

RANCANG BANGUN APLIKASI PENGARSIPAN DOKUMEN MENGGUNAKAN OPTICAL CHARACTER RECOGNITION PADA BADAN KEPEGAWAIAN DAN PENGEMBANGAN SUMBER DAYA MANUSIA KABUPATEN KAPUAS BERBASIS VISUAL

Ahmad Ramadhani¹, Ahmad Shalludin^{2)}**

*¹⁾ Mahasiswa Teknik Informatika STMIK Indonesia Banjarmasin ²⁾Dosen STMIK Indonesia Banjarmasin
Jl Pangeran Hidayatullah, Banua Anyar, Banjarmasin 70000
Email : ahmad@stmik.id¹⁾, roselani1976@gmail.com²⁾*

ABSTRAK

Aplikasi ini dibuat untuk mempermudah pengarsipan dokumen pada Badan Kepegawaian dan Pengembangan Sumber Daya Manusia (BKPSDM) Kabupaten Kapuas, dimana selama ini pengarsipan dokumen selama ini masih menggunakan cara manual, sehingga dirasa zaman teknologi yang serba cepat, pengarsipan manual dirasa lambat dalam pencarian dokumen, penyimpanan, dan kendala daya tahan media kertas yang memiliki umur yang singkat dalam menyimpan data pengarsipan dokumen ini.

Aplikasi ini dibuat menggunakan Visual Studio 2012, dimana bahasa pengembangannya sendiri menggunakan bahasa C#, Menggunakan Library EmguCV x86 2.4.0.1717 untuk Optical Character Recognition (OCR), Tesseract (Software) untuk algoritma dari OCR, Serta PDFRasterizer untuk melakukan konversi file berformat pdf, menjadi file berformat png. Ada beberapa fasilitas yang disediakan dalam program ini berupa penginputan data dokumen masuk, penginputan data dokumen keluar, pendisposisian dokumen, pengeditan, penghapusan, laporan data dokumen masuk, laporan data dokumen keluar, laporan data user, laporan data disposisi, laporan data jabatan, dan laporan jenis dokumen.

Dengan adanya aplikasi ini diharapkan dapat membantu dalam pengarsipan dokumen secara digital di Badan Kepegawaian dan Pengembangan Sumber Daya Manusia (BKPSDM) Kabupaten Kapuas, sehingga lebih efisien dalam melakukan penginputan data, pencarian data, serta meminimalisir kesalahan dan kerusakan pengarsipan dokumen.

Kata kunci: Aplikasi, Microsoft Studio 2012, C#, Optical Character Recognition (OCR), EmguCV, Tesseract (Software), Pengarsipan Dokumen.

ABSTRACT

This application was created to facilitate document filing at the Kapuas District Personnel and Human Resources Development Agency (BKPSDM), where so far document archiving has been using the manual method, so that it is felt that the era of fast paced technology, manual filing is slow in document search, storage, and paper media durability constraints that have a short life in storing this document archiving data.

This application was created using Visual Studio 2012, where the development language itself uses the C # language, Using the EmguCV x86 2.4.0.1717 Library for Optical Character Recognition (OCR), Tesseract (Software) for algorithms from OCR, and PDFRasterizer for converting pdf files into png file format. There are several facilities provided in this program in the form of inputting incoming document data, inputting outgoing document data, document disposing, editing, deleting, incoming document data reports, outgoing document data reports, user data reports, disposition data reports, position data reports, and reports document type.

With this application, it is hoped that it can help in digitally filing documents at the Kapuas District Personnel and Human Resources Development Agency (BKPSDM), so that it is more efficient in data inputting, data search, and minimizing errors and damage to document archiving.

Keywords: Application, Microsoft Studio 2012, C #, Optical Character Recognition (OCR), EmguCV, Tesseract (Software), Document Archiving.

