

INTISARI

PERANGKUMAN RANGKAIAN PERISTIWA OTOMATIS ARTIKEL BERITA BAHASA INDONESIA MENGUNAKAN ALGORITMA PAGERANK

Oleh

Tantri Pandu Pertiwi

12/331355/PA/14618

Jumlah artikel berita untuk suatu peristiwa tertentu di internet yang sangat banyak menyebabkan pembaca berita kesulitan untuk memahami alur dari peristiwa yang rumit dan berkepanjangan. Peran rangkuman rangkaian peristiwa tentu sangat membantu pembaca untuk mengetahui garis besar dari sebuah peristiwa. Situs berita biasa menyajikan rangkuman peristiwa, namun proses perangkuman masih dilakukan oleh manusia secara manual. Hal ini tentu kurang efisien baik dari segi waktu, tenaga, dan biaya, sehingga diperlukan sebuah sistem yang dapat merangkum peristiwa secara otomatis.

Penelitian ini fokus dalam pembangunan sistem yang dapat menghasilkan rangkuman rangkaian peristiwa secara otomatis. Sistem ini bekerja dengan melakukan penilaian terhadap kalimat-kalimat berita, kemudian mengekstrak kalimat dengan skor tertinggi untuk dijadikan rangkuman. Proses penilaian dilakukan secara lokal dan global menggunakan algoritma *PageRank*. Skor lokal dan skor global kemudian dikombinasikan untuk mendapat skor kombinasi kalimat. Kalimat dengan skor kombinasi tertinggi kemudian diekstrak untuk dijadikan rangkuman.

Evaluasi ROUGE terhadap 4 topik berita menunjukkan bahwa rangkuman lokal (rangkuman yang hanya bergantung pada skor lokal kalimat) menghasilkan kualitas yang lebih baik daripada rangkuman global dan rangkuman kombinasi. Hal ini ditunjukkan dengan nilai *F-score* rata-rata rangkuman lokal sebesar 0,297005, sedangkan rangkuman global sebesar 0,281987, dan rangkuman kombinasi sebesar 0,2878084. Semakin tinggi nilai *F-score* berarti kualitas rangkuman semakin baik.

Kata kunci : rangkuman rangkaian peristiwa, perangkuman otomatis, ekstraksi informasi, pagerank

ABSTRACT

AUTOMATIC TIMELINE SUMMARIZATION FOR INDONESIAN NEWS ARTICLES USING PAGERANK ALGORITHMS

By

Tantri Pandu Pertiwi

12/331355/PA/14618

Online news agencies tend to publish news article as soon as a new event happened. This is causing the number of news articles of a story become very large and readers can't follow the development of the story. Timeline summaries are an effective way for readers to keep track of long-lasting news stories. Some online news agencies indeed provide timeline summaries, but most of them are human generated summaries, which are not efficient. An automatic timeline summarization is needed to save time and cost on generating timeline summaries.

This research focuses on generating automatic timeline summaries using PageRank algorithms to score news sentences. This system works by scoring sentences globally and locally, then combine those two scores to get combined score of the sentences. The sentence with highest score will represent the summary of a certain date.

Evaluation is done using ROUGE on four news topic samples. The result shows that local summaries (summaries that generated from local score only) give better quality than global summaries or combination summaries. The F-score from local summaries is 0,297005, while global summaries is 0,281987, and combination summaries is 0,2878084. The higher the F-score means the better the quality of the summaries.

Keywords : timeline summarization, automatic summarization, information extraction, pagerank