

KONVERSI CITRA KARTU NAMA KE TEKS MENGGUNAKAN TEKNIK OCR DAN JARO-WINKLER *DISTANCE*

Dewinta Zulhida Putri¹⁾, Diyah Puspitaningrum²⁾, Yudi Setiawan³⁾

^{1), 2), 3)}Program Studi Teknik Informatika, Fakultas Teknik, Universitas Bengkulu

JL. WR. Supratman, Kandang Limun, Kota Bengkulu 38371A

Email : dewintaputri23@gmail.com¹⁾, diyahpuspitaningrum@gmail.com²⁾, ysetiawan@unib.ac.id³⁾

Abstrak

Kartu nama memiliki data yang dapat disimpan ke dalam database kontak *smartphone*, seperti nama, jabatan, instansi, nomor ponsel, nomor telepon, nomor fax, alamat pertama, alamat kedua dan *e-mail*. Penelitian ini bertujuan membangun aplikasi yang dapat memindahkan data kartu nama dari foto ke database kontak *smartphone* dengan konversi citra kartu nama ke teks dengan melakukan ekstraksi teks dan pengelompokan data. Algoritma yang digunakan untuk ekstraksi teks adalah *Optical Character Recognition (OCR)*. Teks yang sudah diekstraksi kemudian akan dikelompokkan menggunakan algoritma *Jaro-Winkler Distance*. Pengujian sistem yang dilakukan dengan menggunakan data kartu nama yang beralamat di Provinsi Bengkulu sebanyak 50 buah dengan berbagai jenis tulisan dan *background*. Aplikasi dirancang dengan Bahasa pemrograman Java dengan IDE *Android Studio* dan library *Tesseract*. Hasil akhir dari sistem adalah data kontak yang tersimpan di database *smartphone*. Hasil uji coba sistem yang dilakukan didapatkan nilai akurasi rata-rata dari lima kali percobaan adalah 76% untuk data nama, 74% untuk data jabatan, 56% untuk data instansi, 82% untuk data nomor ponsel, 86% untuk data nomor telepon, 78% untuk data alamat, dan 88% untuk data *e-mail*.

Kata Kunci : Konversi Citra, Ekstraksi Teks, Kartu Nama, *Optical Character Recognition (OCR)*, *Jaro-Winkler Distance*

1. Pendahuluan

Latar Belakang

Kartu nama merupakan bagian penting yang sulit dipisahkan dari kehidupan seseorang, terutama orang-orang yang menggeluti dunia bisnis. Kartu nama berisikan identitas seseorang agar orang lain dapat menghubunginya untuk berkomunikasi lebih lanjut di kemudian hari. Kartu nama biasanya berisi nama, nomor-nomor yang dapat dihubungi, alamat perusahaan. Kartu nama juga dapat berisikan informasi lainnya sesuai dengan identitas yang ingin diberikan oleh pemilik kartu nama. Saat menerima kartu nama dari

orang lain, beberapa penerima akan menyimpan kartu nama tersebut di dalam saku, dompet, atau tempat khusus kartu nama. Lalu kemudian kartu nama tersebut akan kembali dicari saat akan menghubungi seseorang yang bersangkutan. Risiko kehilangan kartu nama merupakan batasan dari cara tersebut, sehingga beberapa orang akan langsung memindahkan informasi pribadi yang ada di kartu nama tersebut ke *database handphone* atau *smartphone* agar dapat lebih mudah dicari.

Aktivitas pemindahan data ini tentu menyita waktu dan ketelitian dari penerima kartu nama. Padahal kecanggihan teknologi dapat dimanfaatkan untuk mengatasi kekurangan ini sehingga membuat pemindahan data dari kartu nama ke *database* penerima kartu nama menjadi lebih mudah dan cepat. Salah satu teknologi yang dapat dimanfaatkan adalah dengan melakukan ekstraksi teks yang akan mengkonversi citra kartu nama ke teks. Kartu nama yang telah difoto oleh penerima kartu nama dapat diekstrak langsung menjadi teks sehingga data langsung dapat disimpan tanpa diketik ulang. Kartu nama yang sebagian besar terdiri dari teks dapat dikonversi menjadi teks dengan menggunakan *Optical Character Recognition (OCR)*.

