

ABSTRAK

Nama : Mohamad Iman Solihin Sudrajat
NIM : 41519110007
Program Studi : Teknik Informatika
Judul Laporan Skripsi : Peringkasan Teks Otomatis Abstrak Bahasa Indonesia
menggunakan Model PEGASUS
Pembimbing : Dr.Ir. Eliyani

Dokumen teks besar sulit bisa dipahami dan membutuhkan waktu untuk mengekstrak informasi penting. Salah satu cara mendapatkan ringkasan teks dengan cepat yaitu dengan peringkasan teks otomatis abstrak. Penelitian ini menggunakan dataset indosum yang berisi kumpulan teks berita. Dengan data yang digunakan 2000 sampel dengan ukuran size dokumen yang berkisar 1 paragraf – 22 paragraf. Model algoritma yang digunakan adalah PEGASUS `tunner007/pegasus_summarize` yang disediakan oleh library `huggingface`. Skenario yang dilakukan adalah penggunaan pre-processing dan split data. Skenario 1 implementasikan tanpa stemming dan stopwords removal dengan split data 70:30, lalu skenario 2 implementasikan stemming tanpa stopwords dengan split data 80:20 dan skenario 3 implementasikan stemming sama stopwords removal dengan dataset 90:10. Dari hasil penelitian menunjukkan bahwa skenario 1 memberikan hasil terbaik dengan nilai matriks ROUGE-1 precision 0.581760, recall 0.627699, f1-score 0.602227. lalu ROUGE-2 precision 0.461695, recall 0.498631, f1-score 0.478117 dan ROUGEL-L precision 0.545045, recall 0.588313, f1-score 0.564325.

Kata Kunci : Peringkasan Teks Otomatis, Abstrak, Bahasa Indonesia, PEGASUS, ROUGE

ABSTRACT

Name : Mohamad Iman Solihin Sudrajat
NIM : 41519110007
Study Program : Teknik Informatika
Title Thesis : Automatic Text Summarization of Indonesian Abstracts
Using PEGASUS Model
Counsellor : Dr.Ir. Eliyani

Large text documents are difficult to understand and take time to extract important information. One way to quickly summarize text is with abstract automatic text summarization. This research uses the indosum dataset which contains a collection of news texts. With the data used 2000 samples with document size ranging from 1 paragraph - 22 paragraphs. The algorithm model used is PEGASUS `tunner007/pegasus_summarize` provided by the huggingface library. The scenarios performed are the use of pre-processing and split data. Scenario 1 implement without stemming and stopwords removal with 70:30 split data, then scenario 2 implement stemming without stopwords with 80:20 split data and scenario 3 implement stemming and stopwords removal with 90:10 dataset. The results show that scenario 1 gives the best results with ROUGE-1 precision 0.581760, recall 0.627699, f1-score 0.602227. then ROUGE-2 precision 0.461695, recall 0.498631, f1-score 0.478117 and ROUGEL-L precision 0.545045, recall 0.588313, f1-score 0.564325.

Keywords: Automatic Text Summarization, Abstract, Bahasa Indonesia, PEGASUS, ROUGE