SIMPLE-O, the Essay Grading System for Indonesian Language Using LSA Method with Multi-Level Keywords

Anak Agung Putri Ratna,Universitas Indonesia, Indonesia
Prima Dewi Purnamasari, Universitas Indonesia, Indonesia
Boma Anantasatya Adhi, Universitas Indonesia, Indonesia

The Asian Conference on Society, Education & Technology 2015
Official Conference Proceedings

Abstract
Evaluation is a very important tool in measuring students' level of understanding, whether in e-learning or conventional studying system. Essay is one of the evaluation which to determine students ability where choices are not provided. Students have to answer by sentences, so then it could be various based on their opinion, since it reflects the student's best thoughts of the materials. Today, automatic grading essay is an its development, the system is more efficient and more effective in essay grading especially in maintaining fairness. SIMPLE-O, web based automated essay grading system, has been developed in Electrical Engineering Department, University of Indonesia, using Latent Semantic Analysis (LSA) method for Indonesian Language. This research was conducted by providing modifications with using multi-level key words to increase accuracy when compared with human raters. The results obtained in small classes, agreement with human raters are above 86%

Keywords: latent semantic analysis; essay grading;

Introduction

The prevalence of e-learning in Indonesia especially in Universitas Indonesia has promote a rapid adoption of e-learning system throughout the University (Ratna, 2006). One aspect of e-learning which is often forgotten is evaluation (Valenti, Neri, & Cucchiarelli, 2003). Essay, while being one of the most versatile evaluation method, is rarely used in e-learning system due to the fact that automated grading is not commonly available (Ratna, 2006). Automated essay grading is very important to maintain fairness and efficiency (Lawrence Rudner & Gagne, 2001). Since 2007 has developed the system which Simple-O is web based automated essay grading system developed specifically for Bahasa Indonesia with its unique properties in mind, in contrast with most available solution which is developed for English speaking test takers (Ratna, 2006).

Assessment is very important component in every education system to evaluate student understanding of the whole study materials. There are two main forms of assessment questions: objective form and essay form. Objective type of question is multiple choices with several answers available to choose. Essay is one form of evaluation where the option of answer is not provided, and student must answer in sentence, thus the answers may vary depending on each student's thought. Essay grading also remains as an option for teacher to evaluate student's ability, even though it is not easy to give an objective judgment to every student in essay grading. Essay considered as a powerful tool to achieve studying result by many researcher, also to evaluate the thinking ability in high level such as in synthesis and analysis (Lm Rudner & Liang, 2002).

In e-learning system, an examination can be conducted online, from answering to grading the exam. It gives benefit to the teacher and the student because the on-line examination process becomes more effective & efficient. Another benefit is that evaluation system with the help of computer gives faster and more accurate grading results (Landauer, Foltz, & Laham, 1998; Lawrence Rudner & Gagne, 2001). Moreover, this system can also handle classes with great amount of students.

Online assessment has developed and implemented in Electrical Engineering Department in Universitas Indonesia for quite a long time (Ratna, Hartanto, Ekadiyanto, & Narita, 2002). However automated essay grading in Indonesian language was not available yet.

Bahasa Indonesia

Bahasa Indonesia has a simple structure compared to English and does not require changes in plural, gender, or time. There is no difference in plural and singular objects in Bahasa Indonesia and it only needs to put additional word to identify the plurality. For example: child = *anak,* children = *banyak anak*. The word "*banyak"* identify plural and it is actually can be translated into the word "many". The noun itself, "*anak*" does not change at all. In different time tenses, the verb does not change and it only need to characterize the time by using time marker word, for example yesterday = *kemarin*, tomorrow = *besok*, and next month = *bulan depan.*

Bahasa Indonesia is a national language in Indonesia. This is the language to unite the nation and become a middle way in communication. Indonesia is a country that has thousands of islands and every island have their own characteristics in culture as well

as language. Bahasa Indonesia can be divided into several major categories. The main categories are standard language and daily language. And just like other language, there are writing language and speaking language.

In each islands, there are many small cities. Beside using and learning Bahasa Indonesia, the villagers also have their own mother language. Since elementary school, every student is taught Bahasa Indonesia as one of subject. However, this language is only used in formal situation, while for daily activities villagers usually use their own local language. In several provinces, although they have the same local language, their dialects are sometimes different. From the dialect, people can be distinguished from which city they are coming from.

