

Automatic Text Summarization
Dokumen Bahasa Daerah Menggunakan Metode Term
Frequency – Inverse Document Frequency (TF-IDF) Bebas Web

Nama Mahasiswa : MOCH. GHUFRON
NIM : 13.543.0123
Pembimbing I : Ahmad Izzuddin, S.T., M.Kom
Pembimbing II : Imam Marzuki, S.ST., M.T

ABSTRAK

Membuat dan membaca dokumen berbahasa daerah merupakan salah satu cara melestarikan penggunaan bahasa daerah. Membaca keseluruhan dokumen untuk memahaminya akan sangat melelahkan. Apabila terdapat ringkasan dari sebuah dokumen, maka akan membantu pembaca mendapatkan intisari dari sebuah dokumen dan mempercepat waktu membaca, untuk itu sistem peringkasan otomatis perlu dilakukan karena meringkas secara manual membutuhkan waktu yang cukup banyak. Pada penelitian ini metode yang digunakan adalah TF-IDF, dimana metode ini menggunakan cara pembobotan untuk setiap kata dalam kalimat. Sehingga kata dengan kemunculan terbanyak merupakan sebuah kalimat yang penting. Dari perhitungan tersebut akan menghasilkan kumpulan kalimat yang berisi bagian penting dari keseluruhan dokumen. Dari hasil pengujian sistem yang telah dilakukan dengan memberikan kuesioner kepada seorang ahli bahasa, menghasilkan nilai rata-rata akurasi sebesar 20% untuk dokumen cerpen bahasa Madura pada dengan tingkat kompresi 10%, 20%, 30%, 40% dan 50%, karena isi dari cerpen merupakan kalimat naratif. Dalam dokumen naratif setiap kalimat memiliki makna yang saling berhubungan, apabila beberapa kalimat dihilangkan dari sebuah cerita maka jalan cerita menjadi tidak bagus.

Kata Kunci : Peringkasan Otomatis, TF-IDF, Pembobotan, Kompresi.

***Automatic Text Summarization
Of Vernacular Documents Using Term Frequency
Inverse Document Frequency Method (TF-IDF) Web-based***

By : MOCH. GHUFRON
Student Identity Number : 13.543.0123
Supervisor I : Ahmad Izzuddin, S.T., M.Kom
Supervisor II : Imam Marzuki, S.ST., M.T

ABSTRACT

Creating and reading documents of vernacular is one of the way to preserve of vernaculars. reading totality of document for understanding it will be tried. If there is a summary of a document, it will help the reader get the essence of a document and speed up the reading time, for that the automatic summarization system needs to be done because summarizing it manually requires considerable time. In this study the method used is TF-IDF, where this method uses a weighting method for each word in the sentence. So the word with the most occurrences is an important sentence. From these calculations will produce a collection of sentences that contain an important part of the entire document. From the results of system testing that has been done by giving a questionnaire to a linguist, it produces an average value of 20% accuracy for Madura language short documents with a compression level of 10%, 20%, 30%, 40% and 50%, because the contents from short stories are narrative sentences. In narrative documents each sentence has a meaning that is interconnected, if several sentences are omitted from a story then the storyline is not good.

Keyword : Automatic Summarization, TF-IDF, Weighting, Compression.