Reviewing Techniques For Automatic Response Grading Via Language Processing

Simran Agrawal, Avinash J. Agrawal

Abstract: ***Automatic grading of student answers via natural language processing is a boon for the faculties and educational system in both technical and non-technical fields. It helps the students to get their document evaluation done irrespective of the state of mind of the examiner, and also helps in speeding up the process for grading, and eventually saving a lot of time and effort for the overall examination process. In this paper we have analyzed various methods which are useful in automatic grading of student answers (both long and short), most of them are based on mathematical variations in natural language processing techniques. Via this text we aim to assist researchers to decide which kind of base methods can be used in which kind of document scenarios, so that it helps them in selection of algorithms based on the input type, and speedup their macro-level research.***

Index Terms: Document, grading, response, natural, language, answers

I. INTRODUCTION

In the learning procedure, the evaluation of information assumes a key job for compelling educating. With the scope of evaluation techniques accessible, examinations have ruled the appraisal of understudy learning. Specifically, learning and comprehension in scholarly courses are surveyed by end obviously examinations joined with coursework. Scholarly examinations can be performed utilizing many inquiry types from numerous decision inquiries to free reactions. Manual evaluation is significantly more troublesome for inquiry types, for example, short answer and exposition questions. Understudies are required to give free content reactions for these inquiry types, so every reaction requires literary comprehension and examination rather than reviewing answers with a solitary right answer, for example, numerous decision tests. Consequently, manual scoring takes a lot of time, and arrangement of significant input much more so. Manual scoring of those answers can experience the ill effects of irregularity since the marker must derive importance from the applicants' very own words.

Revised Manuscript Received on May 06, 2019

Simran Agrawal Bachelor of Engineering Degree in Computer Science and Engineering from Shri Balaji Institute of Science and Technology, Rajiv Gandhi Proudyogiki Vishwavidyalaya University, Betul, India 2016.

Dr. Avinash J. Agrawal received Bachelor of Engineering Degree in Computer Technology from Nagpur University, India and Master of Technology degree in Computer Technology from National Institute of Technology, Raipur, India.

Scores on a similar answer may fluctuate from marker to marker. Be that as it may, free content inquiries are a broadly favored appraisal device, utilized all through the learning procedure, because of their viability on creating intellectual aptitudes of understudies and furthermore showing information in short messages. In this way, there is a need to create instruments for moderating these difficulties in evaluation. One methodology is to make programmed scoring instruments and input systems that bolsters markers. This approach was talked about by a few analysts, of late as computational systems become material in this field. Evaluation of common language reactions is a difficult assignment since we can't anticipate that a machine should see free content answers. Be that as it may, advancements in characteristic language handling (NLP) have made halfway or completely robotized scoring of tests conceivable. Programmed evaluating has turned into a mainstream field among specialists because of its advantages on diminishing human slip-ups and time spent. Programmed evaluation can likewise be considered for pre-appraisal of understudy works by giving them backing to improve their work. An understudy could present their work to an input framework and be given data, for example, "Answers like this scored this imprint. So as to improve this you may ...". By following understudy utilization of such a framework and we likewise have the likelihood of mentor comprehension of what is being misconstrued in class, with the open door for eye to eye moderation. This examination presents both regulated and unsupervised methodologies that manage the programmed checking and criticism for short answers. The proposed models depend on the idea of closeness between the model answer and understudy answers, and the revelation of the structure in the corpus of understudy reactions. Our underlying presumption is that given a lot of understudy answers, marks granted by scorers are very reliant on words that the understudy utilized which additionally happen in the model answer. This is on the grounds that practically speaking, grades are regularly granted dependent on how comparable an understudy reaction is to the normal answer (show answer). Utilizing these similitudes, we expect to construct our model to check understudy works. The model ascertains the separation between the model answer and the understudy answer utilizing words in the model response to produce scoring rules.

