# Automated Arabic Essay Scoring (AAES) using Vector Space Model (VSM)

**Dr. Ayad R. Abbas, Ahmed S.Al-qaza**

**Department of Computer Sciences, University of Technology**

ayad_cs@yahoo.com,ahmed_alkzaz243@yahoo.com

## ABSTRACT

**Automated Essays Scoring (AES) may be defined as the ability of computer technology to evaluate electronic essays written by learner according to previous determined essay.Researches and all the former works were applied to essays written in English language and they were applied to essays written in Hebrew, Bahasa Malay, Japanese, Chinese and Swedish. The research paper suggests an Automated Arabic Essays Scoring (AAES) system in web-based learning context based on Vector Space Model (VSM). The suggested system consists of two main processes. The first process deals with applying retrieval information to extract the important information from electronic essays. In the second process, the VSM is applied to find out the similarity degree between the previously written essays by the instructor and the essay written by the learner.**

**The experimental results show that the proposed system provides an electronic assessment closer to instructors' traditional assessment, leading to improve the learning's efficiency, learning performance and to overcome time, cost and reliability.**

**Keywords: Automated essay scoring, Web-based learning, Information retrieval, Vector space model.**

## التقييم الالي للمقالات العربية باستخدام نموذج فضاء المتجهات (VSM)

د. ياد روضان عباس احمد سعد القزاز

قسم علوم الحاسوب -الجامعة التكنلوجية

### المستخلص

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 كتبتباللغة 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). يتكون النظام المقترح من عمليتين رئيسيتين. العملية الأولى يتم فيها تطبيق استرداد المعلومات لاستخراج المعلومات الهامة من مقالات الإلكترونية. الثانية، يتم تطبيق (VSM) 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.

## 1. Introduction

Automated Essay Scoring (AES) defined as the computer technology that evaluates and scores the written prose AES systems mainly used to overcome time, cost, reliability, and generalizability issues in essay scoring. AES system programmed in 1973 but required punch cards and a mainframe computer, making it inaccessible to most instructors. [1] AES continues attracting the attention of public schools, universities, testing companies, researchers and educators. [1] A number of studies conducted to assess the accuracy and reliability of the AES systems. Furthermore, there were several AES studies reported high agreement rates between AES systems and human raters. [2] The vision of having effective algorithms score student essays should be appealing to the teacher, test publisher, and research scientist Teachers freed of the burden of reading and hand-scoring maybe hundreds of student papers and consequently would be more likely to assign written questions and probe for deeper understanding. [1]

Test publishers would be able to score essays for less cost and conceivably provide Higher-quality assigned grades, the computer cannot possibly score an essay the same way, rather, and AES seeks to use the computer's special capabilities. [2]AES continues attracting the attention of public schools, universities, testing companies the vision of having effective algorithms score student essays should be appealing to the teacher, test publisher, and research scientist. Most of the previous works in English. Whereas, the past works are applied in Hebrew, Bahasa Malay, Japanese, Chinese and Swedish. [3-7]

The VSM used to calculate the relation degree between instructor essays and learner's essays. In the VSM, each electronic essay

represented as a vector, and their relevance to the queries submitted by the user measured through appropriate matching functions, The model creates a space in which both electronic essays and queries a represented by vectors .[8]

## 2. Vector Space Model (VSM)

The VSM developed for the SMART information retrieval system, by Gerard Salton and his colleagues. [8]; it is obtained to get the pattern in the document collection, which used to improve the accuracy in information retrieval systems. Its idea is to represent each document in a collection as a point in a space (a vector in a vector space). Points that are close together in this space are semantically similar, and points that are far apart are semantically distant, the user's query represented as appoint in the same space as the documents (the query is a pseudo-document). [9]VSM for information retrieval are just one subclass of retrieval techniques that have been studied in recent years, The taxonomy provided in labels the class of techniques that resemble vector-space models "formal, feature-based, individual, partial match" retrieval techniques since they typically rely on an underlying. Formal mathematical model for retrieval, model the documents as sets of terms that can be individually weighted and manipulated, perform queries by comparing the representation of the query to the representation of each document in the space, and can retrieve documents that do not necessarily contain one of the search terms. Although the vector-space techniques share common characteristics with other techniques in the information retrieval hierarchy, they all share a core set of similarities that justify their own class. [3]

