Subjective Examination Evaluation Based on Spelling Correction and Detection Using Hamming Distance Algorithm

Madhav A. Kankhar(✉) and C. Namrata Mahender

Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India
madhavkankhar.07@gmail.com, cnamrata.csit@bamu.ac.in

Abstract. The usage of Online examination systems in education is not a new concept for the past several years, Objective assessments have been conducted using examination systems. This research examines E-examinations that include an E-assessment system that can be used for subjective questions. The present work aims to investigate the spelling errors, for experiment 12th standard Business studies paper is collected from a CBSC school. The exam was conducted on Microsoft teams. Hamming distance for word matching or spelling mistakes is deployed on one word and one sentence. Types of Error considered while evaluating spell mistakes are Inserting, Missing, Replacement or Substituting and Transposition error or Swap which resulted in a 46.62% correction on the overall result of subjective inspection for spell mistake in answer assessment.

Keywords: Online subjective exam · Question answering system · Hamming distance · Online examination · NLP

1 Introduction

The educational system has experienced various changes in the recent decade, the most notable of which is the shift in learning and examination methodologies. Students' and parents' conceptions of learning are changing as educational institutions steadily move toward online instructional techniques and tests. The need for an automated online subjective testing and evaluation process has grown as a result of the scenario in Covid -19. While designing and evaluating a completely automated Question answering system, phrasing is a challenge [1]. Exams are an important part of a student's education since they evaluate their knowledge and understanding of a subject. As a result, An examination system must include the preparation of a fresh paper for each student as well as follow-up assessments. With the rapid growth of modern education, the notion of an E-learning system was established to improve online course teaching by allowing teachers to administer online assessments through virtual classrooms. Electronic learning addresses a number of problems that students face, including the expensive expense of traditional academic courses [2]. Exam paper preparation and assessment take a lot

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
M. Singh et al. (Eds.): ICACDS 2022, CCIS 1614, pp. 222–234, 2022.
https://doi.org/10.1007/978-3-031-12641-3_19

of time and effort, use up a lot of resources, and place a lot of pressure on course instructors. As a result, E-examination systems are crucial in colleges and institutions since they allow all students in diverse locations to take electronic tests. Electronic assessments for massive open online courses (MOOCs) have been developed by universities such as MIT, Berkeley, and Stanford [3]. E-examination systems can electronically verify and set exam papers, assign scores, and grade answers swiftly and efficiently. These systems require fewer resources and less work on the part of the consumers. Traditional examination systems, on the other hand, necessitate the use of physical resources such as pens and paper, as well as more useful work and time. Exams containing objective questions are now evaluated only by existing electronic-examination systems. However, researchers have recently found the necessity to use this method to examine subjective questions [4]. Manually generated electronic text is full of errors, including spelling and typing errors. Internet search engines, for example, have been chastised for neglecting to spell check the user's query, which would have averted a plethora of pointless searches if the user had misspelled one or more query terms. An approximate word-matching algorithm is required to spot errors in queries with little or no contextual information and give words that are most similar to each misspelled word. [5].

Spelling error correction has been a long-standing Natural Language Processing (NLP) difficulty due to the vast amount of informal and unedited text generated online, such as web forums, tweets, blogs, and emails. It's become particularly significant recently as a result of the several potential applications for the vast volume of unstructured and unedited material generated online, such as web forums, tweets, blogs, and email. Misspellings in such material can cause increased sparsity and inaccuracy in several NLP applications, such as text summarization, sentiment analysis, and machine translation [6].

2 Subjective Examination

Depending on the length of the question, subjective questions can be answered in a few paragraphs or a few pages. As a result of the necessity to use an assessing technique while evaluating those questions, examining subjective replies will take time. Examiners may become frustrated as the action is repeated multiple times and students' comments become increasingly ludicrous. There were mistakes committed to evaluating the subjective questions' responses. As a result, a number of mechanisms are required to modify subjective responses. Human evaluation can be more effective when dealing with subjective difficulties such as sensitive judgments, sophisticated reasoning, and attitude expression. The human evaluator, on the other hand, invested a significant amount of time, sensitivity, and skill in evaluating the responses in order to receive feedback after a delay for scoring an inexperienced examiner would reduce the accuracy with which subjective answers were evaluated and the result in a plethora of redundant processes to correct the evaluation. The SQ&A System efficiently assesses subjective responses to save time. The system also assists examiners in enhancing the accuracy of subjective question judgment by running the algorithms. Aside from that, the strategy may assist professors in keeping track of their students' records, allowing them to create a more accurate graph of their academic status [7].