In some local language, there are subtle and harsh language. When talking about the harsh language and subtle language, we cannot separate them from the history. For example in Java language, long time ago, subtle language used by the royal and his descendants. The intellectuals, aristocrats and local leaders also use subtle language. While the harsh language used by ordinary people. The differences in the language that used by the class differences eventually make a difference in the local language of the particular area. For example, the language that used in Solo or Yogyakarta. In those cities, there were kingdoms that used subtle language in their daily live. Because their ancients use subtle language, people in those cities also used it until now. For some local language, there are also class of language which differentiate the choice of words to use, e.g. when talked to respected people (for example to teacher or parents), to people in the same position (for example friends and siblings), and to people in lower position (for example to a house cleaner).

Aside from the local language, the language commonly used in daily conversation is slang. Slang is the language used by young people to communicate with each other informally. Slang words or styles are derived from many sources, e.g. from foreign, local or a twist from the original standard Bahasa Indonesia. Example for slang that originates from the standard language are *sudah* becomes *udah, hujan* becomes *ujan*, and *pahit* becomes *pait*. Not like standard Bahasa Indonesia, slang words and styles are easily changed over generations. It is unexpected to use slang in formal situation. However, youngsters are often forgotten or unconsciously use it in the classroom.

The challenge for essay grading system for Bahasa Indonesia may not comes from the language structure, because as already mentioned before, structure in Bahasa Indonesia is simple. The challenges are mostly coming from the students' culture and evolution of the language that comes from local, foreign, dialect, style and slang words. The pressure of exam itself gives more challenge, because it often contributes to the non-standard style language used by the students, for example the object and subject becomes interchanged, or the active and passive verbs becomes wrongly used.

Universitas Indonesia

Universitas Indonesia (abbreviated as UI) is one of the top universities in Indonesia, located in Depok, a city belongs to West Java Province but located in the suburbs of Jakarta, the capital city of Indonesia. UI has 14 faculties, 1 vocational school, and postgraduate school. Each faculty has many departments. UI students come from all over Indonesia. About half of them comes from neighboring cities of Depok and

Jakarta, and the rest migrates from their local residence from all over islands and provinces in Indonesia.

The non-local students who migrates from their city/village will normally assemble and form associations that consists of other students who came from the same area. The association is an excellent platform in the process of adaptation in their new environment and their new status as a university student. Many of the seniors who came from the same city helped the junior as a new student in adapting the environment; from finding a place to stay, where to eat or shopping with low budget, until the lectures problem. It certainly makes bonds held around becomes very strong. In communicating within the community, the students from local area often use their local language support. Although they often use their local languages, they still use Indonesian in communicating to other students or faculty.

Online academic system in Universitas Indonesia

In UI, there is a web based online system which serves the whole academic information for students, teachers and administrative, namely SIAK-NG. In this web application system, students are required to fill their academic plans independently for each semester. The teachers upload students grades during and at the end of the semesters. Administrative and teacher can monitor the academic status and achievement of the students. By using this system, students and academic supervisor can also interact and communicate with each other during the academic registration period every beginning of a semester.

In addition to SIAK-NG, there is also an e-learning web application that is used by students and lecturers in the teaching and learning process, namely SCELE. This web application serves as a media for the faculty to share the lecture material, conduct regular tests, or exams and the place for students to upload assignments, and to discuss materials in a forum provided by the teacher. SCELE provided several type of assessment for example multiple choices, true/false, short answer, and essay. Not like the other exam, essay can not be graded automatically by the system. This is unfortunate because for teacher, essay is often the most preferable type of exam. This become the background of our work, to automate the grading of essay.

SIMPLE-O (SIsteM PeniLaian Esai – Otomatis) is a web based automated essay grading system developed specifically for Bahasa Indonesia with its unique properties in contrast with most available solution which is developed for English speakers (Ratna, 2006). SIMPLE-O uses Latent Semantic Analysis (LSA) method (Landauer et al., 1998) to evaluate the students' essay exam answers. LSA compare a text with the words chosen as a reference. LSA represents the words in a text in a semantics matrix, which is Term-Document matrix (TD Matrix). LSA use matrix algebra technique known as Singular Value Decomposition (SVD) to get information from the text. Vector analysis is conducted in reduced space to retrieve similarity between text.

