A Hybrid Scheme for Automated Essay Grading Based on LVQ and NLP Techniques.

Abdulaziz Shehab[1], Mohamed Elhoseny[2], and Aboul Ella Hassanien[3]
[1,2] Faculty of Computers and Information, Mansoura University, Egypt
[3] Faculty of Computers and Information, Cairo University, Egypt
[2,3]The Scientific Research Group in Egypt (SRGE)
Abdulaziz_shehab@mans.edu.eg, Mohmed_elhoseny@mans.edu.eg, aboitcairo@gmail.com

***Abstract*—This paper presents a hybrid approach to an Automated Essay Grading System (AEGS) that provides automated grading and evaluation of student essays. The proposed system has two complementary components: Writing Features Analysis tools, which rely on natural language processing (NLP) techniques and neural network grading engine, which rely on a set of pre-graded essays to judge the student answer and assign a grade. By this way, students essays could be evaluated with a feedback that would improve their writing skills. The proposed system is evaluated using datasets from computer and information sciences college students' essays in Mansoura University. These datasets was written as part of mid-term exams in introduction to information systems course and Systems analysis and design course. The obtained results shows an agreement with teachers' grades in between 70% and nearly 90% with teachers' grades. This indicates that the proposed might be useful as a tool for automatic assessment of students' essays, thus leading to a considerable reduction in essay grading costs.**

***Keywords*—** Essay Grading, Neural Network, Natural Language Processing, Student Writing

I. Introduction

Open-ended questions (essays) have several advantages over traditional multiple-choice, true or false, and fill in blanks assessments but the greatest obstacle for their adoption is the large cost and effort required for scoring. Teachers all over the world spend a great deal of time just assessing students works. Hence, they have to cut down the time they can devote to their other duties. Even doing that, sometimes they do not have enough time to properly assess the big number of students they have. Recently, there exists an extensive work to automate the whole education process using computer and information technology [24], [29]. The impacts of utilizing computers for essay writing have been widely studied for four decades where it has the ability to assess students' work. The best way to improve students writing skills is to write, receive feedback from a teacher, and then repeat the whole process as often as possible for each student in the classroom. Unfortunately, this puts an enormous load on the classroom teacher who is faced with reading and providing feedback for perhaps 40 essays or more every time a topic is assigned. In fact, responding to student papers and carefully tracking them can be a burden for many teachers. Particularly, if they have a large number of students, assign frequent writing assignments, and provide individual feedback to student essays, might be time consuming [13]. As a result, teachers become unable to dictate writing assignments as often as they wish [3]. Therefore, developing systems that can automatically grade these essays can help reduce these costs in a significant way and may facilitate extended feedback for the students. Many years ago, researchers have sought to develop applications that automate the education process [15], [16], specially essay grading and evaluation. Work in automated essay grading began in the early 1960s and has been extremely productive [1]–[6]. Pioneering work in automated feedback was initiated in the 1980s with the Writers Workbench [7]. Detailed descriptions of these systems appear in section 2.

AEGS systems do not actually read and understand essays as humans do. Whereas human raters may directly evaluate various intrinsic variables of interest, such as diction, fluency, and grammar, in order to produce an essay score, AEGS systems use approximations or possible correlates of these intrinsic variables [1]–[3].

Although AEG is a developing technology, the search for better machine scoring is ongoing as investigators continue to move forward in their drive to increase the accuracy and effectiveness of AEG systems. The proposed system combines both automated essay grading and diagnostic feedback. The feedback is specific to the students essay and is based on the kinds of evaluations that teachers typically provide when grading a students writing. The proposed system is intended to be an aid, not a replacement, for grading a large corpus of students' answers. Its purpose is to reduce the teachers load, thereby encouraging the teacher to give students more writing essays tasks.

The rest of the paper is organized as follows: section II presents a literature of different current AEG systems. Section III describes the overall framework of the proposed system. Section IV presents the experimental results and discussion. Finally, conclusion and future work are drawn in section V.

