Automated Arabic Essay Scoring (AAES) Using Vectors Space Model (VSM) and Latent Semantics Indexing (LSI)

Dr. Ayad R. Abbas
Computer Sciences Department, University of Technology/Baghdad.
Email:ayad_cs@yahoo.com
Ahmed S.Al-qazaz
Computer Sciences Department, University of Technology/Baghdad.
Email:ahmed_alkzaz243@yahoo.com

Received on: 30/9/2014 & Accepted on: 29/1/2015

ABSTRACT

Automated Essays Scoring (AES) stands for the ability of computer technologies to evaluate electronic essays written by learner according to previously determined essay. All the previous works and researches were applied to essays written in English language and they were applied to essays written in Bahasa, Hebrew, Malay, Chinese, Japanese, and Swedish. The research paper suggests an Automated Arabic Essays Scoring (AAES) system on web-based learning context based on Vectors Space Model (VSM) and Latent Semantics Indexing (LSI). The proposed system consists of two main processes. The first process, deals with applying information retrieval mechanics to retrieve the significant information from electronic essays. In the second process, VSM and LSI are applying to find out the similarity degree between the previously written essays by the instructor and the essay written by the learner.

The experiments of our results reveal that the proposed system provides an electronic assessment closer to instructors' traditional judgment, leading to improve the learning's efficiency, performance and to overcome.

Keywords: Automated essay scoring, Web-based learning, Information retrieval, Vectors space model, Latent semantics indexing, Truncated singular values decomposition.

التقييم الالي للمقالات العربية باستخدام نموذج فضاء المتجهات (VSM) وفهرسة الدلالات الكامنة (LSI)

الخلاصة:

التقييم الالي للمقالات (AES) هو قدرة تكنولوجيا الكمبيوتر على تقييم المقالات المكتوبة الكترونيا من قبل المتعلم وفقا لمقال محدد مسبقا .البحوث ومعظم الأعمال السابقة طبقت على مقالات مكتوبة باللغة الانكليزية، كما طبقت أيضا على مقالات كتبت باللغة العبرية، ولغة الباهاسا الماليزية، اليابانية، الصينية، السويدية. تقترح هذه الورقة نظام الي لتقييم المقالات المكتوبة باللغة العربية(AAES) في سياق التعلم القائم على الشبكة العالمية باستخدام نموذج فضاء المتجهات (VSM) وفهرسة الدلالات الكامنة (LSI). يتكون النظام المقترح من عمليتين

رئيسيتين. العملية الأولى يتم فيها تطبيق استرداد المعلومات لاستخراج المعلومات الهامة من مقالات الإلكترونية. الثانية، يتم تطبيق (VSM) و (LSI)للعثور على درجة التشابه بين المقالات المعدة مسبقا من قبل المعلم والمقالة المدخلة من قبل الطالب.

النتائج التجريبية تبين أن النظام المقترح يوفر تقييما إلكترونيا قريب التقييم التقليدي للأستاذ، مما يؤدي إلى تحسين كفاءة التعلم والتغلب على عامل الوقت، التكلفة، والموثوقية.

INTRODUCTION

Teachers, especially in the humanities, social sciences, and letters department usually include essay questions in their exams to further measure the current viewpoint of the student regarding a problem, a scenario, or a specific topic. Although essays are one of the most effective ways of measuring students' capabilities, they usually become the main reason for teachers returning the test papers to the students. There are usually quite a lot of students in these subjects at universities require these subjects to be taken up by all the students regardless of their majors. Checking these essay questions is a very tedious task that consumes a lot of time of the teachers. This leads to students, who lack information regarding their current class standing. These shortcomings are very important in the educational setting for these would defeat the purpose of AES systems [1].

Automated Essays Scoring (AES) stands for the computer technologies that scores and evaluates the written essays utilized to overcome cost, time, generalizability and accuracy, and issues in article scoring. AES system programmed in 1973, and interested by many universities, public schools, companies, examiners and researchers. These systems are developed to help instructors in low-stakes assessment of the classroom as well as testing companies in wide-scale. The reliability and accuracy of AES with respect to writing assessment has been studied by a number of researchers. Although AES systems are producing good results on their accuracy and effectiveness; they have been criticized for their lack in human interaction and their vulnerability to cheating , However, most teachers still believe that these tools are inefficient as the current approaches on AES are insensitive to the uniqueness of every teacher [1,2,3].

Most of the previous works are in English. Whereas, the past works are applied in Hebrew, Bahasa `Malay, Japanese, Chinese and Swedish [4, 5, 6, 7, 8].

The VSM used to find the relation degree between instructor essays and learner's essays. In the VSM, each electronic essay is represented as vectors as well as their relevance to the queries posted by the users and evaluated over suitable matching functions [9].

LSI, means analyzing documents to find the underlying/latent, meaning/semantics or concepts, in LSI, instructor essays and learner's essay can have high similarity degrees between them even if they do not share any terms - as long as their terms are semantically similar of those documents [10].

