

PERFORMANCE EVALUATION OF LSA, NMF and ILSA IN ELECTRONIC ASSESSMENT OF FREE TEXT DOCUMENT

ABSTRACT

Aims: To evaluate the performance of an Improved Latent Semantic Analysis (ILSA), Latent Semantic Analysis (LSA), Non-Negative Matrix Factorization (NMF) algorithms in an Electronic Assessment Application using metrics, Term Similarity, Precision, Recall and F-measure functions, Mean divergence and Assessment Accuracy.

Methodology: The three algorithms were separately applied in developing an Electronic Assessment application. One hundred students' responses to a test question in an introductory artificial intelligence course were used. Their performance was measured based on the following metrics, Term Similarity, Precision, Recall and F-measure functions, Mean divergence and Assessment Accuracy.

Results: ILSA outperformed the LSA and NMF with an assessment accuracy of 96.64, mean divergence from manual score of 0.03, and recall, precision and f-measure value of 0.83, 0.85 and 0.87 respectively.

Conclusion: The research observed the performance of an improved algorithm ILSA for electronic Assessment of free text document using Adequacy in Semantic Representation, Retrieval Quality and Assessment Accuracy as performance metrics. The results obtained from the experimental designs shows the adequacy of the improved algorithm in semantic representation, better retrieval quality and improved assessment accuracy.

Keywords: Electronic Assessment, Latent Semantic Analysis, Ant Colony Optimization

1. INTRODUCTION

Electronic Assessment is the use of information technology for any assessment-related activity (Jordan 2013). An E-assessment application is adjudged worthwhile if it generates nearly human grade assessment. There are two basic approaches to Electronic Assessment. These are Information Retrieval (IR) based and Linguistics-based. The IR approach initially captures the semantic content of the documents involved in the assessment in a document-term matrix using co-occurrence statistics, preprocess and reduce the dimension of the semantic representation before comparing the resultant vectorial representation of individual student's response with that of the gold standard essay. Linguistic approaches emphasize the use of structures to decode the semantics. LSA, NMF and ILSA belong to the information retrieval approach used for Semantic representation of the essays. Poor representation of the semantics of the documents as observed in LSA can lead to poor assessment results (Darwish et. al.2019, Zhang et. al, 2014). ILSA is a hybrid algorithm that integrates LSA, NMF and ACO to address the inadequacy associated with LSA. LSA has dimensionality approximation and noise reduction problem (Hoenkamp, 2011; Ayesha et al., 2020). The existing LSA is improved by introducing initialization and optimization of its factors using Non-Negative Matrix Factorization (NMF) and Ant Colony Optimization (ACO). The LSA factors initializes the NMF factors. ACO optimizes these factors by searching for the value of NMF factors that guarantees optimal reduction using the Frobenious norm of the distant matrix as the objective function. The score of a student is determined by computing the similarity value between the Lecturer's vector and the students' responses vectors as captured in the reduced DTM. The performance of the three algorithms were evaluated for

adequate semantic representation and good retrieval quality using Term-Term Similarity value, Precision, Recall and F-measure function respectively as metrics. The best of the three techniques for electronic assessment of free text document was also determined using mean divergence, assessment accuracy and Pearson correlation statistics as metrics.

The rest of the paper is divided into 4 sections. Section 2 reviewed related works revealing the strategies used by researchers, the limitation of work and the results achieved. Section 3 discusses the methodology used in terms of stages involved in the electronic assessment system and the improved algorithm. Section 4 presents the observed comparative experimental results while section 5 concludes the paper and present other areas of application of the improved algorithm.

2. RELATED WORKS

Open ended question is a question whose response is not a single word but requires the composition of free-text document where students give their answers in different pattern and yet may be expressing the same thing. The fact that answer in an open-ended question is free text makes its analysis and comparison with the lecturer marking scheme difficult because we are not looking for a word match but a semantically similar text. Several algorithms have been developed to represent free-text document and check for semantic similarity in an electronic assessment context. A review of these algorithms is presented along with their performance value.

