

INTISARI

OPTIMASI PERINGKASAN TEKS OTOMATIS UNSUPERVISED SECARA EKSTRAKTIF MENGGUNAKAN *INTEGER LINEAR PROGRAMMING*

Oleh

Setyawan Putra Sujana

18/427594/PA/18554

Perkembangan media penyebaran informasi yang begitu masif merubah kebiasaan masyarakat untuk selalu mencari informasi melalui media digital. Dengan banyaknya informasi teks yang tersedia, diperlukan sebuah sistem peringkasan teks otomatis yang akan mempermudah masyarakat dalam mengekstrak informasi dari sebuah teks dengan lebih efektif, yang akan menghemat waktu dan tenaga. Sistem ini secara otomatis akan menghasilkan ringkasan yang mengandung pokok bahasan dari sebuah data teks.

Penelitian ini mencoba mengimplementasikan *Integer Linear Programming* pada berbagai sistem peringkasan teks otomatis ekstraktif dengan pendekatan *unsupervised* yang sudah ada. *INDOSUM* sebagai sebuah penelitian mengenai peringkasan teks otomatis beserta dengan dataset nya akan dijadikan baseline, yang akan dibandingkan performanya, sebelum dan sesudah pengimplementasian *Integer Linear Programming*. Empat algoritma akan diuji pada penelitian ini, yakni *LEAD-3*, *Sumbasic*, *Latent Semantic Analysis* dan *TextRank*.

Evaluasi sistem ini dilakukan menggunakan skor F1 dari metrik ROUGE. Dari pengujian keempat algoritma yang disebutkan diatas, didapatkan rata-rata kenaikan pada sistem yang ditambahkan ILP, berturut-turut sebesar 6.14%, 20.14% dan 7.44% pada ROUGE-1, ROUGE-2 dan ROUGE-L.

Kata Kunci: Peringkasan Teks Otomatis, ILP, Ekstraktif, *Unsupervised*

ABSTRACT

OPTIMIZATION OF UNSUPERVISED EXTRACTIVE AUTOMATIC TEXT SUMMARIZATION USING INTEGER LINEAR PROGRAMMING

By

Setyawan Putra Sujana

18/427594/PA/18554

The development of information dissemination media that is so massive has changed people's habits to always seek information through digital media. With a lot of text information available online, we need an automatic text summarization system that will make it easier for people to extract information from a text more effectively, which will save time and effort. This system will automatically generate a summary containing the subject matter of a data text.

This study tries to implement Integer Linear Programming on various extractive automatic text summarization systems with an existing unsupervised approach. INDOSUM as a study on automatic text summarization along with its dataset, will be used as a baseline against which the performance will be compared, before and after the implementation of Integer Linear Programming. Four algorithms will be tested in this study, namely LEAD-3, Sumbasic, Latent Semantic Analysis and TextRank.

Evaluation of this system is carried out by using the F1 score of the ROUGE metric. From testing the four algorithms mentioned above, the average increase in the system added with ILP was 6.14%, 20.14% and 7.44% in ROUGE-1, ROUGE-2 and ROUGE-L, respectively.

Keywords: Automatic Text Summarization, ILP, Extractive, Unsupervised