II. Literature Review

AEG systems are mainly developed based on English language. There are already a number of systems that assess writing in various languages but less than those found in English [17]. Project Essay Grader (PEG) was developed by Ellis Page in 1966 upon the request of the College Board, which wanted to make the large-scale essay scoring process

9781509028634/16/$31.00 ©2016 IEEE

more practical and effective [18]. PEG uses correlation to predict the intrinsic quality of the essays [20], [21]. Page relies on style analysis of surface linguistic features of a block of text. It has some drawbacks like the need for training, in the form of assessing a number of previously manually marked essays for proxes, in order to evaluate the new essays. In addition it is only rely on linguistic features where neither NLP nor lexical content are taken in account.

Schema Extract Analyse and Report (SEAR) is a software system developed by Christie [22] as a result of his PhD research work. According to [22], the automated grading of essays requires the assessment of both style and content (where appropriate). The main problems faced this system are the lack of style marked essays sets, the confusion that students mistakes in spelling and grammar creates to the system, and the use of many alternative expressions to say just the same.

Automark is a software system developed in pursuit of robust computerized marking of free-text answer to open-ended questions [23]. Automark uses NLP techniques to mark open-ended responses. The grading process geos through a number of stages. First, the incoming text is pre-processed to standardize the input in terms of punctuation and spelling. Then, a sentence analyzer identifies the main syntactic constituents of the text and how they are related. Finally, the feedback module processes the result of the pattern match.

IntelliMetric [25] was created by the company Vantage Learning, after having spent more than ten millions dollars in its development. It is a commercial system whose focus is on emulating the human scorer by grading the content, the style, the organization and the conventions of each response using a 1-4 scale [34]. Its current version is IntelliMetric 9.3 [26]. IntelliMetric requires an initial training phase with a set of manually scored answers in order to infer the rubric and the human graders judgments to be applied by the automatic system. Because it is not an academic product, there is little published information about the techniques that it employs. However, Vantage Learning Technologies has stated that IntelliMetric relies on other of their proprietary systems, the so-called Cogni Search and the Quantum Reasoning Technologies. Moreover, they have claimed that they used an Intelligent Artificial approach, because IntelliMetric uses its intelligence to score the students texts [26].

The Japanese Essay Scoring System (Jess) [27] is the first automated Japanese essay scorer. It has been created in the National Research Center for the University Entrance Exam in Japan. It examines three features in the essay: rhetoric (i.e. syntactic variety), organization (i.e. how ideas are presented and related in the essay) and content (i.e. how relevant is the information provided and how precise and related to the topic is the vocabulary employed). For content, it uses Latent Semantic Analysis (the training is done using editorials and columns taken from the Mainichi Daily News newspaper as learning models).

The Paperless School Marking Engine (PS-ME) [19] is the system presented by Mason and Grove-Stephenson in the Birmingham University in UK in 2002, and it has also become commercially available. Its assessment objective is both summative and formative, with little or no human intervention. Besides, it can be integrated in a learning management system, or be used as a stand-alone application. This system depends on NLP techniques to assess student essays in order to reveal their level for knowledge, understanding and evaluation. PS-ME requires an initial training phase with at least 30 sample of pre-graded essays that could include not only reference texts, but also negative texts with a very low score.

The Semantic Analysis Grader (SAGrader) [?] is the tool offered by the American company Idea Works to assess free-text answer essays. It uses the proprietary QTools developed by the same company to recognize patterns in students essays and compare them with the model answer. In this way, it can give detailed feedback to students. SAGrader has been tested in sociology courses in introductory freshman and sophomore level classes in a university setting. It has been proved its good performance for classes where the primary objective is to assess students knowledge at an introductory or intermediate level. Moreover, students reported how they like the program because it provides immediate detailed feedback any time of the day or night and gives them an opportunity to revise their paper and improve their grade.

III. The Proposed System

This system contains two complementary components: the grading Engine component that is based on neural network which trained on a set of pre-graded essays by human and then compute the similarity between the new essay and the pre-graded essays in order to assign a grade. The second component, writing features analysis, is comprised of a suite of programs that evaluate and provide feedback for errors in grammar, usage, and mechanics, identify the essays discourse structure, and recognize undesirable stylistic features that are based on natural language processing (NLP) tools as shown in Figure 1.

