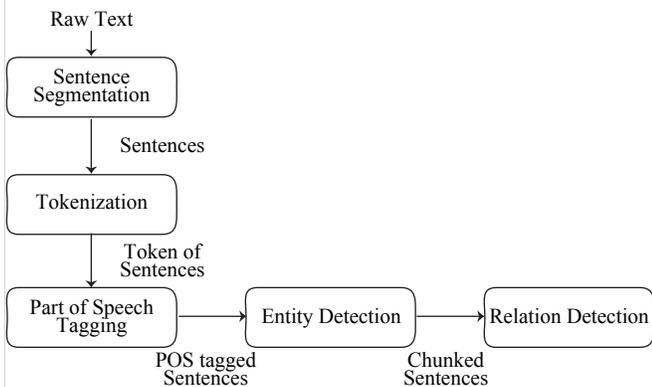

Fig. (3). Diagram explaining step by step process of Entity Recognition.

a) Part of Speech Tagging:

Part of speech is the foundation of the language they describe the sense of how the word is been used in a sentence. In most cases, auxiliary information needed to be associated with each token of text. POS tagging is also known as grammatical tagging is a process which involves making each token labelled corresponding to some part of speech.

It divides every word into some category. It is a supervised learning that is done after tokenization. POS tagging is also useful in providing information about

semantic content before entity detection POS tagging is done. Tag set of Brown Corpus.

Table 2. Tag Set of Brown Corpus used in Parts of Speech Tagging.

LABEL	PART OF SPEECH
CD	Number
JJ	Adjective
NN	Noun
VB	Verb
AT	Article
IN	Preposition

b) Chunking:

It is a technique used for entity finding. Chunking works after POS tagging. The labelled tags are input, and chunks become output. We are interested in digging the noun phrases for this purpose Noun Phrase chunking is considered. For NP chunk, using POS tags, chunk grammar is defined initially. Regular Expression rule is also used.

The rule states that whenever the chunk finds an optional determiner (DT) followed by any number of adjectives (JJ) and then a noun (NN) then the Noun Phrase (NP) chunk should be formed. Using this rule chunk parser is created to extract nouns.

E. Spell & Vocabulary Checker:

After Entity Recognition, for temporary period entities are removed which is further processed using pre-processing techniques for finding spell error and complex vocabulary usage. Below figure 4 illustrate the complete process.

For spell and Vocabulary Checker, we prepared dictionaries one for easy and common words consisting more than around 2000 words and difficult, complex words consisting more than 5000 words. The dictionaries were created using the SAT, GRE, GMAT vocabulary words.
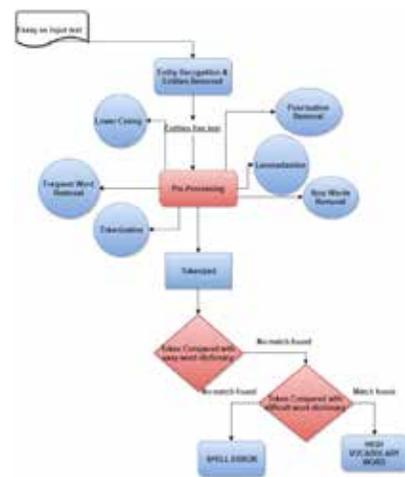

Fig. (4). Flowchart of spell and vocabulary checker.

F. Term Frequency-Inverse Document Frequency (TF-IDF):

We need to quantify the essay this can be done by looking at words.

a) Term Frequency:

Term Frequency is a measurement of a frequently used word, the often occurrence tells the importance of that term. An essay containing 200 words/tokens, a particular token appears 10 times in the whole essay then mathematical method to find term frequency is:

$$\text{TF weight} = \frac{\textit{Number of times a token appeared in essay}}{\textit{Total number of Token it content}}$$

b) Inverse Document Frequency (IDF):

An approach quite different from Term frequency in which commonly used a word the weight is decreased and for not so often word the weight is increased. It works for rare tokens in essays.

$$\text{IDF (t)} = \frac{\textit{Na token appeared in essay}}{\textit{number of essays that containt}}$$

c) TF-IDF:

By combining both TF and IDF to measure the important a word is to the document. It is a weighting factor.
TF-IDF=TF×IDF

G. Keyword Extraction RAKE (Rapid Automatic Keyword Extraction):

Times run out reading the prolonged content. Initially, a person is only interested in knowing the topic of document for this purpose keywords plays a consequential role. RAKE is based on our observation that keywords frequently contain multiple words but rarely contain standard punctuation or stop words, such as the function words and, the, and of, or other words with minimal lexical meaning [7].

Finding keywords is not only bounded to essay writing but it is essential for almost every writing including articles writing, web content writing, publications and finding meaningful keywords is quite a tedious job.

Following are some points which present the advantage of extracting keywords of essay:

• Classification of documents.
• Brief Representation of contents of the essay.
• Help in comparison with other related essays.

Considering the above advantages and to provide ease to assessors we added the feature of Keyword Extraction by implementing Rapid Automatic Keyword Extraction (RAKE) an unsupervised language independent and domain-independent algorithm works on an individual document or text file.

