ABSTRAK

Tahapan ekstraksi kata kunci merupakan salah satu tahapan penting dari beberapa aplikasi *text mining*. Untuk mendapatkan kata kunci yang tepat secara lebih otomatis, berbagai metode ekstraksi kata kuncipun terus dikembangkan dan diuji. Pada artikel ilmiah, ekstraksi kata kunci dibutuhkan untuk memberikan alternatif kata kunci secara lebih sistematis kepada penulis jurnal. Penentuan kata kunci secara manual pada artikel ilmiah tidaklah efektif terutama jika artikel ilmiah yang akan dianalisis kata kuncinya tersebut jumlahnya sangat banyak. Pada penelitian ini ekstraksi kata kunci dikembangkan menggunakan metode textrank untuk mengekstraksi dokumen teks bahasa Indonesia dengan memodifikasi tahapan *preprocessing* pembentukan kandidat kata kunci dari algoritma textrank tersebut menggunakan aturan *multiword expression candidate*. Tahapan keseluruhan metode yang digunakan pada penelitian ini yaitu *preprocessing(text cleaning,tokenizing,case folding,stopword removal,* POS *tagging, candidates multiword extraction),* ekstraksi kata kunci dan tahapan terakhir yaitu *post-processing* untuk pemfilteran kata kunci yang terlalu umum. Hasil akhir dari penelitian ini menunjukkan bahwasanya textrank dengan *multiword expression candidate* memiliki waktu ekstraksi yang lebih cepat dan persentase akurasi *recall* yang sedikit lebih tinggi dibandingkan algoritma textrank biasa pada *top*-15 kata kunci.

Kata kunci : Ekstraksi kata kunci, *Textrank, Preprocessing, Multiword Expression Candidate*

AUTOMATIC KEYWORDS EXTRACTION FROM INDONESIAN TEXT DOCUMENT USING TEXTRANK METHOD

ABSTRACT

Keywords extraction is one of the most important stage in some of text mining applications. To acquire the right keywords more automatically, various methods of keywords extraction continues to be developed and examined. In most scientific articles, keywords extraction is needed to offer alternatives keywords systematically to journal authors. Most of the cases, keywords of scientific articles are offered manually and this is not really effective, especially when many scientific articles contains keywords to be extracted. In this research, keywords extraction is developed by using textrank method to extract Indonesian text document by modifying the preprocessing stage of candidate keywords selection in textrank algorithm using multiword expression candidate rule. The overall stages used in this research are preprocessing (text cleaning, tokenizing, case folding, stopword removal, POS tagging, multiword candidates extraction), keyword extraction and the last stage is post-processing for filter keywords that have common words. The result of this research showed that textrank with multiword expression candidate has a faster extraction time and a slightly higher recall accuracy compared to common textrank algorithm in the top-15 keywords.

Keyword: *Keywords extraction, Textrank, Preprocessing, multiword expression candidat.*