This has the extra advantage of recognizing the confusions and shortcomings of understudies for a point under the presumption that those understudies who have an absence of information are not ready to utilize the majority of the words (or equivalent words) in the model answer. Contingent upon the likenesses between answers, programmed input can be given to make understudies mindful of their dimension of comprehension. Such a framework can be likewise utilized as a regulated learning procedure to foresee an answer's score and mechanize stamping by utilizing a preparation set. In the learning procedure, giving input to understudy likewise assumes a key job, as it enables understudies to comprehend the subjects and improve their learnings and mindfulness. Our second methodology centers around a programmed criticism system. We bunch understudy answers into gatherings to investigate whether reactions given by understudies share comparable qualities. This methodology can reveal regular gatherings of answers having comparative structures that understudies as often as possible use. We discover bunches of comparable answers, and afterward to assess components of these group utilizing both human and computational methods. Along these lines we will furnish instructors with data about the regular answers that understudies give, since understudies by and large answer inquiries in comparative ways. The primary favorable position of this methodology is that another framework can be manufactured that educators can give a typical criticism to understudies having a place with a similar bunch. This should be possible by choosing a model answer(s) from each bunch. Gathering enables educator to give criticism to the model answer, and this input can be alloted to the whole group on the double. It could be additionally utilized in a criticism framework in which understudies submit answers with chronicled input, and utilize this criticism to improve their responses for definite accommodation of assignments. Further input on understudy conduct could be earned from understudy utilization of such a framework. Gathering comparative answers can be executed in the programmed stamping of gatherings. This methodology can altogether eliminate the ideal opportunity for manual stamping, and improve consistency in checking and input. Specifically, when gatherings of answer are distinguished, the framework allocates a typical score to entire gathering by utilizing human stamping. This gathering procedure won't be 100% solid and there might be answers which are increasingly hard to arrange. For this situation human mediation is fundamental. It isn't our longing to expel the human from the checking procedure, just to improve consistency and enable people to apply judgment in the troublesome cases. At long last, we build up a model to anticipate stamps by utilizing separations between the model answer and the understudy answers. We estimate that imprints can be anticipated by this separation between understudy answers and the model answer as imprints are very associated with this separation. The target of this methodology is to demonstrate how removes from model answer can be utilized to check understudy answers. In the outcomes area we will see that the model lies near the normal of the scores of the two markers.

The next section describes various techniques for grading the answers, followed by the conclusion drawn from these techniques and finally some ways to further improve the research in this area.

II. LITERATURE REVIEW

In [1], scientists have utilized different strategies like arrangement of sentences, tokenization and heuristics based coordinating, which has enabled them to get great exactness when contrasted and certain vector based techniques. This strategy can be utilized for long and short content assessments crosswise over various areas. While [2] utilizes quantum registering for short answer examination, the scoring is finished by the factual strategy embracing and incorporating rule-based semantic quantum-based highlights investigation bringing about more exactness. It is in a manner a half breed framework appropriate for short answer type scoring. The exploration done in [3] utilizes ongoing advances in the distinguishing proof of short-content closeness and expands content comparability highlights with key reviewing explicit develops so as to assess short answers, and creates abnormal state of precision when contrasted and standard systems regarding mistake rate and Pearson's coefficient. Utilization of delicate processing is presented in [4], where a methodology dependent on intermittent neural systems to become familiar with the connection between a paper and its doled out score, with no component designing is proposed. The outcomes demonstrate that this framework, which depends on long transient memory systems, beats a solid standard by 5.6% as far as quadratic weighted Kappa. For short and open answers (SOAs), the work in [5] presents a novel methodology with respect to learning evaluation of SOA questions utilizing idea maps. They propose a situation for utilizing this sort of inquiries in internet learning conditions and portray their methodology for consequently assessing this kind of inquiries utilizing idea maps and likenesses measures. In the event that perusers need to assess distinctive techniques for content closeness, at that point they can allude [6], where specialists look at and investigate two comparability measure strategies, cosine similitude and idle semantic examination. The parameters that was utilized to quantify the execution of the techniques are the computational intricacy - estimated by the measure of CPU and memory use, and page load time - and precision - estimated by Pearson Correlation and Mean Absolute Error. The outcomes demonstrated that both calculation expended same measure of memory. They additionally guarantee that cosine has a superior server act so liked to be executed in e-learning programmed article scoring framework. For long replies, analysts in [7], built up a Two-Stage Learning Framework (TSLF) which coordinates the upsides of both component designed and start to finish AES techniques. In tests, they analyze TSLF against various solid baselines, and the outcomes exhibit the adequacy and heartiness of the models. TSLF outperforms every one of the baselines on five-eighths of prompts and accomplishes new best in class normal execution when without negative examples.