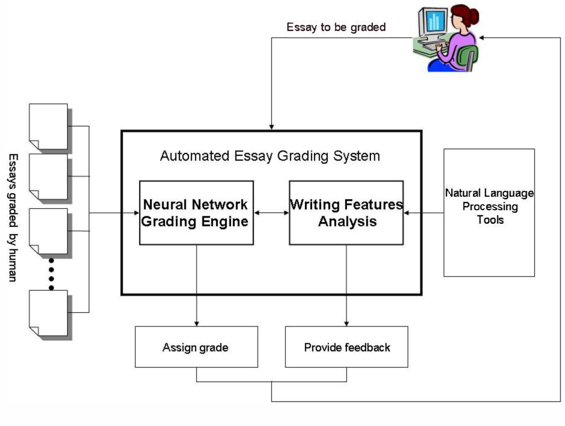
Essay to be graded
Essays graded by human
Automated Essay Grading System
Neural Network Grading Engine
Writing Features Analysis
Natural Language Processing Tools
Assign grade
Provide feedback

Fig. 1. Automated Essay Grading System (AEGS) Framework

A. The neural network grading engine

The neural network grading engine is responsible for evaluating the student's response and assigns it a grade. In order to assign a grade, the neural network grading engine follows the steps shown at Figure 2. The following section briefly describes these steps.

1) Check Spelling: The check spelling module notify the student with the misspelled words found in the essay and provide him/her with suggestions. Then the student has the decision whether to accept the modification or to ignore.

2) Document Tokenization: The second phase is to identify useful features from the student answer. This is done by breaking the stream of characters into words or, more precisely, tokens. This is a fundamental to further analysis. Without identifying the tokens, it is difficult to extracting higher-level information from document. In case under study the white space was taken into consideration as a delimiter in this process.

3) Remove Stopwords: Stop words, or stopwords, is the name given to words which are filtered out prior to, or after, processing of natural language data (text). In this phase the common words (stopwords) that doesn't affect the main idea of the response such as ”a”, ”the”, ”is”, ”was”, ”got”, ”have”, and so on were trimmed. The remaining words treated as the keywords that used to distinguish good essays from bad ones.

4) Stemming to a Root: Stemming is the process for reducing inflected (or sometimes derived) words to their stem, base or root form generally a written word form. The stem need not be identical to the morphological root of the word; it is usually sufficient that related words map to the same stem, even if this stem is not in itself a valid root.

Our stemming efforts in this study were restricted to the more traditional Porter Stemmer which follows the affix removal approach [8]. Stemming process is an optional step such that the system can skip this phase and deal directly with the next step.

5) Learning algorithm: Learning algorithm indicate how to assign a grade to a student's response depending on a set of pre-graded essays Learning vector quantization (LVQ) is a method for training competitive layers in a supervised manner [9].

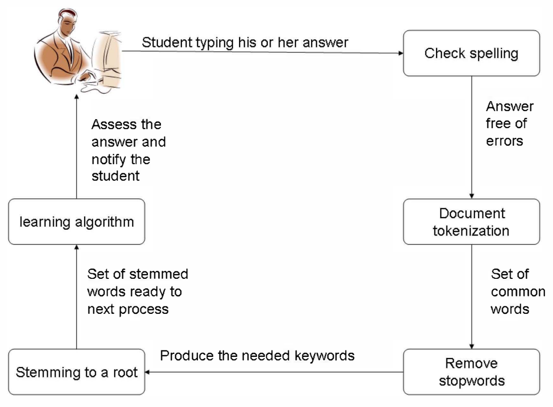
Student typing his or her answer
Check spelling
Answer free of errors
Document tokenization
Set of common words
Remove stopwords
Produce the needed keywords
Stemming to a root
Set of stemmed words ready to next process
learning algorithm
Assess the answer and notify the student

Fig. 2. The inner processes in the Neural Network grading engine

Learning vector Quantization (LVQ) is a neural net that combines competitive learning with supervision. It can be used for pattern classification.