OCR merupakan teknologi yang memungkinkan untuk mengenali karakter melalui mekanisme optik secara otomatis. Diibaratkan pada manusia, mata merupakan mekanisme optik. OCR adalah teknologi yang berfungsi seperti kemampuan manusia membaca. Meskipun OCR tidak mampu bersaing dengan kemampuan membaca manusia. OCR dapat mengenali teks baik tulisan tangan maupun hasil cetakan [1].

Setelah diekstrak maka teks tersebut haruslah dikelompokkan berdasarkan nama, alamat, nomor telepon, nomor ponsel, dan data lainnya yang ingin disimpan. Data dikelompokkan berdasarkan ciri-ciri yang telah ditentukan, misalnya alamat pasti mengandung data jalan. Ini berarti diperlukan algoritma pencocokan *string* untuk mengelompokkan data-data tersebut. Salah satu algoritma yang dapat digunakan untuk mencocokkan *string* adalah *Jaro-Winkler Distance*. *Jaro-Winkler Distance* sangat cocok dan memberikan hasil terbaik pada pencocokan dua *string* singkat, misalnya nama orang [2].

Berdasarkan hal tersebut, maka penulis mengusulkan proposal penelitian untuk menciptakan aplikasi yang dapat melakukan Konversi Citra Kartu Nama Ke Teks Menggunakan Teknik *Optical Character Recognition*

(OCR) Dan Jaro-Winkler *Distance* dengan pengimplementasi ada *smartphone*.

Tujuan

Penelitian ini diharapkan dapat menghasilkan aplikasi yang akan memudahkan pemindahan dan pengelompokan data dari kartu nama ke *database* kontak *smartphone* berdasarkan nama, alamat, nomor ponsel, nomor telepon, dan email.

Metodologi Penelitian

Pengumpulan data dalam penelitian ini menggunakan beberapa metode, yaitu:

1. Studi Pustaka

Studi pustaka dilakukan dengan mengumpulkan data dan informasi yang akan digunakan dalam perancangan dan pengembangan aplikasi ekstraksi teks dari foto kartu nama. Data yang digunakan diperoleh dari buku-buku, jurnal-jurnal, laporan penelitian, artikel ilmiah, skripsi, tesis, dan sumber tertulis lainnya yang berkaitan dengan kartu nama, *Optical Character Recognition* (OCR), Jaro-Winkler *Distance*, *Data Flow Diagram* (DFD), OpenCV, Android, dan basis data dengan SQLite.

2. Dokumentasi

Dokumentasi merupakan catatan peristiwa yang sudah lalu. Dokumentasi dapat berbentuk dokumen gambar, tulisan, atau karya-karya monumental dari seseorang. Penelitian ini menggunakan metode pengumpulan data dokumentasi yang menghasilkan dokumen foto-foto.

Tinjauan Pustaka

1. Kartu Nama

Kartu nama merupakan sebuah kartu yang berisikan informasi seseorang berupa nama, nama perusahaan, dan nomor telepon., terdapat beberapa tujuan seseorang memiliki dan memberikan kartu nama kepada yang lainnya, seperti:

- Kartu nama yang dirancang dengan baik menandakan bahwa pemberinya merupakan seseorang yang bertanggungjawab dan mudah dijangkau.
- Saat seseorang memasarkan dirinya dalam mencari pekerjaan, maka kartu nama "harus menjadi bagian integral dari gudang pemasaran." [3]
- Kartu nama adalah bagian penting untuk menyampaikan citra profesional [4].

2. Konversi Citra ke Teks

Citra juga dapat dikonversi menjadi bentuk teks dengan memanfaatkan proses pengenalan teks dari citra. Pengenalan teks adalah metode yang dikembangkan untuk membuat sistem yang mampu memberi pengertian atau deskripsi mengenai objek teks pada citra. Pengenalan teks merupakan bagian dari teknik analisa citra. Elemen dasar citra yang dianalisa pada teknik pengolahan teks adalah elemen bentuk. Proses pengenalan teks atau pendeteksian teks pada media gambar berarti sistem dapat mengenali/membaca suatu objek teks pada media gambar dan menuliskannya ke bentuk

string. String merupakan defenisi karakter berdasarkan kode ASCII [5].