### 3. Similarity Measure Estimation

**The VSM is used to calculate the relation degree between essay-based examination (between instructor essays and learner's essays). In the VSM, each essay is represented as a vector, and their relevance to the queries submitted by the user is measured through appropriate matching functions. The model creates a space in which both essays and queries represented by vectors, For a fixed collection of essays, an M-dimensional vector generated for each essay and each query from sets of terms (i.e., keywords, vocabularies) with associated weights, where M is the number of unique terms in the essays collection. In VSM, weights associated with the terms calculated based on the following four numbers: [9] [10]**

**1- $tf_k$ is the term frequency (term counts) or number of times a term k occurs 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.**

**4- IDF is invers of the essay frequency or number of essays containing term**

**The coordinate of the (i) th essay in the direction corresponding to the (k) th linguist term (the weight of the term in the specific essay) can be determined the follows:**

$$w_{ik} = tf_{ik} \times \log\left[\frac{D}{Df_{ik}}\right] \times IDF \text{.....(Equation 1)} \quad [10]$$

**Then, a vector similarity function, such as the inner product, can be used to compute the similarity between essay and a query:**

$$r_{ij} = \cos(\theta) = \frac{\sum_{h=1}^{m} w_{ih} w_{jh}}{\sqrt{\sum_{h=1}^{m} w_{ih}^{2} \sum_{h=1}^{m} w_{jh}^{2}}} \text{.....(Equation2)} \quad [10]$$

**Where $c_i=\{w_{i1}, w_{i2},..., w_{ik},..., w_{im}\}$ and $c_j=\{w_{j1}, w_{j2},..., w_{jk}, ..., w_{jm}\}$ respectively, rij denotes the concept relation degree between the ith and jth essays.[9] [10]**

**4. Term-document matrix**

**To apply linear algebra to information retrieve, we first need to transform the problem into mathematical form; it is a sparse matrix whose rows correspond to terms and whose columns correspond to documents. [11]**

**5. Overview of the System Architecture**

**Figure 2 illustrates the proposed system architecture, which provides various services, such as a learner's interface service, essay-based examination management 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.**

