STM JOURNALS

Journal of
Multimedia Technology &
Recent Advancements

ISSN: 2349-9028
Volume 11, Issue 3, 2024
September–December
DOI (Journal): 10.37591/JoMTRA

https://journals.stmjournals.com/jomtra

Review JoMTRA

Integrated Qualitative Response Assessment System

Prasad A. Lahare[1,*], Smita K. Thakare[1]

Abstract

Examinations for universities and yearboards are traditionally administered offline, with a significant number of students opting for subjective exams. This preference stems from the labor-intensive nature of evaluating subjective responses, which requires considerable time and effort from educators. Additionally, subjective grading can be influenced by the evaluator's mood, leading to inconsistencies. In contrast, multiple-choice and objective questions are prevalent in entrance and competitive exams due to their ease of automated grading, which is error-free and resource-efficient. Currently, there are no systems available to automate the evaluation of descriptive (subjective) questions, which necessitates manual grading. Automating the evaluation process for descriptive responses could transform the education sector by addressing these challenges. An automated system could utilize technologies such as natural language processing (NLP), machine learning, and optical character recognition (OCR) to evaluate and rate written content. This approach would significantly reduce the time and effort required for grading, ensure consistent and unbiased evaluations, and handle large volumes of exam papers efficiently. Implementing such a system involves several steps: collecting a large dataset of graded papers, developing and training machine learning models, integrating OCR and NLP technologies, and rigorously testing the system for accuracy and reliability. Overcoming challenges such as the complexity and subjectivity of responses, as well as ensuring fairness, is crucial. Ultimately, automated evaluation of descriptive exam responses would enhance the efficiency, consistency, and scalability of the grading process, benefiting both educators and students.

Keywords: Machine learning, natural language processing (NLP), computer-based evaluation, natural language toolkit (NLTK), training phase, support vector machine (SVM)

INTRODUCTION

A prevalent practice in the assessment of student learning processes is the use of computer-based evaluation systems. Integrating computers into educational processes has fundamentally transformed learning systems. Initially, these computer-assisted assessment systems were designed to evaluate single-word responses such as those found in multiple-choice questions. They can also assess paragraph responses, for example, descriptive answers, based on keyword matching. However, a major limitation of these systems is that students may not recognize their errors and, thus, may not attempt to correct them. To address this issue and enhance student proficiency in English and grammar, a novel approach using artificial neural networks (ANNs) and natural language processing (NLP) algorithms was proposed.

This new approach aims to assess students' descriptive responses more effectively. The standard procedure for evaluating students' learning processes involves computer-based assessment of their responses. Initially designed for single-word

***Author for Correspondence**
Prasad A. Lahare
E-mail: prasadlahare7@gmail.com

[1]Assistant Professor, Department of Computer, Pune Vidyarthi Griha's College of Engineering & S. S. Dhamankar Institute of Management, Nashik, Maharashtra, India

Received Date: June 28, 2024
Accepted Date: September 13, 2024
Published Date: October 18, 2024

Citation: Prasad A. Lahare, Smita K. Thakare. Integrated Qualitative Response Assessment System. Journal of Multimedia Technology & Recent Advancements. 2024; 11(3): 18–24p.

and keyword-matching assessments, these systems evolved to include more sophisticated methods. Academic institutions and organizations that administer competitive exams can benefit from this technology. In practice, students write their responses on an answer sheet, which is then scanned and pre-processed. The system extracts text from the scanned answers as input. The model response sets provided by the evaluator were used to train the system.

The evaluation system was based on three criteria: keywords, grammar, and question-specific items. However, traditional keyword-matching approaches often fail to inform students of their mistakes, leading to repeated errors in subsequent tests and an inadequate improvement in their knowledge. Therefore, employing ANNs and NLP could significantly improve the evaluation process. In this approach, instructors create questions and answers for text mining, which is performed using the WordNet techniques and NLP. Computational linguistics and artificial intelligence have focused on the interactions between computers and human languages. WordNet organizes English terms into synsets, collections of synonyms with brief descriptions, and usage guidelines. To extract crucial keywords from student responses, POS taggers (part-of-speech taggers) are used in conjunction with AI. These keywords were categorized into technical, subordinate, and obligatory groups. WordNet provides synonyms for literal words in subordinate terms. Instructors can then supply these phrases to the system for student evaluation. This study aimed to evaluate student descriptive answers using ANN and NLP algorithms, developing a tool that utilizes the ANN algorithm for comparing standard answers and generating grades while employing the NLP algorithm for grammatical checking.

