

INTISARI

PERINGKASAN TEKS OTOMATIS MENGGUNAKAN METODE *MAXIMUM MARGINAL RELEVANCE* BERBASIS *SEMANTIC NETWORKS* DAN *CORPUS STATISTICS*

WINDA YULITA
16/403721/PPA/05238

Peringkasan teks otomatis bertujuan meringkas teks menjadi lebih pendek tanpa menghilangkan makna dan isinya. Salah satu metode peringkasan teks otomatis adalah metode *Maximum Marginal Relevance* (MMR). Hasil ringkasan dengan metode MMR memiliki kelemahan yaitu terdapat bagian-bagian yang terpisah satu sama lain dalam hasil ringkasan. Hal ini disebabkan karena metode tersebut tidak menerapkan teknik pemahaman bahasa alami. Selain itu, metode MMR dengan pencocokan leksikal yang sederhana didalam *Cosine Similarity* belum cukup untuk mempresentasikan jarak antara dua dokumen (kalimat). Oleh karena itu, penelitian tentang peringkasan teks dengan metode MMR yang berbasis *semantic* dan *corpus statistics* sangat diperlukan.

Pada penelitian ini, metode yang digunakan adalah metode MMR berbasis *semantic networks* dan *corpus statistics* yang memanfaatkan *WordNet* Bahasa dan *corpus* dalam proses peringkasan teks. Data yang digunakan berupa 50 teks berita *online* dengan panjang teks rata-rata 400 kata. Hasil ringkasan dengan metode tersebut dibandingkan dengan hasil ringkasan menggunakan metode MMR berbasis *non-semantic* yang memanfaatkan TF-IDF. Pengujian hasil ringkasan teks dilakukan dengan menggunakan *toolkit ROUGE (Recall-Oriented Understanding for Gisting Evaluation)*.

Hasil penelitian menunjukkan bahwa hasil ringkasan metode MMR berbasis *semantic networks* dan *corpus statistics* memiliki *f-score* terbaik yaitu 0,570 dalam waktu 10 menit 11 detik, sedangkan MMR berbasis *non-semantic* memiliki nilai *f-score* terbaik yaitu 0,598 dalam waktu 1 menit 3 detik. Nilai tersebut dihasilkan dengan menambahkan proses *preprocessing* berupa *stemming* dan pemampatan hasil ringkasan 30%. Selain itu, nilai λ yang digunakan adalah 0,9, sedangkan nilai δ yang digunakan adalah 0,95. Perbedaan nilai *f-score* disebabkan oleh terdapat kata yang tidak sesuai dengan KBBI dalam data penelitian.

Kata kunci: peringkasan teks otomatis, metode MMR, *semantic*, *non-semantic*

ABSTRACT

AUTOMATIC TEXT SUMMARIZATION USING MAXIMUM MARGINAL RELEVANCE METHOD BASED ON SEMANTIC NETWORKS AND CORPUS STATISTICS

Winda Yulita

16/403721/PPA/05238

Automatic text summary aims to summarize the text to be shorter without losing its meaning and content. One method of automatic text summary is Maximum Marginal Relevance (MMR) method. The result of the summary with the MMR method has a disadvantage that there are separate parts of each other in the summary results. This is because the method does not apply natural language comprehension techniques. In addition, the MMR method with simple lexical matching in Cosine Similarity is not enough to represent the distance between two documents (sentences). Therefore, research on text summarization with the MMR method based on semantic and corpus statistics is needed.

In this study, the method used is the MMR method based on semantics networks and corpus statistics which utilizes WordNet Language and corpus in the text summarization process. The data used is 50 online news texts with an average text length of 400 words. Summary results with these methods are compared with summary results using non-semantic MMR-based methods that utilize TF-IDF. Testing the results of the summary text is done using the ROUGE (Recall-Oriented Understanding for Gisting Evaluation) toolkit.

The results showed that the summary results of the MMR method based on semantic networks and corpus statistics had the best f-score of 0.570 in 10 minutes 11 seconds, while the non-semantic MMR had the best f-score value of 0.598 in 1 minute 3 seconds. This value is generated by adding a preprocessing process in the form of stemming and compression of 30% summary results. In addition, the value of λ used is 0.9, while the value of δ used is 0.95. The difference in the f-score value is caused by a word that is not in accordance with the KBBI in the research data.

Keywords: automatic text summary, MMR method, semantic, non-semantic