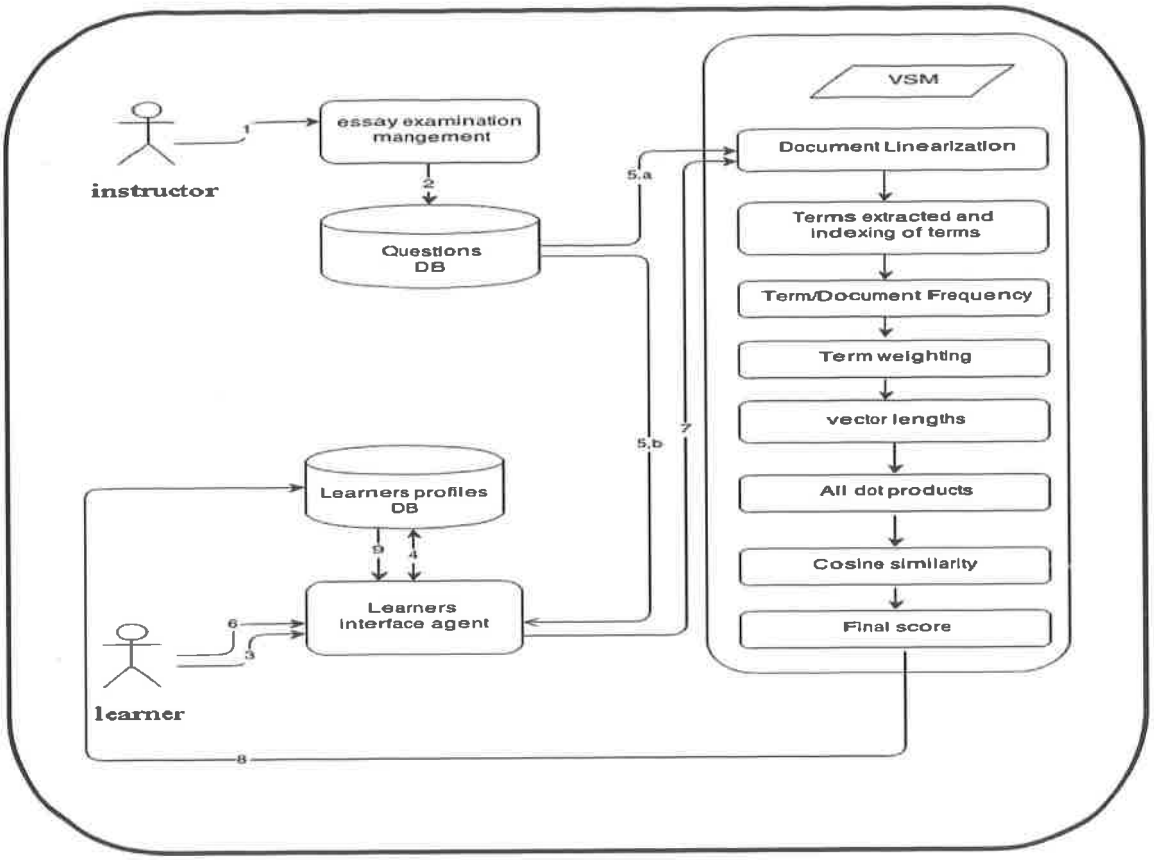


**Figure 2: The proposed system architecture**

**The operation procedure of the proposed system:**

**m architecture summarized as follow:**

**Steps 1-2: Instructor accesses the essay-based examination management service to access, activate and manageessay-based examination. Moreover, they can assign best essay-based examinationto the learner and AAES service.**

**Steps 3-7: The learner accesses the learner's interface agent, and then the system gets his or her learning profile from the learner profile database. Moreover, learning profile service receives specific electronic essay exam from questions database and then sends the learner answers to the AAES service.**

**Steps 8: implement (VSM) to find the final score.**

**Steps 9: Finally, the learner's interface agent presents the electronic essay grades to the learners by connected with AAES service.**

## 6. System Results and Implementation:

**The proposed system was successfully implemented using Microsoft Windows 7 with Apache Web Server version 2.2.6, Java script, PHP script language version 5.2.5, MySQL database version 5.0.45 and Matlab-mathworks version R2013a.in order to provide the proof of feasibility study and verify the system performance for the proposed method, 30essays has been applied to the system.**

**The results of tests the proposed system by enter 4-predefind essays in questions database (table 1), and 1-learners essay in learners interface agent (table 2), are shown in (table3) and the following operations.**

| No | essay |
|---|---|
| 1 | الشبكات هي هيكلية تنظيمية لربط حاسبتين او مجموعة من الحاسبات تنتشرفي 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 . |
| 2 | الشبكات هي ربط حاسوبين او مجموعة من الحواسيب تنتشر في مكان واحد او اماكن مختلفة وترتبط ببعضها بواسطة جهاز اتصال او اجهزة الاتصال المختلفة سلكيا او لاسلكيا من اجل تبادل البيانات بينها ومشاركة الاجهزة وموارد الحاسوب مثل الالة الطابعة او البرامج التطبيقية او الالعابو كذلك تسمح بالتواصل المباشر بين الاشخاص. |
| 3 | 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 وغيرها من الاجهزة لغرض تبادل المعلومات والتواصل مع الاخرين. |
| 4 | 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 وتسمح بتبادل البيانات وموارد الكمبيوتر مثل الطابعات وتسمح للمستخدمين بالتواصل بشكل فوري . |

**Table 1: predefined essays**

| Essay | no |
|---|---|
| هي الحاسبات و الاجهزة المحمولة المربوطة مع بعضها لإرسال الرسائل و الصور و البيانات و تصفح الأنترنتو التواصل مع الأصدقاء | 1 |

**Table 2: learner essay**

| t e r m s | Quer y | Documen t 1 | Documen t 2 | Documen t 3 | Document 4 | df | idf | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| اتصال | 0 | 0 | 1 | 0 | 0 | 1 | 0.6021 | 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 |
| ..... | ... | ... | ... | ... | .. | .. | ..... | ... | .... | .. | ... |

**Table 3: the proposed system's retrieval results**

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 (document 1, document 2, document 3, document 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 invers 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 (1)

The (VSM) score finds by compute the similarity between predefined essays and a query by using equation 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 take the highest score as the VSM score=0.274

The proposed system suggests a score range between (0-10) for this reason the system mapped the final score to the suggests score by multiply the final score by 10, the final score now becomes 2

**Comparison of results between traditional human scoring and the Proposed AAES system scoring for 4-predefind essays and 30-learners essay are shown in (figure 3)**

