International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2019 IJSRCSEIT | Volume 5 | Issue 2 | ISSN : 2456-3307
DOI : 10.32628/CSEIT1952236

Automatic Content Analyzer

Piyush Mishra[1], Ronit Parikh[1], Pallavi Sharma[1], Romit Parikh[1], Dhananjay Joshi[2]
[1]Computer Engineering, SVKM NMIMS MPSTME, Shirpur, Dhule, Maharashtra, India
[2]Assistant Professor Computer Engineering, SVKM NMIMS MPSTME, Shirpur, Dhule, Maharashtra, India

ABSTRACT

Essays and short answers are crucial testing tools for assessing academic achievement, integration of ideas and ability to recall, but are expensive and time consuming to grade manually. Manual grading of essays takes up a significant numbers of instructors' valuable time, and hence is an expensive process. Automated grading, if proven to match or exceed the reliability of human graders, will significantly reduce costs. The work done in our project on Content Analyzer System analyzes the subjective type answers and grade them based on the features of a written text such as language, grammar, organization and content. Our system automatically grades the essays or short answers based on the above-mentioned features and provides the user with essay statistics which includes word count, sentence count, paragraph count and the overall weighted score which is the mean of scores of each feature.

Keywords : Content Analyzer, MsNLP, Electronic Essay Rater, Latent Semantic Analysis, LSA, BOW, POS, NLP, Feature Extraction, Word Similarity

I. INTRODUCTION

1.1 PURPOSE

One of the key roadblocks to teaching and evaluating critical thinking and analytical skills is the expense associated with scoring tests to measure those abilities. For example, tests that require “constructed responses” (i.e., written answers, written essays) are useful tools, but they typically are hand scored, commanding considerable time and expense from public agencies. So, because of those costs, standardized examinations have increasingly been limited to using “bubble tests” that deny us opportunities to challenge our students with more sophisticated measures of ability.

1.2 SCOPE

The product has scope in departments of education where developing new forms of testing and grading methods, to assess the new common core standards. For example, we know that essays are an important expression of academic achievement, but they are expensive and time consuming for states to grade them by hand. So, we are frequently limited to multiple-choice standardized tests. We believe that automated scoring systems can yield fast, effective and affordable solutions that would allow states to introduce essays and other sophisticated testing tools.

Benefits:

- Human time and effort is saved.
- Coherence in evaluation of all the scripts present
-

Objective:

- To evaluate and assign a score for a short answer/essay without human intervention.
- To provide visualization of a scores for group of students.

II. RELATED WORK

Many systems have been developed in this field for either commercial use or as a result of some research in this area. Some of these are discussed in this section.

2.1 Educational Testing Service (ETS)

Lexical semantic techniques are used here for building a scoring methodology, which uses small data sets. Training data set is used to build concept based grammar and lexicon which are specific to the domain[1][2]. Microsoft NLP (MsNLP) tool is used here for parsing the training data essays, where all suffixes and a few stop words are removed manually. This produces lexicons (a total stock of words and word elements that carry a meaning).

The list of words and terminology present in the lexicon continues to be invariable and at the same time as the features related with each entry are modular, hence they can be replaced as obligatory [1]. Manual classification of some words as metonyms of each other is necessary. For each category of answers, the system creates grammar rules manually (these individual sets should consist of all the paraphrases for the plausible result) with the help of syntactic parses of individual sentences from the training dataset being used as well as the lexicons.

2.2 Electronic Essay Rater (E- Rater)

It has been built to use a blend of NLP and statistical techniques to mine the essay for semantic features which are then evaluated. The essays are assessed based on a training dataset of human evaluated essays[2]. Here, the essays which stay relevant to the topic and contain a well-organized structure are more likely to receive a higher score on the six point scale. Some of the major features included is the study of the discourse structure, the syntactic structure and of the vocabulary of the essay i.e. domain analysis[3]. The way it functions is by applying a corpus based method to build a model which uses actual sample essays as data for analyzing the features of the student essays[5].

The design of the application is such to recognize the features in the response that contain merits which would be in a human scored essay and currently composes of five modules[4]. Three of those classify features which could be used for scoring the syntactic variety, the way ideas are organized and the vocabulary being used in the essay[6]. The fourth module is used for selecting and weighing features which can be used for scoring the essay. The last module is used for calculating the final score.

2.3 LATENT SEMANTIC ANALYSIS (LSA)

Similarity measured using LSA are considered equal to the human meaning similarities for words and texts. Other than that, it is able to successfully imitate human word selections and categorical judgments [7]. The main philosophy on which it works is that the passage under consideration is dependent on the words it comprises and even changing one of those can result in having the whole meaning of the passage being changed. On the other hand, two passages having different words can have similar meaning [7].

