

IT.ENG.5

## PLAGIARISM DETECTION ON ELECTRONIC SUBMISSIONS OF TEXT BASED ASSIGNMENTS

**M. A. C. Jiffriya<sup>1</sup>, M. A. C. Akmal Jahan<sup>1</sup>, Roshan G. Ragle<sup>2</sup>,  
Sampath Deegalla<sup>2</sup>**

<sup>1</sup>*Post Graduate Institute of Science, University of Peradeniya*

<sup>2</sup>*Department of Computer Engineering, Faculty of Engineering,  
University of Peradeniya*

With the advancement in technology that enabled preparation and submission of assignments electronically, plagiarism is one of the growing issues in the academic field. Plagiarism is known as illegal use of others' part of or the whole work as one's own work. Plagiarism on text based assignment is a current concern in Sri Lankan University Systems. Due to plagiarised submissions of assignments, academic staff face difficulties in marking students' assignments with higher degree of judgment.

Our research focuses on creating a simple and effective tool for plagiarism detection of text based digital assignments to detect and therefore to minimize plagiarism in Universities. Our plagiarism detection tool named *AntiPlag* was developed using a simple tri-gram sequence matching technique. A set of text based assignments were tested by *AntiPlag* on GNU/Linux based environment and the results were compared against an exiting commercial plagiarism detection tool known as *Plagiarism Checker X*.

*AntiPlag* showed better results (with lower false positives) in proper plagiarism detection than *Plagiarism Checker X* due to effective pre-processing steps performed in *AntiPlag*. In addition, for the dataset tested *AntiPlag* was three times faster than *Plagiarism Checker X*. To improve on the detection latency of *AntiPlag* further, a data clustering approach was applied on assignments to create appropriate clusters. The clustering technique applied will group similar assignments together into a single cluster and dissimilar assignments into different clusters. Then the plagiarism detection was applied only on clusters making the detection latency even lower. The clustering approach improved the execution time of *AntiPlag* by another twenty times making *AntiPlag* 60x faster than *Plagiarism Checker X* for the assignments we considered.

*AntiPlag*, a simple and effective plagiarism detector proposed here, could be used to isolate plagiarized text based assignments from non-plagiarised assignments easily. The detection process of *AntiPlag* was optimized and enhanced through the clustering approach. *AntiPlag* is fast and capable of comparing all plagiarised pairs of assignments automatically at once. We have proved that the tool is simple, small in size, user friendly and effective. Therefore, *AntiPlag* is a simple and effective tool for plagiarism detection on text based electronic assignments.