A training set consisting of Q training vector - target output pairs are assumed to be given

$$\left\{\mathbf{s}^{(q)} : \mathbf{t}^{(q)}\right\}, \qquad q = 1, 2, \ldots, Q,$$

where $\mathbf{s}^{(q)}$ are N dimensional training vectors, and $\mathbf{t}^{(q)}$ are M dimensional target output vectors. M is the number of classes, and it must be smaller than Q. The target vectors are defined by

$$t_i^{(q)} = \begin{cases} 1, & \text{if, } \mathbf{s}^{(q)} \text{ belongs to class } i \\ 0, & \text{otherwise.} \end{cases}$$

The LVQ is made up of a competitive layer, which includes a competitive subnet, and a linear layer. In the first layer (not counting the input layer), each neuron is assigned to a class. Different neurons in the first layer can be assigned to the same class. Each of those classes is then assigned to one neuron in the second layer. The number of neurons in the first layer, Q, will therefore always be at least as large as the number of neurons in the second layer, M.

In the competitive layer, neurons in the first layer learns a prototype vector which allows it to classify a region of the input space. Closeness between the input vector and any of the weight vectors is measured by the smallness of the Euclidean distance between them. A subnet is used to find the smallest element of the net input

$$\mathbf{n}^{(1)} = \left[\ \|\mathbf{x} - \mathbf{W}^{(1)}_{\cdot 1}\| \quad \|\mathbf{x} - \mathbf{W}^{(1)}_{\cdot 2}\| \quad \ldots \quad \|\mathbf{x} - \mathbf{W}^{(1)}_{\cdot Q}\|\ \right]$$

and set the corresponding output element to 1, indicating that the input vector belongs to the corresponding class, and set all others to 0. The action of this subnet is represented as a vector-valued vector function

$$\mathbf{a}^{(1)} = \text{compet}(\mathbf{n}^{(1)}).$$

Since some of these classes may be identical, they are really subclasses. The second layer (the linear layer) of the LVQ network is then used to combine subclasses into a single class. This is done using the $\mathbf{W}^{(2)}$ weight matrix which has element

$$w_{ij} = \begin{cases} 1, & \text{if the } i \text{ neuron belong to a subcass of } j, \\ 0, & \text{otherwise.} \end{cases}$$

Once $\mathbf{W}(2)$ is set, it will not be altered.

On the other hand, the weights, $\mathbf{W}(1)$, for the competitive layer have to be trained using the Kohonen LVQ rule.

At each iteration one of the training vector is presented to the network as input $\mathbf{x}$, and the Euclidean distance from the input vector to each of the prototype vector (forming columns of the weight matrix) is computed. The hidden neurons compete. Neuron $j*$ wins the competition if the Euclidean distance between $\mathbf{x}$ and the $j*$ prototype vector is the smallest. The $j*$ element of $\mathbf{a}^{(1)}$ is set to 1 while others are set to 0. The activations $\mathbf{a}^{(1)}$ is then multiplied by $\mathbf{W}^{(2)}$ on its right to get the net input $\mathbf{n}^{(2)}$. This produces the output of the entire network $\mathbf{a}^{(2)} = \mathbf{n}^{(2)}$, since the transfer function of the output neurons is an identity function. $\mathbf{a}^{(2)}$ also has only one nonzero element $k*$, indicating that the input vector belongs to class $k*$.

Two phases required to grade the essays. Training phase: the teacher is responsible for feeding the net with a set of pre-graded answers and then the motivation for the algorithm is to find the output unit (grade) that is closest to the input vector (answer) until reaches the weights that have the ability to correctly classify the different types of students' answers. Testing phase: consists of a set of students' answers (not used in the training phase) of the un-graded essays. Then the system follows the pre-mentioned steps until converting the answer into an input vector. Finally the input vector presented to the LVQ net in order to classify the answer to a winner a class which represent a specific grade.

