

CHAPTER 6

CONCLUSION

6.1 Conclusion

From the testing with 3 section of test and analysing of winnowing algorithm to detect plagiarism in text document, so taken some following conclusion :

1. More higher value of window then the accuration percentage similarity of text is decrease.
2. More smaller value of kgram then the accuration percentage similarity of text is increase.
3. Using value kgram 1 will giving high result of similarity but is uneffective because the text will split into only alphabet.
4. In range kgram 1 till 10, using value kgram 1,2 and 3 is give more higher result similarity of text.
5. Value of kgram and window on winnowing algorithm are very affect with the percentage similarity of text.
6. Basic prime number in rolling hash process on winnowing algorithm is affect with the percentage similarity of text
7. Based from testing of 3 similarity method, the highest mean percentage similarity result is using Sorensen Dice Similarity Coefficient and the lowest mean percentage similarity is using Andberg Similarity Coefficient.
8. If sorted the 3 similarity method by number, here is the result :
 - 1.) Sorensen Dice Similarity Coefficient
 - 2.) Jaccard Similarity Coefficient
 - 3.) Andberg Similarity Coefficient
9. The difference of result percentage similarity in 3 similarity method is caused of the different formula in each similarity method on calculating and choose document fingerprint.

6.2 Further Research

There is still lot of method could be implemented and combine with winnowing algorithm in further research. Such stemming and tokenization to increase quality of the text. So the winnowing algorithm can more maximize and accurate to detect plagiarism in text document.