Vector Space Model (VSM)

The VSM developed for the **SMART (System for the Mechanical Analysis and Retrieval of Text) Information Retrieval System** [1], by Gerard Salton and his colleague. It is obtained to get the pattern in the document collection, which used to improve the accuracy in information retrieval systems. VSM stands for representing each document into a collection as vectors or points in a space. Vectors that are near together in its space are semantically identical, where the vectors that are not close are semantically far, the user's query represented as the documents in the same space. VSM for information retrieval are one subordinate class of retrieval mechanisms that have been used and studied in recent years. The vector-space models "individual, feature-based, formal, partial match" are reassembled by taxonomy provided on labels the class of techniques. The documents are modeled as sets of terms by an information retrieval mathematical model, that can be individually manipulated, weighted, and performed queries by comparing and analyzing the query representation to the of each document in the space, and the documents that do not necessarily include one of the search terms can be retrieved. However, the prevalent characteristics with another mechanism in the hierarchy of an information retrieval and a central set of similarities that setting their class are both shared by VSM techniques [3, 9, 11].

These methods are simple, have limited requirements and don't need the training data. They focus on non-linguistic text features such as position of a keyword and the frequency both of a term and inverse document. The statistical information about the words can be applied to distinguish the keywords in the document [12].

Similarity Measure Estimation [13,14]

The VSM is used to find the relation degree between essay-based examination (between instructor essays and learner's essays). In the VSM, each essay is represented as vectors, and suitable matching functions measure their relevancy to the queries passed by the users. A space in both essays and queries represented by vectors are created by the model, for a fixed collection of essays, an M-dimensional vector created for each query and essay from sets of terms (i.e., keywords, vocabularies) with corresponding weights, where M denotes the number of singular terms in the collection of an essay. The following four steps are used to calculate the weights associated with the terms:

1- tf_k is the term counts (term frequency) or number of times a term k exists in an essay.
2- .D_{fk} is the essay frequency or number of essays containing term k.
3- D is the number of essays in a database.

[1] . The SMART (System for the Mechanical Analysis and Retrieval of Text) Many vital ideas in data recovery were created as a major aspect of exploration on the SMART framework, including the vector space model, significance input, and Rocchio Classification. Gerard Salton drove the gathering that created SMART.

4- IDF is the inverse of the essay frequency or number of essays containing a term The coordinates of the (i) the essay in the direction corresponding to the (k) the linguistic term (the weight of the term in the specific assay) can be determined the gallows

5-

$$w_{ik} = tf_{ik} \times \log\left[\frac{D}{Df_{ik}}\right] \times IDF \quad ...(1)$$

Then, the inner product is an example of a similarity of a vector function can be used to find the similarity between a query and an essay:

$$r_{ij} = \cos(\theta) = \frac{\sum_{h=1}^{m} w_{ih}\ w_{jh}}{\sqrt{\sum_{h=1}^{m} {w_{ih}}^2 \sum_{h=1}^{m} {w_{jh}}^2}} \quad ...(2)$$

Where

$c_i=\{w_{i1}, w_{i2},..., w_{ik},..., w_{im}\}$ and $c_j=\{w_{j1}, w_{j2},..., w_{jk}, ..., w_{jm}\}$ respectively, rig is the correlated degree between the ith and jth essays.

Latent Semantic Indexing (LSI)

LSI is a space of vector based method, which attempts to capture associations and the hidden structure of topics by reducing the dimensionality of the system, using techniques from linear algebra. It assumes that words in an essay have similar meaning will exist in similar places of essays. Similar essays will have a similar occurrence of terms. A term-document matrix is built from a large piece of sentence and singular value decomposition is utilized to reduce the number of columns while saving the similarity structure among rows. The cosine angle between two vectors established by any two rows finds the correlation degree between terms. Values are closer to one are very similar words while values closer to zero are very dissimilar words [10].

LSA is a statistical or mathematical technique for inferring relations and contextual usage extracting of the words in passages. First step, convert the text to a matrix in which each row represents a unique word, each column represents the text passage while each cell contains word frequency. Second, the cell entries are subjected to a transformation. Third, LSA utilizes Singular Value Decomposition (SVD) by converting it to the matrix. However, SVD decomposes a rectangular matrix in the production of three other matrices. The original row entities described by one composed matrix as vectors of derived the values of an orthogonal coefficient, while in the same way, another describes the entities of an original column, and the third is a diagonal matrix contains values of a measurement..

LSI holds the following constraints of queries:

1- The synonymy multiple words that have the same meanings.

2- The polysemy words include more than one meaning.

Term-document matrix [15]

To apply linear algebra to information retrieve, we first need to transform the problem into mathematical form. LSI utilizes a term-essay based matrix, which describes the existence of terms in essays; it is a matrix where columns represent essays and rows represent terms.

Truncated SVD

In many applications, the SVD is an important method, or tool which has been studied. However, SVD has utilized to find the variable values using resulting normal equations [16].