String sering kita temui pada piranti lunak pengolahan kata atau pada program penyunting teks. Teks pada media citra adalah bukan merupakan string, karena teks tersebut bukan merupakan perwakilan dari kode – kode ASCII, tapi merupakan objek yang terbentuk dari susunan piksel. Metode dasar yang digunakan untuk mengenali objek tersebut sebagai teks adalah dengan mencirikan bentuk dari karakter tersebut masing – masing [5].

3. Optical Character Recognition

Optical Character Recognition (OCR) adalah proses yang dapat mengkonversi gambar yang berisikan teks menjadi karakter ASCII yang dapat dikenali oleh komputer. Berbagai pendekatan yang dapat digunakan pada *Optical Character Recognition*, yaitu matrix matching, fuzzy logic, ekstraksi fitur, analisa struktural, dan jaringan syaraf [1].

4. Tesseract

Tesseract adalah mesin OCR *open-source* yang dikembangkan di HP (Hewlett-Packard) antara tahun 1984 dan 1994. Tesseract muncul sebagai proyek penelitian PhD di HP Labs, Bristol yang tersedia di <http://code.google.com/p/tesseract-ocr> [6]. Tahapan proses *Optical Character Recognition* oleh Tesseract adalah:

- Page Layout Analysis*
- Blob Finding*
- Find Text Line and Words*
- Recognition Word Pass 1*
- Recognition Word Pass 2*
- Fuzzy Space and x-height Fix-up*

5. Jaro-Winkler Distance

Jaro-Winkler *distance* adalah Jaro *distance* dengan Winkler modification. Jaro *distance* digunakan untuk komputasi antara dua string (A dan B) dan berdasarkan pencocokan dan transposisi karakter [7]. Tahapan algoritma ini adalah:

- Menghitung panjang string,
- Menemukan jumlah karakter yang sama di dalam dua string
- Menemukan jumlah transposisi
- Menghitung jarak (dj) antara dua string yaitu s1 dan s2 dengan menggunakan rumus 2.1.

$$dj = \frac{1}{3} \times (m/|s1| + m/|s2| + (m-t)/m) \quad (2.1)$$

Keterangan:

m = jumlah karakter yang sama persis

|s1| = panjang string 1

|s2| = panjang String 2

t = jumlah transposisi

- Jaro-Winkler *distance* (dw) akan didapatkan dengan persamaan 2.2.

$$dw = dj + (lp (1 - dj)) \quad (2.2)$$

dimana :

dj = Jaro *distance* untuk strings s1 dan string s2

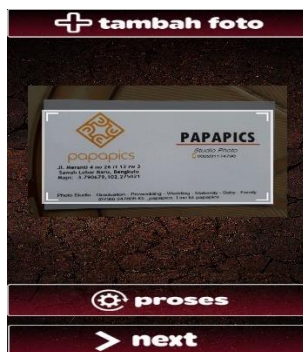
l = panjang prefiks umum di awal string nilai maksimalnya 4 karakter (panjang karakter yang sama sebelum ditemukan ketidaksamaan max 4)
 p = konstanta scaling factor. Nilai standar untuk konstanta ini menurut Winkler adalah $p = 0,1$.