3 Related Work

Maram et al. [8] present an Arabic-language AEE stands for Automatic Evaluation of an Essay. The system makes use of a hybrid approach that combines the LSA and the RST (rhetorical structure theory) algorithms. The LSA method aids in the semantic analysis of the essay, whereas the RST method assesses the writing method as well as the essay's cohesion. Even if two texts do not include similar words, the LSA approach calculates their similarity ratio. A training phase and a testing phase are used by the system to process the input essay. The LSA method is used after calculating the average number of words in each essay, determining the top ten visible terms on a given topic, and determining the average number of words in each essay. The following steps are included in the testing phase: 1) Calculate the LSA distance 2) Counting how many words a vernacular has. 3) Counting the number of times a sentence is repeated. 4) Determining the length of the essay. 5) Counting how many spelling mistakes there are 6) Using the RST method 7) Examining the essay's overall coherence on the subject. The final score is then calculated using two phases and the LSA cosine distance between the input and training essays. The system assigns grades to schoolchildren's essays based on three criteria: 40% for writing approach, 50% for essay cohesion, and ten percent for spelling and grammar errors.

An automatic evaluation approach for descriptive English answers with several phrases is proposed by Anirudh et al. [9]. For questions in professional courses, the system evaluates the student's response using an answer key. Among the natural language processing methods used are Wu and Palmer's Longest Common Substring (LCS), LSA Cosine Similarity, and Pure PMI-IR. The similarity scores are then extracted from algorithms and blended using logistic regression to get a score that the instructor recommends. Each word in the student's answer is compared to each word in the answer key using the Wu-Palmer method. If both words are found in the English dictionary, the Wu-Palmer approach calculates a similarity score. Otherwise, if both terms aren't in the dictionary, the edit distance is used to compare them. LCS was used to compare both the student's answer and the answer-key phrases. Using the similarity matrix method, the LCS similarity score was combined with a Wu-Palmer approach similarity score. To determine the degree of similarity between the student's answer and an answer key, the algorithms compare them.

Ishioka and Kameda [10] propose the jess system, which is an automated Japanese essay rating method. In Japan, the system is used to grade essays for university entrance exams. The essay is graded on three criteria: eloquence, content, and organization Rhetoric is a syntactic variety that evaluates readability, lexical diversity, the number of large words, and the percentage of passive sentences. The process of presenting and integrating concepts in an essay is referred to as "organization." Jess evaluates the document's logical structure and looks for clear conjunctive sentences for the organization assessment. Content refers to material that is relevant to the issue, such as the precise information presented and the word used. Jess uses a technique called LSA to analyze content, which can be used to determine whether the contents of a written essay are appropriate for the essay topic. Jess uses learning models based on editorials and articles from the Mainichi Daily News newspaper.

In [11] To analyse online descriptive type students' replies, proposes utilising the Hyperspace Analog to Language (HAL) methodology and the Self-Organizing Map (SOM) method. To assess a learner's response, the student writes it down and sends it to HAL as input From an n-word vocabulary, HAL generates a high-dimensional semantic matrix. A method for building a matrix that involves the corpus motivating a window of length "1" by incrementing one word at a time. HAL disregards punctuation and sentence breaks, transforming each word into numeric vectors that express information about its meanings. Inside the window, "d" denotes the distance between two words, whereas "(l − d + 1)" denotes the weight of a word association. The terms in this matrix are displayed in order of their lexical co-occurrence. Every word in the row vector appears based on the co-occurrence data for words that come before it, and every word in the column vector appears based on the co-occurrence data for words that come after it. The SVD function is used to turn the matrix into a singular value. The HAL-generated vector is fed into the Self-Organizing Map as an input (SOM). SOM stands for neural technique. SOM creates a document map using vectors. The document will then be shared with neighboring neurons. Other clustering algorithms such as Farthest First, Expectation-Maximization (EM), Fuzzy c-Means, k-Means, and Hierarchical were compared to SOM's results. They came to the conclusion that SOM rewards exceptional performance.

