

ABSTRAK

Skripsi merupakan tugas akhir yang sangat penting bagi seorang mahasiswa, karena skripsi sebagai salah satu syarat lulus mahasiswa di perguruan tinggi. Mahasiswa sering mengambil jalan praktis dalam pengerjaan skripsi untuk mencari kemudahan. Dalam pengerjaan skripsi sering kali mahasiswa mengutip atau menggunakan beberapa atau keseluruhan kata – kata, kalimat, gagasan, ide, dan bagian – bagian lain hasil skripsi orang lain tanpa mencantumkan secara jelas sumber kutipan yang didapat. Hal ini sangat dilarang keras oleh setiap universitas , sebab mahasiswa yang melakukan tindakan plagiarisme tidak mencerminkan sikap kreatif sebagai kalangan terpelajar. Maka dari itu skripsi ini bertujuan membuat suatu sistem untuk mendeteksi tindakan plagiarisme dokumen skripsi yang dilakukan oleh mahasiswa Universitas Sanata Dharma.

Proses untuk melakukan pendeteksian plagiarisme dokumen skripsi diawali dengan melakukan proses *preprocessing*, yaitu tokenisasi, *stopword removal*, *stemming*, dan *splitter* kalimat. Kemudian dengan menggunakan metode *overlap measure function* yang akan menghasilkan nilai *similarity* (nilai bobot) dari setiap dokumen *query* yang dibandingkan dengan dokumen asli, nilai *similarity* dokumen tersebut yang akan menentukan deteksi plagiarisme yang disesuaikan dengan nilai *threshold* plagiarisme.

Skripsi ini menggunakan data 50 dokumen skripsi mahasiswa Teknik Informatika untuk pengujian deteksi plagiarisme, sebelum dideteksi terlebih dahulu harus melakukan pencarian nilai *threshold*, pencarian nilai *threshold* plagiarisme menggunakan persentase 50, 40 , dan 30 dari rata – rata jumlah *similarity* dari dokumen – dokumen yang benar – benar sama(dokumen dengan dirinya sendiri). Setelah dilakukan pengujian persentase nilai *threshold* ,persentase yang paling baik adalah 50 dibandingkan dengan 30 dan 40. Persentase 50 memiliki nilai *precision* 94,8%, sedangkan persentase 40 dengan *precision* 92,4% dan persentase 30% dengan *precision* 70%. Maka dari itu untuk mendeteksi plagiarisme pada sistem ini digunakan nilai *threshold* dengan persentase 50% yaitu 869,3 dan sistem ini berlaku untuk mendeteksi dokumen dengan panjang kalimat 100 sampai 500 kalimat dan panjang kata 5 sampai dengan 50 kata.

ABSTRACT

Thesis is the final task which is very important for a college students, because of thesis is one of conditions for graduating college students. College Students often take a practical way to make things easy in doing their thesis. Sometimes they using all of the words, phrases, ideas, and the rest of the thesis of others without citation shall state clearly the source derived. It is strictly prohibited by any university, for students who commit acts of plagiarism does not reflect their creative attitude as the educated. This thesis aims to create a system to detect plagiarism action thesis documents carried by students at the University of Sanata Dharma.

Process to conduct thesis document plagiarism detection process begins with preprocessing, namely tokenisasi, stopword removal, stemming, and sentence splitter. Then by using the method of overlap measure function that will return the value of similarity (weighted value) of each document query is compared to the original document, the document similarity values that will define plagiarism detection threshold value is adjusted to plagiarism.

This thesis uses 50 informatics engineering thesis student for plagiarism detection training, before the first detectable threshold value must do a search, the search for plagiarism using the percentage threshold values 50, 40, and 30 from the average number of similarity of the documents which is - is the same (with the document itself). After testing the percentage threshold value, the best percentage was 50 compared with 30 and 40. The percentage of 50 has a value of 94.8% precision, while the percentage of 40 with a precision 92.4% and the percentage of 30% with 70% precision. Therefore to detect plagiarism in these systems are used threshold value with the percentage of 50% is 869.3 and the system is applicable for the detection of long documents with a sentence of 100 to 500 sentences and word length from 5 to 50 words.