2. Pembahasan

A. Aplikasi Pemindai Kartu Nama

Aplikasi ini terdiri dari tiga halaman, yaitu halaman utama, halaman proses, dan halaman simpan. Halaman yang akan melakukan segala proses utama dalam aplikasi tersebut, yaitu OCR dan pengelompokan data. Halaman proses dapat dilihat pada Gambar 1. Di dalam halaman proses terdapat tiga tombol, yaitu tambah, proses dan *next*. Tombol tambah dapat dipilih pengguna untuk menambahkan foto yang ingin diproses dan diambil datanya. Saat memilih tombol tambah pengguna akan diminta untuk memilih sumber foto yang ingin diproses, yaitu melalui kamera atau *gallery smartphone* pengguna. Setelah memilih foto kartu nama yang ingin diproses, foto tersebut kemudian akan ditampilkan di bagian tengah halaman. Guna memproses foto tersebut, pengguna dapat memilih tombol proses. Pertama-tama aplikasi akan melakukan OCR terhadap foto tersebut dan mengambil bagian teksnya saja. Setelah teks didapatkan, lalu hasil OCR tersebut akan dibersihkan dengan membuang tanda baca yang tidak diinginkan.

Setelah teks hasil OCR dibersihkan, selanjutnya data tersebut akan dikelompokkan menjadi nama, alamat, nomor ponsel, nomor telepon dan email menggunakan Jaro-Winkler *Distance*. Hasil dari pengelompokan tersebut akan disimpan di dalam *database*. Semua proses ini akan dilakukan ketika pengguna memilih tombol proses. Di bagian bawah aplikasi terdapat tombol *next* yang akan membawa pengguna ke halaman berikutnya, yaitu halaman simpan.



Gambar 1. Halaman Proses Aplikasi

B. Pengujian Black Box Sistem

Metode ini berhasil melakukan pengecekan terhadap fungsi yang tidak benar atau hilang masalah antarmuka, kesalahan dalam struktur data atau akses database eksternal, dan kesalahan kinerja. Pada pengujian ini telah didapatkan hasil 100% kebenaran sistem, mulai dari tidak ada kesalahan antar muka, sistem dan lain sebagainya.

C. Pengujian *Optical Character Recognition*

Data uji coba yang digunakan sebanyak 50 data. Pengujian dilakukan sebanyak 5 kali. Hasil dari pengujian OCR terhadap kartu nama dapat dilihat pada Tabel 1 untuk kartu nama Gambar 2.



Gambar 2. Contoh Data Kartu Nama

Hasil ekstraksi menunjukkan *library* Tesseract sudah dapat mengekstraks teks dari citra kartu nama. Namun, hasil menunjukkan terdapat karakter-karakter yang tidak terdapat di kartu nama juga ikut terkestraksi. Banyaknya huruf-huruf atau karakter yang tidak terdapat di kartu nama tapi juga ikut diekstraksi disebabkan oleh adanya *figure* atau grafik di dalam kartu nama tersebut yang dapat menghasilkan teks atau karakter kacau pada bagian tersebut [8].

Terlihat hasil yang dihasilkan menjadi satu kolom dari kartu nama yang sebelumnya terdiri dari dua kolom. Hal ini karena *library* yang digunakan, yaitu Tesseract tidak mempunyai *analysis layout* untuk kolom, sehingga tidak dapat mendeteksi gambar yang multikolom [8].

Pada proses ekstraksi di Tabel 1 masih terdapat beberapa karakter yang berbeda antara data asli dengan data hasil ekstraksi. Perbedaan ini dikarenakan Tesseract gagal mengenali karakter yang dimaksudkan. Kegagalan ini dapat disebabkan oleh banyaknya *noise* yang terdapat di dalam gambar, sehingga proses analisis *connected component* memberikan hasil yang salah. Analisis *connected component* merupakan proses awal dari Tesseract Tess-Two sehingga jika proses ini mengalami kegagalan, maka akan mempengaruhi proses berikutnya [6].

Percobaan yang dilakukan selama lima kali terhadap masing-masing kartu nama dapat menghasilkan data hasil ekstraksi yang berbeda setiap kali proses ekstraksi dilakukan. Tapi, hal ini tidak menjamin semakin lama hasilnya semakin bagus karena hasil terbaik tidak pada

percobaan terakhir, tapi dapat terjadi pada percobaan kedua atau ketiga. Hal ini karena Tesseract dapat mengenali teks dengan baik pada percobaan kedua atau ketiga. Setelah itu hasil dari proses Tesseract dapat sama dengan proses sebelumnya atau menjadi lebih buruk. Ini dikarenakan analisis tata layout Tesseract yang buruk [9].

