**Introduction**

Automated essay scoring (AES) is defined as the computer technology that evaluates and scores the written prose. Nowadays it is becoming more and more popular for a number of reasons. First of all, AES simplifies teacher’s work. If we take into account the percentage of students per teacher, essay scoring may take a long time. And this process can be accelerated through implementation of AES systems, which are capable of identifying and correcting most of the obvious mistakes and also giving a feedback about the complexity of the structure, grammar, and vocabulary of an essay. AES systems can be very useful because they can provide the student with a score as well as feedback within seconds. Moreover, AES may help to make up for a prejudiced attitude on the part of a teacher. In order to get an objective grade, an essay is usually checked by two teachers, and sometimes even by three, while an AES system is able to take on their duties, saving human resource and employers’ money. Well-designed AES programs can also be useful for language learners who can not have a teacher checking their preparational essays for any reason. An automated grading system which analyses the text and suggests its potential improvements is a viable option in this case.

Nevertheless, AES systems also have a number of drawbacks that may be accounted for by the novelty of this subject. First of all, computers can not assess an essay as human raters do because the computer would do “what it is programmed to do” and it wouldn’t “appreciate” an essay.Another criticism is the construct objections. That is, the computer counts variables that might not be “truly” important in essay grading, i.e., focusing on formal aspects rather than organizational ones.

**Part 1. Research background**

A number of various AES programs have been developed and utilized for various purposes. We reviewed some of the most significant and widely used systems as part of our project, focusing on their engines and distinctive features.

Arguably the most important of all the automated rating systems is the Project Essay Grader, or PEG™, which is the first AES program ever created. PEG™ analyses a number of distinct external properties such as word length, punctuation, word rarity and so on. Each of these text features is evaluated and turned into variables referenced as proxes. There is no in-depth information about the methods of the proxes calculation, though, because the source code of this commercial product is closed. These proxes are then compared to reference values which are obtained from a number of manually graded essays of the same type, which are rated by experts beforehand. PEG™’s focus is directed towards the assessment of general writing ability.

The success and the general acception of PEG™ approach proves that morphological analysis is applicable for essays grading. Furthermore, our idea of counting verb forms amount and comparing it with reference values is also inspired by the PEG™ experience. Not only it is one of the central contexts of our project, but it is also reported that the vast majority of AES systems compare the variables obtained by analyzing the aspects of inputted text with reference. There are also other pieces used for essays rating, such as Intelligent Essay Assessor (IEA), IntelliMetric™, E-rater et cetera. However, these projects are also commercial and their code is also closed. Since they are based on semantics and discourse analysis, which are the methods we were not aiming to implement, they were of minor interest for us within this project.

**Part 2. Research**

The aim of our research is to determine some criteria for automated essay scoring which could be used for automatic essays evaluation to help students preparing for IELTS writing test identify their weaknesses by giving them feedbacks. These functions are then to be implemented in REALEC Inspector, the program created for the purposes of specific learner corpus - REALEC.

Our hypothesis was that the mark of an essay correlates with the amount of  the following forms of the verbs used in it: gerund, to+infinitive, and the Past Participle form. In order to test the hypothesis, we used students’ essays of NRU HSE which were divided into two groups: best and worst rated, based on the grades they were given by teachers. For the research we took 40 best rated and 40 worst rated essays, with the average marks for both groups being 7.7 and 4, accordingly. Firstly, we manually looked through all the essays in both groups in which we marked all the aforementioned verbs’ forms. Then we created an algorithm which purpose was to automatically highlight and count the selected verbs forms in inputted text. Our algorithm first tags all the text using Stanford POS tagger, then certain rules are applied to determine the forms of our interest within the essay. By continuous comparing the algorithm processing results with manually obtained data, we improved the automatic outputs to be as close to ours as possible. In terms of rules, we ended up with making the program to highlight the following patterns:

1. to followed by VB to select infinitives;
2. VBG, which is not preceded by any form of to be to distinguish gerunds;
3. VBN not preceded by any form of been to determine past participle forms.

The overall result shows that in the best essays the mean amount of these three forms is 12.7, while in the worst essays it is 3.06. According to the resulting statistics, we can say that the chosen forms could be used as an assessing indicator for the complexity of essays submitted to REALEC Inspector.

**Conclusion**

Currently,  REALEC Inspector is still under development. In the future we plan to implement the criteria found within this project into the list features that are already included for the examination of essays, so that HSE students and other users could have an opportunity to assess their essays quickly and identify their weak points  with the feedback from the REALEC Inspector.

One of our further goals is to keep improving our algorithm. Given that every automatic tagger has a margin of error, we have already coped with major issues (e. g. by applying *“to”* and *“to be”* conditions). Subsequently, as lesser problems unfold, we are planning to introduce more conditions in order to get better recognition of forms.

We also intend to get a more in-depth look into how parts of speech present in the text relate with the essay grade. One of the possible strategies is to further research representation of the selected groups, for instance, by dividing the past participle group into passives, perfect forms and participles, and analyzing their effects. In order to achieve this, we plan to include further criteria into our algorithm.