

Abstract

Plagiarism is one of the illegal acts that harm others because it means stealing someone's copyright. Plagiarism occurs due to the lack of ability to innovate and create an original work. Plagiarism usually happens in education field especially in the college or school. Most frequent cases of plagiarism is copying other's report writing, assignments, paper or student thesis. It would be difficult to manually measure the similarity of documents by simply reading the documents and comparing each other, plus if the documents that would be checked is more than two. Therefore, an automatic system to detect plagiarism (text reuse detection) is needed. In this final project, an automatic system to detect plagiarism has been created using Text Alignment by parsing N-gram for each input. Then, the system would find the string using Rabin Karp Algorithm. Building Text Alignment would need 3 phases which are Seeding, Extension and Filtering. This method yields 82,20% Micro Average F1-Score for case evaluation level, and 49,64% Micro Average F1-Score for character evaluation level. This system is ranked 17th and 16th for Precision and Recall in the Overview of 5th International Competition on Plagiarism Detection PAN 2013.

Keywords : text alignment, plagiarism detection, text reuse detection, similarity, seeding, extension, filtering.