LITERATURE SURVEY

Wang J., Dong Y. Measurement of text similarity: a survey. Information. 2020. [1]. This study provides a comprehensive overview of text similarity measurements and offers a systematic examination of the field's research. Wang and Dong explored various approaches and strategies to measure text similarity, including their benefits and limitations. They discussed knowledge-based approaches that use pre-existing word similarity metrics to assess verbal and noun similarities, as well as fundamental word overlap techniques for adjective and adverb similarities. In addition, methods such as Dynamic Time Warping consider the semantic closeness of phrases. Machine translation metrics are used to evaluate the semantic similarity between a system's output and a set of reference translations. The integration of ontology in evolutionary algorithms for text clustering is also covered, illustrating evolving approaches to text similarity measurement. This comprehensive review highlights the complexity and diversity of measuring text similarity.

Yun W., Gao C., Jiang J., Zhang X., Yuan X., Han M. A survey on the techniques, applications, and performance of short-text semantic similarity [2]. This literature review thoroughly analyzes the methods, use cases and performance assessment strategies for short-text semantic similarity. The authors explored various methods used in NLP, which are crucial for tasks such as sentiment analysis, information retrieval, and text categorization. This research clarifies the diverse approaches and strategies for determining the semantic similarity between short sentences. Understanding the latest developments in this field is essential for effectively applying semantic similarity methodologies in NLP tasks. Proficiency in these methodologies and their practical applications is crucial for leveraging semantic similarity approaches in various NLP applications.

Evaluating student descriptive answers using natural language processing by Patil and Patil [3]. This study advances the field of educational evaluation by demonstrating the value of NLP in assessing students' descriptive responses. Patil and Patil employed NLP to analyze linguistic variations in student essays using multidimensional analysis to evaluate the corpus across different scores. Their approach provides valuable insights into an essay's language usage, topic, structure, vocabulary, and mechanics, highlighting areas for improvement in student writing. This study shows how NLP can enhance the assessment of written work.

Subjective answer evaluation using machine learning [4]. In recent years, there has been growing interest in machine learning techniques for subjective response assessments. These techniques are increasingly being applied in grading, decision-making, and scheduling tasks across various industries, including education. Machine learning has been used to assess programming evaluations through unit testing, grading short and long essays, and classifying images of breast cancer histology. Additionally, machine learning algorithms are integrated into systematic review platforms to aid decision-making by organizing and prioritizing references based on relevant comments. The widespread use of these algorithms in various tools and platforms underscores their versatility and applicability.

Document plagiarism detection using a new concept similarity in formal concept analysis. Plagiarism, or presenting someone else's work as one's own, is a significant issue in both academic and professional settings. The rapid increase in information accessibility owing to the digital era has made the detection of document plagiarism more critical. This study explores innovative solutions for plagiarism detection using new concept similarity in formal concept analysis. The authors' approach addressed the challenges of identifying instances of document plagiarism in the context of the growing availability of information.

AIM AND OBJECTIVE

1. The primary goal of the project was to provide an institution with more interactive and user-friendly software.
2. Because every student employs the same inference method, evaluating our system will guarantee consistency in marking.

MOTIVATION

One prevalent practice in many student learning process assessment domains is computer-based evaluation of student responses. The brilliant concept of integrating computers into the educational process has fundamentally altered the field of learning systems [6].

The computer-assisted assessment system is designed to grade multiple-choice questions that require single-word answers [7]. Additionally, it is possible to assess paragraph responses, for example, descriptive responses using keyword matching. Academic institutions can make extensive use of this method to verify answer sheets. It is also applicable to other organizations that administer competitive examinations. On the answer page, students write their responses [8]. Once the answer is prepared, the system extracts text from a scanned copy of the response as input. The moderator or evaluator supplies the model response sets. Next, the model responses were prepared. Three criteria form the basis of this system: keywords, grammar, and question-specific items [9].