1. Pendahuluan

Arsip mempunyai peran penting dalam perusahaan atau sebuah instansi, yaitu sebagai sumber informasi dan pusat ingatan, begitu juga dengan kasus yang ada pada Badan Kepegawaian dan Pengembangan Sumber Daya Manusia (BKPSDM) Kabupaten Kapuas. Arsip bagi Badan Kepegawaian dan Pengembangan Sumber Daya Manusia (BKPSDM) Kabupaten Kapuas diantaranya berisi informasi yang berguna dalam pengambilan keputusan, sebagai alat bukti bila terjadi masalah, sebagai alat pertanggungjawaban, manajemen serta dapat dijadikan alat transparansi dalam instansi. Pentingnya arsip bahkan dibuktikan dengan pemberlakuan beberapa peraturan perundangan yang mengatur tentang kearsipan.

Arsip akan bermanfaat secara optimal apabila dikelola dengan tertib dan teratur, namun sebaliknya apabila arsip tidak dikelola dengan secara baik akan menimbulkan masalah. Apalagi dengan perkembangan zaman yang semakin pesat diharapkan pengelolaan arsip di Badan Kepegawaian dan Pengembangan Sumber Daya Manusia (BKPSDM) Kabupaten Kapuas dapat beralih dari sistem arsip secara manual ke sistem yang terotomasi, karena dengan sistem manual, lama kelamaan dapat menimbulkan serangkaian kendala seperti kesulitan penemuan kembali secara cepat dan tepat karena sistem penyimpanan yang kurang sistematis, dan bertambahnya secara terus-menerus arsip tanpa penyusutan sehingga terjadi penumpukan. Kendala lainnya yaitu Tata kerja dan Fasilitas pengelolaan arsip yang tidak memadai dan tidak mengikuti perkembangan teknologi kearsipan modern. Maka dari itulah setiap perusahaan atau instansi dituntut untuk melakukan otomasi arsip.

Dengan teknologi yang semakin maju juga memberi dampak yang positif bagi dunia kearsipan, Pengelolaan arsip secara modern atau tata kearsipan otomatis merupakan sistem kearsipan yang menggunakan sarana pengolahan data secara elektronik. Potensi teknologi yang serba canggih telah memberikan peluang untuk melakukan kegiatan otomasi arsip, penggunaan media otomasi arsip bukan saja menjamin efisiensi, tetapi juga mampu mengurangi atau mengembangkan kebutuhan duplikasi apabila hal itu diperlukan, Pengiriman, pemrosesan, penyimpanan dan penemuan kembali informasi dapat dilakukan dengan cepat.

Ditambah lagi dengan pemanfaatan teknologi Optical Character Recognition (OCR), yang dapat

membantu melakukan input arsip secara otomatis, diharapkan dapat membuat kegiatan ini lebih efektif dan efisien dengan adanya implementasi Optical Character Recognition (OCR) yang terintegrasi didalamnya.

Optical Character Recognition (OCR) sendiri merupakan salah satu aplikasi pengolahan citra yang berfokus pada pengolahan input citra menjadi output karakter yang dibaca infomasinya dan dimasukkan ke dalam sistem komputer. Hasil dari citra inilah yang akan dimasukkan kedalam database aplikasi pengarsipan dokumen, sehingga diharapkan arsip yang awalnya di scanning masih dalam bentuk Image dapat diambil informasi di dalamnya untuk mengisi beberapa hal penting dalam dokumen secara otomatis seperti, nomer dokumen, tanggal dokumen, perihal dokumen, dan atribut lainnya.

Implementasi Optical Character Recognition (OCR) pada aplikasi pengarsipan dokumen pada Badan Kepegawaian dan Pengembangan Sumber Daya Manusia (BKPSDM) Kabupaten Kapuas sendiri diharapkan dapat membantu pekerjaan pengarsipan yang masih manual, sehingga ketika beberapa arsip seperti dokumen surat masuk, dapat melakukan input lebih efisien.

