ABSTRACT

Read the whole document takes a long time, let alone a document that is read is the article Sundanese. To read the entire article and understand it will be very tiring. But if there is a short summary of an article, it will assist the reader in understanding the contents of the article Sundanese. Make a summary manually would take a long time, therefore, to address the problem of time reading, will be made systems peringkas automatic text that will summarize the entire contents of a document to get the gist of the document Sundanese language, so as to understand the contents of a document without having to read the entire document.

In this study, the method used for automatic text peringkasan is TF-IDF (Term Frequency-Inversed Document Frequency). And some of the process undertaken to summarize the article, namely Case folding, tokenizing, Filtering, Stemming Algorithm Nazief-Andriani and Pattern Discovery / Analysis.

This level of accuracy obtained from 5 respondents and 5 Sundanese language articles with a value of 1 to 10 for each - each length of the summary obtained maximum value of 78% and the average value of 66% *accuracy. While the level of accuracy obtained from 5 respondents and 5 Sundanese language articles relevant to determining whether or not the summary results obtained maximum value of 5 (100%) and the average value of accuracy of 73%*

Keyword : summary, auto text summarization, text mining, tf-idf

ABSTRAK

Membaca keseluruhan dokumen membutuhkan waktu yang lama, apalagi dokumen yang dibaca adalah artikel berbahasa Sunda. Untuk membaca keseluruhan artikel dan memahaminya akan sangat melelahkan. Namun jika terdapat ringkasan pendek dari sebuah artikel, maka akan membantu pembaca dalam memahami isi artikel berbahasa Sunda. Membuat ringkasan secara manual akan membutuhkan waktu yang lama, oleh karena itu untuk mengatasi masalah waktu baca, akan dibuat sistem peringkas teks otomatis yang akan meringkas keseluruhan isi dari sebuah dokumen untuk mendapatkan intisari dari dokumen berbahasa Sunda, sehingga dapat memahami isi dokumen tanpa harus membaca keseluruhan dokumen.

Pada penelitian ini metode yang digunakan untuk peringkasan teks otomatis adalah TF-IDF (*Term Frequency-Inversed Document Frequency*). Dan beberapa proses yang dilakukan untuk meringkas artikel yaitu *Case folding, Tokenizing, Filtering, Stemming* dengan Algoritma Nazief-Andriani dan *Pattern Discovery / Analysis.*

Tingkat akurasi yang didapat dari 5 responden dan 5 artikel berbahasa sunda dengan memberikan nilai 1 sampai 10 untuk masing - masing panjang ringkasan diperoleh nilai maksimal 78% dan nilai rata-rata akurasi 66%. Sedangkan tingkat akurasi yang didapat dari 5 responden dan 5 artikel berbahasa sunda dengan menentukan relevan atau tidaknya hasil ringkasan diperoleh nilai maksimal 5 (100%) dan nilai rata-rata akurasi 73%

Kata kunci: ringkasan, peringkas teks otomatis, *text mining, tf-idf*