For the evaluation of the overall quality of an essay, the LSA is trained on the texts that best represent the writing prompt. After this step characterization of the essay is to be done by using a mathematical representation of it called the LSA vectors and at last the conceptual content and the significance of the essay is compared to other texts. On being compared to the factors related to content of the essay such as arguments, style, etc. the mechanical and syntactical features can be easily separated from these. The reason being that these content related factors are easily affected by word choices.

Here the text is represented as a matrix. Each row stands for a unique word, whereas each column stands for the context. Frequency of the word is contained in each cell. The frequency of each and every cell is considered by a feature that represents the contextual importance of the word as well as the degree to which the information is carried by the word type in the domain discourse. Based on the occurrence of the word we can verify its semantics. The semantic space is also determined by the number of times each word is encountered in the text. Example, 400 paragraphs and 1000 words provide a 400X 1000 matrix. Here, while each word is Represented by a 400-dimentional vector, each paragraph is represented by a 1000-dimentional vector.

LSA includes semantic similarities between words by reducing these dimensions. The representation of word meaning is permitted through the context of their occurrence which makes this reduction critical. Another important aspect here is the number of dimensions. If the number is too large limited dependencies will be drawn between the vectors and if the dimensions are too small some of the information might be lost. Hence as stated by this method the semantic information is determined by the occurrence of these words in large quantity of texts.

2.4 INTELLIGENT ESSAY ASSESSOR (IEA)

IEA can effectively evaluate not only the creative narratives, but also the content-based essays. This system is required to be trained on a specific set of domain- representative texts in order to judge the quality of any essay. For example, an English literature book can be used to assess an English literature essay. Here three methods are used to assess the essays:

1. Previously scored essays of other students,
2. Model essays and knowledge source materials like books,
3. And internal comparison of some unscored sets of essays[7]

With the help of these methods, IEA can be used to compare student essays with comparable texts in terms of content quality [7][8][10]. First, IEA compares the similarities between essays written by the students and some other essays on the similar topics that are already graded by human evaluators and determines the proximity between them.[7][9][10]It can then, predict the total score by adding "corpus-statistical writing-style" and mechanics Irrespective of switching of synonyms, rephrasing, or reshuffling of sentences, the two essays will be alike with LSA[8]. Detecting plagiarism here is a crucial feature since this sort of dishonesty is very hard for human evaluators to detect, mostly because the number of essays to be graded are huge [7].

III. PROPOSED SOLUTION

The problems faced above can be solved by automating the answer grading system. The aim is to develop an automated system which will provide results instantly and promise to remove human errors that commonly occur during manual checking. Thereby reducing human efforts and saving time and resources.

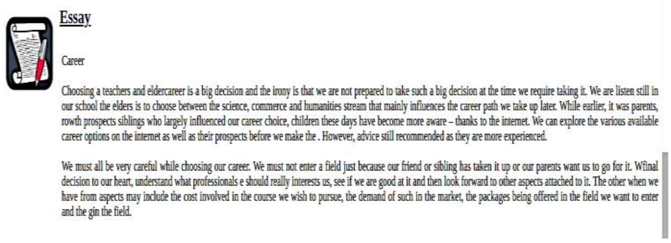
The workflow for our proposed approach is as follows: First we extract the features from each essay. Bag of Words (BOW), Parts of Speech (POS) count, number of simple features such as word count, sentence count, average sentence length, paragraph count. These features represent the fluency and dexterity of the writer. This feature is extracted using natural language toolkit (NLTK) part of speech tagger. The essay is first tokenized into sentences before the tagging process. The statistics will be used to score the essay based on the evaluation criteria used by GRE/GMAT.

In our project we will develop a system which will input the subjective answer in text file format and generate the report of graded essay score in an Html file. The system will process the input file and evaluate the overall quality of essay by taking into account general skill areas such as language, grammar, content, organization. The evaluating and scoring criteria will be on the lines of GRE/GMAT.

IV. PROPOSED WORK

Subjective answer will be given as input to ACA system in text file format. ACA will use Natural Language Toolkit, or more commonly NLTK, which is a suite of libraries and programs for symbolic and statistical natural language processing for English written in the Python programming language. Answer statistics will be calculated using libraries and functions designed by us. The answer statistics will be further used for scoring the quality of answer based on individual parameters which include spell check and grammatical structure analysis. These individual parameters are scored using the same methodology which is also used in GRE/GMAT for grading essays. Overall score is calculated by taking the mean of the scores of individual parameters. The report containing the essay statistics and overall score will be displayed to user in the Html format.

