

ABSTRAK

Pesatnya perkembangan internet menyebabkan peningkatan jumlah dokumen ilmiah seperti jurnal-jurnal ilmiah. Besarnya jumlah dokumen menyebabkan pemrosesan informasi membutuhkan waktu yang lama. Dengan mengekstrak kata kunci yang merepresentasikan isi dokumen, pemrosesan keseluruhan informasi pada dokumen tidak lagi diperlukan. Namun, ekstraksi kata kunci secara manual tidak efektif dan efisien dari segi waktu maupun sumber daya. Otomasi ekstraksi kata kunci diperlukan untuk mengatasi keterbatasan tersebut.

Pada Tugas Akhir ini akan diimplementasikan *Conditional Random Field* (CRF) model dengan menganalogikan proses ekstraksi kata kunci sebagai proses pelabelan *term* pada dokumen. Pemodelan ini membutuhkan proses pelatihan untuk menghasilkan parameter pendukung fitur yang optimal. Pengujian dilakukan untuk mengetahui pengaruh jumlah fitur, pendekatan ekstraksi fitur, jumlah data latih, serta keterlibatan eliminasi *stopwords* terhadap performansi proses ekstraksi kata kunci. Pengujian dilakukan terhadap dua kelompok dokumen dengan tingkat homogenitas berbeda (dokumen kedokteran dan kesehatan masyarakat) untuk mengetahui performansi ekstraksi kata kunci pada kedua kelompok dokumen tersebut.

Hasil pengujian menunjukkan bahwa performansi ekstraksi kata kunci terbaik pada dokumen kedokteran memberikan nilai *precision* 0.3892, *recall* 0.714, dan *f-measure* 0.4648 sementara performansi ekstraksi kata kunci terbaik pada dokumen kesehatan masyarakat memberikan nilai *precision* 0.2833, *recall* 0.692, dan *f-measure* 0.3901. Penambahan jumlah fitur dan keterlibatan eliminasi *stopwords* menyebabkan penurunan performansi ekstraksi kata kunci.

Kata kunci : Ekstraksi kata kunci, kata kunci, *Conditional Random Field* (CRF) Model