B. Writing Features Analysis

AEGS was always based on a large number of features that were not individually described or linked to intuitive dimensions of writing quality. Many systems [10] are based on hundreds of undisclosed features.

The writing analysis tools identify five main types of grammar, usage, and mechanics errors agreement errors, verb formation errors, wrong word use, missing punctuation, and typographical errors [12]. The writing analysis tools also highlight aspects of style that the writer may wish to revise, such as the use of passive sentences, as well as very long or very short sentences within the essay. Another feature of undesirable style that the system detects is the presence of overly repetitious words, a property of the essay that might affect its rating of overall quality [11].

The purpose of developing automated tools for writing instruction is to enable the student to get more practice writing. At the same time, it is essential that students receive accurate feedback from the system with regard to errors, comments on undesirable style, and information about discourse elements and organization of the essay. If the feedback is to help students improve their writing skills, then it should be similar to what an instructors comments might be. The feedback obtained on these writing features improve performance, in part, because they better reflect what teachers actually consider when grading student writing.

Finally after the pre-mentioned errors in writing were detected and displayed to the student, a common feedback on his/ her answer might be something like one of the following:

- *Your essay does not resemble others being written on this topic.*
- *Your essay might not be relevant to assigned topic.*
- *Your essay appears to be restatement of the topic with a few additional concepts.*
- *Compared to other essays written on this topic, your essay contains more repetition of words.*
- *Your essay shows less development of a theme than other essays written on this topic.*

In this study, a small set of meaningful and intuitive features were used. The system relies on four modules shown in figure 3 in order to provide feedback to the student. Below is a brief description of these modules.

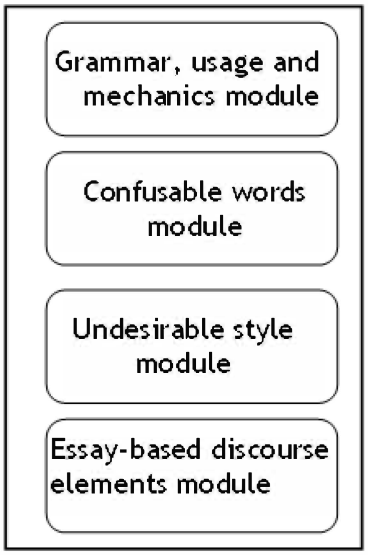
Grammar, usage and mechanics module
Confusable words module
Undesirable style module
Essay-based discourse elements module

Fig. 3. Writing feature analysis modules

1) Grammar, usage and mechanics module: The approach to detecting violations of general English grammar is corpus based and statistical, and can be explained as follows. The system is trained on a large corpus of edited text, from which it extracts and counts sequences of adjacent word and part-of-speech pairs called bigrams. The system then searches student essays for bigrams that occur much less often than would be expected based on the corpus frequencies [12]. The expected frequencies come from a model of English that is based on 30-million words of newspaper text. Every word in the corpus is tagged with its part of speech using a version of the MXPOST [13] part-of-speech tagger that has been trained on student essays.

To detect violations of general rules of English, the system compares observed and expected frequencies in the general corpus. The statistical methods that the system uses are commonly used by researchers to detect combinations of words that occur more frequently than would be expected based on the assumption that the words are independent.

2) Confusable Words Module: Some of the most common errors in writing are due to the confusion of homophones, words that sound alike. The Writing Analysis Tools detect errors among their/there/theyre, its/its, affect/effect and hundreds of other such sets. The context consists of the two words and part-of-speech tags that appear to the left, and the two that appear to the right, of the confusable word [14]. For example, a context for effect might be a typical effect is found, consisting of a determiner and adjective to the left, and a form of the verb BE and a past participle to the right. For affect, a local context might be it can affect the outcome, where a pronoun and modal verb are on the left, and a determiner and noun are on the right.