For a m*n matrix A with rank k there exists a decomposition such that [10]

$$A = U\sum V^T \qquad \text{…….. (3)}$$

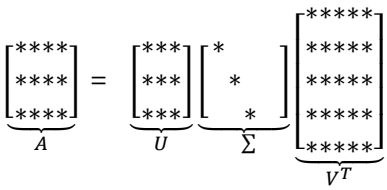

Figure (1): graphical representation of SVD of a matrix A

Where

U is m*m matrix .

Σ is m* n matrix.

V^T is n*n matrix.

The columns of U are orthogonal eigenvectors of AA^T the columns of V are orthogonal eigenvectors of A^TA [10].

SVD has pleasant hypothetical properties; however SVD contains a great deal of information, most likely more than is essential for this application. It may contain noises that will influence the recovery efficiency, We can discover a diminished rank close estimation (truncated SVD) to A by setting everything except the first k biggest singular values equivalent to zero and utilizing just the first k sections of U and V. Truncated SVD utilizes k biggest singular values approximates the first term-by-essay matrix A by Ak. As indicated by the optimality property, SVD gives the closest rank k close estimation of a matrix [10].

Query [10]

A query is a vector that has the same shape as a document. That is, if the term-document matrix is an m by n matrix, then each of the columns (m by 1 vector) represents an essay and a query has the same size. A query q is likewise mapped into the k-space by:

$$\hat{q} = q^T U_K \sum_k^{-1} \qquad \text{…….(4)}$$

Where

$\boldsymbol{q^T U_K}$ represents the sum of these k -dimensional term vectors, $\sum_{\boldsymbol{k}}^{\boldsymbol{-1}}$ Is the differential weight of the separate dimensions [10].

An inquiry is done by contrasting the likeness between an essay and query in the new k-space. Measure of closeness between an essay and query is characterized by the angle cosine between essay vector and query vector. An expansive cosine implies that the angle between the two vectors is little while a little cosine implies the inverse:

$$COS(\overrightarrow{a_j}, \vec{q}) = \frac{\overrightarrow{\boldsymbol{a_j}}, \boldsymbol{q}}{\|\boldsymbol{a_j}\|_2 \, \|\boldsymbol{q}\|_2} \qquad \ldots\ldots (5)$$

Where

$\overrightarrow{\boldsymbol{a_j}}$: Documents in the new k-space

Updating Database [15]

If more documents or terms need to be added, we have to upgrade the approximated term-essay matrix. There are two approaches to upgrade the approximated term-essay matrix. The first one is the recalculate SVD for the completely new matrix. The second one is to use a folding-in algorithms to add the terms and documents. The first method is straightforward, Every time there is a change in research profile; a new term-document matrix needs to be built from scratch. The obvious disadvantage of this method is the inefficiency of rebuilding the whole term-documents since it is usually a large sparse matrix. The alternative is to use a folding-in algorithm to reconstruct the matrix. To fold in a new n x 1 essay vector, d, into a current LSI model, a projection $\hat{\boldsymbol{d}}$ of d onto the span of the current term vectors (columns of Uk) is processed by:

$$\hat{\boldsymbol{d}} = \boldsymbol{d^T U_K} \sum_{\boldsymbol{k}}^{\boldsymbol{-1}} \qquad \ldots\ldots.(6)$$

Overview of the Proposed System Architecture

Figure 2 shows the proposed framework architecture, which gives various services, such as a learner's interface service, essay-based examination service and AAES service. The learner's interface service presents an essay-based examination for learners to interact with the electronic tests, and it provides the ability for learners to activate navigation events. The essay-based examination management service allows the teacher to access, activate and manage essay-based examination; its key functions are creating, updating or deleting of essay-based examination. Finally, AAES service as a similarity measure process.