Tabel 1. Data Hasil Ekstraksi

No	Data Asli Pada Kartu Nama	Hasil Percobaan 1	Hasil Percobaan 2
1.	HOTEL SANTIKA BENGKULU Hospitality from the Heart Jl. Raya Jati No. 45, Sawah Lebar, Bengkulu 38228, INDONESIA Ph. (62-736) 25858, Fax: (62-736) 25854 E-mail : bengkulu@santika.com www.santika.com	Hotel Santlka B E N G K U L U Hospitality from the Heart Jt. Raya Jan No. 45, Sawah Lebar. Bengkulu 38228, INDONESIA Ph. (62—736) 25858. Fax (62736) 25854 Email bengkulu@sannka.com www.santika.com	Hotel Santlka B E N G K U L U Hospitality from the Heart Jt. Raya jam No. 45, Sawah Lebar. Bengkulu 38228, INDONESIA Ph. (62—736) 25858. Fax (62—736) 25854 Email bengkulu@san tlka.com www.santika.com

Hasil Percobaan 3	Hasil Percobaan 4	Hasil Percobaan 5
Hotel Santlka B E N G K U L U Hospitality from the Heart Jt. Raya Jan No. 45, Sawah Lebar. Bengkulu 38228, INDONESIA Ph. (62—736) 25858. Fax (62736) 25854 Email bengkulu@sannka.com www.santika.com	Hotel Santlka B E N G K U L U Hospitality from the Heart Jf. Raya Jatl No. 45, Sawah Lebar. Bengkulu 38228, INDONESIA Ph. (62736) 25858. Fax (62736) 25854 Emaii bengkulu@sannka.com www.santika.com	Hotel Santlka B E N G K U L U Hospitality from the Heart Jf. Raya Jatl No. 45, Sawah Lebar. Bengkulu 38228, INDONESIA Ph. (62736) 25858. Fax (62736) 25854 Emaii bengkulu@sannka.com www.santika.com

D. Pengujian Jaro Winkler

Pengelompokan dilakukan dengan menghitung nilai Jaro-Winkler Distance setiap baris hasil ekstraksi dengan data pembandingan yang terdapat di dalam

database. Contoh perhitungannya dapat dilihat pada Tabel 2.

Tabel 2. Contoh perhitungan Jaro Distance

No	s1	s2	m	T	dj
1	62736 = 5	+62736 = 6	5	0	0,94444444 44
2	62736 = 5	0736 = 4	3	0	0,78333333 33
3	62736 = 5	08 = 2	0	0	~
4	62736 = 5	+628 = 4	2	0	0,63333333 33
5	62736 = 5	Kementeria n = 11	0	0	~

Nilai string pertama didapat dari data hasil ekstraksi kartu nama. String pertama ini akan dibandingkan dengan nilai-nilai pembandingan yang akan menjadi string 2. Nilai pembandingan yang dibandingkan sebanyak empat nilai pembandingan untuk ponsel, lima nilai pembandingan untuk alamat, 18 nilai pembandingan untuk perusahaan dan 28 nilai pembandingan untuk nomor telepon berdasarkan ciri-ciri yang telah disebutkan pada bagian poin 4.A.

Tabel 3. Contoh perhitungan Jaro-Winkler Distance

No	Dj	I	p	dw
1	0,9444444444	0	0,1	0,9444444444
2	0,7833333333	0	0,1	0,7833333333
3	~	0	0,1	~
4	0,6333333333	0	0,1	0,6333333333
5	~	0	0,1	~

Dilihat dari contoh perhitungan pada Tabel 3 dapat dilihat bahwa nilai Jaro-Winkler Distance terbesar adalah nilai untuk string pembandingan '+62736' dengan nilai 0,9444444444. Jika nilai Jaro-Winkler Distance lebih besar dari 0.9 untuk string pembandin tertentu, maka data yang mengandung string pertama akan masuk ke kelompok data string pembandingan tersebut, karena semakin mendekati nilai 1 maka semakin mirip kedua string tersebut [7]. Pada contoh perhitungan pada Tabel 2 dan 3 nilai Jaro-Winkler Distance terbesar merupakan milik string pembandingan '+62736' yang masuk ke ciri-ciri data telepon, sehingga data yang mengandung string pertama '62736' akan masuk ke data telepon.