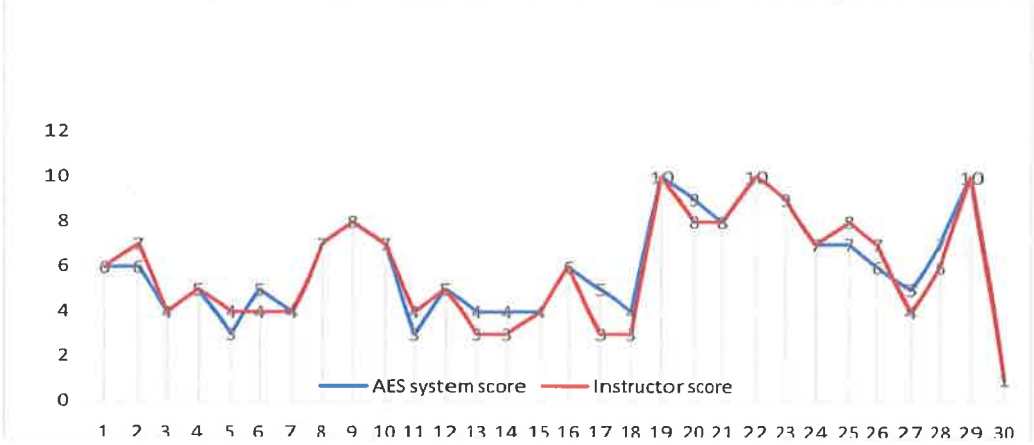


**Figure 3: Comparison of results between traditional human scoring and the Proposed AAES system scoring**

## 7. Conclusions and future work

**The major contribution of this research is an innovative approach to designing and implementing an AAES system using VSM. This Model has been successfully applied to find degree of similarity between electronic essays. The experimental results show that the proposed system can exactly provide an electronic assessment closer to instructors' traditional assessment; 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. 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 apply some of our methods to other languages and other data sets, for instance of second language learners. Since our system is quite general, all that would be needed to adapt it to another domain is a training corpus of graded essays.**

## References


1- **Shermis, M. D., Burstein, J., Higgins, D., & Zechner, K. (2010), "*Automated essay scoring: Writing assessment and instruction*", In E. Baker, B. McGaw & N. S. Petersen (Eds.), International Encyclopedia of Education, Volume 4,Issue 1,Page 20-26,Oxford, UK: Elsevier.**

2- **Semire Dikli, (2010) , *"The Nature of Automated Essay Scoring Feedback"*, CALICO Journal,Volume 28, Issue 2, Page 99-134.**

3- **Vantage Learning, (2001),"*A Preliminary study of the efficacy of IntelliMetric™ for use in scoring Hebrew assessments*, Newtown, PA: Vantage Learning.**

4- **Vantage Learning, (2002), "*A study of IntellimetricTM for responses in scoring Bahasa Malay*", Newtown, USA, Vantage Learning publisher.**

5- **Kawate-Mierzejewska, M, (2003.),"*E-rate software*", Paper presented at the Japanese Association for Language Teaching, Tokyo, Japan.**

6- **Xingyuan Peng , Dengfeng Ke ,Zhenbiao Chen & Bo Xu,(2010). "*Automated Chinese Essay Scoring using Vector Space Models*", Universal Communication Symposium (IUCS), Beijing, china, IEEE publisher.**

7- **Robert Östling, André Smolentzov, Björn Tyrefors Hinnerich and Erik Höglin, (2013) ", *Automated Essay Scoring for Swedish*", appear in The 8th Workshop on Innovative Use of NLP for Building Educational Applications, Atlanta, Georgia, US.**

8- **Peter D. Turney, Patrick Pantel ,(2010*),"From Frequency to Meaning: Vector Space Models of Semantics*" , Journal of Artificial Intelligence Research ,Volume 37,Issue 1, Page 141-188**

**9- Salton, G., Wong, A., & Yang, C.-S. (1975), "*A vector space model for automatic indexing* ", Communications of the ACM, Volume 18, Issue 11, Page: 613-620.**

**10- Zhen Yu; Xing she Zhou, (2009),"*Combining Vector Space Model and Category Hierarchy Model for TV Content Similarity Measure*",Presented at the Third International Conference of Multimedia and Ubiquitous Engineering (MUE), Zhangjiajie , China.**

**11- Landauer, T. K., Foltz, P. W., & Laham, D. ,(1998),*Introduction to Latent Semantic Analysis*, Discourse Processes journal, Volume 25,Issue 2,Page:259-284.**