Amalia, Gunawan, Fithri and Aulia, (2019) applied LSA in automating assessment in an essay written in Bahasa Indonesia. The essays were initially subjected to pre-processing step which ensured the removal of punctuations and irrelevant symbols, conversion of the entire text to lower case, breaking the sequences of strings to minimal meaningful units, stop word removal and reducing the word content to their root form. A document-term matrix which has the document label as its rows and pre-processed term as the columns. Its entries are the frequency of occurrence of the term in the document. The entries of the matrix were weighted, after which Single Value Decomposition was used to decompose the matrix. The resultant matrix was compared with the lecturer row vector using cosine similarity rule to obtain the score of the students. The LSA technique was evaluated against manual assessment by Amalia et al (2019) and the result was 83.3%. This work demonstrates the portability of LSA across languages. However, the accuracy level needs to be improved.

The work of (Mokhtari-Fard, 2013) is a practical demonstration of the application of Natural Language Processing methods. In this algorithm, the questions and correct answers are separately received by the system in the form of natural language. Then, the accurate answer for each question is converted into objects which represent the object-based representations of the input texts. In the subsequent step, the user inputs the relevant answer for each question through Graphics Users Interface (GUI). The system converts the input text for each question into distinctive objects. For analyzing the accuracy level of the answers, the created objects are compared with each other and the answering grades are calculated. On the other hand, it is hard to accomplish and very difficult to port across languages. Other systems that use the Natural Language approach are: C-rater and Paperless School free text Marking Engine (PS-ME). Another approach used in NLP is to grade the student essay by summarizing it so that only the relevant information is taken into account and this minimizes the presence of noise (Burstein and Marcu, 2000).

Darwish et al (2019) worked on automated essay evaluation by applying Latent Semantic Analysis and Fuzzy Ontology. The LSA was used in checking the semantic of the essays involved. While the Fuzzy Ontology was used to check the essays for consistency and coherence thereby resolving the problem of vagueness in language. The system scores the syntax of the essay, measures his semantic coherence and provides feedback to students about their mistakes. However, further work needs to be done to improve the semantic attributes representation and the feedback algorithm

3. MATERIAL AND METHODS

ILSA is a hybrid algorithm that incorporates LSA, NMF and ACO to improve the existing LSA algorithm by modifying its dimension reduction function using SVD-NMF-ACO for adequate semantic representation that guarantees improved assessment result. Figure 1 shows the block diagram for the ILSA electronic assessment system. The steps in the algorithm are

Document Collections

Relevant documents which comprise of the scripts to be graded and the lecturer marking scheme that will be used for the grading were collected as the data sets as shown in Table 1. These documents were used for training and extraction of terms using Syntactic Analyzer.

Table 1: Documents used in Constructing the Document-Term Matrix

SN	ANSWERS
1	Artificial Intelligence: This is the process of making robots and machine act like what we watch like movies
2	Artificial Intelligence can be described as a process of solving
3	Artificial Intelligence is the ability to find a problem, create the solution to the problem and solve the problem with the solution
4	Artificial Intelligence can be defined as the course that is concerned with the study and the creation of a computer system that behave in the form of intelligence
5	Artificial Intelligence: Is the science or exhibition of intelligence using a system or machine.
6	Artificial Intelligence: This is the intelligence exhibited by machine or software
7	Artificial Intelligence: Is a computer base representation of human intelligence and it's the first formal logic representation of human intelligence
8	Artificial Intelligence is a simulation of human intelligence process by machine especially computer system and the process include the following such as learning, reason and self-correction
9	Artificial Intelligence can be defined as a process or science of making computer to do things that requires when done by human.
10	Artificial Intelligence is the use of computer to do human task
11	Artificial Intelligence is the study and creation of computer and machines that exhibit some form of intelligence. it is concerned with solving tasks where require complex knowledge and reasoning process.
12	Artificial intelligence is defined as intelligence exhibited by an artificial entity, such as a system is assumed to be a computer
13	Artificial intelligence can be defined as the science and engineering of making intelligent machine especially intelligent computer programs.
14	Artificial Intelligence can be defined as intelligence exhibited by an artificial entity. Such system was regarded as computer. e.g. ability to recognize what you have seen before
	•
	•
92	Artificial Intelligence is the branch of computer science that deals with modelling of human-like intelligent system
93	Artificial Intelligence can be define as intelligence demonstrated by machines in contrast to the natural intelligence deployed by humans
94	Artificial Intelligence is the science of making computer do things that require intelligence when done by human
95	Artificial Intelligence is a branch of science which deals with human interaction and manipulations of machines in a behavioural ways, i.e. artificial-man-made intelligence thinking
96	Artificial Intelligence is concerned with getting computer to do task that require human intelligence
97	Artificial intelligence is getting a computer to do the task that requires human intelligence
98	Artificial Intelligence: is getting computer to do the task that requires human intelligence
99	Artificial Intelligence: Is a process of getting computer to carry out task that requires human intelligent
100	Artificial Intelligence is the science that make computer to do what require intelligence when done by human