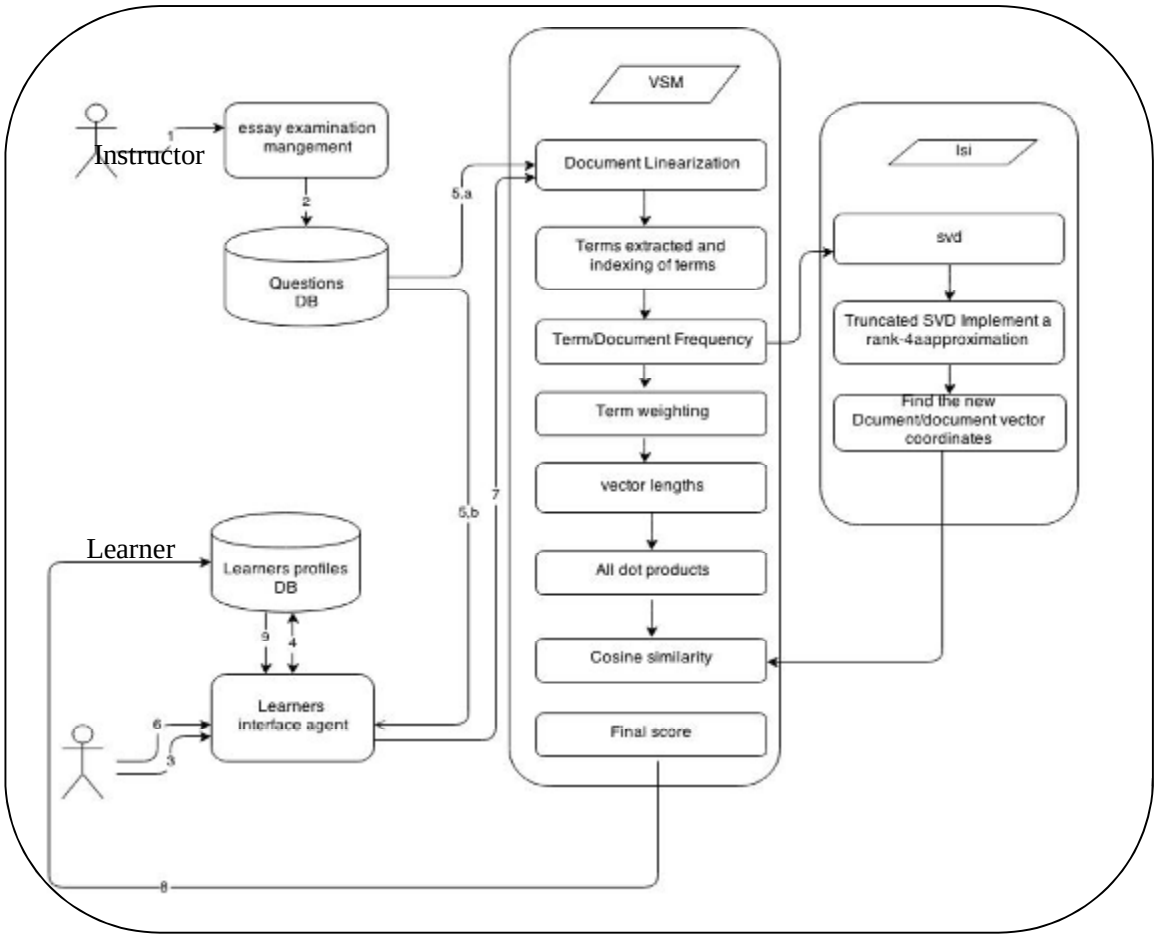

Figure (2): The proposed system architecture

The operation technique of the proposed framework is summarized as follow:

Step 1: Instructor accesses exam management service to access, activate and manage questions and best answers (model answers) of the questions.

Step 2: Questions and model answers of the questions, inserting into the questions database.

Step 3: The learner accesses the learner's interface agent.

Step 4: The proposed system gets his/ her learning profile from the learner profile database.

Step 5.b: Exam management service sends specific question to learning profile service from the questions database to display it to the learners.

Step 5.a: Exam management service sends specific model answers to the AAES service from the questions database.

Step 6: Learning profile service receives the learner answer.
Step 7: Learning profile service sends the learner answer to the AAES service.
Step 8: AAES service receives model answers, learner answer and implement (VSM), (LSI) to find the final score, which sends to the Learning profile service.
Step 9: Finally, the learner's interface agent presents the electronic essay grades to the learner.

10. System Implementation, Results and Test:

The proposed system was successfully implemented using Microsoft Windows 8 with Apache Web Server version 2.2.6, HTML, PHP script language version 5.2.5, MySQL database version 5.0.45 and Matlab-Mathworks version R2013a , we test the proposed system by let the instructor enters 1 question(table 1), and 4 model answers(table 2) ,30-learners answer.

The results of tests the proposed system by enter 1- learner answer in (table 3), are shown in (table 4) and the following operations:

Table 1: Instructor Question

Question
ماهي شبكات الحاسوب؟

Table 2: Model Answers

Model Answers	No
الشبكات هي هيكلية تنظيمية لربط حاسبتين او مجموعة من الحاسبات تنتشر في مواقع مختلفة وترتبط ببعضها بواسطة ادوات الاتصال المختلفة من اجل تبادل المعلومات والموارد والبيانات بينها و مشاركة الاجهزة و الموارد مثل الالة الطابعة او البرامج التطبيقية و كذلك تسمح التواصل المباشر بين المستخدمين .	1
الشبكات هي ربط حاسوبين او مجموعة من الحواسيب تنتشر في مكان واحد او اماكن مختلفة و ترتبط ببعضها بواسطة جهاز اتصال او اجهزة الاتصال المختلفة سلكيا او لاسلكيا من اجل تبادل البيانات بينها و مشاركة الاجهزة و موارد الحاسوب مثل الالة الطابعة او البرامج التطبيقية او الالعاب و كذلك تسمح بالتواصل المباشر بين الاشخاص.	2
الشبكات هي حاسوبين او أكثر متصلة مع بعضها البعض عن طريق مجموعة من الوسائل مثل خطوط الهاتف او الكابلات او الاقمار الصناعية وترتبط سلكيا او لاسلكيا وتسمح لمستخدميها بمشاركة البيانات والبرمجيات والاجهزة المتصلة بالشبكة كالطابعات و غيرها من الاجهزة لغرض تبادل المعلومات و التواصل مع الاخرين.	3
الشبكات هي مجموعة من الحواسيب المرتبطة مع بعضها بطريقة ربط معينة عبر وسائط تتبع في ذلك لمعايير مختلفة تتكون شبكة الحاسب من جهازين متصلين او أكثر ببعضهما بطريقة سلكية او لاسلكية ويقومان بتبادل الملفات و تسمح بتبادل البيانات و موارد الكمبيوتر مثل الطابعات و تسمح للمستخدمين بالتواصل بشكل فوري .	4

