

Abstraksi

Berbagai informasi dalam jumlah besar dapat diperoleh dengan mudah dari Internet. Tetapi, tidak mudah untuk mencari informasi yang berguna diantara sekian banyak sumber informasi yang kita dapatkan. Beberapa *search engine* dapat melakukan pencarian teks lengkap, tetapi hasilnya kurang memuaskan. Efisiensi pencarian dari jumlah dokumen yang sangat banyak sangat dipengaruhi oleh kualitas dari kata kunci yang diberikan oleh setiap dokumen. Oleh karena itu, perlu dibutuhkan kata kunci yang berkualitas yang diambil secara otomatis diekstrak dari dokumen-dokumen agar mendapatkan nilai efisiensi yang tinggi. Pada Tugas Akhir ini, akan membahas mengenai "Ekstraksi Kata Kunci dengan menggunakan Conditional Random Fields". Conditional Random Fields (CRF) adalah model probabilistik untuk segmentasi dan pelabelan data sekuens. CRF menggunakan teknik yang berbeda dalam melakukan *preprocessing* dimana kalimat-kalimat yang ada akan disegmentasi dan diberi label, sehingga akan memberikan pengaruh terhadap hasil dari ekstraksi kata kunci tersebut. Pengujian dilakukan dengan menghitung nilai *precision*, *recall* dan *F-Measure* untuk mengetahui performansi sistem yang akan dibuat. Hasil pengujian menunjukkan bahwa ekstraksi kata kunci dengan CRF memiliki akurasi yang lebih baik pada jenis dokumen yang bersifat umum (topik umum) atau dengan kata lain memiliki variasi data yang tinggi. Sedangkan dari sisi fungsi fiturnya, fitur POS (tanpa normalisasi) memberi pengaruh yang cukup signifikan terhadap performansi sistem, sedangkan fitur Len tidak memberikan pengaruh yang signifikan terhadap performansi sistem. Selain itu, berdasarkan jumlah dokumen trainingnya, ekstraksi kata kunci dengan CRF bisa diterapkan secara efektif walaupun jumlah data trainingnya sedikit.

Kata Kunci : Ekstraksi kata kunci, Conditional Random Fields, *preprocessing*, fitur POS, fitur Len, *F-Measure*.