Raheel and Christopher [12] provide a one-of-a-kind solution for automatically marking short answer questions. The authors describe the system's architecture, which is comprised of three phases that address the student's response and compute the grade for the student's response. The first step is to use an Open Source spell checker like JOrtho to verify and correct your spelling. 2) Parsing the student's response with the Stanford Parser. This statistical parser can generate extremely precise parses. The parser outputs the following findings, which are part of the speech tagged text and design-dependent grammatical relationships between singular words. 3) The third part of the processing answer is a comparison of the tagged text with syntactical structures provided by writers in Question and Answer Language. This phase is managed by the syntax analyzer. The design also incorporates a grammatical relation analyzer, which examines the grammatical relations in the student's response to the examiner's grammatical relations. The final responsibility in the comparison phase is to pass the data aggregated from the syntax analyzer and the grammatical relation analyzer to the marker, who calculates the final grade of the answer.

For a student's answer test of a short essay, Mohd et al. [13] developed an automatic marking method. The system was used to process sentences written in the Malay language, which required the use of technology. Grammatical Relations (GR) from Malay sentences are represented using the syntactic annotation and dependency group approaches proposed in [12]. Tokenizing, recognizing, collocating, and extracting the GRs to process the sentences from the marking scheme and the students' answers are all entries to the Computational Linguistic System (CLS). The system incorporates a database with a table of Malay words and their Parts of Speech to assist the CLS (POS). To compute the grade for the student's answer, compare the GR received from the students' responses to the GR for the marking scheme. Compare the following sentence components: subject to subject, verb to verb, object to object, and phrase to phrase, to put it another way. The authors put the system through its paces to evaluate how it stacks up against human-awarded grades. To assign a grade to each question, they picked Malaysian teachers with prior experience in marking the scheme. The test conditions had been established. The test examines if the system can provide grades that are comparable to those provided by professors.

In their Indonesian essay evaluation, Lahitani et al. [14] used the TF-IDF method and the cosine similarity methodology to determine the degree of similarity. The test datasets consisted of ten student documents obtained from an e-learning source that were examined for similarity to the documents provided by their five experts. The best degree of cosine similarity was 0.39, and it was calculated using a ranked-based methodology.

4 Types of Spelling Error

Error patterns, or patterns of spelling errors, were used to develop techniques. As a result, various research on the types and trends of spelling errors have been done. Damerau's research is the most well-known of all of them. There are two types of spelling problems, according to his research. Typographical errors and cognitive errors [15].

4.1 Typographical Errors

This error occurs when the correct spelling of a word is known yet the term is mistyped by accident. These errors do not comply with any linguistic criterion because they are usually always related to the keyboard. Damerau's research found that 80% of typographic errors fit into one of four categories.

1) Inserting a single letter, such as "Obsolete for Obpsolete".
2) Deleting a single letter, such as "Obsolete for Osolete".
3) Substituting a single letter, such as "Obsolete for Obselete".
4) Transposition of two adjacent letters, such as typing "Obsolete for Oboslete".

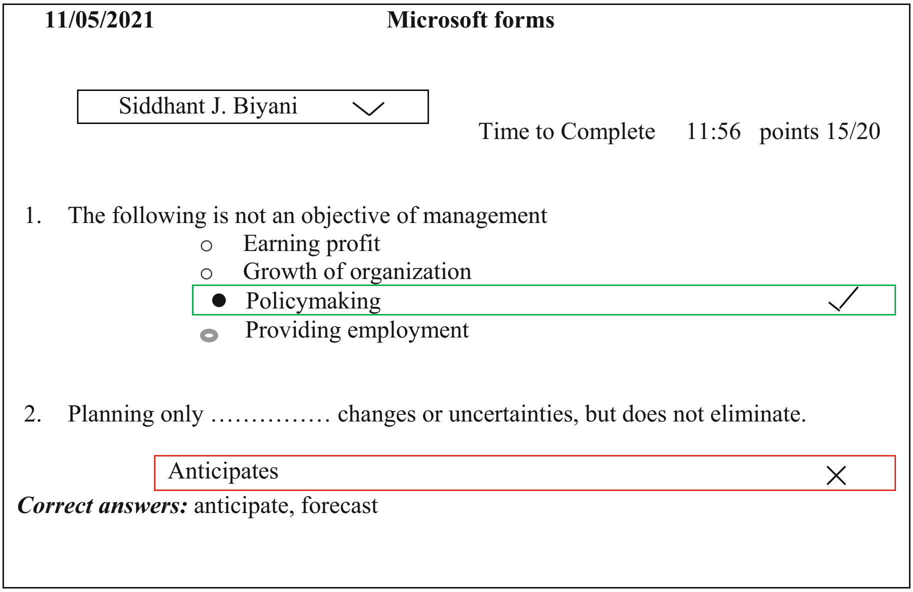
Single errors can be generated by any of the editing operations listed above.

