Abstrak

Pendeteksian dan penghitungan kesamaan dokumen yang saat ini luas digunakan adalah pendeteksian dan penghitungan kesamaan berbasis kata. Dua dokumen yang membicarakan hal yang berbeda di lingkungan yang sama bisa saja menggunakan kata- kata yang sama. Dengan menggunakan pendeteksian dan penghitungan kesamaan berbasis kata kedua dokumen ini bisa dianggap mengandung informasi yang sama. Tetapi dengan memperhatikan frasa yang digunakan oleh masing masing dokumen maka dapat lebih terlihat perbedaan dan kesamaan informasi di antara kedua dokumen tersebut.

Tugas akhir ini membangun sebuah aplikasi yang dapat mendeteksi dan menghitung kesamaan antar dokumen dengan cara membandingkan dua dokumen pada satu waktu. Dokumen akan direpresentasikan dengan menggunakan konsep graph sehingga dapat dideteksi dan dihitung kesamaannya. Selain mendeteksi dan menghitung kesamaan frasa, juga dihitung kesamaan gabungan (gabungan kesamaan frasa dan kata) . Dokumen inputan adalah dokumen *.txt berbahasa Indonesia

Terdapat 4 Langkah pada proses pendeteksian dan penghitungan kesamaan dokumen yaitu preprocessing, pembentukan graph, pendeteksian kesamaan dan penghitungan kesamaan..

Keluaran dari aplikasi ini adalah nilai kesamaan antar tiap dokumen yang menjadi inputan..

Kata kunci: *Similarity*, Dokumen, Graph, Bahasa Indonesia, txt