Preprocessing

It reveals words that actually depict the meaning of a sentence. It involves stop-words removal, stemming and lemmatization. Stopwords are words that occur too frequently in a document and they contribute less to the semantic meaning of a document while stemming reduce a word to its root form by removing affixes.

Document-Term Matrix Construction

The extracted terms and the documents from which they were extracted, were used to create a document-term matrix, where documents tagged as student1, student2, student3 (based on the number of students) serve as the matrix row headings and the terms as the column headings. The entries to the matrix were the frequencies of occurrence of a term in a particular document as shown in the initial DTM in Table 2.

Term-Weighting

The entries were weighted using Term Frequency - Inverse Document Frequency (TF-IDF) weighting scheme in order to give emphasis to terms with higher semantic value.

Dimension Reduction using ILSA

The weighted matrix was subjected to dimension reduction using a combination of Latent Semantic Analysis (LSA), Non-Negative Matrix Factorization and Ant Colony Optimization Techniques in order to filter out noise and words with less semantic contribution. (Rufai et.al., 2021)

The LSA Algorithm was modified through its Dimension Reduction function to further minimize noise in its DTM by introducing initialization and optimization of its factors using Non-Negative Matrix Factorization (NMF) and Ant Colony Optimization (ACO). The LSA factors initializes the NMF factors. ACO optimizes these factors by searching for the value of NMF factors that guarantees optimal reduction using the Frobenious norm of the distant matrix as the objective function. The product of the optimal factors, W and H gives the reduced DTM. Algorithm 1 shows the dimension reduction process of the ILSA

This dimension reduction problem was modelled as a bound optimization problem of the form

$$\text{Min } f(W,H), W \geq 0, H \geq 0 \tag{1}$$

And the objective function is the frobenious norm function given in Equation 2

$$f(W,H) = \frac{1}{2} \|Z - WH\|_F^2 \tag{2}$$

Where Z is a non-negative Document Term Matrix and W and H are the factors of NMF. The ACO optimization algorithm seek the values of W and H that minimizes the objective function in Equation 2

W and H represent the search space of the optimization problem. The upper and lower bound values of the optimization problem are the initialized minimum and maximum values of W and H from the factors of LSA i.e. $W^{m,p} = |U|$, $H^{p,n} = |SV^T|$

Ant Colony optimization for continuous variable was used in iteratively searching for values of W and H that optimally minimizes the objective function.

Similarity Measurement

The similarity value between the lecturer Marking scheme and the students' responses was determined by evaluating similarity of their vectors using cosine similarity rule.