2. METODE PENELITIAN

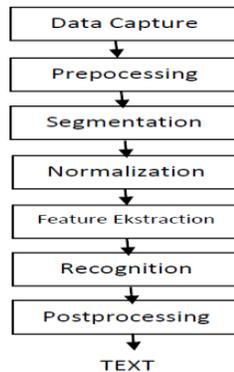
2.1 Optical Optical Character Recognition.

Menurut Priyatma, dkk (2004) pengenalan karakter dengan menggunakan alat optik (*Optical Character Recognition*) adalah sebuah sistem komputer yang dapat membaca karakter termasuk (huruf), baik yang berasal dari sebuah alat pencetak (printer atau mesin ketik) maupun yang berasal dari tulisan tangan. Adanya sistem pengenal huruf ini akan meningkatkan fleksibilitas ataupun kemampuan dan kecerdasan sistem komputer.

Adanya sistem pengenal huruf yang cerdas akan sangat membantu usaha besar-besaran yang saat ini dilakukan banyak pihak yakni usaha digitalisasi informasi dan pengetahuan, misalnya dalam pembuatan koleksi pustaka digital, koleksi sastra kuno digital, dll. Dengan adanya pengenal huruf juga akan memudahkan penanganan pekerjaan yang memakai input tulisan seperti penyortiran surat di kantor pos, pemasukan data buku di perpustakaan, dll. Selain itu adanya sistem pengenal huruf otomatis maka user dapat lebih leluasa memasukkan data karena user tidak harus memakai papan ketik tetapi bisa menggunakan

pena elektronik untuk menulis sebagaimana user menulis di kertas. (Priyatma dkk, 2004)

Dalam pengenalan pola otomatis, sistem pengenalan pola mencoba mengenali apakah citra masukan yang diterima cocok dengan salah satu citra yang telah ditentukan. Sistem ini misalnya dipakai untuk mendeteksi sidik jari, tanda tangan,



bahkan wajah seseorang. Menurut Sukmawan

(2008) OCR dapat dipandang sebagai bagian dari pengenalan otomatis yang lebih luas yakni pengenalan pola otomatis (*Automatic Pattern Recognition*). Ada banyak pendekatan yang dapat dipakai untuk mengembangkan pembuatan pengenalan pola otomatis antara lain memakai pendekatan numerik, statistik, sintaktik, neural dan, aturan produksi (*rule-based*). Secara umum metode-metode tersebut dapat digolongkan menjadi dua kelompok metode yakni metode berbasis statistik dan metode berbasis struktur.

Dalam metode yang berbasis statistik, setiap pola ditransformasi ke dalam vektor yang memakai ukuran dan karakteristik tertentu. Karakteristik ini seringkali lebih bersifat statistik misalnya distribusi pixel ataupun jarak piksel. Sedang dalam metode yang berbasis struktur, setiap pola yang diproses dinyatakan sebagai gabungan beberapa struktur elementer. Pengenalan selanjutnya dilakukan dengan mencocokkan komposisi struktur elementer dengan struktur yang sudah disimpan memakai aturan tertentu misalnya memakai pendekatan teori bahasa formal dan automata. (Sukmawan, 2008)

Ilustrasi Proses OCR

Prinsip kerja dari aplikasi OCR menurut Manik (2010) adalah sebagai berikut:

Gambar 2.1 Ilustrasi Proses OCR

1. Mengambil objek berupa teks menggunakan kamera sehingga didapat sebuah file citra. File citra tersebut diproses menggunakan perangkat lunak aplikasi

pengenalan teks, di melakukan proses pengenalan terhadap karakter-karakter yang ada pada file citra tersebut.

2. Keluaran dari perangkat lunak aplikasi pengenalan teks ini berupa file teks yang berisi karakter-karakter yang telah dikenali dan siap untuk diolah lebih lanjut.