ALGORITHMS

SVC (Support Vector Classifier)

SVC is a specific implementation of the Support Vector Machine algorithm designed specifically for classification tasks [10].

TF-IDF

Term frequency-inverse document frequency (TF-IDF) is a widely used statistical method in NLP and information retrieval [11].

NLTK

The natural language toolkit (NLTK) is a powerful library in Python for working with human language data. In the context of a subjective answer grading system, NLTK can be beneficial for various tasks related to NLP [12].

1. Text preprocessing
2. Stopword removal
3. Feature extraction

SYSTEM ARCHITECTURE

Figure 1 depicts a flowchart with two main phases: (1) the training phase and (2) the recognition phase. Below is the description of each phase [13].

1. *Training phase*
   i. *Past labeled data:* This block represents the input data that have been previously labeled, meaning that the correct answers or grades are already known.
   ii. *Data preprocessing:* This step involves cleaning and preparing the data for analysis, which may include tasks such as eliminating noise, normalizing the data, and dividing them into training and testing sets.
   iii. *Feature extraction:* This block involves extracting significant features from pre-processed data. These features are critical for making predictions and classification.
   iv. *Prediction (classifier):* In this step, a classifier is trained using the features extracted from previously labeled data. A classifier learns to predict the output based on these features.
   v. *Output:* The final block in the training phase, where the trained model produces the output, which can be the predicted grade or score.
2. *Recognition phase*
   i. *Login:* This process begins with a login step, where users are required to authenticate themselves.
   ii. *Authentication:* This step involves confirming the identity of the user. This is connected to the “Check Password” decision block.
   iii. *Check password:* A decision block where the system checks the entered password.
   iv. *No:* If the password is incorrect, the system returns to the authentication step.
   v. *Yes:* If the password is correct, proceed to the next step.
   vi. *Select question:* After successful authentication, the user selects a question to answer.
   vii. *Write answers:* The user writes the answers to the selected question.
   viii. *Feature extraction:* The written answers underwent feature extraction to identify the key elements required for evaluation.
   ix. *Prediction (classifier):* The features from the written answers are fed into the trained classifier during the training phase to predict the score or grade.
   x. *Output:* The final output is generated, which can be the predicted grades or scores for the answers written by the user.