In the wake of adding some papers to the first datasets, TSLF beats the highlights designed and start to finish baselines, all things considered, and demonstrates extraordinary heartiness. Article appraisal takes quite a while whenever redressed physically. Consequently, looks into on programmed article scoring have been developing quickly as of late. The strategy that is generally utilized for programmed exposition scoring is Cosine Similarity by using pack of words as the component extraction. Nonetheless, the element extraction by utilizing pack of words did not consider to the request of words in a sentence. In the mean time, the request of words in an article has an essential job in the appraisal. In [8], a programmed paper scoring framework dependent on n-gram and cosine similitude was proposed. N-gram was utilized for highlight extraction and changed to part by word rather than by letter with the goal that the word request would be considered. In light of assessment results, this framework got the best connection of 0.66 by utilizing unigram on inquiries that don't think about the request of words in the appropriate response. For inquiries that think about the request of the words in the appropriate response, bigram has the best relationship esteem by 0.67. In [9], analysts have built up a module where understudy reaction and right answer will be prepared by at first separating them into token for example words. Later on thing expression and action word gathering will be alloted to every single word with the assistance of Part-Of-Speech (POS) tagger. This undertaking is cultivated by NLP system. Every single expression of understudy reaction is contrasted and right answer. On the off chance that definite match is found in word just as POS tag and word position in sentence the scores are doled out. After score task Final scores are determined by making summation of appointed scores everything being equal. This technique is useful for both long and short answers. In [10], exposition based model is proposed with the name "Goodness" which utilizes a decency score to discover the benefit of coordinating and dependent on this score the reviewing is finished. Intricacy measures, similar to the Flesch Reading Ease Score (FRES) are additionally utilized as highlights in the framework. Notwithstanding those, parse tree highlights, similar to the normal parse tree profundity and the quantity of subordinate provisos (SBAR) in the content are utilized for score assessment. This strategy is useful for assessment of long messages. Another correlation is done in [11], where different methodologies for computerized appraisal of elucidating answers and reasoned that Text-Cmap examination based methodology is more powerful than Text-Text and Graph-Graph. The principal, third and fourth quartile in box and bristle diagram of TextCmap approach, shows that in a portion of the cases human and PC assessors distributes practically comparable imprints or with the least divergence. In any case, the issues like culmination of area explicit idea map, significance of a hub weighted hub, scientific count of level of closeness and revealed covered up semantic importance are the difficulties to be tended to. This paper can be utilized as a point of concentrate for long content answers.

The framework in [12], contains stages, for example, question order, answer grouping and answer assessment for the appropriate responses given by the understudy and grade them with suitable score. A grammatical connection based element extraction method is proposed for programmed assessment of elucidating type answers. The framework has additionally received a psychological based methodology where the understudy answers are made a decision for its accuracy dependent on the expressions utilized for addressing the inquiries. The score and input are given to make mindful of the understanding dimension of the subject. The exploratory investigation indicates 0.85% higher accuracy and review when contrasted with the before frameworks. Which probably won't be much, yet is still some improvement over the past frameworks. In [13], an investigation of different measures for coordinating the content is done, perusers can allude it to watch different measurements for score assessments. In [14] and [25], an answer for reviewing of papers of hypothesis based subjects is gotten where in Automatic Paper Grading will be performed utilizing Natural Language Processing. AI methods like Semantic Analysis is embraced. As a solitary answer can be displayed in various ways by various understudies, coordinating catchphrases is wasteful. That is the reason, utilizing metaphysics, extraction of words and their equivalent words identified with the area is done which makes the assessment procedure all encompassing as nearness of catchphrases, equivalent words, the correct word mix and inclusion of ideas would now be able to be checked. The previously mentioned strategies is executed with Ontology and is tried on regular information comprising of specialized answers. The outcomes are broke down and an impartial, high precision robotized reviewing framework for a hypothesis based subject will be gotten with next to no mistake rate which is practically identical to a differential human-to-human blunder rate. The calculation is planned dependent on the reactions gathered amid the overview directed among educators with respect to their parameters while rectifying papers physically. Specialists in [15] guarantee that one of the regular learning exercises utilized in instructive dimensions and trains is exposition composing. The issues of the exposition composing exercises are tedious, worries in creating quick outcome as well as input from educators to understudies, and the instructors will in general be emotional in evaluating the article exercises. The investigation in [15] means to apply the starter approach for consequently creating the area idea metaphysics in articles utilizing OntoGen and connected common language preparing calculations utilizing NLTK (Natural Language Tool Kit) that upgrade the instructors paper reviewing. The outcomes got appear to be very reassuring particularly the programmed calculation of the likeness among records. They demonstrated in their speculation that the quantity of words or vocabulary the understudies can display add to the score of the exposition. Automatic Essay Scoring (AES) is one of the improvement frameworks for deciding a score consequently from content archive source to encourage the adjustment and scoring by using applications that keep running on the PC. AES process is utilized to assist the teachers with scoring proficiently and viably. Other than it can decrease the subjectivity scoring issue.