The core idea behind RAKE is the keywords have no punctuation and no stop word.
For a given text RAKE algorithm works as follow.

Consider the example of the short text "Pakistan is land of peaceful people."

1　Split the text, breaking at word delimiter (punctuation and space) resulting in some sequence of words.
["Pakistan ", "is"," land" "of", " peaceful ", " people "]

2　The sequence is split, breaking at stop words (Common words used in a language example: the, a, an, is, that, was etc.). Avoiding the stop words by skipping. The update the sequence is the "Candidate Keyword.
["Pakistan ", " peaceful ", "people"]

3　Find the score of each word in candidate keyword list using

$$\textit{Degree of work}\Big/\textit{frequency of word}$$

Frequency ("Pakistan ") = 1
Frequency ("peaceful ") = 1
Frequency ("people ") = 1

4　In this step after calculating score of each word in the list, candidate keyword score is find by adding the scores of word to its constituent word.
word_score("Pakistan ") = 2/1 = 2 word_score("peaceful ") = 1/1 = 1 word_score("people ") = 2/1 = 2

5　Score ("Pakistan ") = word_score("Pakistan ") = 2
score ("people ") = word_score ("people ") = 2

6　Final keywords are extracted by taking as one third the number of highest scoring words.
So, the highest scoring extracted keyword is from the sentence "Pakistan is land of peaceful people." is "Pakistan", "people"

H. Automatic Summary Extraction:

Summarization is the feature we have introduced to facilitates the teachers. The essay writing most of the times become lengthy with respect to the topic, 300-500 is considered as the normal range. It can become an arduous

task to go through each student's essay. For saving time and for the sake of efficiency of our system we provide automatic summarization as part of the feedback.

Why we need automatic text summarization:
1. Time reduction for reading.
2. summaries make the selection process easier when researching important points
3. Effectiveness of indexing.
4. Less biased as compared to summarizers generated by humans.

Text Summarization aims to generate a concise and compressed form of original documents. With text mining, the information to be extracted is clearly and explicitly stated in the text. Text mining summarizes salient features from a large body of text, which is a subfield of text summarization [8].

There exist two approaches for automatic summarization. Extractive and abstractive. This is because of the fact that abstractive summarization methods cope with problems such as semantic representation, inference and natural language generation which is relatively harder than data-driven approaches such as sentence extraction. As a matter of fact, there is no completely abstractive summarization system today.

Existing abstractive summarizers often rely on an extractive pre-processing component to produce the abstract of the text. [5]

Text Summarization is implemented using textRank, unsupervised based on weighted graphs algorithm. Which is [9]

1. The given text is preprocessed.
2. A graph is constructed and each vertex presents the sentence. Following is the given method by Rada Mihalcea and Paul Tarau. [9]

a) *Conversion of Text into Graph:*

- Identify text units that best define the task at hand, and add them as vertices in the graph.
- Identify relations that connect such text units, and use these relations to draw edges between vertices in the graph. Edges can be directed or undirected, weighted or unweighted.
- Iterate the graph-based ranking algorithm until convergence.
- Sort vertices based on their final score. Use the values attached to each vertex for ranking.

- Connect every sentence to every other sentence by an edge. The weight of the edge is how similar the two sentences are.

3. Rank algorithm applied to the graph.

b) *The Textrank Model:*

An algorithm based on graph ranking is a method for determining the importance of a vertex in a graph. The basic idea implemented by a graph-based ranking model is that of "voting" or "recommendation". [10] Linking of one vertex to another is actually assumed as casting a vote for that particular vertex. The number of votes given to a vertex determines the importance of vertex. Hence, the score associated with a vertex is determined based on the votes that are cast for it, and the score of the vertices casting these votes" [10]. The score of a vertex is calculated using the below formula:

$$S(Vi) = (1 - d^* \sum_{j \in \text{In } (Vi)} \frac{1}{|o\ ut(Vj)|}\ S(Vj)$$

Suppose a directed graph, G= (V, E) having vertices V. For a vertex Vi In (Vi) is the set of vertices pointing to its predecessors and Out (Vi) set of vertices that Vi points.

- Pick the vertices(sentences) with the highest PageRank score.

c) *Sentence Extraction Using Textrank:*

TextRank is applied on a graph of text. For each sentence is given in text a vertex is added to the graph, vertex represents the sentence. To rank the sentence in a graph. The relationship between sentences is determined.

The relation here is similarity relation which is measured as a function of their content overlap. Such a relation between two sentences can be seen as a process of "recommendation": a sentence that addresses certain concepts in a text, gives the reader a "recommendation" to refer to other sentences in the text that address the same concepts, and therefore a link can be drawn between any two such sentences that share common content. [10]

The overlap property between two sentences is found by comparing a similar number of tokens or syntactic filters. After going through the whole process the graph we acquired is highly connected with weights on each edge. The textRank Model, the ranking formula is applied which assign the score or rank to sentences, after the ranking process is completed on a graph. Sentences are sorted in reversed order of their score, and the top ranked sentences are selected for inclusion in the summary.