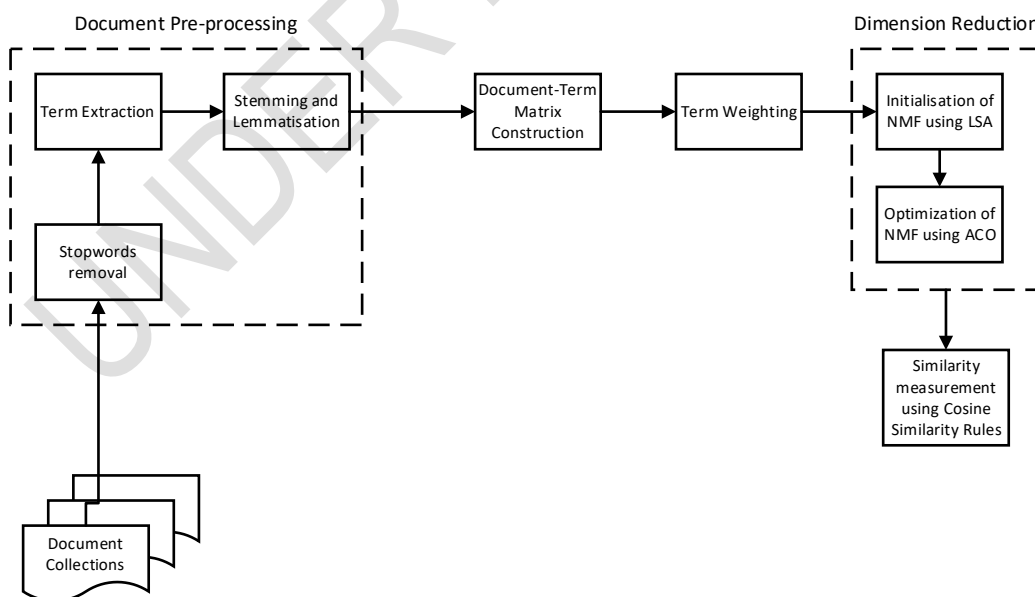


Figure 1: Block Diagram of The ILSA Algorithm

Algorithm 1: The ILSA Algorithm that Integrates LSA, NMF and ACO

- Step 1: Compute the rank k of factorization such that $k < \frac{mn}{m+n}$
Step 2: Decompose Z using SVD-LSA in order to obtain $Z=U\Sigma V^T$ with a rank of P
Step 3: Initialise NMF factors with SVD-LSA factors as $W=|U|$ and $H=|\Sigma V^T|$
Step 4: Update W and H using the multiplicative Update equation

$$H = H \times \frac{(WTZ)}{(WTWH + \varepsilon)}$$
$$W = W \times \frac{(ZHT)}{(WHHT + \varepsilon)}$$

- Step 5: Compute the Distance Matrix (D) as $D=Z - WH$
Step 6: Compute the row-wise Frobenius Norm of D as
$$\|D\|_F^{RW} = (\sum_{i=1}^m |d_i^r|^2)^{1/2}$$

Step 7: Identify the rows of D with the highest norm and look for the corresponding rows of W that minimizes $D=\|z_i^r - w_i^r H\|_F$ using ACO
Step 8: Identify the columns of D with the highest norm and look for the corresponding columns of H that minimizes $D=\|z_j^c - Wh_j^c\|$ using ACO
Step 9: Multiply the minimized rows of W with the minimized column of H to obtain \hat{Z} which is the reduced dimension of Z
Step 10: Compute the similarity value between the first column of \hat{Z} and its other columns using the cosine similarity rule expressed as:

$$cosSIM(\vec{A}, \vec{B}) = \frac{\vec{A} \cdot \vec{B}}{\|\vec{A}\| \|\vec{B}\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n A_i^2} \times \sqrt{\sum_{i=1}^n B_i^2}}$$

Where A is the first column of \hat{Z} and B is used interchangeably for any other column and it represent the students answer.

Information Retrieval

The Algorithms (i.e. LSA, NMF & ILSA) were applied in an information retrieval application to further demonstrate its suitability for information retrieval. One hundred and two documents were used as the search space and 11 query texts. Relevant documents were sought for in the search space.

The procedures followed were:

1. Represent the query text as a vector of the terms in the Term Document matrix.
2. Convert the vector to a scaled, weighted sum of component term vectors using

$$q = q^T V_k S_k^{-1} \quad 4$$

3. Compute the cosine similarity between the query vector and each document in the collection to determine relevant documents and level of relevance
4. Compute Precision, Recall and F-Measure using Equation 5, 6 and 7 respectively

Precision (P) is the fraction of retrieved documents that are relevant given as:

$$P = \frac{\text{number of relevant documents retrieved}}{\text{number of retrieved documents}} \quad 5$$

Recall is given as:

$$R = \frac{\text{number of relevant documents retrieved}}{\text{number of relevant documents}} \quad 6$$