In any case, usage of AES relies upon numerous components and cases, for example, language and system of scoring process particularly for article scoring. Various techniques executed for weighting the terms from archive and achieving the answers for taking care of relative dimension between records answer and master's report still characterized. In [16], analysts executed the weighting of Term Frequency – Inverse Document Frequency (TF-IDF) technique and Cosine Similarity with the estimating degree idea of similitude terms in a report. Tests did on various Indonesian content based reports that have experienced the phase of pre-handling for information extraction purposes. This procedure results is in a positioning of the record weight that have closeness coordinate dimension with master's report. In [17], scientists present an alternate unsupervised methodology which manages understudies' answers comprehensively utilizing content to content comparability. Distinctive String-based and Corpus-based likeness measures were tried independently and afterward joined to accomplish a greatest connection estimation of 0.504. The accomplished relationship is the best esteem accomplished for unsupervised methodology Bag of Words (BOW) when contrasted with some standard strategies. While, [18] is centered around the improvement of programmed short answer scoring. Some programmed scoring frameworks utilized on long answer have demonstrated ideal outcomes in giving a score on the understudies answer. Programmed long answer frameworks utilize the data recovery technique to gauge comparability between understudies answer and references answer. Programmed short answer scoring does not give the best outcome yet. Short answer has a restricted word in each answer. Each answer comprises of one expression to three sentences. Appraisal of the short portrayal that has set number of words requires uncommon taking care of, particularly in the weighting procedure. With the confinements of the way toward weighting the word, it is impossible with recurrence show, on the grounds that the words event is uncommon. This examination attempts to think about a few techniques that apply the covering strategies to decide the level of closeness between the references answer and understudies answer. From the examination in [18] it demonstrates that the strategy Cosine Coefficient has preferable outcomes over the Dice and Jaccard Coefficient strategies.