4.2 Cognitive Error

These are mistakes that arise when the correct spelling of a term is unknown. In the instance of cognitive errors, the misspelled word's pronunciation is identical or almost identical to the intended right word's pronunciation [16].

5 Method

Microsoft team was used to perform the test. The question paper for 12th Business studies was prepared which had a combination of objective (MCQ) and subjective questions, here subjective questions were limited to one word to one sentence responses. The response of 63 students was recorded for performing the experiment. Figure 1 shows the sample of the student answer sheet.


11/05/2021 **Microsoft forms**

Siddhant J. Biyani

Time to Complete 11:56 points 15/20

1. The following is not an objective of management
   - Earning profit
   - Growth of organization
   - Policymaking
   - Providing employment

2. Planning only …………… changes or uncertainties, but does not eliminate.

Anticipates

Correct answers: anticipate, forecast

Fig. 1. Sample of student answer sheet

6 Proposed System

The question paper is provided, together with a model answer, and is used as a template for evaluating the student's answer sheet. The system's working model is depicted as a block diagram in Fig. 2.

Question file along with the module answer act as a template, which is stored separately, and the responses of students are collected at one place for evaluation of objective done, objective is done auto, while the one word and one sentence are evaluated using our proposed method as follows.

a) For one word the word is checked for spelling mistakes, case sensitivity is removed before performing hamming distance.
b) For one sentence, the sentence POS tagged & it's grammatically checked & then only it passed to our spell checker as discussed above in 'a'.