Table(3): Learner Answer

Learner Answer	no
هي الحاسبات و الاجهزة المحمولة المربوطة مع بعضها لإرسال الرسائل و الصور و البيانات و تصفح الأنترنت و التواصل مع الأصدقاء	1

Table (4): the proposed system's retrieval results

Terms	query	Document 1	Document 2	Document 3	Document 4	df	idf	w_{1k}	w_{2k}	w_{3k}	w_{4k}
اتصال	0	0	1	0	0	1	0.601	0	0.1156	0	0
اجل	0	1	1	0	0	2	0.3011	0.0602	0.0578	0	0
اجهزة	0	0	1	0	0	1	0.6021	0	0.116	0	0
ادوات	0	1	0	0	0	1	0.6021	0.1204	0	0	0
أكثر	0	0	0	1	1	2	0.3011	0	0	0.0602	0.0587
الاتصال	0	1	1	0	0	2	0.3011	0.0602	0.0578	0	0
الاجهزة	1	1	1	2	0	3	0.1250	0.0249	0.0240	0.0396	0
الاخرين	0	0	0	1	0	1	0.6021	0	0	0.1204	0
الاشخاص	0	0	1	0	0	1	0.6021	0	0.1156	0	0
الاقمار	0	0	0	1	0	1	0.6021	0	0	0.1209	0
الالة	0	1	1	0	0	2	0.3011	0.0602	0.0578	0	0
الالعاب	0	0	1	0	0	1	0.6020	0	0.1156	0	0
البرامج	0	1	1	0	0	2	0.3011	0.0602	0.0578	0	0
البرمجيات	0	0	0	1	0	1	0.6021	0	0	0.1204	0
البعض	0	0	0	1	0	1	0.6021	0	0	0.1204	0
…...	…	…	…	…	..	..	…….	…	….	..	…

1. The first column (terms) represents the terms of predefined essay.
2. The second column (query) represents the term frequency (term counts) or number of times a term k occurs in an essay in the learner's essay.
3. The (third, fourth, fifth, sixth) columns (essay 1, essay 2, essay 3, essay 5) represent the term frequency (term counts) or number of times a term k occurs in an essay in the learner's essay.
4. The sixth column (DF) represents the essay frequency or number of essays containing term.
5. The seventh column (idf) represents the inverse of essay frequency or number of essays containing term.
6. The (eighth, ninth, tenth, elven) columns (weight 1, weight 2, weight 3, weight 4) represents the weight of the term in the specific essay which can be found using equation (3.1).

The (VSM) score finds by computing the similarity between predefined essays and a query by using equation (3.2).
VSM_score_query_essay1=0.274
VSM_score_query_essay2=0.009

VSM_score_query_essay3=0.102
VSM_score_query_essay4=0.038

The proposed system takes the highest score as the VSM score=0. 274

Then, in order to calculate the LSI score the proposed system first applying the equation (6.1) to decompose the original term-by-document matrix.

$$\underset{U}{\begin{bmatrix} -0.0701 & 0.0408 & -0.0568 & -0.1840 \\ -0.1332 & 0.1017 & -0.1125 & -0.0013 \\ -0.0701 & 0.0408 & -0.0568 & -0.1840 \\ -0.063 & 0.06090 & -0.0557 & 0.1827 \\ -0.1015 & -0.1265 & 0.1542 & 0.0262 \\ -0.1332 & 0.1017 & -0.1125 & -0.0013 \\ -0.2313 & 0.13706 & 0.2328 & 0.0007 \\ -0.0490 & 0.0176 & 0.172 & 0.0010 \\ -0.0701 & 0.0408 & -0.0568 & -0.1840 \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}} \underset{\sum}{\begin{bmatrix} 8.432 & 0 & 0 & 0 \\ 0 & 6.1462 & 0 & 0 \\ 0 & 0 & 5.235 & 0 \\ 0 & 0 & 0 & 3.837 \end{bmatrix}} \underset{V^T}{\begin{bmatrix} -0.5322 & 0.3743 & -0.2916 & 0.7011 \\ -0.5912 & 0.2509 & -0.2973 & -0.7064 \\ -0.4135 & 0.1085 & 0.9039 & 0.0041 \\ -0.44291 & -0.8860 & -0.0966 & 0.0966 \end{bmatrix}}$$

Then, applying the equation (7.1) to query q to mapped the query into the k-space

$\hat{\boldsymbol{q}}$ =-0.0881 0.020380.0936 0.1090

Finally, The (LSI) score finds by compute the similarity between predefined essays and a query by using equation (7.2)

LSI_score_query_essay1=0.7286
LSI_score_query_essay2=-0.2805
LSI_score_query_essay3=0.6105
LSI_score_query_essay4=0.1323

The proposed system takes the highest score as the LSI score=0.7286,The final score finds by computing the average of the two scores (VSM AND LSI) = 0.274+0.7286/2=0.5013.