Latent Semantic Analysis (LSA)

LSA is a technique to extract and represent sentences with mathematical or statistical calculation. The strength of the LSA technique lies in the syntax structure insensitive, thus the words processed are words from a bag of words ignoring the sequence of the sentences. LSA express the ideas about the meaning of the word, where words are

occupying a position in semantic space and the meaning is the relationship between one sentences to another [3].

Assumptions that underlie LSA are that the similarities and differences in the meaning of words can be influenced by the similarities and differences in the overall context in which the word is there or not. Conversely, in the meaning of the sentence can be verbal outline of the combination (in the form of mathematics) from the words within it. This assumption implies that usually the dominant verbal meaning is based on the selection and combination of words in the speech. And for various purposes, the order in a text can be ignored in the sense to estimate similarity with only slight loss of accuracy. Thus the assessment of any text with the LSA more emphasis on the words contained in any text without notice a linguistic characteristics, such as grammar, how to write, and order of words in a sentence, therefore a sentence does not require a good rhetoric.

To apply to the basic assumption in computing systems will require a model where the word is a representation of mathematics function as a set of the observed linguistic context, and representation of linguistic context is a mathematical function of the words that are in it. In LSA, the linier function between word and meaning of text and linier factorization are used to construct text into the form of high dimension vectors. In LSA, the text is a combination of the vectors containing the words, and words are the meaning of the vectors from the observed text. Computation form that is used in the LSA is an algebra method of Singular Value Decomposition (SVD) matrix, continued with the dimensional reduction.

Singular Value Decomposition

Singular Value Decomposition (SVD) is a mathematical matrix decomposition technique. Matrix formed from whether there is or there isn't a specific word appears in a text (usually already defined as keywords). This matrix by Singular Value Decomposition divided into 3 (three) matrices [3]. In reconstructed matrix from decomposed matrix using SVD will be seen strong correlation between topics or sentences joined in a specific group.

After the 3 matrices are obtained, the next process is reducing matrix dimension by reducing the second matrix dimension, a diagonal matrix. Reducing the dimension of a diagonal matrix is performed by setting all diagonal values of the second matrix into zero (0) except for certain chosen dimension. And if the three matrices components are multiplied, it will produce other reconstruction matrix for desired correlation value purpose.

Mathematically, a matrix can be well decomposed if it has small factor value compared to smallest dimension of the original matrix. Thus, the best matrix reconstruction will be obtained when the factor value is smaller than the sum of factor used.

Modification for LSA in SIMPLE-O

The original LSA was made for long essay or document and it is tested for English. In our research, we modify LSA so that it can be used for analyzing and grading short essay as in question-answer essay that is often preferable in exam. And also we tried to modify LSA so that it can be useful for analyzing answers in Bahasa Indonesia.

For the first modification, instead of using a large corpus for building term document matrix, multiple matrices with similar dimension is used, so that it has lower calculation overhead. Secondly, key term is used instead of including all term in TD matrices. By doing this approach, the sensitivity for short essay is increased. Furthermore, we also use multilevel key term to increase accuracy. The complete proposed algorithm is as follows:

1. A lecturer may create a question and define the "golden answer" which is then submitted to the system.
2. The answer then is preprocessed and a statistical information for each term is gathered
3. Preprocessing includes punctuation marks removal, stop word removal, similarity check, etc.
4. Each unique term occurrence is counted and has its location noted, then this information is stored in database.
5. Key-terms are any terms appear in the "golden answer" which are considered important by the lecturer.
6. A key-term may not be a member of stop words.
7. Most important key-terms are member of ordinary key-terms which are very important.
8. The "golden answer" vector, which is the output of the SVD process, is stored in the database. This pre-calculated vector is then compared with all student answer's vector.
9. Each student score is then stored in the database for future use.

There are 3 main entity of the systems, which are user, teacher and the system itself. Therefore to see the interaction between each entity can be seen through activity diagram in Fig 1.

In the current implementation, the algorithm is entirely implemented in PHP. This is to ensure portability of the program. The SVD calculation is done in Java Matrix port on PHP. The ADOdb is used as a database abstraction layer which further improves the portability. MySQL is the default database choice. The current implementation will run on PHP version 5.0 or higher, Apache version 2.0 or higher and any database server supported by ADOdb 5. For the Graphical User Interface (GUI) we design the system based on Bootstrap framework. The example of SIMPLE-O GUI can be seen in Fig. 2.