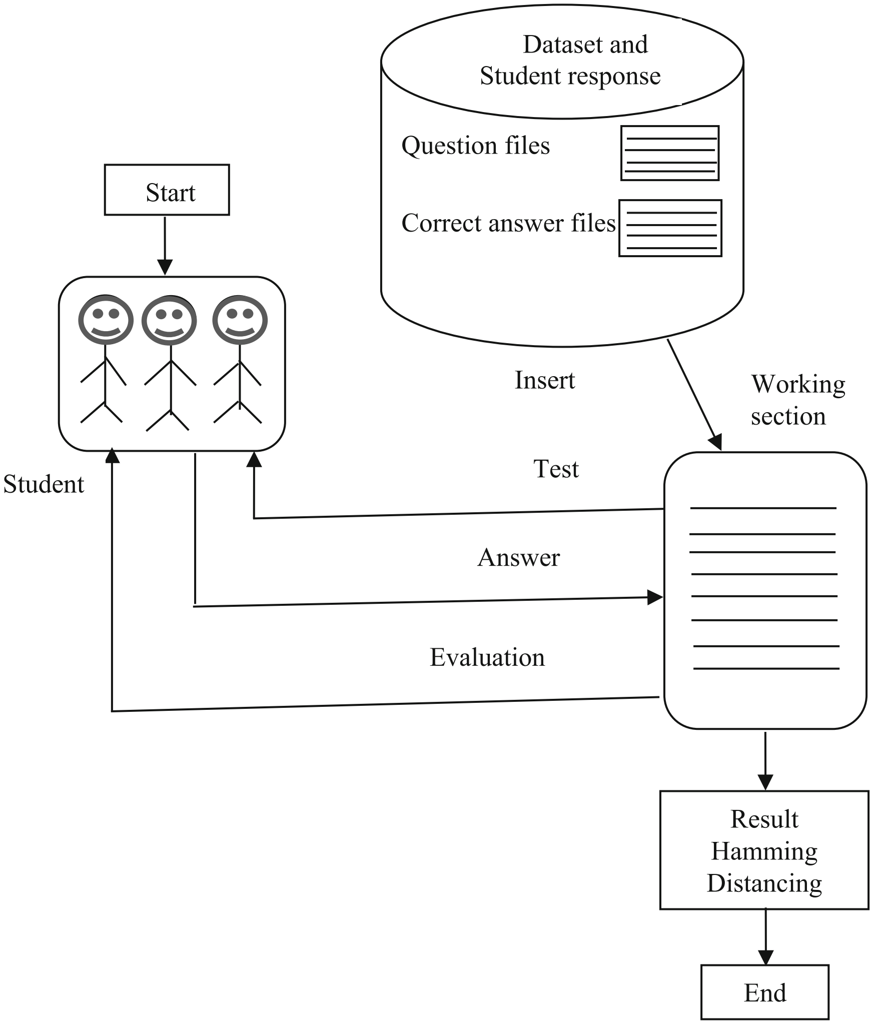


Fig. 2. Block diagram of the working model of the system

7 Hamming Distance

The Hamming Distance metric examines the similarity of any two texts of the same length, where the Hamming distance is the number of differences between the corresponding letters. To grasp the idea of hamming distance, consider any two texts.

The abbreviations "ABCDEF" and "ABCDDSQ" stand for "ABCDEF" and "ABCDDSQ," respectively. The distance between character A in the first position of the text "ABCDEF" and character A in the first position of the text "ABCDSQ" is zero. In the second, third, and fourth locations in both texts, the characters "BCD" are identical, and the Hamming distances are zero. The characters E and F of the first text differ from the characters S and Q of the second text at the same fifth location, hence the Hamming distances are 1. The Hamming distance is used to calculate the distance between binary vectors in binary texts. The Hamming distance is used for error repair and detection in intra-network data transmissions [17].

Using hamming distance we have calculated four types of error

1) Insert: - Means one extra character add here.
2) Missing: - It means some character missing there.
3) Replacement/Substituting: - It means any particular character replaced with another word.
4) Transposition error/Swap: - It means that the character position is changed from the actual alphabet position in a word.

The percentage error calculated for the responses collected is as follows, the percentage error rate is shown in Table 2.

Example:

See Table 1.

Table 1. Shows the example four types of error

Word	Insert	Missing	Replacement	Swap
Obsolete	Obsoleete	Obslete	Obxolete	Obsolete
Supervision	Supervviision	Supervison	Supercision	Supervisino

Table 2. The percentage error for subjective examination

Types of error	Percentage (%)
Inserting	27.72
Missing	11.97
Replacement/Substituting	12.6
Transposition error/Swap	6.3

While performing spell check we came across that position of the character wrongly inserted or swapped or substituted makes a difference while evaluating the answers for example color or colors may be taken as correct answers for the expected answers colors. Table 4 shows the error at the position in the word.

Due to misspelled alphabet change the meaning of actual word or meaning less word showing in Table 3.

Example:

Table 3. Show changes meaning of the word

Sr. No.	Word	Change meaning
1	Desert	Dessert
2	Heal	Heel
3	Mail	Male
4	Accept	Except

Table 4. Number of characters making error

Error position (in no. of character) in word	Percentage of words
1	40.95
2	8.19
3	3.15
4	1.89

8 Evaluation and Result

Table 5. Shows the overall evolution of spell mistake

Sr. no.	Question	Total question	Right answer	Wrong answer	Spell mistake	Blank answer
1	Anticipate, forecast	63	1	61	1	0
2	Market orientation	63	38	17	3	6
3	Responsibility	63	39	13	5	1
4	Supervision	63	27	35	0	1
5	Threat	63	8	14	37	5
6	Unity Of command	63	29	27	2	5
7	Obsolete	63	11	36	2	14
8	Motivation	63	57	3	3	0

(*continued*)

Table 5. (*continued*)

Sr. no.	Question	Total question	Right answer	Wrong answer	Spell mistake	Blank answer
9	Casual callers	63	38	9	15	1
10	Compensation	63	19	43	1	0
11	Workforce analysis	63	21	33	5	4
Overall Calculation		**756**	**288**	**291**	**74**	**37**