The proposed system suggests a score range between (0-10) for this reason the system mapped the final score to the suggests score by multiplying the final score by 10, the final score now becomes 5.

Comparison of results between traditional human scoring and the Proposed AAES system scoring for 1 question, and 4 model answers, 30-learners answer are shown in (table 5).

Table (5): Comparison of results between traditional human scoring and the Proposed AAES systemscoring.

Answer no.	Instructor's score	AAES system scores	Accuracy
1	1	1	100%
2	3	3	100%
3	3	3	100%
4	3	4	90%
5	4	4	100%
6	4	5	90%
7	4	4	100%
8	4	4	100%
9	4	4	100%
10	4	4	100%
11	5	5	100%
12	5	5	100%
13	5	5	100%
14	5	5	100%
15	6	6	100%
16	6	6	100%
17	6	7	90%
18	7	6	90%
19	7	7	100%
20	7	7	100%
21	7	7	100%
22	7	6	90%
23	8	8	100%
24	8	9	90%
25	8	8	100%
26	8	7	90%
27	9	9	100%
28	10	10	100%
29	10	10	100%
30	10	10	100%

The execution of a system for scoring essays can be assessed by measuring what number of the mechanized score closer to the instructor's score. More nearly the computerized scores closer to the instructor's score are more precise. (Table 6) demonstrates the mapping between educator's score to the scores reviewed by the AAES framework. Here, out of 30 understudy's submitted answers 5 essays have been scored 7 by instructor's score while the AAES framework scored 3 essays for 7 and 2 essays have been have scored 6 . Also, we see that framework has been missed a few responses to score effectively.

Significance measure have tried the framework by true positive, true negative, false positive and false negative. We have characterized this measure as bellow:

True positive: If a test outcome shows positive result that is truly positive is called genuine positive. In this test if AAES framework gives an essay 10 evaluation for which the instructor's score is 10 then the result is genuine positive.

True negative: If a test outcome shows negative result that is truly negative is called genuine negative. In our examination if AAES framework does not give score 1 where the instructor's score 1 is not introduce in the current article set then it is called genuine negative.

False positive: If a test outcome shows positive result that is truly negative is called false positive. In our examination if AAES framework gives score 1 for an answer where the instructor's score 1 is not set for that essay then it is called false positive.

False negative: If a test outcome shows negative result that is truly positive is called false positive. In our test if AAES framework gives 1 for an answer where the human review 10 is allocated for that answer then it is called false negative.

Missed: The term missed denots to the quantity of answers to which instructor's score (and not by the AAES System) denotes each one score.

Spurious: The term spurious demonstrates the quantity of answers to which the AAES System (and not by instructor's score) denotes each one score.

Table (6): Performance of proposed system

Instructor score	**Number of answers**		**AAES System score**								
			10	9	8	7	6	5	4	3	1
10	3	→	3	0	0	0	0	0	0	0	0
9	1	→	0	1	0	0	0	0	0	0	0
8	4	→	0	1	2	1	0	0	0	0	0
7	5	→	0	0	0	3	2	0	0	0	0
6	3	→	0	0	0	1	2	0	0	0	0
5	4	→	0	0	0	0	0	4	0	0	0
4	6	→	0	0	0	0	0	1	5	0	0
3	3	→	0	0	0	0	0	0	1	2	0
1	1	→	0	0	0	0	0	0	0	0	1
total	30	→	3	2	2	5	4	5	6	2	1

(Table 7) shows true and false positive, and negative, respectively from the results acquired by AAES framework.

Table (7): True and False Positive, and Negative, Respectively of AAES system

Instructor score	Number of answers	No. of Essay Correctly Scored by AAES system	Missed	Spurious	True Positive	True Negative	False Positive	False Negative
10	3	3	0	0	100%	0%	0%	0%
9	1	1	0	1	100%	0%	100%	0%
8	4	2	2	0	50%	0%	0%	50%
7	5	3	2	2	60%	0%	40%	40%
6	3	2	1	2	66.7%	100%	40%	33.3%
5	4	4	0	1	100%	0%	25%	0%
4	6	5	1	1	83.3%	0%	16.7%	16.7%
3	3	2	1	0	66.7%	0%	0%	33.3%
1	1	1	0	0	100%	0%	0%	0%
total	30	23	7	7	81%	0%	28%	19%

(Table 8) demonstrates the results got by the AAES while calculating in semantic revision.