Tingkat keberhasilan dari perangkat lunak pada aplikasi pengenalan teks ini sangat bergantung dari sejumlah faktor berikut:

1. Kualitas gambar teks yang ada pada dokumen yang dibaca serta tingkat kompleksitasnya (ukuran, format, teks, warna, latar belakang).
2. Kualitas perangkat lunak aplikasi pengenalan teks itu sendiri.
3. Kualitas alat optik yang dipakai (kamera).

2.2 Tesseract

Google Tesseract merupakan mesin *Optical Character Recognition*(OCR) untuk berbagai sistem operasi yang awalnya dikembangkan oleh *Hewlet Packard* pada tahun 1985 hingga 1995(Smith, 2007). Tesseract dimasukkan dalam test akurasi OCR tahunan UNLV keempat sebagai “OCR Lab HP” tetapi kode-nya telah banyak berubah sejak saat itu, termasuk konversi ke *Unicode* dan *retraining*(latihan ulang) (Smith, 2007). Setelah tahun 2006, Tesseract dilepas oleh HP untuk digunakan secara bebas. Sejak saat itu, tesseract dikembangkan secara luas oleh Google dan dirilis di bawah lisensi Apache 2.0 serta dapat diakses di <http://code.google.com/p/tesseract-ocr> (Smith, 2007). Tesseract mungkin menjadi mesin OCR paling akurat saat ini. Tesseract dikombinasikan dengan *Leptonica Image*

Processing Library yang dapat membaca berbagai format gambar dan mengkonversikannya ke teks di lebih dari 60 bahasa. Dalam konferensi ICDAR, Smith mengungkapkan Tesseract mengkombinasikan kedua algoritma *matrix matching* dan ekstraksi fitur. Tesseract hanya memerlukan sedikit data pelatihan dan menggunakan kedua *static classifier* dan *adaptive classifier* sehingga memungkinkan mesin dapat berlatih sendiri pada dokumen yang dianalisis(Smith, 2007).

2.3 Emgu CV

OpenCV (*Open Source Computer Vision*) adalah sebuah library fungsi pemrograman *real time* untuk *computer vision*. Emgu CV adalah wrapper .Net untuk OpenCV. Dengan EmguCV, fungsi-fungsi dalam OpenCV bisa dipanggil melalui bahasa pemrograman yang compatible dengan .NET seperti C#, VB, dan VC++. Selain itu, Emgu CV juga cross

platform sehingga dapat di-compile lewat Mono dan dijalankan di atas sistem operasi Linux atau Mac OS.

Dari pengertian di atas telah diberikan deskripsi dari kedua *open source* tersebut. OpenCV merupakan *library* yang cukup terkenal di dunia *Computer Vision*. *Computer Vision* adalah salah satu bidang di teknologi informasi yang fokus pada pemrosesan *images* atau gambar yang diperoleh dari dunia nyata untuk diekstrak dan diinterpretasikan informasinya. Untuk mempermudah *developer* dalam mengembangkan aplikasi yang menggunakan teknologi *computer vision*, digunakanlah *library* seperti VXL, Camellia, OpenCV, dan lainnya.

Maka dari itu EmguCV berperan untuk menjembatani C# dan OpenCV. EmguCV adalah wrapper .Net untuk OpenCV. Keuntungan menggunakan EmguCV yang paling utama adalah *library* ini sepenuhnya ditulis dengan bahasa pemrograman C# yang mana lebih aman karena pembuatan *object* atau pun *reference* di-manage oleh *garbage collector*.

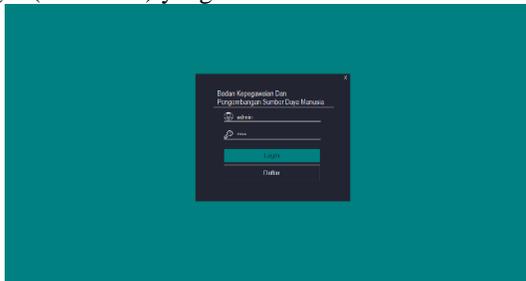
Ada dua konsep penting yang perlu diketahui terlebih dahulu sebelum menggunakan EmguCV. Pertama mengenai *layer* pada EmguCV. EmguCV terdiri dari 2 *layer*, yaitu *basic layer* dan *second layer*. *Basic layer* mengandung fungsi, struktur, dan enumerasi yang secara langsung merefleksikan apa yang ada di OpenCV. Dengan adanya *layer* inilah kita bisa memanggil fungsi-fungsi pada OpenCV dengan bahasa pemrograman C#. Sedangkan *second layer* mengandung kelas-kelas yang memanfaatkan keunggulan teknologi .NET.