No	Field	Data Asli	Hasil 1	Hasil 2	Hasil 3	Hasil 4	Hasil 5
1	Nama	Hotel Santika	-	-	-	-	-
2	Instansi	Hotel Santika	Hotel Santika	Hotel Santika	Hotel Santika	Hotel Santika	Hotel Santika
3	Jabatan	-	-	-	-	-	-
4	Nomor Ponsel	-	-	-	-	-	-
5	E-Mail	bengkulu@antika.com	bengkulu@antika.com	bengkulu@antika.com	bengkulu@antika.com	bengkulu@antika.com	bengkulu@antika.com

Berdasarkan data hasil pengelompokan pada tabel di atas Jaro-Winkler sudah dapat mengelompokan data kartu nama berupa nama, nomor ponsel, telepon 2/fax, email, alamat pertama, alamat kedua, jabatan dan instansi yang terdapat di dalam kartu nama. Namun, kesalahan masih ditemui pada bagian pengelompokan untuk kategori instansi dan jabatan. Kesalahan pengelompokan pada *field* instansi karena beberapa kartu nama menggunakan logo sebagai tanda yang menunjukkan instansi, sedangkan aplikasi ini tidak dapat mengenali logo instansi tersebut jika tidak ada tulisan secara horizontal mengenai instansi tersebut. Hal ini karena Jaro-Winkler *Distance* merupakan algoritma pencocokan string yang hanya akan menganalisis data jenis string [2]. Beberapa kartu nama yang memakai

logo instansi, seperti BNN (Badan Narkotika Nasional), BPS (Badan Pusat Statistik), Radio Papeja, Polda, dan Radio B-One.

Berdasarkan hasil ujicoba yang dilakukan selama lima kali proses ekstraksi dan pengelompokan didapatkan nilai akurasi seperti pada Tabel 4.

Table 4. Data Akurasi

No	Perco baan	Nilai Akurasi (dalam %)						
		N	J	I	NP	NT	A	E
1.	1	76	74	56	82	86	78	86
2.	2	76	74	56	82	86	78	90
3.	3	76	74	56	82	86	78	90
4.	4	76	74	56	82	86	78	88
5.	5	76	74	56	82	86	78	86
Rata – rata		76	74	56	82	86	78	88

Keterangan :

N = Nama

J = Jabatan

I = Instansi

NP = Nomor Ponsel

NT = Nomor Telepon

A = Alamat

E = E-mail

Nilai akurasi merupakan persentase data yang benar berbanding dengan keseluruhan data sehingga rumus menghitung akurasi seperti pada rumus 5.1.

$$akurasi = \frac{data\ yang\ benar}{keseluruhan\ data} \times 100\ % \tag{5.1}$$

3. Kesimpulan

Berdasarkan hasil dan pembahasan yang telah dijabarkan sebelumnya, maka kesimpulan yang dapat diambil adalah sebagai berikut:

1. Aplikasi konversi teks telah dapat melakukan konversi teks kartu nama dengan melakukan ekstraksi teks dari foto kartu nama menggunakan *Optical Character Recognition* (OCR) dengan *library Tesseract Tess-Two* berbasis mobile android.
2. Aplikasi ekstraksi teks telah dapat melakukan pengelompokan data hasil ekstraksi nama, jabatan, instansi, nomor ponsel, alamat pertama, alamat kedua, nomor telepon pertama, nomor telepon kedua atau fax dan e-mail dengan menggunakan algoritma Jaro-Winkler *Distance* dengan menggunakan *library Apache Commons Lang* yang keseluruhan data yang telah dikelompokan dapat disimpan di dalam *database smartphone*.