The work in [19], plans to build up a robotized exposition appraisal framework by utilization of AI systems by grouping a corpus of printed elements into modest number of discrete classifications, comparing to conceivable evaluations. Straight relapse procedure is used for preparing the model alongside utilizing different arrangements and grouping methods. Specialists expect to prepare classifiers on the preparation set, influence it to experience the long and short answer datasets, and after that measure execution our dataset by contrasting the acquired qualities and the dataset values. While their methodology is great, they have not utilized any correlations with demonstrate their outcomes. When alluding to long replies, an investigation in [20] utilizes a blended technique to break down review results. Likert scale questions were quantitatively dissected while open finished inquiries were subjectively investigated. Topical investigation was utilized to break down the subjective information the specialist gathered through four open-finished inquiries. In the initial step the analyst acquainted himself with the information by perusing and rehashing the reactions. Next beginning codes were made utilizing the respondents claim language. These codes were then translated and fell into general subjects that developed as the information was examined. As a last advance the topics were checked on to find designs and to inspect the all the more intriguing subjects with regards to more detail. This examination doesn't arrive at any resolutions on the strategies, yet is great from purpose of study. In [21] a vector based framework is proposed, this framework utilizes different sentence portrayal strategies and wide scope of comparability measures are looked at lastly a framework for short answer reviewing is proposed. The framework either beats the cutting edge frameworks on various informational indexes or accomplishes equivalent outcomes. Proposed framework manages the short answer reviewing issue as content similitude task. A little portrayal of each calculation or portrayal utilized is given nearby an illumination of the investigation setting to ease duplicating the announced outcomes. The vector-based proposed framework accomplished new cutting edge results on four informational indexes with two dialects and accomplished outcomes that not a long ways behind the Inter annotator understanding. The framework proposed consolidates different sorts of closeness with principle reliance on word vector portrayal. In [22], the analysts investigated the information utilized in Automatic Short Answer Grading (ASAG) look into, thinking about numerous perspectives from the language to number of inquiries, answers, and so on. At that point, they took a gander at which regular language handling and AI strategies are the most utilized in the field. From that point forward, they exhibited the center of the examination, how answers are displayed so as to extricate highlights that can anticipate their scores. Lastly, they indicated how analysts assessed their frameworks and how they can (or can't) be contrasted with one another. All introduced outcomes demonstrates the pith and development of ASAG investigate utilizing AI techniques. Open datasets are accessible for a really long time prior, and examine in the field is available to new methods, datasets and exceptionally to profound realizing, that has been as of late adding to a variety of regions and still very underexplored in ASAG. In [23], a methodology on long answer assessment utilizing lexical and semantic comparability measure has been introduced. The objective of this work is to present a framework which automatically assesses the long replies from the examinee and henceforth diminish the time and exertion of human intercession just as make the assessment method unbiased to the whole client. In this work, first the client answers (examinee answer) are coordinated with standard answers (analyst answer) utilizing lexical comparability measure. In testing stage, five arrangements of inquiry answers have been viewed as where each set contains a solitary inquiry from a subject space and its five unique answers.

The framework steadfast a worthy precision as indicated by human choice. In the following period of the work, both the appropriate responses have been analyzed utilizing semantic comparability measure. In this stage, the synonymous expressions of the watchwords from both the appropriate responses are recovered from the semantic lexicon WordNet to expand the needful and significant cover between the appropriate responses. Applying this semantic likeness estimating procedure on a similar inquiry answer sets, the precision of assessment has been expanded which is approved by a specialist. Another survey in [24], features on the plan and execution of a robotized assessment framework for short specialized replies by utilizing Machine Learning, alongside a near investigation of the different methodologies that have been utilized till current date, is given. The predominant framework has its very own respite as far as volume, staffing, variety, rightness inside the methods for evaluating. The proposed framework is an assessment framework which can recognize the coordinating catchphrases in the literary answers and assessing marks for the equivalent dependent on some past learning obtained by the machine. A linguistic connection based component extraction procedure is proposed for programmed assessment of enlightening kind answers. Understudies are additionally profited with a criticism framework that can enable them to improve their score. It likewise gives a stage to scholarly establishments to upgrade their framework which can give better outcomes in task assessment. While trying to explore the significance of lexical highlights in AES frameworks, another all-encompassing list of capabilities is created in [26] by joining famously known highlights. The joined list of capabilities contains 22 includes that catches five unique parts of composing characteristics. The significance of each element in the joined list of capabilities is tried by disposing of each element independently. It was discovered that utilizing the quantity of things in the article somewhat debases the AES framework execution. The noteworthiness of the joined list of capabilities is looked at against three best in class AES business frameworks and its execution was discovered equivalent. Additionally in [27], the analysts give examination of 21 cutting edge approaches for mechanized exposition assessment and feature their shortcomings and open difficulties in the field. We finish up with the discoveries that the field has created to the point where the frameworks give important criticism on understudies' composition and speak to a helpful supplement (not substitution) to human scoring. The way that an ever increasing number of traditional instructive methodologies has been automatized utilizing PCs raises concerns. Gigantic open online courses (MOOC) have progressed toward becoming piece of the instructive frameworks and are supplanting the customary educator - understudy connection and raise doubt about the instructive procedure in the study halls. While PC evaluating of different decision tests has been utilized for quite a long time, PC scoring of increasingly abstract material like papers is presently moving into the scholastic world. Computerized article assessment is playing one of the key jobs in the present improvement of the robotized instructive frameworks, including MOOC. Every one of these leaves many open inquiries in regards to the supplanting of human educators with PC, which ought to be mulled over later on and be replied with the further improvement of the field. In [28], the scientists proposed a way to deal with check the level of learning of the understudy/student, by assessing their spellbinding test answer sheets. By speaking to the engaging answer as diagram and contrasting it and standard answer are the key strides in our methodology. The student's distinct answer and standard answer is changed over into its graphical structure and afterward, to apply a portion of the similitude estimates, for example, string match, wordNet and spreading process for the computation of comparability score are the real strides in the proposed calculation. The calculation gives an answer for the mechanization of graphic answer assessment process.