3. HASIL DAN PEMBAHASAN

Berikut tampilan aplikasi pengarsipan dokumen menggunakan *optical character recognition* pada badan kepegawaian dan pengembangan sumber daya manusia kabupaten Kapuas:

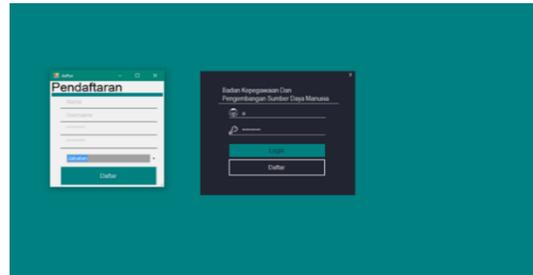
1. Login

Form login adalah form yang digunakan untuk keamanan program, berikut tampilan dari form login (Gambar 1) yang ada :



Gambar 1. Login

2. Daftar



Daftar Adalah Form yang digunakan untuk mendaftar pada calon user, berikut tampilan dari form daftar(Gambar 2):

Gambar 2. Daftar

3. Form Menu Utama

Form Menu Utama Adalah form yang digunakan untuk menuju semua form yang ada didalam aplikasi, berikut tampilan dari form menu utama(Gambar 3):



Gambar 3. Form Menu Utama

4. Form Beranda

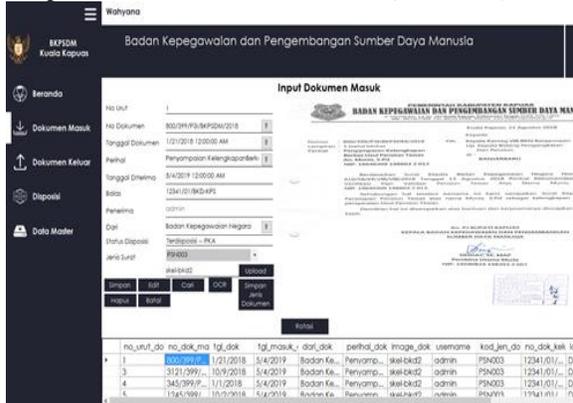
Form Beranda adalah form yang berfungsi untuk menampilkan informasi di halaman depan, sehingga informasi itu dapat di lihat semua user, berikut tampilan dari form beranda(Gambar 4):



Gambar 4. Form Beranda

5. Form Dokumen Masuk

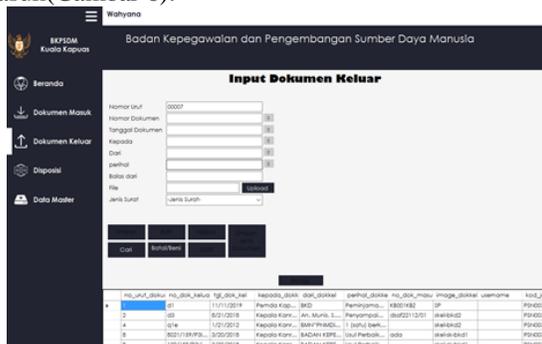
Form Dokumen Masuk adalah form yang berfungsi untuk melakukan input dokumen yang masuk ke Badan Kepegawaian dan Pengembangan Sumber Daya Manusia Kabupaten Kuala Kapuas, berikut tampilan dari form dokumen masuk (Gambar 5):