3. Ujicoba 50 data sebanyak lima kali proses ekstraksi dan pengelompokan menghasilkan nilai akurasi rata-rata untuk data nama sebesar 76%, data jabatan sebesar 74%, data instansi sebesar 56%, data nomor ponsel sebesar 82%, data nomor telepon sebesar 86%, data alamat sebesar 78%, dan data email sebesar 88%.

Berdasarkan analisa dan perancangan sistem, implementasi, dan pengujian sistem, maka saran yang dapat diberikan untuk pengembangan penelitian selanjutnya adalah sebagai berikut:

1. Pada penelitian berikutnya dapat menggunakan *library* OCR versi yang berbeda agar dapat mengidentifikasi teks berdasarkan kolomnya.
2. Pada penelitian ini *database* ciri-ciri instansi untuk mengenali instansi masih terbatas sehingga masih terdapat beberapa instansi yang belum dikenali, pada penelitian berikutnya peneliti dapat merancang antarmuka pengguna yang dapat memungkinkan pengguna untuk memasukan database untuk instansi.
3. Pada penelitian ini beberapa nama jalan yang gagal dikenali 100% saat proses ekstraksi membuat data alamat seringkali tidak terbaca, pada penelitian berikutnya peneliti dapat menambah database nama-nama jalan yang ada di kota Bengkulu.
4. Peneliti dapat menambah data training dan metode pengenalan logo-logo instansi pemerintah atau swasta agar pengenalan instansi dapat lebih mudah dikenali saat proses ekstraksi dan pengelompokan data.
5. Pada penelitian ini, peneliti tidak membuat proses *preprocessing* terhadap citra uji kartu nama dan hanya menggunakan *tools tesseract* yang terdapat di dalam *library* Tesseract Tess-Two sehingga ketika dilakukan proses *recognition* untuk mengekstraksi teks, beberapa teks masih tidak dikenali, terutama bagi kartu nama yang mengandung banyak gambar. Pada penelitian berikutnya disarankan dapat dilakukan *preprocessing* terlebih dahulu, seperti *thresholding*, *skewing*, *rotation*, atau proses lainnya untuk memperbaiki kualitas citra uji dengan menggunakan *tools leptonica* yang juga terdapat di dalam *library* Tesseract Tess-Two atau menggunakan *tools* lainnya.

Daftar Pustaka

- [1] R. Mithe, S. Indalkar and N. Divekar, "Optical Character Recognition," *International Journal of Recent Technology and Engineering (IJRTE)*, pp. 72-75, 2013.
- [2] A. Kurniawati, S. Puspitodjati and S. Rahman, "Implementasi Algoritma Jaro-Winkler Distance untuk Membandingkan Kesamaan Dokumen Berbahasa Indonesia," Universitas Gunadarma,

Depok, 2012.

- [3] D. Ratliff, "Using Business Cards to Market Yourself While Job-Hunting," 1999. [Online]. Available: <http://www accuratewriting.com/businesscard.shtml>.
- [4] H.-C. Jeske, W. Lederer, I. Lorenz, C. Kolbitsch, J. Margreiter, J. Kinzl and A. Benzer, "The Impact of Business Cards on Physician Recognition After General Anesthesia," *ECONOMICS AND HEALTH SYSTEMS RESEARCH*, pp. 1262-1264, 2001.
- [5] I. D. P. Manurung, A. Hidayatno and B. Setiyono, "Pengenalan Teks Cetak Pada Citra Teks Biner," Universitas Diponegoro, Semarang, 2011.
- [6] R. Smith, "An Overview of the Tesseract OCR Engine," in *Proc 9th Int. Conf. on Document Analysis and Recognition*, 2007.
- [7] X. D. Coster, C. D. Groote, A. Destine and P. Deville, "Mahalanobis distance, Jaro-Winkler distance and nDollar in UsiGesture," Université catholique de Louvain, Louvain-la-Neuve, 2011.
- [8] A. Kay, "Tesseract: an Open-Source Optical Character Recognition Engine," *Linux Journal*, 2007.
- [9] M. Heliński, M. Kmieciak and T. Parkoła, "Report on the Comparison of Tesseract and ABBYY FineReader OCR Engines," National Library of the Netherlands, Den Haag, 2011.