3) Undesirable style module: The identification of good or bad writing style is subjective; what one person finds irritating another may not mind. The Writing Analysis Tools highlight aspects of style that the writer may wish to revise, such as the use of passive sentences, as well as very long or very short sentences within the essay. Another feature of undesirable style that the system detects is the presence of overly repetitious words, a property of the essay that might affect its rating of overall quality [11].

4) Essay-based discourse elements module: A well-written essay should contain discourse elements, which include introductory material, a thesis statement, main ideas, supporting ideas, and a conclusion. Full details are presented in [14], [28].

IV. Experimental Results

The two courses selected for this study were introduction to information systems (1^{st} year level course) and systems analysis and design (a 3^{rd} year level course). In the spring of 2014, we collected two datasets from Mansoura University student's essays (student responses) written as a part of a mid-term exam in introduction to information systems course. Every dataset contains 400 essays. Also, another four datasets in systems analysis and design course were collected. Every dataset contains 200 essays. Table 1 shows the questions that were used in this study.

TABLE I
THE QUESTIONS THAT WERE USED IN THIS STUDY

Subject	Question
Introduction to Information System	1- Desine the term ”Information System”
	2- What is Data mining?
Systems Analysis and Design	1- Write short notes about prototype
	2- Write short notes about JAD
	3- Write short notes about RAD

Two human graders manually graded all essays in the six datasets. The first grader was the course teacher and the second was the teaching assistant for the course. Both the teacher and the teaching assistant assess (grade) the responses using the same aspects of the model answer provided by the course teacher. Each dataset is divided into training and test sets. Both the first dataset and the second dataset used in the 1st course (introduction to information systems) were divided into 50 training and 350 testing samples. For the rest of the datasets in the 2nd course (systems analysis and design), the corpus was divided into 50 training and 150 testing samples. The grade assign for each response was range from 0 (low) to 2 (high) step 0.5 such that we have five output groups. We took a 10 sample from each group for total 50 sets used to train the network on the various aspects of the students' responses. The ANN grade on each response is compared to its teacher's grade and teaching assistant's grade. Figure 4 and Figure 5 show the grades assigned by teacher, teaching assistant, and the machine on a sample of 50 students' response to two different questions. The correlation coefficient between human and automatic system was calculated. Table II shows correlation coefficients between the system (machine grader) and human graders. S_1 refers to Introduction to Information System where S_2 refers to System Analysis and Design. T_g, M_g, and TA_g refer to the teacher grader, the machine grader, and the teaching assistant grader respectively. As noticed from the table II, the correlation coefficients obtained either between teacher and machine or between teaching assistant and machine are close to those obtained between teacher and teaching assistant. As English is a second language for those who write these essays, there were multiple errors in their writing. Finally all what we want to say now is that the accuracy of the system is highly dependent on the datasets that the system was trained on.

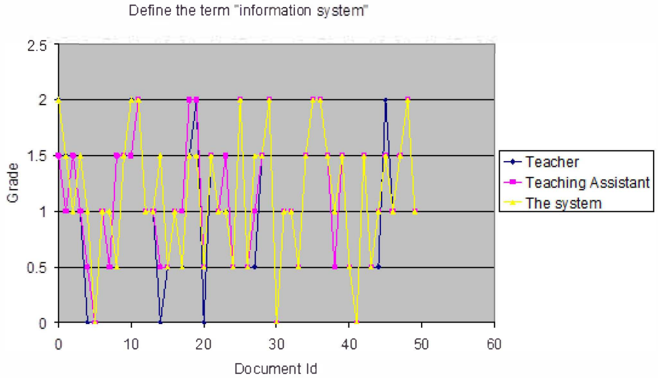

Fig. 4. Grades assigned by teacher, teaching assistant, and the machine on a sample of 50 students' response (case 1)