4.1 DATA COLLECTION



Essay

Career

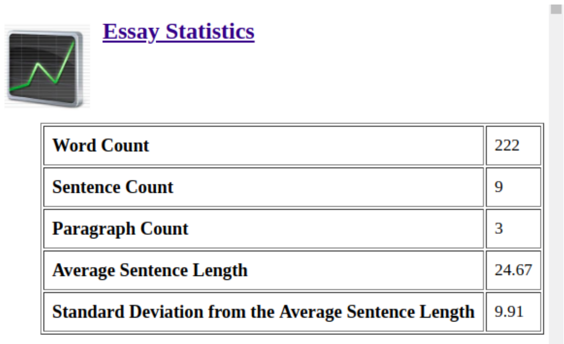
Choosing a teachers and eldercareer is a big decision and the irony is that we are not prepared to take such a big decision at the time we require taking it. We are listen still in our school the elders is to choose between the science, commerce and humanities stream that mainly influences the career path we take up later. While earlier, it was parents, rowth prospects siblings who largely influenced our career choice, children these days have become more aware – thanks to the internet. We can explore the various available career options on the internet as well as their prospects before we make the . However, advice still recommended as they are more experienced.

We must all be very careful while choosing our career. We must not enter a field just because our friend or sibling has taken it up or our parents want us to go for it. Wfinal decision to our heart, understand what professionals e should really interests us, see if we are good at it and then look forward to other aspects attached to it. The other when we have from aspects may include the cost involved in the course we wish to pursue, the demand of such in the market, the packages being offered in the field we want to enter and the gin the field.

Figure 1. INPUT ESSAY

The above Essay will be input to the system in text file format. The essay will be on the lines of GRE/GMAT essays. Essays will be in simple format without any bullets and numbering.

4.2 DATA PRE-PROCESSING



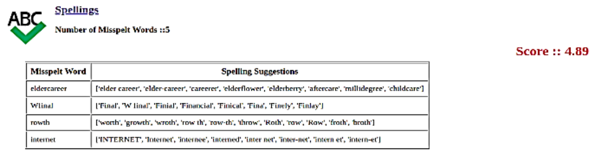
Essay Statistics

Word Count	222
Sentence Count	9
Paragraph Count	3
Average Sentence Length	24.67
Standard Deviation from the Average Sentence Length	9.91

Figure 2. ESSAY STATISTICS

The above presented Essay Statistics will be included in the Html report displayed to user. Word Count and sentence count is calculated by tokenizing the essay based on the space and full stop respectively. The paragraphs in the essay, if it contains multiple paragraphs, will be separated by a blank line. The essay statistics will further be used for scoring individual parameter based on which the essay is to be graded.

4.2 FEATURE SELECTION



Spellings

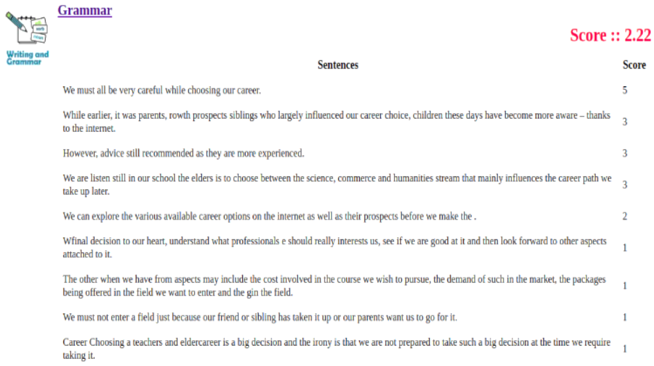
Number of Misspelt Words ::5

Score :: 4.89

Misspelt Word	Spelling Suggestions
eldercareer	['elder career', 'elder-career', 'careerer', 'elderflower', 'elderberry', 'aftercare', 'millidegree', 'childcare']
Wfinal	['Final', 'W final', 'Finial', 'Financial', 'Finical', 'Fina', 'Finely', 'Finlay']
rowth	['worth', 'growth', 'wroth', 'row th', 'row-th', 'throw', 'Roth', 'row', 'Row', 'froth', 'broth']
internet	['INTERNET', 'Internet', 'internee', 'interned', 'inter net', 'inter-net', 'intern et', 'intern-et']

Figure 3. SPELL CHECK

The above segment displays the misspelt words and there total count. Spelling suggestions provide the words which could be related to the misspelt word. The essay is scored based on this spell check parameter.