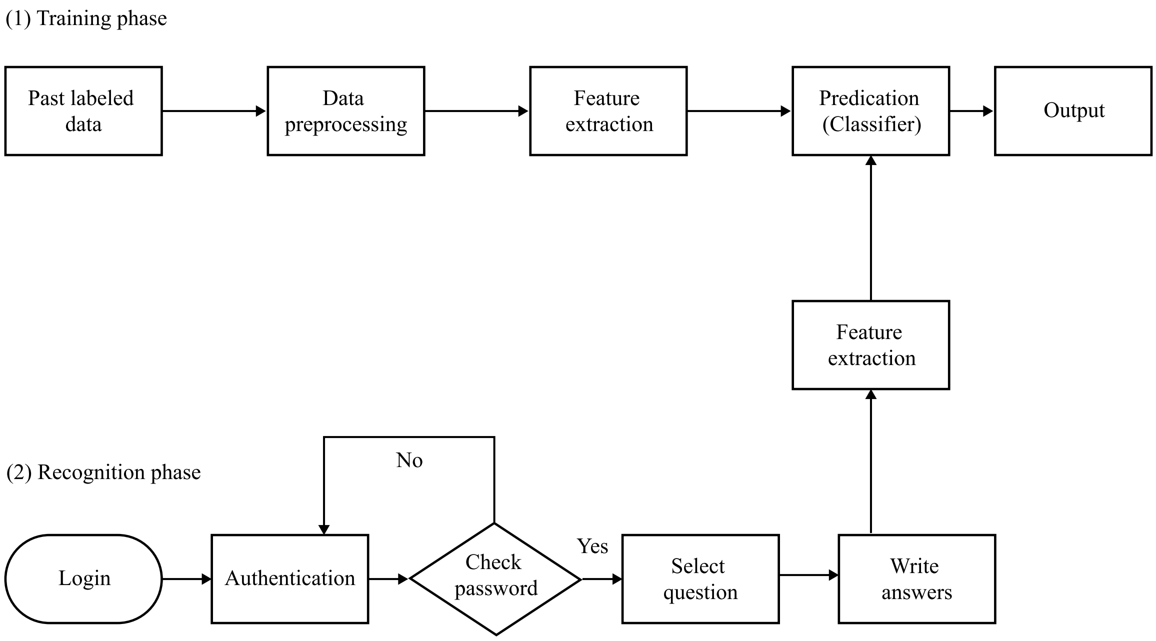

Figure 1. System architecture diagram.

APPLICATION

- *Automated essay scoring (AES):* This method ensures fairness and reduces teacher workload by efficiently and reliably grading large numbers of essays [14].
- *Personalized learning:* Examine students' open-ended replies to determine their areas of strength and weakness so that you can adjust learning paths and offer focused feedback. Adaptive assessments yield more accurate results and better learning outcomes by dynamically adjusting the complexity of the questions based on student performance [15].
- *Early feedback:* Give students early feedback on their essays and open-ended questions so that they may recognize and correct any misunderstandings.

RESULTS

This article explores the transparency and accessibility of the examined materials using a novel approach. This study investigates how this platform displays question papers and answer sheets in an image format, following the declaration of examination results [16]. Unlike traditional methods that often limit access to textual content, this website provides scanned images of both question papers and answer sheets, offering a unique perspective on examination transparency, as shown in Figures 2–6.

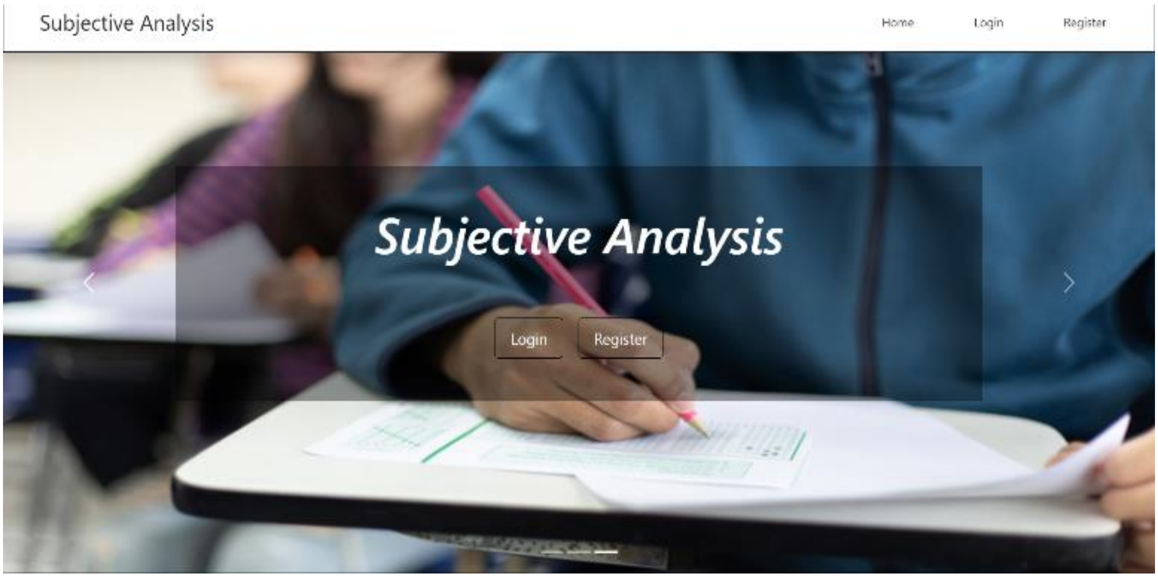

Figure 2. Subjective analysis login page.