Gambar 5. Form Dokumen Masuk

6. Form Dokumen Keluar

Form Dokumen Keluar adalah form yang berfungsi untuk melakukan input dokumen yang akan keluar dari Badan Kepegawaian dan Pengembangan Sumber Daya Manusia Kabupaten Kuala Kapuas, berikut tampilan dari form dokumen masuk (Gambar 6):



Gambar 6. Form Dokumen Keluar

7. Form Input Kordinat

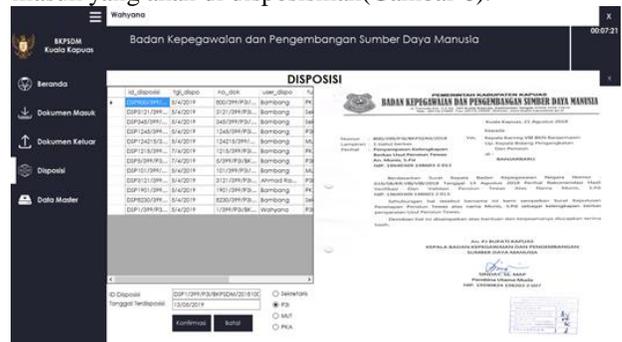
Form input kordinat berfungsi untuk melakukan input kordinat di bagian mana proses OCR di butuhkan, proses ini berfungsi agar proses OCR yang di jalankan tidak melakukan konversi citra ke huruf pada seluruh citra, melainkan hanya pada bagian yang di tentukan oleh user saja yang akan di lakukan proses OCR, berikut tampilan dari form input kordinat (Gambar 7):



Gambar 7. Form Input Kordinat

8. Form Disposisi

Form disposisi berfungsi untuk memberikan informasi disposisi dokumen, serta untuk memberikan konfirmasi penempatan dokumen masuk yang akan di disposisikan (Gambar 8):



Gambar 8. Form Disposisi

9. Form Data Master

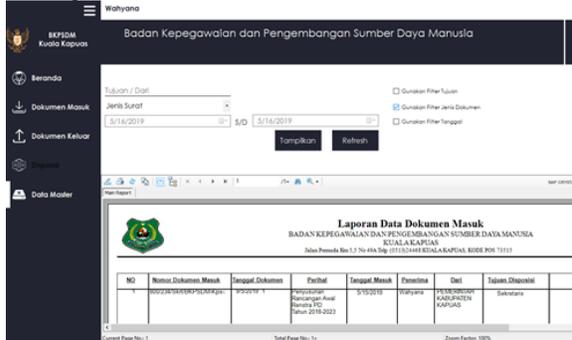
Form data master merupakan form yang berfungsi untuk menampilkan laporan dan juga dapat mengedit data laporan tersebut sesuai jabatan user yang aktif (Gambar 9):



Gambar 9. Form Data Master.

10. Form Data dan Laporan Dokumen Masuk

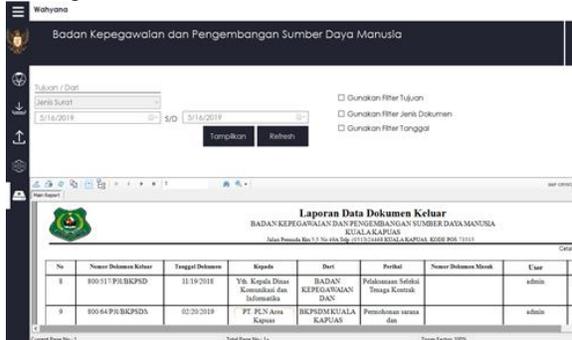
Form Data dan Laporan Dokumen Masuk merupakan form yang membuat rekapitulasi atau laporan dari dokumen masuk, form ini terdiri dari 3 checkbox yang memiliki fungsinya masing-masing(Gambar 10).



