

INTISARI

PERINGKAS TEKS OTOMATIS BERITA EKONOMI BERBAHASA INDONESIA DENGAN MENGGUNAKAN 5W+1H DALAM KONSIDERSI

Jota Fauzan Hutomo

16/398515/PA/17476

Seiring perkembangan jaman, akses informasi menjadi lebih mudah. Dengan mudahnya akses informasi dan jumlah orang yang berbagi informasi lebih banyak, untuk seseorang bisa memahami semua informasi – informasi tersebut menjadi lebih susah. Untuk membantu masalah tersebut, rangkuman bisa dibentuk oleh manusia untuk pembaca bisa membaca terlebih dahulu dan membuat pilihan untuk lanjut membaca atau tidak. Namun rangkuman tidak bisa seterusnya dibentuk oleh manusia karena jumlah informasi untuk dirangkum juga sangat banyak. Jadi dibentuk peringkass teks otomatis untuk mempercepat proses pembuatan rangkuman.

Akan diteliti perangkuman teks berita yang memiliki topik ekonomi. Digunakan konsep 5W+1H sebagai fitur klasifikasi. Pilihan untuk mengambil konsep 5W+1H dilakukan karena konsep tersebut adalah dasar dari bagaimana membuat berita yang baik dan benar, serta merupakan elemen utama dari sebuah berita. Untuk klasifikasi berita yang memiliki aspek 5W+1H tersebut digunakan klasifier *Naïve Bayes*. Serta untuk mengurangi jumlah kalimat lebih banyak, digunakan penilaian teks berbasis TF-IDF.

Dalam penelitian ini, evaluasi dihitung dengan ROUGE-1 dan ROUGE-2. Kombinasi klasifier *Naïve Bayes* dengan penilaian teks TF-IDF dapat menghasilkan rata – rata *precision* ROUGE-1 sebesar 0,65 dan rata – rata *precision* ROUGE-2 sebesar 0,56 pada rata – rata kompresi teks artikel sebesar 40% dari teks asli.

Keyword: *Natural Language Processing, Naïve Bayes, TF-IDF, 5W+1H*

ABSTRACT

AUTOMATIC TEXT SUMMARIZATION OF INDONESIAN ECONOMIC NEWS WITH 5W+1H IN CONSIDERATION.

Jota Fauzan Hutomo

16/398515/PA/17476

Over time, access to information has become easier. With how easy it is to access new information, and more people have access to share their information, information has been abundant. Understanding every information someone gets has been more challenging than ever. To help the problem, summarization can be generated by other people for readers to be able to understand what the text is about and to continue reading or not. However, humans cannot make summarizations forever, as there is much information to summarize.

Automatic text summarization has been studied to accelerate the process of summarization generation. Automatic text generation will be researched for news text with the economy as its topic. The concept of 5W+1H will be used as a feature for classification. The choice for using the concept of 5W+1H is because that concept is the basis of how excellent and coherent news should be made and is the main element for news. For the classification of news that has the 5W+1H aspect, a Naïve Bayes classifier will be used. Moreover, to further reduce the number of sentences generated, sentence scoring using TF-IDF will be used.

In this research, evaluation is calculated with ROUGE-1 and ROUGE-2. The combination of Naïve Bayes classifier and TF-IDF sentence scoring produces a machine-generated summary with average ROUGE-1 precision of 0,65 and average ROUGE-2 precision of 0,56 with average compression of 40% from its original text.

Keyword: *Natural Language Processing, Naïve Bayes, TF-IDF, 5W+1H*