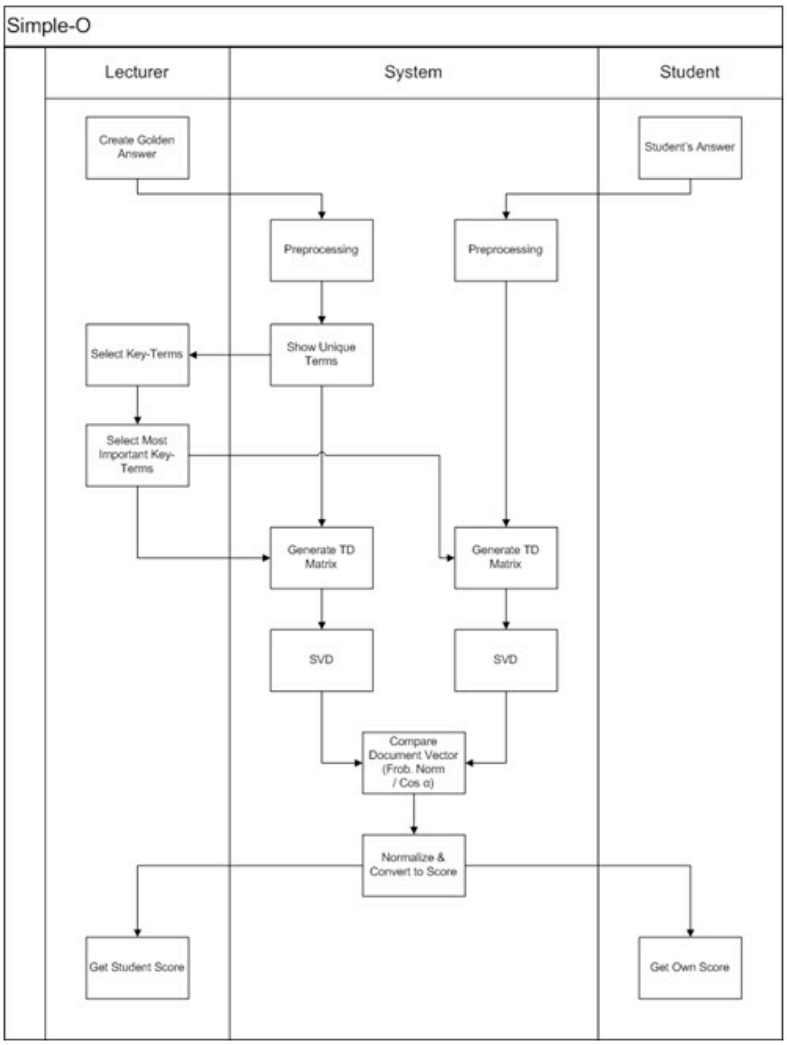

Fig. 1. Activity diagram for SIMPLE-O

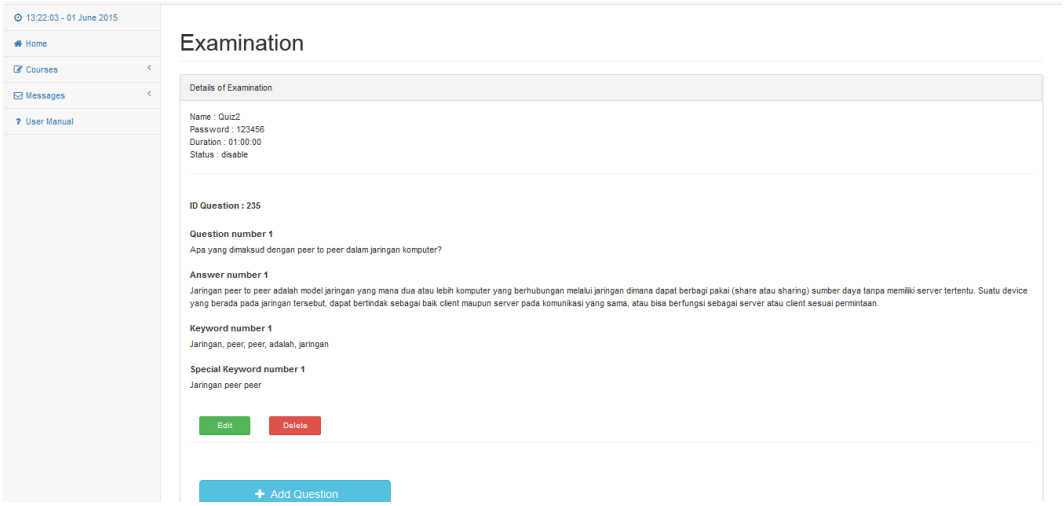

Fig. 2. Example of SIMPLE-O graphic interface for the teacher when editing question and answer