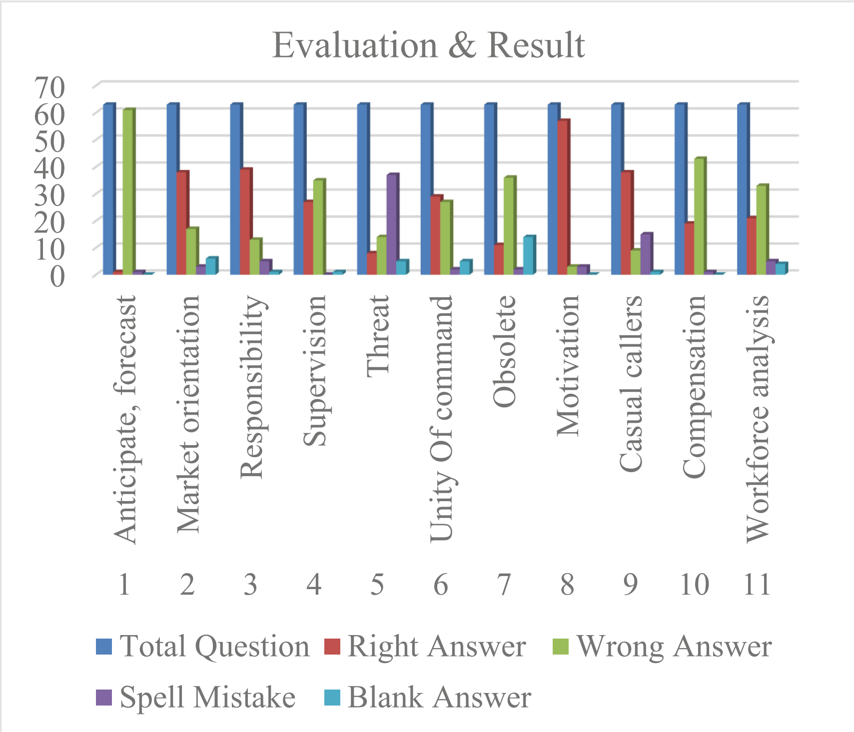


Fig. 3. Evaluation & result in column

The above Table 5 shows the result of subjective evaluation, in our system, we have constructed the exam like one sentence, one-word answer, multiple-choice question (MCQ), filling the blanks. We only use for evaluation in a survey that is an objective question all over 11 questions are there. Anticipate, forecast question attempted by total candidate 63, 1 student-written the right answer, 61 students given the wrong answer, 1 student given the spelling mistake. Market Orientation question attempted by a total of candidates 63, 38 students write the right answer, 17 students given the wrong answer, 3 students given the spelling mistake, 6 students do not answer that question. Responsibility question attempted by a total of candidates 63, 39 students written the right answer, 13 students gave the wrong answer, 5 students given the spelling mistake, 1 student does not answer that question. Supervision question attempted by a total of candidates 63, 27 students write the right answer, 35 students given the wrong answer, 1 student does not answer that question. Threat question attempted by a total of candidates 63, 8 students wrote the right answer, 14 students given the wrong answer, 37 students given the spelling mistake, 5 students do not answer that question. Unity of Command question attempted by total candidate 63, 29 students written the right answer, 27 students given the wrong answer, 2 students given the spelling mistake, 5 students do not answer that question. Obsolete question attempted by total candidate 63, 11 students write the right answer, 36 students given the wrong answer, 2 students given the spelling mistake, 14 students do not answer that question. Motivation question attempted by a total of candidates 63, 57 students write the right answer, 3 students gave the wrong answer, 3 students given a spelling mistake. Casual Callers question attempted by a total candidate 63, 38 students wrote the right answer, 9 students gave the wrong answer, 15 students given a spelling mistake, 1 student does not answer that question. Compensation question attempted by a total of candidates 63, 19 students write the right answer, 43 students gave the wrong answer, 1 student given a spelling mistake. Workforce Analysis question attempted by a total of candidates 63, 21 students wrote the right answer, 33 students gave the wrong answer, 5 students given a spelling mistake, students do not answer that question. Overall attempted this exam student is 756, total right question attempted by a student is 288, the total wrong answer attempted by a student is 291, total spell mistake question attempted by a student is 74, and blank answer question student not attempted by a student that is 37 (Fig. 3).

Correct spellings are very important for word-based word-based well as one sentence, misspelled words if have a more than 2 to 3 positional changes of the alphabet from the original word then might affect in (− negative) or fewer marks to the evaluation. As this paper has only taken care of topological error, the range of alphabet replacement & one alphabet insertion has shown more frequently occurring in the answer sheets, which cost to the overall performance of students.

As cognitive error has not been part of the present study still the cause of wrong answers can be termed as cognitive & it is around 38%.