Table (8): PRCISION, RECALL AND F1 OF AAES system for 30 submitted answers

Instructor score	Number of answers	No. of Essay Correctly Scored by AAES system	Missed	Spurious	Precision	Recall	F1
10	3	3	0	0	100%	100%	100%
9	1	1	0	1	50%	100%	67%
8	4	2	2	0	100%	50%	67%
7	5	3	2	2	60%	60%	60%
6	3	2	1	2	50%	66.70%	57%
5	4	4	0	1	80%	100%	89%
4	6	5	1	1	83%	83.30%	83%
3	3	2	1	0	100%	66.70%	80%
1	1	1	0	0	100%	100%	100%
total	30	23	7	7	80%	81%	78%

In the above table we describe the eight columns as follows:

1. A test score, we have given to the answers (instructor's score.
2. The quantity of answers that teacher physically assigned to each one score.
3. The quantity of answers accurately assessed by AAES framework.
4. The quantity of answers to which the teacher (and not by the AAES) allocated each one score.
5. The quantity of answers to which the AAES (and not the instructor) gave each one score.

6. The last three columns show accuracy, review and F1 values.
In this connection, we characterized accuracy, review and F1 as takes after:
Precision : Precision is the quantity of answers accurately scored by AAES divided by the aggregate number of answers scored by AAES.
Recall: Recall is the quantity of answers accurately scored by AAES divided by the aggregate number of answers scored by the instructor.
F1: The F1 score (additionally called F-measure) is a measure of a test's precision. It's a join measure of accuracy and review is the symphonious mean of precision and review. In this setting, we characterized these measures as takes after [17]:

$$precision = \frac{\text{number of answers correctly scored by AAES}}{\text{the total number of answers scored by AAES}}$$

$$\text{recall} = \frac{\text{number of answers correctly scored by AAES}}{\text{the total number of answers scored by the instructor}}$$

$$F1 = \frac{2 * precision * \text{recall}}{precision + \text{recall}}$$

From (Table 8) the results reveal 78% precision is accomplished by AAES. A few essays have missed and spurious by AAES and those made a few blunders, from table (6), (7) and (8) we have observed that proposed framework results are closer to the instructor's results.
Fig. 3 demonstrates the graphical presentation of the correlation between traditional human scoring (instructor's score) and Proposed AAES framework scoring.

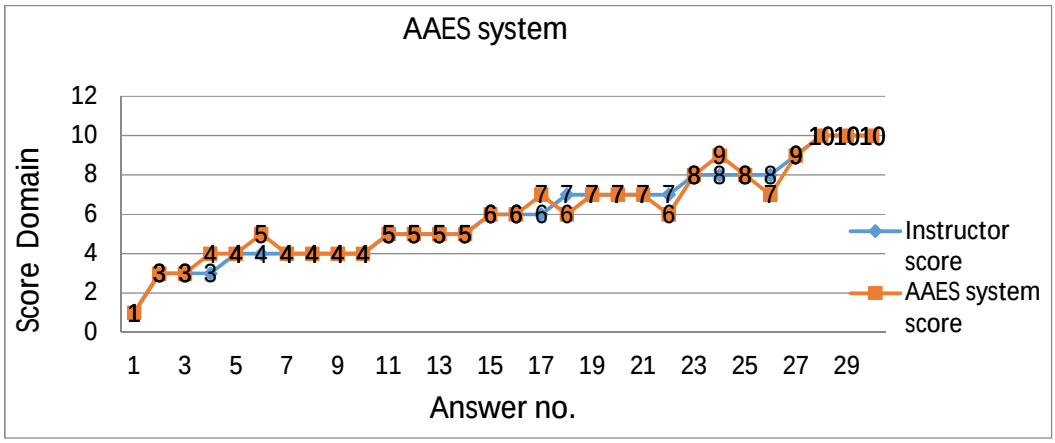

Figure (3): Comparison of results between traditional human scoring and the Proposed AAES system scoring.

In Fig. 3 the blue lines represent the score of the instructor for a particular answer and the orange lines represents the score of AAES for a particular answer. For each answer,

the orange lines tend to equal with blue lines. We have observed that the vast majority of AAES scores are equivalent to the instructor's scores which make the straight line. A percentage of the instructor's scores are missed by the framework, yet at the same time near to the straight line. Along these lines, we see that the proposed framework's score is near to instructor's score for each one answer.

For the quality measure of the proposed framework and in the assessment stage, the score of the answer by AAES has been compared with the teacher score. For correlation, we have processed the mean of errors by averaging the greatness each one machine score deviated from its relating instructor score. Also, we likewise processed the standard deviation of errors [17].

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$SD = \frac{\sqrt{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + (x_n - \bar{x})^2}}{n}$$

When

$\bar{x}$ Is the arithmetic mean from all errors.

x_i Is an absolute value of an error between instructor score and AAES score.

n Is the number of data set where n =30.

From the results of table 6 we have calculated the standard deviation (SD) error. (Table 9) shows the Mean of Errors, Standard Deviation of Errors, Correlation Degree and Accuracy of the proposed system.

