

CHAPTER 1

INTRODUCTION

1.1 Background

Cases of plagiarism in recent years has been an issues. The level of plagiarism was increased year by year (news.okezone.com-2014). Document and text was copied without any references. However it's happen because effect of some factor. One of the factor is copying the text without giving the source of the text as written in Journal Socio-cultural Semarang University with title "Kesadaran dan Perilaku Plagiarisme di Kalangan Mahasiswa". Not just people who do a plagiarism that should be blame. One of the reason is because easy access and free access works of other people in internet, as written in Journal Faculty Information Technology YARSI University with title "Analisis dan Pencegahan Plagiarisme di Kalangan Mahasiswa". Especially in this time to do control act of plagiarism is way really difficult. Because of that issues, some various method and algorithm was make to solve plagiarism such rabin-karp algorithm, manber algorithm and winnowing algorithm. Those algorithm is giving a value of percentage of level similarity in document text with an different ways. The problem is the accuration of those some algorithm to check a plagiarism in a text.

In this final study project, will analysing the accuration one of plagiarism algorithm, that is the winnowing algorithm with implementing 3 different method similarity. Winnowing algorithm is plagiarism detection algorithm based on fingerprinting. This algorithm are more purpose to a document text such an article, journal, and text file. Basic step of winnowing algorithm was input value of k-gram and input value of window to find value of fingerprinting and calculate the similarity fingerprinting of document text with the jaccard similarity coefficient, sorensen dice similarity coefficient, and andberg similarity coefficient.

From that step of winnowing algorithm, percentage of plagiarism is recorded and next step is analyse the results plagiarism in each similarity method. That

analyse was based on the testing of different value of k-gram, different value of window and compared the results each method. From this analyse, expecting an some new conclusion of winnowing algorithm.

1.2 Scope

Based a background above, there is arranged a scope of this research :

1. The different percentage similarity of winnowing algorithm with a different value of k-gram and window with range 1 - 10 ?
2. Effect value of k-gram and window in winnowing algorithm
3. Analysing three similarity method which is jaccard similarity coefficient, sorensen dice similarity coefficient, andberg similarity coefficient
4. The different percentage each similarity method to detect plagiarism in text ?

Also constraint for this research is :

1. Text format is (.txt) and input a text manually
2. Using a winnowing algorithm
3. Calculate the similarity fingerprint with jaccard similarity coefficient, sorensen dice similarity coefficient, andberg similarity coefficient
4. Text sample that used in this research is history computer which is abbreviated in 3 paragraph (wikipedia/Sejarah_komputer)
5. Testing text that used in this research is one of the cases plagiarism in Indonesia as written in news (kabar24.bisnis.com/diduga-plagiat-ini-perbandingan-artikel-anggito-abimanyu-hotbonar-sinaga,2014).

1.3 Objective

The goal of this research is to analyse percentage of plagiarism in text with different value of k-gram and window with 3 different similarity method. And giving conclusion based on analysing of winnowing algorithm .