III. RESULT ANALYSIS AND CONCLUSION

From the review we can observe that the techniques like vector based processing and graph mining are best for short answers, while the techniques like Two-Stage Learning Framework (TSLF) can be used for long answers. Thus, out of many options it is still the researchers choice on which method to choose based on the selected dataset, and referring the review on this text will assist them to do that task with utmost ease.

IV. FUTURE WORK

As a future work, researchers can explore various techniques like artificial intelligence and real time learning from previous question answer sets in order to re-adjust the weights of the algorithm and get a better chance of improving the accuracy of the existing assessment systems.

REFERENCES

1. Selvi, P., and A. K. Bnerjee. "Automatic short-answer grading system (ASAGS)." arXiv preprint arXiv:1011.1742 (2010).Walia, Tarandeep Singh, Gurpreet Singh Josan, and Amarpal Singh. "Quantum-based automated short answer scoring system." Modern Physics Letters B 32.33 (2018): 1850403.
2. Walia, Tarandeep Singh, Gurpreet Singh Josan, and Amarpal Singh. "Quantum-based automated short answer scoring system." Modern Physics Letters B 32.33 (2018): 1850403.
3. Sultan, Md Arafat, Cristobal Salazar, and Tamara Sumner. "Fast and easy short answer grading with high accuracy." Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2016.
4. Taghipour, Kaveh, and HweeTou Ng. "A neural approach to automated essay scoring." Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. 2016.
5. Salem, Safa Ben, Lilia ChenitiBelcadhi, and Rafik Braham. "Using concept maps in knowledge assessment of short and open answer questions." 2013 17th International Conference on System Theory, Control and Computing (ICSTCC). IEEE, 2013.Pribadi, FeddySetio, et al. "Automatic short answer scoring using words overlapping methods." AIP Conference Proceedings. Vol. 1818. No. 1. AIP Publishing, 2017.
6. Pramukantoro, Eko Sakti, and M. Ali Fauzi. "Comparative analysis of string similarity and corpus-based similarity for automatic essay scoring system on e-learning gamification." 2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS). IEEE, 2016.

