

INTISARI

Peringkasan Teks Otomatis Ekstraktif pada Berita Berbahasa Indonesia Menggunakan *Maximal Marginal Relevance* dan *Non-Negative Matrix Factorization*

Oleh

Inggar Riyandi Musyaffanto
15/381066/PA/16746

Pesatnya perkembangan internet memunculkan banyak situs berita *online* dan tidak jarang beritanya memiliki judul yang tidak sesuai dengan isi berita, sehingga pembaca harus membaca keseluruhan teks untuk memahami teksnya. Masalah tersebut bisa diatasi secara otomatis dengan *automatic text summarization*.

Pada penelitian ini teks berita akan dilakukan pra-pemrosesan data berupa *sentence segmentation*, *case folding*, *stopword removal*, dan *stemming*. Setelah itu digunakan *Maximal Marginal Relevance* (MMR) untuk merangkum teks secara otomatis. Pada pembentukannya terdapat dua macam *query* yaitu yang berasal dari judul dan yang berasal dari hasil *Non-Negative Matrix Factorization* (NMF) untuk mengatasi judul yang tidak sesuai dengan isi berita.

Berdasarkan hasil evaluasi dengan ROUGE-1 pada rangkuman sistem yang hanya menggunakan MMR didapatkan nilai *recall* sebesar 62.7%, *precision* sebesar 69.59%, dan *f-measure* sebesar 63.58% serta untuk ROUGE-2 didapatkan nilai *recall* sebesar 53.11%, *precision* sebesar 58.61%, dan *f-measure* sebesar 55.21%. Sedangkan untuk rangkuman sistem yang menggabungkan MMR dan NMF pada pengujian ROUGE-1 didapatkan nilai *recall* sebesar 67.15%, *precision* sebesar 69.39%, dan *f-measure* sebesar 67.73% serta untuk ROUGE-2 didapatkan nilai *recall* sebesar 58.50%, *precision* sebesar 58.40%, dan *f-measure* sebesar 58.05%.

Kata Kunci : peringkasan, ekstraktif, *maximal marginal relevance*, *non-negative matrix factorization*, berita

ABSTRACT

Automatic Extractive Text Summarization for Indonesian News Articles Using Maximal Marginal Relevance and Non-Negative Matrix Factorization

by

Inggar Riyandi Musyaffanto
15/381066/PA/16746

The rapid development of internet has led to many online news sites and the title of the news is frequently misleading, so readers have to read all the text to understand the meaning of the text. This problem can be overcome automatically with automatic text summarization.

This research did data preprocessing for the news text such as sentence segmentation, case folding, stopword removal, and stemming. After that, Maximal Marginal Relevance (MMR) is used to summarize the text automatically. To form the summaries, there are two types of query. The first one is from the title of the news and the second one is from the Non-Negative Matrix Factorization (NMF)'s result to overcome the misleading title.

Based on evaluation using ROUGE-1, system summary that only use MMR got 62.70% for recall, 69.59% for precision, and 63.58% for f-measure also based on evaluation using ROUGE-2, it got 53.11% for recall, 58.61% for precision, and 55.21% for f-measure. While in system summary that combine MMR and NMF, based on evaluation using ROUGE-1 got 67.15% for recall, 69.39% for precision, and 67.73% for f-measure also based on evaluation using ROUGE-2, it got 58.50% for recall, 58.40% for precision, and 58.05% for f-measure.

Keywords : summarization, extractive, maximal marginal relevance, non-negative matrix factorization, news