Gambar 10. Form Data dan Laporan Dokumen Masuk

11. Form Data dan Laporan Dokumen Keluar

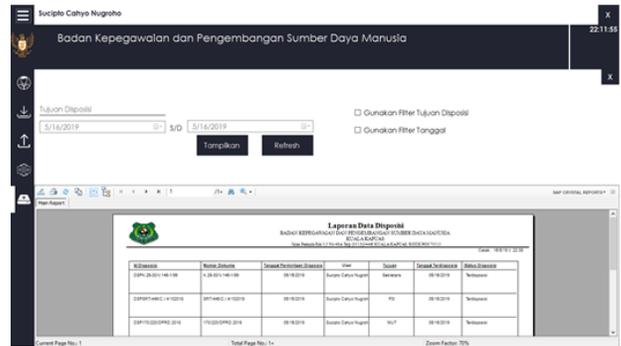
Form Data dan Laporan Dokumen Keluar merupakan form yang membuat rekapitulasi atau laporan dari dokumen keluar, form ini terdiri dari 3 checkbox yang memiliki fungsinya masing-masing(Gambar 11).



Gambar 11. Form Data dan Laporan Dokumen Keluar

12. Form Data dan Laporan Disposisi

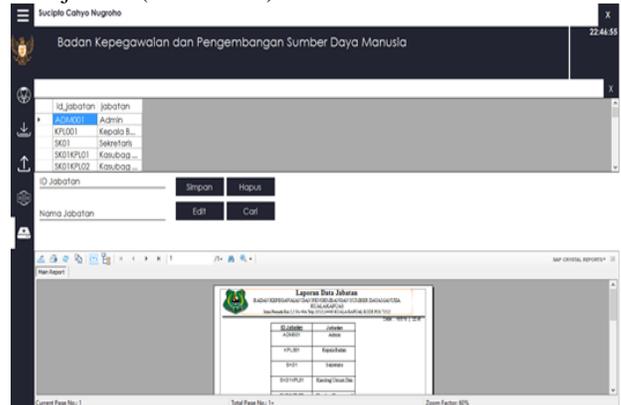
Form Data dan Laporan Disposisi merupakan form yang berfungsi untuk menampilkan data dan laporan dari data disposisi, pada form ini terdapat 2 buah checkbox yang memiliki fungsi masing-masing(Gambar 12).



Gambar 12. Form Data dan Laporan Disposisi

13. Form Data dan Laporan Jabatan

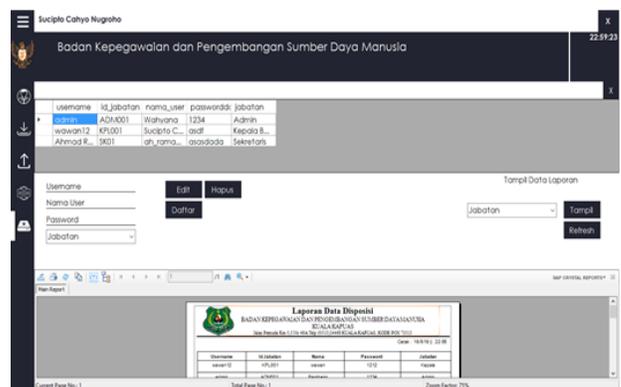
Form data dan laporan jabatan berfungsi untuk melakukan penambahan, penyuntingan, menghapus data jabatan(Gambar 13).



Gambar 13. Form Data dan Laporan Jabatan

14. Form Data dan Laporan User

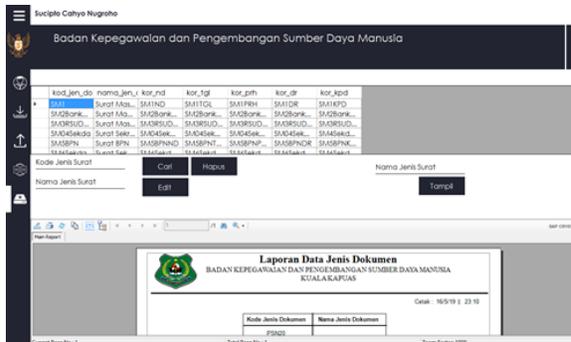
Form data dan laporan data user berfungsi untuk melakukan penambahan, menghapus, dan penyuntingan data user serta pengguna yang aktif juga dapat melakukan pencetakan laporan sesuai jabatan sebagai parameter penyaringan data(Gambar 14).