7. Liu, Jiawei, Yang Xu, and Lingzhe Zhao. "Automated Essay Scoring based on Two-Stage Learning." arXiv preprint arXiv:1901.07744 (2019).
8. Fauzi, M. Ali, et al. "Automatic Essay Scoring System Using N-Gram and Cosine Similarity for Gamification Based E-Learning." Proceedings of the International Conference on Advances in Image Processing. ACM, 2017.
9. Patil, Shweta, and SonalPatil. "Intelligent tutoring system for evaluating student performance in descriptive answers using natural language processing." International Journal of Science and Research 3.9 (2014): 779-783.
10. Mathias, Sandeep, and Pushpak Bhattacharyya. "Thank "Goodness"! A Way to Measure Style in Student Essays." Proceedings of the 5th Workshop on Natural Language Processing Techniques for Educational Applications. 2018.
11. Kaur, Amarjeet, and M. Sasikumar. "A comparative analysis of various approaches for automated assessment of descriptive answers." 2017 International Conference on Computational Intelligence in Data Science (ICCIDS). IEEE, 2017.
12. Nandini, V., and P. Uma Maheswari. "Automatic assessment of descriptive answers in online examination system using semantic relational features." The Journal of Supercomputing (2018): 1-19.
13. Djamel, Bennouar, and DjellatouRachda. "Evaluating Short Answer Using Text Similarity Measures."
14. Rokade, Amit, et al. "Automated Grading System Using Natural Language Processing." 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT). IEEE, 2018.
15. Contreras, Jennifer O., ShadiHilles, and Zainab BintiAbubakar. "Automated Essay Scoring with Ontology based on Text Mining and NLTK tools." 2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE). IEEE, 2018.
16. Lahitani, AlfirnaRizqi, Adhistya Erna Permanasari, and Noor Akhmad Setiawan. "Cosine similarity to determine similarity measure: Study case in online essay assessment." 2016 4th International Conference on Cyber and IT Service Management. IEEE, 2016.
17. Gomaa, Wael H., and Aly A. Fahmy. "Short answer grading using string similarity and corpus-based similarity." International Journal of Advanced Computer Science and Applications (IJACSA) 3.11 (2012).
18. Pribadi, Feddy Setio, et al. "Automatic short answer scoring using words overlapping methods." AIP Conference Proceedings. Vol. 1818. No. 1. AIP Publishing, 2017.
19. Ramalingam, V. V., et al. "Automated Essay Grading using Machine Learning Algorithm." Journal of Physics: Conference Series. Vol. 1000. No. 1. IOP Publishing, 2018.
20. Lewis, John. "Ethical Implementation of an Automated Essay Scoring (AES) System: A Case Study of Student and Instructor Use, Satisfaction, and Perceptions of AES in a Business Law Course." Satisfaction, and Perceptions of AES In a Business Law Course (2013).
21. Magooda, Ahmed Ezzat, et al. "Vector based techniques for short answer grading." The Twenty-Ninth International Flairs Conference. 2016.
22. Galhardi, Lucas Busatta, and Jacques DuílioBrancher. "Machine Learning Approach for Automatic Short Answer Grading: A Systematic Review." Ibero-American Conference on Artificial Intelligence. Springer, Cham, 2018.
23. Goswami, Riya, et al. "A knowledge based approach for long answer evaluation." 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT). IEEE, 2017.
24. Moray, Shivam. "AUTOMATED SHORT-ANSWER GRADING SYSTEM USING MACHINE LEARNING: A REVIEW." Journal of Global Research in Computer Science 10.1 (2019): 13-16.
25. Shweta, M. Patil, and SonalPatil. "Evaluating Student Descriptive Answers using Natural Language Processing." Internation Journal of Engineering Research and Technology (IJERT) (2014).
26. Eid, Soha M., and Nayer M. Wanas. "Automated essay scoring linguistic feature: Comparative study." 2017 Intl Conf on Advanced Control Circuits Systems (ACCS) Systems & 2017 Intl Conf on New Paradigms in Electronics & Information Technology (PEIT). IEEE, 2017.
27. Zupanc, Kaja, and Zoran Bosnic. "Advances in the field of automated essay evaluation." Informatica 39.4 (2016).
28. Kaur, Amarjeet, et al. "Algorithm for Automatic evaluation of single sentence descriptive answer." International Journal of Inventive Engineering and Science 1.9 (2013): 112-121.

AUTHORS PROFILE

Simran Agrawal received Bachelor of Engineering Degree in Computer Science and Engineering from Shri Balaji Institute of Science and Technology, Rajiv Gandhi Proudyogiki Vishwavidyalaya University, Betul, India 2016 and persuing Master of Technology degree in Computer Science and Engineering from Shri Ramdeobaba college of Engineering and Management, Nagpur, India.

Dr. Avinash J. Agrawal received Bachelor of Engineering Degree in Computer Technology from Nagpur University, India and Master of Technology degree in Computer Technology from National Institute of Technology, Raipur, India in 1998 and 2005 respectively. He completed his Ph.D. from Visvesvaraya National Institute of Technology, Nagpur. His research area is Natural Language Processing and Databases. He is having 20 years of teaching experience. Presently he is Associate Professor in Shri Ramdeobaba college of Engineering and Management, Nagpur,India.