F measure, which is the weighted harmonic mean of precision and recall is given as:

$$F = 2 \times \frac{P \times R}{P + R} \quad 7$$

4. RESULTS AND DISCUSSION

The developed ILSA algorithm is compared with its component algorithms (i.e. LSA and NMF) to reveal its better performance using the following metrics:

b. Semantic Adequacy

It is a measure of how adequately the semantic space capture the semantic content of the documents involved. In this research we assessed the semantic representational adequacy by using two methods which are the Term-Term Cosine Similarity measure and The Precision, Recall and F-Measure Functions.

i. Term Similarity

Term-Similarity is a measure of how semantically close a term is to another term. Our approach uses Term similarity to confirm two naturally similar terms and two naturally dissimilar terms. The cosine angle was computed on the Document Term Matrix also known as the semantic space of LSA, NMF and ILSA to show their respective semantic adequacy. The similarity between two terms is computed using Equation 3 and the results is shown in Table 3 and Table 4

$$\text{cosSIM}(\vec{t}_1, \vec{t}_2) = \cos(\theta) = \frac{\vec{t}_1 \cdot \vec{t}_2}{\|\vec{t}_1\| \|\vec{t}_2\|}$$

3

Table 3: Similarity Value of Synonyms

Term1	LSA	NMF	ILSA	Term2
Computer	0.80	0.79	0.84	Hardware
Create	1.00	0.95	0.70	Make
Decision	1.00	1.00	0.96	Logic
Internet	0.80	0.79	0.84	Web
Laptop	0.80	0.79	0.84	Computer
Learn	0.78	0.77	0.74	Study
Learn	1.01	1.00	0.98	Knowledge
Learn	1.00	1.00	0.99	Acquire
Behaviour	1.00	1.00	0.88	Manner
Machine	0.89	0.86	0.89	Device
Develop	0.73	0.73	0.69	Design
Task	0.78	0.76	0.54	Work
Theory	1.00	0.99	0.92	Study
Think	1.00	1.00	0.93	Logic
Man	0.93	0.93	0.91	Human

The result of Table 1 shows the similarity values between Term1 and Term2 in the semantic space of LSA, NMF and ILSA. A similarity value of ≥ 0.5 confirms similarity while < 0.5 shows dissimilarity. The words compared in Table 1 are similar terms that can be used interchangeably in a sentence without altering the sentence meaning which is the characteristics of synonyms.

Table 4: Word Pairs with low Similarity values

Term1	LSA	NMF	ILSA	Term2
Human	0.37	0.33	0.15	Machine
Hardware	0.42	0.42	0.41	Software
Natural	0.75	0.7	0.33	Artificial
Solution	0.46	0.43	0.38	Problem
Input	0.29	0.29	0.29	Output

Table 4 shows the low similarity recorded for naturally dis-similar terms by the tree algorithms. However, LSA and NMF erred in their similarity values between “natural” and “artificial” which may be the consequence of their poor noise handling mechanism.

ii. Precision, Recall and F-Measure

Table 4 shows the Precision, Recall and F-Measure results using the three algorithms. Precision shows the proportion of relevant documents that are retrieved while Recall shows the proportion of retrieved document that are relevant. It is expected that technique with better semantic capturing shows a better Precision, Recall and F-Measure result which can be observed in the result of ILSA. ILSA outperforms LSA, NMF in terms of retrieval quality. Figure 2 shows the graphical representation of the result of Table 5 and Figure 2.