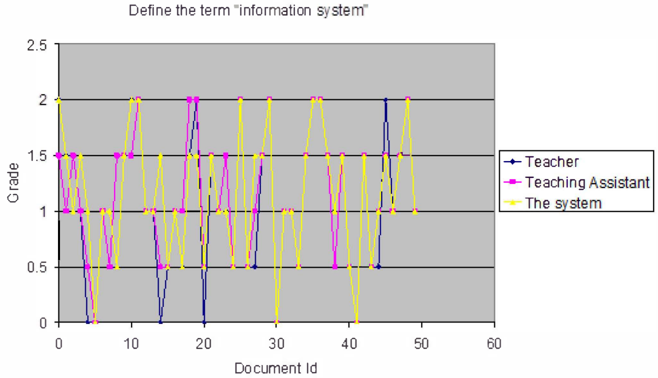

Fig. 5. Grades assigned by teacher, teaching assistant, and the machine on a sample of 50 students' response (case 2)

TABLE II
THE CORRELATION COEFFICIENT OBTAINED IN THIS STUDY

Subject	Dataset	T_g against M_g	TA_g against M_g	T_g against TA_g
S_1	1	0.8712	0.8343	0.8954
	2	0.8667	0.7823	0.8701
S_2	1	0.8457	0.7864	0.8716
	2	0.8112	0.8	0.8571
	3	0.7355	0.7211	0.7665

V. Conclusion and Future Work

This paper has presented a framework for the automatic grading of student essays which based on both ANN and NLP. The neural network was trained on a set of pre-graded essays by the teacher of the subject. As soon as the neural network has been trained, it can grade an answer of the unseen sets. The natural language processing tools has the task of providing the feedback to the students. The feedback is necessary especially for the students who speak English as a foreign language. It is noteworthy that the architecture of the essay grading system does not take the human reader out of the loop. Indeed, because the system requires initial training sets of manually graded essays. A drawback for this system is the requirement of 50 pre-graded responses that are needed to train the network. In some cases, the system may need more than 50 essays in order to build a reliable network be able to correctly grade the answers. This would obstacle in the datasets that are small. We are currently experimenting with an interactive system that can be used to improve, or extend, the model automatically even when the data sets are small. This will make the system building more flexible and this in turn should make it possible to apply the automatic grading not just for large datasets but also for small. Although most of currently AEG systems are limited to assess only English texts, a movement to take into account more languages is being started. Fragile research attempts have been conducted on Arabic essays. In future studies, we hope having a sufficient dataset to build a system capable of assessing Arabic writing.

References

[1] Page, E. B, ”The Imminence of Grading Essays by Computer”. Phi Delta Kappan, 48:238-243, 1966.

[2] Burstein, J., Kukich, K., Wolff, S., Lu, C., Chodorow, M. Braden-Harder, L., and Harris M. D., ”AutomatedScoring Using A Hybrid Feature Identification Technique”.Proceedings of 36th Annual Meeting of the Association for Computational Linguistics, 206-210. Montreal, Canada, 1998.

[3] Foltz, P. W., Kintsch, W., and Landauer, T. K. ”Analysis of Text Coherence Using Latent Semantic Analysis”, Discourse Processes 25(2-3):285-307, 1998.

[4] Larkey, L., ”Automatic Essay Grading Using Text Categorization Techniques”, Proceedings of the 21st ACMSIGIR Conference on Research and Development in Information Retrieval, 90-95. Melbourne, Australia, 1998.

[5] Elliott, S., ”Intellimetric: From Here to Validity”, In Shermis, M., and Burstein, J. eds. ”Automated essay scoring: A cross-disciplinary perspective”. Hillsdale, NJ: Lawrence Erlbaum Associates, 2003.

[6] Shermis, M., and Burstein, J. eds., ” Automated EssayScoring: A Cross-Disciplinary Perspective”, Hillsdale, NJ:Lawrence Erlbaum Associates, 2003.

[7] MacDonald, N. H., Frase, L. T., Gingrich P. S., and Keenan, S.A., ”The Writers Workbench: Computer Aids for Text Analysis”, IEEE Transactions on Communications 30(1):105-110, 1982.

[8] Porter, M. F., ”An algorithm for suffix stripping”, Program, 14(3), 1980

[9] Kohonen, T., ”Self-Organization and Associative Memory, 2nd Edition”, Berlin: Springer-Verlag, 1995

[10] Elliot, S. M. , ”IntelliMetric: From here to validity”, Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA, april.2010.