Grammar

Score :: 2.22

Sentences	Score
We must all be very careful while choosing our career.	5
While earlier, it was parents, rowth prospects siblings who largely influenced our career choice, children these days have become more aware – thanks to the internet.	3
However, advice still recommended as they are more experienced.	3
We are listen still in our school the elders is to choose between the science, commerce and humanities stream that mainly influences the career path we take up later.	3
We can explore the various available career options on the internet as well as their prospects before we make the .	2
Wfinal decision to our heart, understand what professionals e should really interests us, see if we are good at it and then look forward to other aspects attached to it.	1
The other when we have from aspects may include the cost involved in the course we wish to pursue, the demand of such in the market, the packages being offered in the field we want to enter and the gin the field.	1
We must not enter a field just because our friend or sibling has taken it up or our parents want us to go for it.	1
Career Choosing a teachers and eldercareer is a big decision and the irony is that we are not prepared to take such a big decision at the time we require taking it.	1

Figure 4. GRAMMAR CHECK

Essay is scored based on its syntactic structure using link grammar parser.

4.3 .GRADING

Overall Score

GRADE (0-5)	3.37
Spelling(0-5)	4.89
Grammar(0-5)	2.22
Coherence(0-5)	3.00

Figure 5. OVERALL SCORE

An overall score computed by taking the weighted mean of individual scores of spelling, grammar and coherence which are the parameters on which the quality of an essay is decided.

V. CONCLUSION

ACA system developed by us solved the problems that were caused by manual checking of the answer sheet by automating the complete process. The system calculates the score and provide results instantly. It removes human errors that commonly occur during manual checking. Thus, the system reduces human efforts and saves time and resources. The methodology used by us for grading the subjective answers is the same which is used in GRE/GMAT to score the essays. ACA system has a simple user interface which inputs the answer in txt file format and generates the analysis report in Html file which contains answer statistics, an overall score computed by taking the weighted mean of individual scores of the parameters used for assessing the quality of a subjective answer.

VI. REFERENCES

[1]. Shristi Drolia., et al. “Automated Essay Rater using Natural Language Processing” International Journal of Computer Applications (0975 - 8887) Volume 163 - No 10, April 2017

[2]. Shermis, Mark D., and Jill C. Burstein, eds. Automated essay scoring: A cross-disciplinary perspective. Routledge, 2003

[3]. Burstein, Jill, Martin Chodorow, and Claudia Leacock. "Automated essay evaluation: The Criterion online writing service." Ai Magazine 25.3 (2004): 27.

[4]. Rudner, Lawrence M., and Tahung Liang. "Automated essay scoring using Bayes' theorem." The Journal of Technology, Learning and Assessment 1.2 (2002).

[5]. Powers, Donald E., et al. "Stumping E-Rater: Challenging the validity of automated essay scoring." ETS Research Report Series 2001.1 (2001).

[6]. Dikli, Semire. "An overview of automated scoring of essays." The Journal of Technology, Learning and Assessment 5.1 (2006).

[7]. Landauer, T. K., Laham, D., & Foltz, P. W. (2003). Automated Essay Scoring: A Cross Disciplinary Perspective. In Mark D. Shermis and Jill C. Burstein (Eds.), Automated Essay

[8]. Scoring and Annotation of Essays with the Intelligent Essay Assessor. Mahwah, NJ:Lawrence Erlbaum Associates.

[9]. Landauer, T. K., Laham, D., & Foltz, P. W. (September/ October, 2000). The Intelligent Essay Assessor. In Marti A. Hearst (Ed), The debate on automated essay grading. IEEE Intelligent systems, 27- 31.

[10]. Rudner, L. & Gagne, P. (2001). An overview of three approaches to scoring written essays by computer (ERIC Digest number ED 458 290).

[11]. Streeter, L., Psotka, J., Laham, D., & MacCuish, D. (2004). The credible grading machine: essay scoring in the DOD [Department of Defense].

[12]. Burstein, Jill, Martin Chodorow, and Claudia Leacock. "Automated essay evaluation: The Criterion online writing service." Ai Magazine 25.3 (2004): 27.

[13]. Rudner, Lawrence M., and Tahung Liang. "Automated essay scoring using Bayes' theorem." The Journal of Technology, Learning and Assessment 1.2 (2002).

[14]. Semire DIKLI, “Automated Essay Scoring”, Turkish Online Journal of Distance Education-TOJDE January 2006 ISSN 1302-6488 Volume: 7