Table 5: Precision, Recall and F-Measure Evaluation for LSA, NMF and LSA-NMF-ACO

	Precision	Recall	F-Measure
LSA	0.75	0.95	0.83
NMF	0.77	0.97	0.85
ILSA	0.79	0.98	0.87

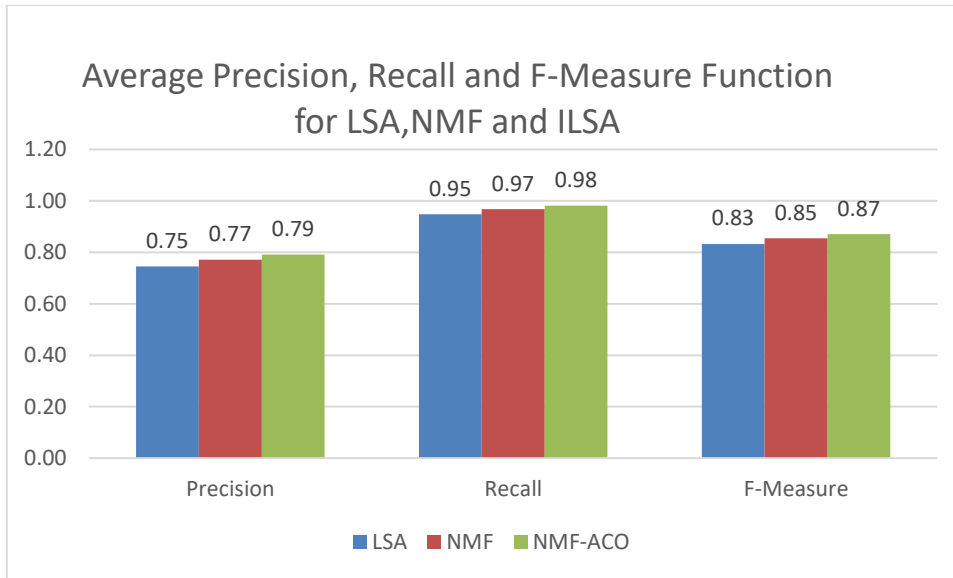


Figure 2: The Relative Performance of LSA, NMF and ILSA for Precision, Recall and F-Measure function

c. Mean divergence and Measurement of Accuracy

Divergence measures the difference between the machine generated score and the manual score at \pm value. The machine score will be acceptable if the difference between it and the human score is minimal (Adesiji *et al.* 2016). The divergence variance V of result of a question q for n students is given in equation 8 and 9.

$$DF_q = |S - M|_q$$

$$V_q = \frac{\sum_i^n DF_q}{n}$$

8

9

where DF is set of score differences, M is score obtained from Human score, S is score obtained from machine i represents distinct student in set n .

$\sum_i^n DF_q$ is the sum of the differences between the machine score and the human score

$$\text{Accuracy} = 100 - (100 * V_q)$$

The result of the divergence and accuracy of the various algorithms is given in Table 6

Table 6: Mean Divergence and Assessment Accuracy of the Various Algorithm

	Mean Divergence	Assessment Accuracy
LSA	0.10	89.52
NMF	0.09	91.13
ILSA	0.03	96.63

CONCLUSION

The research observed the performance of an improved algorithm ILSA for electronic Assessment of free text document using Adequacy in Semantic Representation, Retrieval Quality and Assessment Accuracy as performance metrics. The experimental results obtained shows Precision, Recall and F-Measure values 0.75,0.95,0.83 for LSA; 0.77,0.97,0.85 for NMF and 0.79,0.98,0.87 for ILSA which is an indication of a better performance of ILSA in terms of retrieval quality. The assessment accuracy observed for LSA, NMF and ILSA were 89.52, 91.13 and 96.63 respectively, while mean divergence from manual were 0.10, 0.09, 0.03 respectively and Pearson Correlation coefficient were 0.643, 0.730, 0.959 respectively. The improved algorithms returned term-similarity value of >0.5 for similar words and <0.5 for dissimilar words which confirms its adequacy in semantic representation

ILSA outperformed LSA and NMF in terms of retrieval quality and Assessment Accuracy which is an indication that ILSA actually improves the existing LSA in terms of Assessment Accuracy, Retrieval Quality and Semantic Space representation. Further evaluation on the basis of noise reduction can be carried out on the improved algorithm. The developed algorithm can be applied in Document Summarization, text classification, text categorization and other text-based applications.

The work can be adopted by Examination conducting bodies and educational institutions for mass marking of theoretical questions. However, future work can be geared towards evaluating the performance of these algorithms in terms of noise reduction and extent of approximation of the initial semantic space.