[11] Burstein, J. and Wolska, M., ”Toward evaluation of writing style: Overly repetitious word use in student writing”, In Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics. Budapest, Hungary, 2013.

[12] Bridgeman, B., Trapani, C., and Attali, Y., Comparison of human and machine scoring of essays: Differences by gender, ethnicity, and country. Applied Measurement in Education, 25, 2740.2012

[13] Ratnaparkhi, A., ”A Maximum Entropy Part-of- Speech Tagger”, In Proceedings of the Empirical Methods in Natural Language Processing Conference, University of Pennsylvania, 1996.

[14] Burstein, J., Marcu, D., and Knight, K., ”Finding the WRITE Stuff: Automatic Identification of Discourse Structure in Student Essays”, IEEE Intelligent Systems: Special Issue on Natural Language Processing 18(1), pp. 32-39, 2003.

[15] Mahmud K., Ahmed H., Aziza A., and Mohamed I., UML analysis for Quality Assurance management System for Higher education, International Journal of Engineering Science and Technology, 2(4), 417-432, 2010

[16] Ahmed H., and Mohamed I., Designing quality e-learning environments for higher education, Educational Research, 1(6):186-197, 2010

[17] Shermis, M. D. and Burstein, J. ,”Automated Essay Scoring: A cross disciplinary perspective”., Mahwah, NJ: Lawrence Erlbaum Associates, 2003.

[18] Page, E. B. Project Essay Grade: PEG. In M. D. Shermis and J. Burstein (Eds.), ”Automated essay scoring: A cross-disciplinary perspective”, (pp. 4354). Mahwah, NJ: Lawrence Erlbaum Associates, 2003.

[19] Mason, O. and Grove-Stephenson, I. , ” Automated free text marking with paperless school” in ’Proceedings of the 6th International Computer Assisted Assessment Conference, 2002.

[20] Rudner, L. and Gagne, P., ”An overview of three approaches to scoring written essays by computer”, ERIC Digest number ED 458 290, 2001.

[21] Robert W., ”Automated essay grading: An evaluation of four conceptual models”, School of Information Systems, Curtin University of Technology, Retrieved August, 26, 2008

[22] Christie, J. R., ”Automated essay marking-for both style and content. In M. Danson (Ed.), Proceedings of the Third Annual Computer Assisted Assessment Conference, Loughborough University, UK, 1999.

[23] Mitchell, T., Russel, T., Broomhead, P., and Aldridge N. , ”Towards robust computerized marking of free-text responses”, In M. Danson (Ed.), Proceedings of the Sixth International Computer Assisted Assessment Conference, Loughboroug University, Loughborouh, UK., 2002.

[24] Hazem M, Alla R., Aziza A., Mohamed I, Ahmed H., Mahmod K., and Nikos M., Design and Implementation of Total Quality Assurance Management System for Universities, International Conference of recent advances in business administration, UK, 2010

[25] Vantage Learning., ”A study of expert scoring and IntelliMetric scoring accuracy for dimensional scoring of Grade 11 student writing responses” Newtown, PA: Vantage Learning, 2000.

[26] Rudner, L., Garcia, V., and Welch, C., ”An Evaluation of Intellimetric Essay Scoring System Using Responses to GMAT AWA Prompts”, GMAC Research report number RR-05-08, 2005.

[27] T. Ishioka and M. Kameda., ” Automated Japanese Essay Scoring System: Jess”, in Proceedings of the 15th International Workshop on Database and Expert Systems, pages 48,2004.

[28] Ahmed Hassan and Mohamed Ibrahim, ”Designing quality e-learning environments for higher education”, Educational Research, 1:6 , pp. 186-197, 2010

[29] Mahmoud K., Ahmed H., Aziza A., and Mohamed I., Prototype of Web2based system for Quality Assurance Evaluation Process in Higher education Institutions, International Journal of Electrical & Computer Sciences, 10 (2), 2010