

## MEMBANDINGKAN KEMAMPUAN KONVERSI OCR SOFTWARE

<sup>1</sup>Angga A, Angga P, Andhika, Rini I, Marzuk  
<sup>2</sup>I Wayan S. Wicaksana

<sup>1</sup>Program Studi Teknik Informatika, Universitas Gunadarma ( [{angga\\_a,angga\\_jihadco,  
andhika\\_rina,rolly} @student.gunadarma.ac.id](mailto:{angga_a,angga_jihadco,andhika_rina,rolly}@student.gunadarma.ac.id))

<sup>2</sup>APusat Studi Teknologi Sistem Informasi ([iwayan@staff.gunadarma.ac.id](mailto:iwayan@staff.gunadarma.ac.id))

### ABSTRAK

*Permasalahan output dari scanner adalah merupakan file dengan model grafik, untuk pengolahan data digital akan lebih mudah dikelola dalam model text. Dalam perkembangan teknologi scanner saat ini, OCR sebagai salah satu perangkat yang fungsinya mengkonversi format gambar menjadi teks. Sehingga pengidentifikasian character yang ada di dalam file digital. Permasalahan adalah bagaimana memilih OCR yang memiliki resolusi dan kecepatan pemrosesan yang optimal. Untuk mengetahui kualitas software OCR, kami menguji dengan membandingkan beberapa software OCR yang kami dapatkan. Dalam pengujian beberapa software tersebut, kami menentukan beberapa parameter yang umum digunakan. Diantaranya waktu pemrosesan konversi grafik ke teks oleh software OCR dengan melihat parameter waktu konversi, kemudahan dalam instalasi serta penggunaan dan tingkat kesalahan software OCR dengan melihat hasil konversi dari berbagai jenis format grafik model file input. Hasil dari penelitian ini sangat membantu pada pengguna yang akan menggunakan OCR dalam melakukan konversi dokumen cetak ke digital dalam format teks. Karena dengan memanfaatkan metadata teks akan relatif lebih mudah dalam pengelolaan database dibandingkan dengan file grafik saja..*

*Kata Kunci: konversi, OCR, perbandingan*

### 1. PENDAHULUAN

Seiring perkembangan teknologi scanner sebagai salah satu OCR, yang fungsi sebelumnya hanya memindai dokumen cetak menjadi gambar, menjadi dapat memindai suatu document cetak menjadi gambar dan mengidentifikasi character yang ada di dalam gambar tersebut untuk di jadikan text dan disimpan kedalam computer atau media penyimpanan digital, sehingga dokumen cetak tersebut tidak memakan tempat, tidak mengalami kerusakan akibat rayap dan tidak mengalami kesulitan dalam memeriksa dan mencari file tersebut yang kemudian di sebut sebagai dokumen digital.

Namun dalam prakteknya tidak semua hasil pemindaian dari dokumen cetak ke text memberikan hasil yang baik, di sebabkan banyaknya software-software OCR untuk merubah gambar hasil scanning ke bentuk teks yang berbeda-beda kemampuannya, dimana setiap software-software OCR tersebut memiliki perbedaan dalam kecepatan dalam memindai dokumen cetak ke bentuk gambar lalu merubahnya ke bentuk teks, keakuratan memindai dari gambar ke dalam bentuk teks, jenis-jenis file input dan output yang dapat di identifikasikan oleh software OCR tersebut, dan spesifikasi apa yang di butuhkan computer agar dapat menginstall software OCR tersebut

Untuk itu kami melakukan pengujian beberapa software OCR terhadap permasalahan diatas dengan landasan teori cara kerja software OCR yang kami dapat sebagai berikut : Fauziah dalam tulisannya tentang Teknologi Scanner menuliskan “cara kerja OCR adalah teknik yang disebut matrix atau pencocokan pola dimana OCR akan menyimpan data tentang bentuk tiap karakter pada programnya”. menurut Resmana Lim [1] dalam tulisannya tentang APLIKASI OCR DENGAN METODE TEMPLATE MATCHING UNTUK PENDUKUNG BERMAIN GITAR menuliskan “dimana cara kerja metode Template Matching adalah melakukan pattern recognition pada karakter yang ingin dikenali dan membandingkan antara input pattern dengan template yang disimpan.