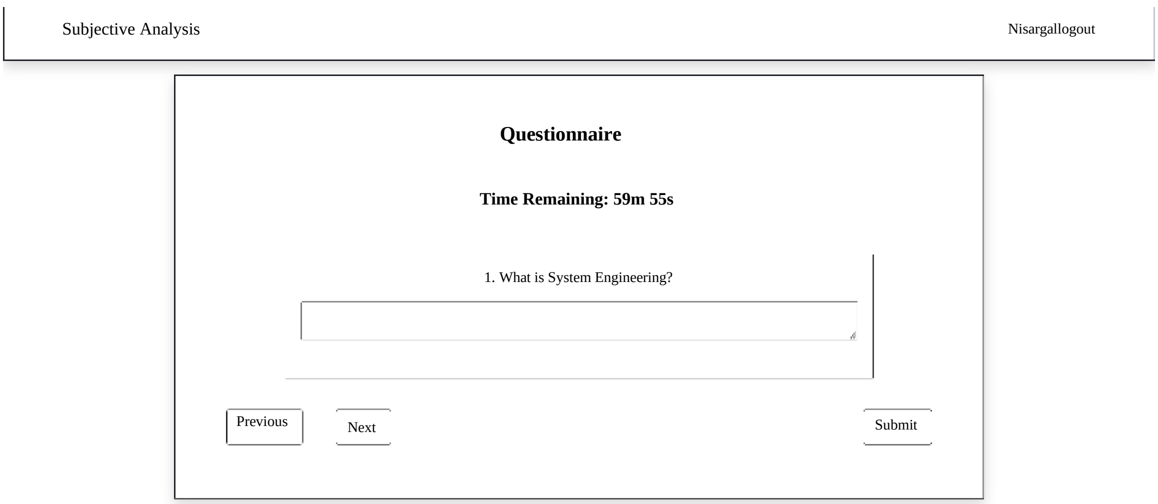

Figure 3. Questionnaire form.

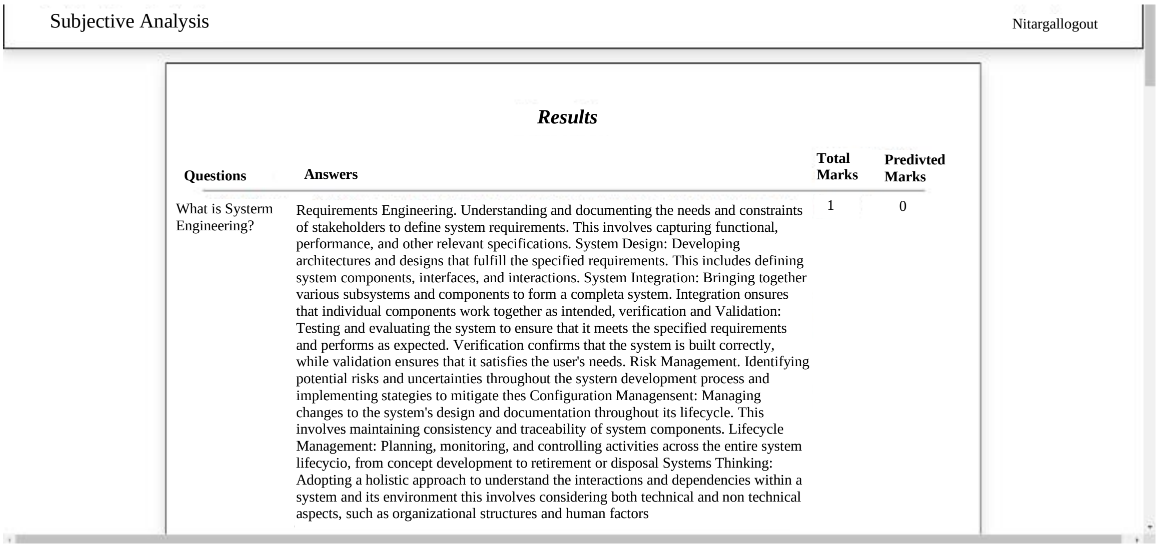

Figure 4. Answer sheet 1.