Experiment and Results

In order to measure the performance of the system developed, we conduct 3 experiments and testing, which are:

1. Human-rater agreement. In this experiment, we measures the average correlation between system score and human raters score. This can also be seen as the accuracy of the system, which is a quantitative assessment of the system.
2. Processing time. In this experiment, we measure the average time needed by the system required to grade a student answer. This is also a quantitative assessment.
3. User Experience. In this experiment, we measure the satisfaction of the system user. This experiment is a qualitative assessment for the system.

All experiments were conducted in a computer lab as an online exam scenario with 40 student as participant and 3 lecturer who gives scores as a benchmark scores.

As a result for the first experiment, Fig. 3 shows that for varied number of keywords, 5, 6, 7 and 8, the agreement with human rates are stable, about 88.13 – 86%. The usage of multi level keywords is proven to succeed in increasing the accuracy into 93.94 – 95.88%.

As the result for the second experiment, the number of words in the answer provided by te students are contributing to the longer processing time by the system.

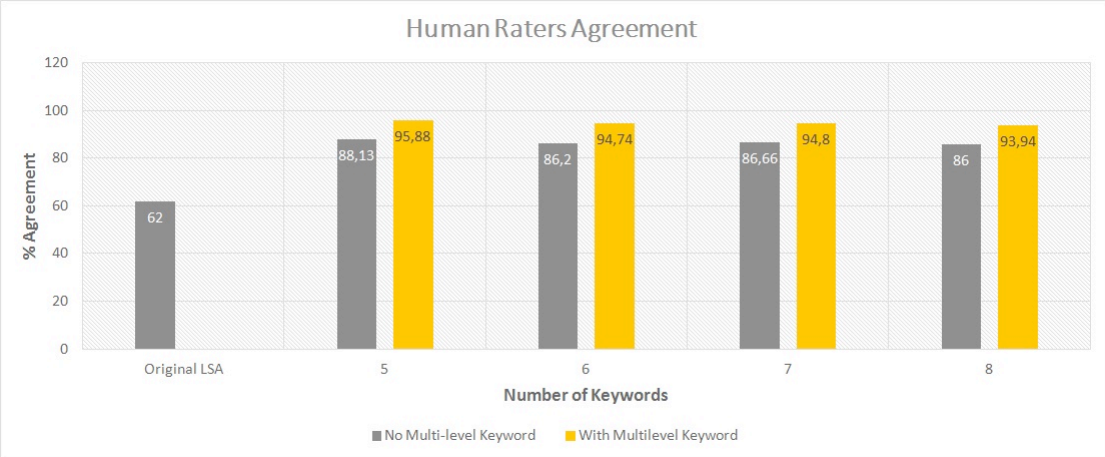

Fig. 3. The effect of number of keywords to human raters agreement

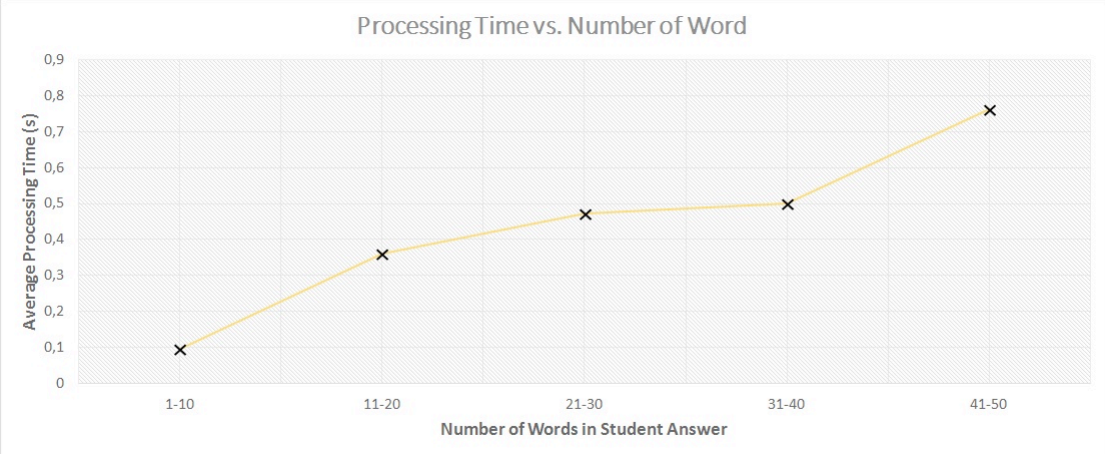

Fig. 4. The effect of number of words to processing time

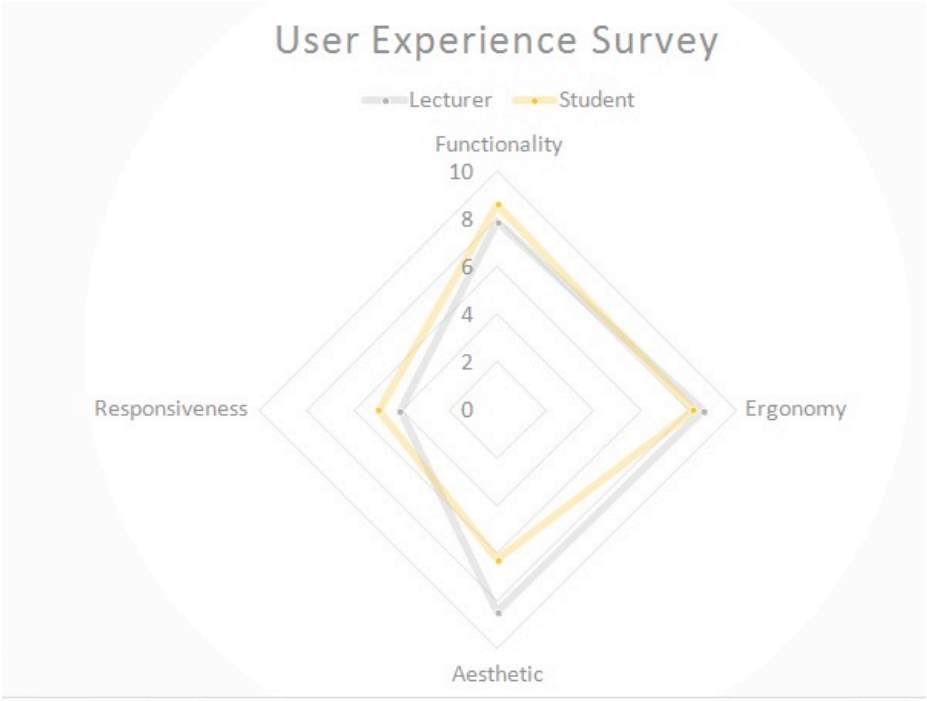

Fig 5. SIMPLE-O user experience survey