VI. EVALUATION

Following are the developed system screenshots from figure 5-7. In the analyze page the system asks a user to mention the topic word limit and enter essay or upload a file.

Fig. (5). Submission Page.

After the submission is done, the system prompts the result page.

Fig. (6). Result Page.

Fig. (7). Result Page.

Table 3. Expected Accuracy dependency of system module.

MODULE	ACCURACY DEPENDANCY
Entity Recognization	Depend on POS tagging.
Spell & Vocabulary Checker	Depend on the provided dictionary.
Keyword Extraction	Depend on the Generation of stop list.
Automatic Summary Extraction	The number of keyword selection.
Topic Modeling	Depend on Number of topics, lengths of a document, lad parameters.

a) AEES Feedback Properties:

AEE systems are capable to detect specific language errors such as spell, punctuation, grammatical which comes under the category of syntactic errors. The feature of generating automatic feedback on each error to correction is also provided.

AEES is supposed to release the pressure and load from teachers and promoting the essay writing activities by also providing general feedback on the content of the essay. The systems whose work also includes the suggestions and feedback are aid and helping tool, not a replacement technology. The current limitation of the feedback is that its content is limited to the completeness or correctness of the syntactic aspect of the essay. Some attempts have been made to include also semantic evaluation, but these approaches are not automatic and work only partially. [1]

VII. CONCLUSION & FUTURE WORK

In this paper, we had conducted a detailed study on automatic evaluation system and its possible approaches by proposing the framework. Aiming to fulfil the need and improving the educational standards of Pakistan. Pakistan is country of diverse educational backgrounds some are granted with high-end facilities, updated coursework and where some underground classes lack, due to this gap in our existing education system, the difference in the standard of education is present, affecting the quality of education system. Many students lack writing skills, proper resources which impact on their academic life as well as professional life. By mean of this system students as well the teacher s will enjoy many valuable advantages which directly promote interest and build writing skills in students and as well as release the high burden on teachers. Less failure rate in international tests.

The limitations and challenges exist which includes text semantic, it can be addressed for making AEE systems more efficient and trustworthy. To assess the essay a wide variety

of attributes are used such as grammar, spellchecking, vocabulary usage, content and lexical style. But there is a lack of text semantics which is considered as main weakness. Some systems use latent semantic analysis (LSA), latent Dirichlet allocation (LDA), and content vector analysis (CVA) to evaluate the semantic of the essay. [1] The major limitation of LSA is that it only retains the frequency of words by disregarding the word sequence, and the syntactic and semantic structure of texts. [3]

Experiments suggest that syntactic and semantic structural information can significantly improve the performance of the models for automated essay evaluation. However, only two existing systems use approaches that partially check if the statements in the essays are correct. Despite the efforts, these systems are not automatic, as they require manual interventions from the user. None of the existing systems is therefore capable of assessing the correctness of the given common-sense facts. [1]

Furthermore, in the future, the scope of AEE systems can be extended by working on limitations text semantics, addressing the subjectivity of essay topics, and can come up with better approach in conjunction with deep learning.

REFERENCES

[1] K. Zupanc and Z. Bonsic, "Advances in the Field of Automated Essay Evaluation", *Informatica*, vol. 39, pp. 118-132, 2016.

[2] M. D. Shemis and J. C. Burstein, *Automated Essay Scoring: A Cross-Disciplinary Perspective, Routledge*, 2003.

[3] S. Ghosh and S. S. Fatima, "Design of an Automated Essay Grading (AEG) System in Indian Context", *International Journal of Computer Applications*, vol. 11, pp. 72-77.

[4] Waegel, "The Development of Text-Mining Tools and Algorithms". unpublished, 2006.

[5] L. Kumar and P. Bhatia, "Text Mining: Concepts, Process and Applications", *Journal of Global Research in Computer Science*, vol. 4, pp. 36-39, 2013.

[6] A. O. Ade-Ibijola, I. Wakama and J. C. Amadi, "An Expert System for Automated Essay Scoring (AES) in Computing using Shallow NLP Techniques for *Inferencing," International Journal of Computer Application*, vol. 51, pp. 37-45, 2012.

[7] S. Rose, N. Cramer, W. Cowley and D. Engel, "Automatic Keyword Extraction from Individual Documents", in *Text Mining Applications and Theory*, M. W. Berry and J. Kogan, Eds. John Wiley & Sons. 2010, pp: 1-20.
DOI: 10.1002/9780470689646.ch1

[8] M. P. Uddin, A. S. Asa, S. Akter M. D. Hossein, S. K. Roy and M. I. Afjal, "A Comprehensive Survey on Extractive Text Summarization Techniques," *American Journal of Engineering Research*, vol. 6, no. 1, pp: 226-239.

[9] Y. Chali and S. A. Hasan, "On the Effectiveness of Using Syntactic and Shallow Semantic Tree Kernels for Automatic Assessment of Essays," In *Proceedings of Joint Conference on Natural Language Processing*, 2013, pp: 767–773.

[10] R. Mihalcea and P. Tarau, "TextRank: Bringing Order into Texts", In *Proceedings of 2004 Conference on Empirical Methods in Natural Language Processing*, 2004, 404–411.