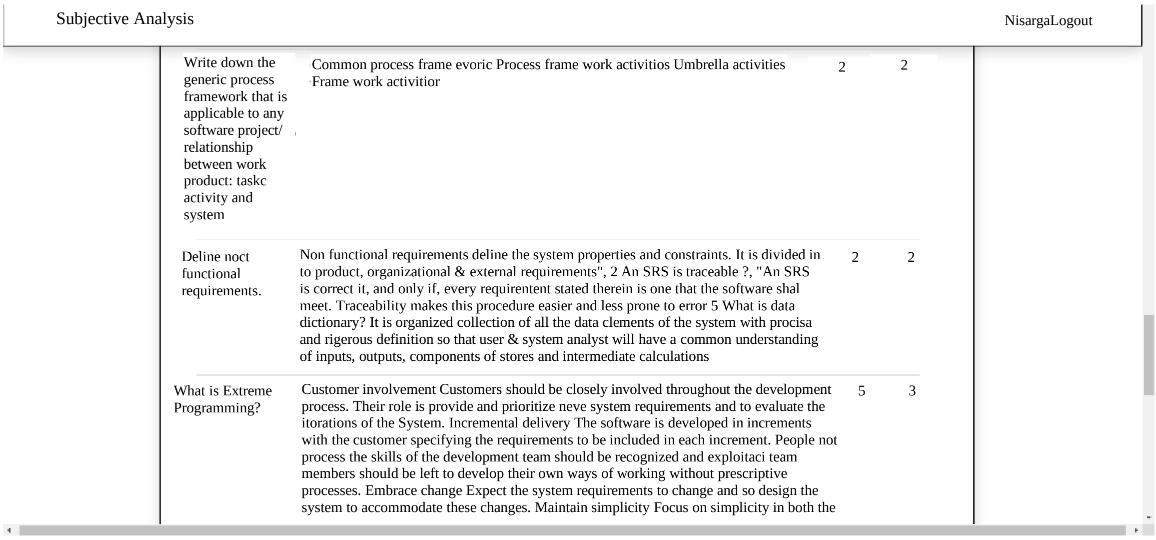

Figure 5. Answer sheet 2.

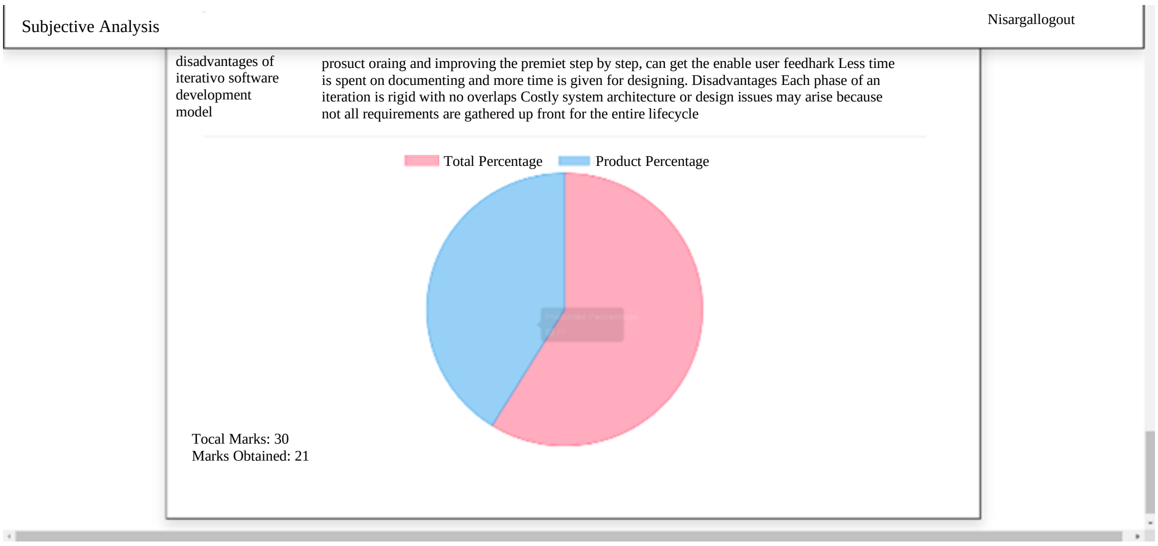

Figure 6. Result declaration.