In the last experiment, we measure the user acceptance and satisfaction of this system. By looking at Fig. 5, it can be seen that For the functionality and ergonomy, this system is considered good. However in responsiveness and aesthetic the system should be improved.

Conclusion

This work in progress has demonstrated the possibility to build an automated essay grader which is optimized for Bahasa Indonesia. More than 86% human raters' agreement can be achieved with Simple-O. We still have some future work, which is to speed up computation process by streamlining the algorithm, improve human raters agreement, and explore the system behavior on many different kind of test.

References

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2-3), 259–284. http://doi.org/10.1080/01638539809545028

Ratna, A. A. P. (2006). *Latent Semantic Analysis Algorithm Enhancement on Automatic Essay Grading*. Universitas Indonesia.

Ratna, A. A. P., Hartanto, D., Ekadiyanto, A., & Narita, S. (2002). Daskom On-Line: Implementation Distance Learning On Basic of Computer Course. In *Web Net Proceedings* (pp. 1600–1602).

Rudner, L., & Gagne, P. (2001). An overview of three approaches to scoring written essays by computer. *Practical Assessment, Research & Evaluation*, *7*(26), 3–7. Retrieved from http://pareonline.net/getvn.asp?V=7&n=26

Rudner, L., & Liang, T. (2002). Automated essay scoring using Bayes' theorem. *The Journal of Technology, Learning and …*, *1*(2). Retrieved from http://napoleon.bc.edu/ojs/index.php/jtla/article/view/1668

Valenti, S., Neri, F., & Cucchiarelli, A. (2003). An Overview of Current Research on Automated Essay Grading. *Journal of Information Technology Education*, *2*, 3–118. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.101.5757&rep=rep1&type=pdf