Table (9): Mean of Errors, Standard Deviation of Errors, Correlation Degree and Accuracy of the proposed system

Experiment	PROPOSE AAES SYSTEM
Mean of Errors	0.233
Standard Deviation of Errors	0.077
Correlation Degree	0.978
Accuracy	90-100%

From the previous tests, the experimental reveals great results, The exploratory results demonstrate a significant connection between instructor's score and AAES score. This can prompt the modifying of automatic scoring systems' not just focused around Multiple Choice Question(MCQ), yet rather on semantic of unlimited essays. This system can also be used in the distance learning systems where students can connect to the system and freely submit the answers and obtained the score in about 10 seconds only, thus leading to improve the learning's efficiency, learning performance and overcome time and cost, AAES system is objective, based on the model answers and free of instructor's mood this leading to improve the learning's reliability.

Conclusions and future work

The objective of this research is an innovative approach to designing and implementing an AAES system using VSM and LSI. This model has been successfully applied to find the degree of similarity between electronic essays. The experimental results when applied the 4 model answers and the 30-learners answer shows that the proposed system can exactly provide an electronic assessment closer to instructors' traditional assessment in about only 10 seconds only; resulting in facilitates learning efficiency, learning performance and to overcome time, cost and reliability by evaluating of a large number of essays and tests answers with short time (about 10 seconds). Moreover, it provides the possibility of holding exams at any time or place on the Internet without the needs for the existence of users in the educational institution and at a specific time, which saves time and effort and money and this leads to the lifting of the educational process.

It would be interesting to use the proposed AAES framework to different information sets, for example of not native language. Since proposed AAES framework is truly general, all that eventual expected to adjust it to an alternate space is a preparation corpus of reviewed essays.

REFERENCES

[1] .Reinald Kim T. Amplayo, Sean Carlo C. Bermejo & Michael John N. Pedros, "A Teacher-Guided Automated Essay Grading Tool using Gradient Descent Algorithm and Reflective Random Indexing", International Journal of Advanced Research in Computer Science and Software Engineering, vol. 4, no. 3, pp. 23-29, 2014.

[2] .Shermis, M. D., Burstein, J., Higgins, D., & Zechner K., "Automated essay scoring: Writing assessment and instruction", International Encyclopedia of Education, vol. 4, no. 1, pp. 20-26, 2010.

[3] .S. Dikli, "The Nature of Automated Essay Scoring Feedback", CALICO Journal, vol. 28, no. 2, pp. 99-134, 2010.

[4] .V. Learning, A Preliminary study of the efficacy of IntelliMetric™ for use in scoring Hebrew assessments, Newtown,USA: PA: Vantage Learning, 2001.

[5] .V. Learning, A study of IntellimetricTM for responses in scoring Bahasa Malay, Newtown, USA: PA: Vantage Learning, 2002.

[6] .Kawate-Mierzejewska, "E-rate software", in Paper presented at the Japanese Association for Language Teaching, Tokyo, Japan, 2003.

[7].Xingyuan Peng ,Dengfeng Ke ,Zhenbiao Chen & Bo Xu, "Automated Chinese Essay Scoring using Vector Space Models", in Universal Communication Symposium (IUCS), Beijing, china, 2010.

[8] .Robert Östling, André Smolentzov, Björn Tyrefors Hinnerich and Erik Höglin, "Automated Essay Scoring for Swedish", in appear in The 8th Workshop on Innovative Use of NLP for Building Educational Applications, Atlanta, Georgia, USA, 2013.

[9] .Peter D. Turney, Patrick Pantel, "From Frequency to Meaning: Vector Space Models of Semantics", Journal of Artificial Intelligence Research, vol. 37, no. 1, pp. 141-188, 2010.

[10] .C. Yu, "Latent Semantic Indexing for an Online Research Interest Matching System", in Presented at the International Conference on Advanced Information Engineering and Education Science (ICAIEES), Beijing, China, 2013.
[11] .Salton G., Wong A., & Yang C.-S, "A vector space model for automatic indexing", Communications of the ACM, vol. 18, no. 11, pp. 613-620, 1975.
[12] .Abdul Monem S. Rahma , Suhad M. Kadhem & Alaa Kadhim Farhan, "Finding the Relevance Degree between an English Text and its Title", Engineering & Technology Journal, vol. 30, no. 9, pp. 1625-1640, 2012.
[13] .A. R. Abbas, "Designing a Personallsed and Recommended E-learning System",Ph.D. dissertation,School of Computer Science, Wuhan University ,Wuhan,P.R China, 2009.
[14] .Zhen Yu , Xing She Zhou, "Combining Vector Space Model and Category Hierarchy Model for TV Content Similarity Measure", in Presented at the Third International Conference of Multimedia and Ubiquitous Engineering (MUE), Zhangjiajie , China, 2009.
[15] .Landauer T. K., Folt P. W. & Laham D., "Introduction to Latent Semantic Analysis", Discourse Processes journal, vol. 25, no. 2, pp. 259-284, 1998.
[16] .Dr. Amaal A. Mohammed & Dr. Sudad K. Ibraheem, "Approximated Solution of Higher–Order Linear", Engineering & Technology Journal, vol. 28, no. 14, pp. 4722-4729, 2010.
[17] .Md. Monjurul Islam, A. S. M. Latiful Hoque, "Automated Essay Scoring Using Generalized", Journal of Computers, vol. 7, no. 3, pp. 616-626, 2012.