

ABSTRAK

Optical Character Recognition (OCR) adalah sebuah sistem komputer yang digunakan secara otomatis mengenali serangkaian karakter yang berasal dari mesin ketik, mesin cetak ataupun tulisan tangan. Dengan kata lain OCR adalah proses pengalihan dokumen teks menjadi file komputer tanpa harus pengeditan ulang, setiap karakter baik huruf, kata, kalimat dapat dikenali secara tepat dan dibaca oleh perangkat lunak yang lain, tanpa harus pengetikan ulang dan editing. Pada penelitian sebelumnya, teknik OCR dikembangkan dengan menggunakan pendekatan vektor dan region pada ekstraksi cirinya, untuk mengidentifikasi suatu file gambar (bmp) yang berisi karakter yang berasal dari pemindaian *hardcopy* atau dari sumber lainnya hanya sampai pada tahap pengenalan karakter yang mempunyai tingkat keakuratan karakter rata-rata sekitar 86%.

Maka pada proyek akhir ini dibuat suatu aplikasi yang dapat menambah akurasi pada proses sistem OCR yang dikembangkan dengan menggunakan pendekatan vektor dan region pada ekstraksi cirinya, dengan sistem *Auto Correct*. *Auto Correct* bekerja dengan cara mengidentifikasi susunan karakter pada artikel teks hasil sistem OCR menjadi sebuah kata. Kemudian mengoreksi ejaan kata tersebut, lalu membandingkan masing-masing kata pada database. Database merupakan kumpulan kata yang dibuat seperti kamus yang disusun berdasarkan abjad. Lalu sistem akan menghitung akurasi secara otomatis.

Untuk mengevaluasi performansi dari *Auto Correct* dengan menggunakan metode tersebut, dilakukan pengujian terhadap beberapa sampel masukan baik yang berasal dari dokumen *hardcopy* maupun yang berasal dari sumber lainnya. Hasil analisis menunjukkan bahwa sistem *Auto Correct* ini mampu menaikkan akurasi kata sebesar 21.53 % untuk font yang sudah dilatihkan oleh sistem OCR.

Kata kunci: OCR, *autocorrect*.