

ABSTRACT

Eko Prasetyo

(150411100103)

Dosen Pembimbing I : Yoga Dwitya Pramudita, S.Kom.M.Cs.

Dosen Pembimbing II : Husni, S.Kom., MT.

Reading the news is a social activity to obtain information. And today, people can read the news anytime and anywhere. With the rapid development of technology today, the number of news that can be read is becoming more and more widespread on the internet, so reading the news and getting information from it will take longer if you have to read the entire contents of the news document. To overcome this, an automatic text summarization system was created (Automatic Text Summarization) so that readers can more quickly and easily obtain information from the news document. This system uses the K-Means Clustering method to group text data into several different groups. Where the first preprocessing will be done including sentence solving, case folding, filtering, tokenizing, and stopword removal. The results of the preprocessing will be weighted using TF-IDF. Then the last clustering process will be carried out using the K-Means Clustering method. In this study, the data used were 33 text documents taken from web sources on the internet. For testing the summary results of this research, 3 expert respondents will manually correct the summary results from the system. From the results of the tests that have been carried out, the accuracy is 69.95%.

Keywords: automatic text summarization ,K-means, TF-IDF

ABSTRAK

Eko Prasetyo

(150411100103)

Dosen Pembimbing I : Yoga Dwitya Pramudita, S.Kom.M.Cs.

Dosen Pembimbing II : Husni, S.Kom., MT.

Membaca berita merupakan kegiatan masyarakat untuk memperoleh informasi. Bahkan pada saat ini masyarakat dapat membaca berita dilakukan kapanpun dan dimanapun. Dengan berkembang pesatnya teknologi saat ini jumlah berita yang bias dibaca menjadi semakin banyak dan tersebar luas diinternet, sehingga untuk membaca berita dan mendapatkan informasi darinya akan membutuhkan waktu yang lebih lama jika harus membaca keseluruhan isi dokumen berita tersebut. Untuk mengatasi hal tersebut dibuatlah sistem peringkasan teks secara otomatis (*Automatic Text Summerization*) agar pembaca dapat lebih cepat dan mudah untuk mendapatkan informasi dari dokumen berita tersebut. Sistem ini menggunakan metode *K-Means Clustering* untuk mengelompokkan data teks menjadi beberapa kelompok yang berbeda. Dimana yang pertama akan dilakukan *Preprocessing* diantaranya yaitu pemecahan kalimat, *case folding*, *filtering*, *tokenizing* dan *stopword removal*. Hasil dari preprocessing tersebut akan dilakukan pembobotan menggunakan *TF-IDF*. Maka yang terakhir akan dilakukan proses *clustering* menggunakan metode *K-Means Clustering*. Pada penelitian ini digunakan data berjumlah 33 teks dokumen yang diambil dari sumber web diinternet. Untuk pengujian hasil ringkasan dari penelitian ini dilakukan oleh 3 responden ahli yang akan mengkoreksi secara manual hasil ringkasan dari sistem. Dari hasil pengujian yang telah dilakukan tersebut mendapatkan akurasi sebesar 69,95 %.

Kata kunci : peringkasan teks otomatis, *k-means*, *TF-IDF*