References

1. Adesiji KM, Agbonifo OC, Adesuyi AT, Olabode O. Development of an automated descriptive text-based scoring system. *Journal of Advances in Mathematics and Computer Science*. 2016 Oct 28:1-4.
2. Amalia A, Gunawan D, Fithri Y, Aulia I. Automated Bahasa Indonesia essay evaluation with latent semantic analysis. In *Journal of Physics: Conference Series* 2019 Jun 1 (Vol. 1235, No. 1, p. 012100). IOP Publishing. Amalia, A., Gunawan, D., Fithri, Y., & Aulia, I. Automated Bahasa Indonesia essay evaluation with latent semantic analysis. In *Journal of Physics: Conference Series* IOP Publishing.2019; 1235(1):012100.
3. Ayesha S, Hanif MK, Talib R. Overview and comparative study of dimensionality reduction techniques for high dimensional data. *Information Fusion*. 2020 Jul 1;59:44-58. Ayesha, S., Hanif, M. K., & Talib, R.. Overview and comparative study of dimensionality reduction techniques for high dimensional data. *Information Fusion*, 2020;59:44-58.
4. Burstein J, Marcu D. Towards using text summarization for essay-based feedback. In *La 7e Conference Annuelle sur Le Traitement Automatique des Langues Naturelles TALN* 2000 Oct. Burstein, J., and Marcu, D. Towards using text summarization for essay-based feedback. In *La 7e Conference Annuelle sur Le Traitement Automatique des Langues Naturelles TALN*. 2000
5. Darwish SM, Mohamed SK. Automated Essay Evaluation Based on Fusion of Fuzzy Ontology and Latent Semantic Analysis. In *International Conference on Advanced Machine Learning Technologies and Applications* 2019 Mar 28 (pp. 566-575). Springer, Cham. Darwish, S. M., & Mohamed, S. K. (2019, March). Automated Essay Evaluation Based on Fusion of Fuzzy Ontology and Latent Semantic Analysis. In *International*

Conference on Advanced Machine Learning Technologies and Applications (pp. 566-575). Springer, Cham.

6. Darwish, S. M., & Mohamed, S. K. (2019, March). Automated Essay Evaluation Based on Fusion of Fuzzy Ontology and Latent Semantic Analysis. In International Conference on Advanced Machine Learning Technologies and Applications (pp. 566-575). Springer, Cham.
7. Hoenkamp E. Trading spaces: on the lore and limitations of latent semantic analysis. In Conference on the Theory of Information Retrieval 2011 Sep 12 (pp. 40-51). Springer, Berlin, Heidelberg.
8. Hoenkamp, E. Trading spaces: on the lore and limitations of latent semantic analysis. In Conference on the Theory of Information Retrieval. Springer, Berlin, Heidelberg. 2011; 40-51. Janecek A, Tan Y. Iterative improvement of the multiplicative update nmf algorithm using nature-inspired optimization. In 2011 Seventh International Conference on Natural Computation 2011 Jul 26 (Vol. 3, pp. 1668-1672). IEEE.
9. Jordan S. E-assessment: Past, present and future. *New Directions in the Teaching of Physical Sciences*. 2013(9):87-106. Jordan, S. E-assessment: Past, present and future. *New Directions in the Teaching of Physical Sciences*, 2013;9:87-106.
10. Mokhtari-Fard I. Natural Language Understanding for Grading Essay Questions in Persian Language. In Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data 2013 Oct 10 (pp. 144-153). Springer, Berlin, Heidelberg. Mokhtari-Fard, I. Natural Language Understanding for Grading Essay Questions in Persian Language. In Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data. Springer, Berlin, Heidelberg. 2013; 144-153
11. Rufai M. M, Afolabi A, Fenwa O. D, & Ajala F.A. (February,). An Improved LSA Model for Electronic Assessment of Free Text Document. *International Journal of Innovative Technology and Exploring Engineering Blue Eyes Intelligence Engineering and Sciences Publication*. 2021;10(4):152-159
12. Zhang M, Hao S, Xu Y, Ke D, Peng H. Automated Essay Scoring Using Incremental Latent Semantic Analysis. *JSW*. 2014 Feb 1;9(2):429-36.