CONCLUSION

The last phase of the project report, the Automatic Answer Checker, is now underway. The program was created for every potential mistake. We were thinking about making the system highly dependable and effective. Because of the exceptional robustness of the application, there are several possibilities for improvisation. The application will be authorized, verified, and put into use. To detect all syntactic problems in our keywords, the following work will involve developing an algorithm for the assessment. Next, we look at high-performance and high-equality approaches to fix them.

REFERENCES

1. Wang J, Dong Y. Measurement of text similarity: A survey. Information. 2020;11:421. DOI: 10.3390/info11090421.
2. Han M, Zhang X, Yuan X, Jiang J, Yun W, Gao C. A survey on the techniques, applications, and performance of short text semantic similarity. Concurr Comput Pract Exp. 2021;33. DOI: 10.1002/cpe.5971.
3. Patil MS, Patil MS. Evaluating student descriptive answers using natural language processing. Int J Eng Res Technol (IJERT). 2014;3.
4. Patil P, Patil S, Miniyar V, Bandal A. Subjective answer evaluation using machine learning. Int J Pure Appl Math. 2018;118:1–3.
5. Muangprathub J, Kajornkasirat S, Wanichsombat A. Document plagiarism detection using a new concept similarity in formal concept analysis. J Appl Math. 2021;2021:1–10. DOI: 10.1155/2021/6662984.
6. Hu X, Xia H. Automated assessment system for subjective questions based on LSI. In: 2010 Third International Symposium on Intelligent Information Technology and Security Informatics. 2010. p. 250–4. DOI: 10.1109/IITSI.2010.76.
7. Kusner M, Sun Y, Kolkin N, Weinberger K. From word embeddings to document distances. In: International Conference on Machine Learning; 2015. p. 957–66. PMLR.
8. Ouahrani L, Bennouar D. Ar.-ASAG an Arabic dataset for automatic short answer grading evaluation. In: Proceedings of the Twelfth Language Resources and Evaluation Conference; 2020. p. 2634–43.
9. Xia LZ, Ye JF, Luo DA, Guan MX, Liu J, Cao XM. Short text automatic scoring system based on Bert-BiLSTM model. J Shenzhen Univ Sci Eng. 2022;39:349–54. DOI: 10.3724/SP.J.1249.2022.03349.
10. Saha S, Dhamecha TI, Marvaniya S, Sindhgatta R, SenGupta B. Sentence level or token level features for automatic short answer grading?: Use both. In: Artificial Intelligence in Education. Proceedings, Part I: 19th International Conference, AIED 2018, London, UK, June 27–30, 2018, Vol. 19. Springer International Publishing; 2018. p. 503–17.
11. Jagadamba G. Online subjective answer verifying system using artificial intelligence. In: 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC). 2020. p. 1023–7.
12. Seshathriaathithyan S, Sriram MV, Prasanna S, Venkatesan R. Affective—Hierarchical classification of text—An approach using NLP toolkit. In: 2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT). 2016. p. 1–6. DOI: 10.1109/ICCPCT.2016.7530228.
13. Zhang K. Research on the optimizing method of question answering system in natural language processing. In: 2019 International Conference on Virtual Reality and Intelligent Systems (ICVRIS). 2019. p. 251–4. DOI: 10.1109/ICVRIS.2019.00069.
14. Tereshchenko G, Gruzdo I. Overview and analysis of existing decisions of determining the meaning of text documents. In: 2018 International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T). 2018. p. 645–53. DOI: 10.1109/INFOCOMMST.2018.8632014.
15. Ying L, Huidi L. Review of text analysis based on deep learning. In: 2020 International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI). 2020. p. 384–8. DOI: 10.1109/ICHCI51889.2020.00087.
16. Han H, Wang Q, Chen C. Policy text analysis based on text mining and fuzzy cognitive map. In: 2019 15th International Conference on Computational Intelligence and Security (CIS). 2019. p. 142–6. DOI: 10.1109/CIS.2019.00038.