9 Conclusions

The Question Answering System is the most effective technique for keeping track of valid and correct replies to candidate questions that are answered in natural language rather than via a query. The basic purpose of QAS, like the other queries submitted by users, is to receive accurate answers. Each individual's grasp of the subject is influenced by their expressive power, language used, and comprehension of the issue, and all of these courses have significant differences in subject and writing style, i.e. the reaction varies from person to person. The work focuses on the Question Answering System's spelling error. Error patterns, or patterns of spelling errors, were used to develop techniques. As a result, various research on the types and trends of spelling errors have been done. According to this experiment, the overall mistake corrected is 46.62% when using hamming distance for word matching or spelling mistakes.

Acknowledgment. The authors would like to acknowledge thanks to Chattrapati Shahu Maharaj Research Training & Human Development Institute (SARTHI), Pune they awarded fellowship, CSRI DST Major Project sanctioned No. SR/CSRI/71/2015(G), and also thanks to Computational and Psycholinguistic Research Lab Facility supporting this work and Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India.

References

1. Kankhar, M.A., Mahender, C.N.: Word level similarity auto-evaluation for an online question answering system. J. Eng. Res. (Kuwait) (2021)
2. Alrehily, A.D., Siddiqui, M.A., Buhari, S.M.: Intelligent electronic assessment for subjective exams. ACSIT, ICITE, SIPM, pp. 47–63 (2018)
3. Dreier, J., Giustolisi, R., Kassem, A., Lafourcade, P., Lenzini, G., Ryan, P.Y.: Formal analysis of electronic exams. In: 2014 11th International Conference on Security and Cryptography (SECRYPT), pp. 1–12. IEEE, August 2014
4. Kudi, P., Manekar, A., Daware, K., Dhatrak, T.: Online examination with short text matching. In: 2014 IEEE Global Conference on Wireless Computing & Networking (GCWCN), pp. 56–60. IEEE, December 2014
5. Hodge, V.J., Austin, J.: A comparison of standard spell-checking algorithms and a novel binary neural approach. IEEE Trans. Knowl. Data Eng. **15**(5), 1073–1081 (2003)
6. Farra, N., Tomeh, N., Rozovskaya, A., Habash, N.: Generalized character-level spelling error correction. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, vol. 2, Short Papers, pp. 161–167, June 2014
7. Kankhar, M.A., Sayyed, S.N., Mahender, C.N.: Challenges in Online Subjective Examination Systems: An Overview (2021)
8. Al-Jouie, M.F., Azmi, A.M.: Automated evaluation of school children's essays in Arabic. Procedia Comput. Sci. **117**, 19–22 (2017)
9. Kashi, A., Shastri, S., Deshpande, A.R., Doreswamy, J., Srinivasa, G.: A score recommendation system towards automating assessment in professional courses. In: 2016 IEEE Eighth International Conference on Technology for Education (T4E), pp. 140–143. IEEE, December 2016
10. Ishioka, T., Kameda, M.: Automated Japanese essay scoring system: jess. In: Proceedings of 15th International Workshop on Database and Expert Systems Applications, 2004, pp. 4–8. IEEE (2004)
11. Meena, K., Raj, L.: Evaluation of the descriptive type answers using hyperspace analog to language and self-organizing map. In: 2014 IEEE International Conference on Computational Intelligence and Computing Research, pp. 1–5. IEEE, December 2014
12. Siddiqi, R., Harrison, C.: A systematic approach to the automated marking of short-answer questions. In: 2008 IEEE International Multitopic Conference, pp. 329–332. IEEE, December 2008
13. Ab Aziz, M.J., Dato'Ahmad, F., Ghani, A.A.A., Mahmod, R.: Automated marking system for short answer examination (AMS-SAE). In: 2009 IEEE Symposium on Industrial Electronics & Applications, vol. 1, pp. 47–51. IEEE, October 2009

14. Lahitani, A.R., Permanasari, A.E., Setiawan, N.A.: Cosine similarity to determine similarity measure: study case in online essay assessment. In: 2016 4th International Conference on Cyber and IT Service Management, pp. 1–6. IEEE, April 2016
15. Greene, D., Parnas, M., Yao, F.: Multi-index hashing for information retrieval. In: Proceedings 35th Annual Symposium on Foundations of Computer Science, pp. 722–731. IEEE, November 1994
16. Geisler, J.: Error Detection & Correction. Taylor University, Computer and System Sciences Department (2005)
17. Al-azani, S.A., Mahender, C.N.: Improve Hamming character difference based-on derivative lexical similarity and right space padding