

## INTISARI

### AUTOMATIC SUMMARIZATION BAHASA INDONESIA MENGGUNAKAN TEXTRANK

Oleh

Nurdian Nugraha  
12/331417/PA/14672

Seiring dengan perkembangan teknologi di masa ini, penyebaran informasi menjadi hal yang sangat mudah dilakukan. Kemudahan dalam penyebaran informasi ini diiringi dengan bertambah banyaknya informasi yang tersedia, seperti dokumen. Namun, bertambah banyaknya informasi yang tersedia tidak selalu memberikan kemudahan bagi pembaca. Hal ini dikarenakan setiap orang tidak mempunyai cukup waktu untuk membaca keseluruhan informasi yang tersedia. Apalagi setelah dibaca, ternyata informasi yang tersedia pada suatu dokumen tidak sesuai dengan informasi yang diinginkan pembaca. Oleh karena itu, dibutuhkan sebuah sistem peringkasan dokumen otomatis untuk memberikan pembaca gambaran umum dari sebuah dokumen sebelum membacanya.

Pada penelitian ini, akan diimplementasikan sistem peringkasan dokumen otomatis menggunakan algoritma TextRank dan Semantic Networks and Corpus Statistic. Penggunaan TextRank ini memungkinkan ekstraksi kalimat-kalimat utama dari sebuah dokumen yang selanjutnya akan digunakan sebagai kalimat pada ringkasan keluaran. Peringkasan dokumen pada TextRank ini terdiri dari beberapa proses, yaitu tokenisasi kalimat, pembentukan graf, perhitungan nilai *edge* dengan menggunakan algoritma Semantic Networks and Corpus Statistic, perhitungan nilai *vertex*, pengurutan nilai *vertex*, dan pembentukan ringkasan.

Pengujian dilakukan dengan menghitung nilai *recall*, *precision*, dan F-Score dari ringkasan keluaran sistem dengan ringkasan idealnya menggunakan metode ROUGE-N untuk mengukur kualitas dari ringkasan keluaran sistem. Dari hasil pengujian yang dilakukan, diketahui bahwa kualitas ringkasan dipengaruhi oleh gaya penulisan dokumen, pemilihan kata-kata dan simbol pada dokumen, serta panjang ringkasan keluaran sistem. Panjang ringkasan optimal yang memberikan nilai F-Score terbesar dari hasil pengujian adalah 10% dari panjang dokumen dengan nilai F-Score 0,1635 dan 150 kata dengan nilai F-Score 0,1623.

Kata kunci: TextRank, Peringkasan, Semantic Networks, Corpus Statistic, WordNet, ROUGE

## **ABSTRACT**

### **AUTOMATIC SUMMARIZATION FOR INDONESIAN LANGUAGE USING TEXTRANK**

by

Nurdian Nugraha  
12/331417/PA/14672

With the advancement of technology in this era, spreading information becomes an easy thing to do. This ease of information dissemination is accompanied by an increase in available information, such as document. However, increasing the amount of available information does not always make it easy for the reader. This is because not everyone have enough time to read through all available information. Especially after reading through all available information, it turns out that the information is not in accordance with the desired information. Therefore, an automatic summarization system is needed to give the reader general image of a document before reading it.

In this research, automatic summarization system is implemented using TextRank, and Semantic Networks and Corpus Statistic algorithm. The use of this TextRank allows the system to extract the important sentences from a document which is used as the sentences for the output summary. Summarization within TextRank consists of several processes, namely sentence tokenization, graph creation, edge value calculation using Semantic Networks and Corpus Statistic algorithm, vertex value calculation, sorting the vertex value, and summary creation.

Testing is done by calculating the recall, precision, and F-Score of the system summary with its ideal summary using ROUGE-N method to measure the quality of system summary. Based on the test results, it is known that the quality of the summary is affected by the writing style, the selection of words and symbols in the document, as well as the length of the output summary. The optimal length of the summary that gives the largest F-Score value from the test results is 10% of the document length with an average F-Score value of 0.1635 and 150 words with an average F-Score value of 0.1623.

**Keywords:** TextRank, Summarization, Semantic Networks, Corpus Statistic, WordNet, ROUGE