## 2. LANGKAH PENGUJIAN

Pengujian yang kami lakukan berdasarkan pada permasalahan berikut yaitu berapa lama software OCR dalam memproses suatu dokumen cetak ke bentuk text digital, Tingkat kesalahan yang terjadi dalam pengidentifikasian character, Menguji jenis format gambar apa saja yang di input Software OCR tersebut dan bentuk format output nya, Keakuratan dari software OCR

### Alat yang digunakan :

- Pc pentium 4. 3Ghz
- Ram 1 Gb
- VGA 256 Mb
- Scanner Canon D646u ex

### Software yang di uji

- FreeOCR.Net
- TopOCR
- SimpleOCR

- Omnipage Pro 14.0

### Bahan pengujian :

- Dokumen text full 1 halaman
- Dokumen text full 1 halaman berbentuk kolom
- Dokumen text & gambar 1 halaman
- Dokumen text & gambar 1 halaman berbentuk kolom
- Dokumen text dan table 1 halaman
- Dokumen text dan table 1 halaman kolom
- Dokumen text,table dan gambar 1 halaman
- Dokumen text,table dan gambar 1 halaman berbentuk kolom

Cara pengujian yaitu: penghitungan kecepatan proses software akan dilakukan dengan penghitung waktu dengan satuan detik, pengujian tingkat kesalahan akan dilakukan penghitungan ,dimana tiap berapa character software OCR melakukan kesalahan dalam memindai karakter yang ada di dokumen, menguji format apa saja yang dapat di input dan di output oleh software OCR, akan dilakukan penyimpanan hasil scanning OCR dan dilihat format apa saja yang digunakan dari software OCR,Dalam pengujian keakuratan akan dilihat perbedaan penempatan dengan membandingkan urutan yang ada pada dokumen cetak atau gambar dengan hasil dari software OCR.

## 3. HASIL PERCOBAAN

Berikut ini hasil percobaan dari beberapa masalah terhadap bahan uji seperti yang disebutkan diatas dapat dilihat di tabel 1.

Pengujian Software OCR		FreeOCR				TopOCR				SimpleOCR				OmniPage			
		Kecepatan	formata output	Tingkat kesalahan	Keakuratan	Kecepatan	format output	Tingkat kesalahan	Keakuratan	Kecepatan	format output	Tingkat kesalahan	Keakuratan	Kecepatan	format output	Tingkat kesalahan	Keakuratan
File Uji	Text	51.68	Doc/txt	40	60	53.22	Pdf, Txt	50	50	34.93	Doc/txt	60	40	46.64	Rtf	20	80
	Text & Gambar	51.68	Doc/txt	80	20	53.22	Pdf, Txt	60	40	34.93	Doc/txt	80	20	46.64	Rtf	40	60
	Text & Table	51.68	Doc/txt	80	20	53.22	Pdf, Txt	75	25	34.93	Doc/txt	85	15	46.64	Rtf	15	85
	text,table dan gambar	51.68	Doc/txt	60	40	53.22	Pdf, Txt, Rtf, Htm	70	30	34.93	Doc/txt	60	40	46.64	Rtf	10	90

File Uji	text full kolom	50.65	Doc/txt	90	10	47.30	Pdf, Txt	45	55	46.27	Doc/txt	50	50	46.23	Rtf	20	80
	Text, gambar, kolom	50.65	Doc/txt	90	10	47.30	Pdf, Txt	85	15	46.27	Doc/txt	85	15	46.23	Rtf	30	70
	Text, table, kolom	50.65	Doc/txt	90	10	47.30	Pdf, Txt	85	15	46.27	Doc/txt	80	20	46.23	Rtf	15	85
	Text, tabel, gambar, kolom	50.65	Doc/txt	90	10	47.30	Pdf, Txt, Rtf, Htm	90	10	46.27	Doc/txt	90	10	46.23	Rtf	15	85

#### 4. KESIMPULAN

Dari hasil penelitian diatas disimpulkan bahwa kualitas Software OCR di tentukan dari kemampuannya dalam menganalisa file gambar untuk menjadi dokumen teks digital. Hasil menunjukan bahwa proses document kertas menjadi document digital mengalami tingkat kesalahan yang sangat tinggi pada document berbentuk kolom.

Dari percobaan yang dilakukan Top OCR software memiliki kecepatan tertinggi. Sementara OmniPage software memberikan akurasi tertinggi, walau pun dari sisi kecepatan bukan yang terbaik.

#### Daftar Pustaka

- [1] Lim, Resmana, dkk. "Aplikasi OCR dengan metode Template Matching untuk pendukung bermain gitar", Universitas Kristen Petra, 2004, <http://www.telekom.ee.itb.ac.id/tssa2004/papers-Abstract/resmana-ocr-gitar-TSSA2004.doc>, 27 Maret 2008.
- [2] Ananta, Adityo, dkk, "Penggunaan Metode Thinning Pada Pemrosesan Gambar", 2003. [www.cs.ui.ac.id/webkuliah/citra/2003/thinning/kelompok10.doc](http://www.cs.ui.ac.id/webkuliah/citra/2003/thinning/kelompok10.doc), diakses pada 27 Maret 2008 2.55 p.m
- [3] Hassane, OCR(Optical Character Recognition), 2007, [www.libararycorner.org](http://www.libararycorner.org), diakses pada 27 Maret 2008 3.20 p.m