Gambar 14. Form Data dan Laporan User

15. Form Data dan Laporan Jenis Dokumen

Form data dan laporan jenis surat berfungsi untuk melakukan penambahan, menghapus, dan penyuntingan data user serta pengguna yang aktif juga dapat melakukan pencetakan laporan sesuai nama jenis dokumen sebagai parameter penyaringan data (Gambar 15).



Gambar 15. Form Data dan Laporan Jenis Dokumen.

3. Kesimpulan

Berdasarkan penelitian yang penulis kerjakan dan berdasarkan dari rumusan masalah yang ada yaitu, bagaimana merancang aplikasi pengarsipan dokumen menggunakan Optical Character Recognition (OCR) pada Badan Kepegawaian dan Pengembangan Sumber Daya Manusia Kabupaten Kapuas Berbasis Visual. Maka dapat di ambil beberapa kesimpulan diantaranya :

1. Perancangan dan pembuatan aplikasi pengarsipan dokumen menggunakan Optical Character Recognition (OCR) pada Badan Kepegawaian dan Pengembangan Sumber Daya Manusia Kabupaten Kapuas dibuat untuk mempermudah petugas pemerintah kantor Badan Kepegawaian dan Pengembangan Sumber Daya Manusia Kabupaten Kapuas untuk melakukan pengarsipan dokumen agar lebih efisien dalam proses input data, dan pembuatan laporan.
2. Dengan adanya aplikasi pengarsipan dokumen menggunakan Optical Character Recognition (OCR), ini diharapkan dapat meminimalisir waktu memasukkan data dokumen, agar proses input data lebih cepat.
3. Kemiringan dan kualitas citra sebagai data dalam proses Optical Character Recognition (OCR) ini sangat berpengaruh terhadap hasil pengenalan karakter tersebut.
4. Tulisan tangan mampu berpengaruh dan mengubah hasil pengenalan karakter, sehingga besar kemungkinan citra karakter yang di proses akan

tidak tepat dan hasil tidak sesuai dengan karakter pada citra.

5. Kerapatan dan tebal karakter pada citra berpengaruh pada hasil pengenalan karakter yang di lakukan oleh sistem, dimana sistem akan tidak tepat jika suatu kalimat yang pindah sangat rapat dan tebal.
6. Ketajaman citra dan kejelasan karakter dapat mempengaruhi hasil ketepatan dari pengenalan suatu karakter pada citra, semakin tajam citra dan kejelasan karakter maka akan semakin tepat sistem mengenali suatu karakter yang sedang di pindai.
7. Berhubungan dengan point ke 6 (enam), citra yang berasal dari hasil fotokopi keberhasilan pengenalan suatu karakter tidak terlalu akurat akibat ada bagian karakter yang terhapus mengakibatkan salahnya proses pengenalan karakter pada citra.

Daftar Pustaka

- [1] Permata Sari, Indah. 2013. Pengertian Rancang Bangun Dan Konsep Sistem Informasi. [Online] Available at: <http://indahpermata6.blogspot.com/2013/06/pengertian-rancang-bangun-dan-konsep.html>. Diakses: 4 Agustus 2019.
- [2] Setiawan, Samhis. 2019. Dokumen” Pengertian Menurut Para Ahli & (Jenis – Sifat – Literatur – Kepentingan – Fungsi). [Online] Available at: <https://www.gurupendidikan.co.id/dokumen-pengertian-menurut-para-ahli-jenis-sifat-literatur-kepentingan-fungsi/>. Diakses: 4 Agustus 2019.
- [3] Pranata, Yuliana, 2017. Optical Character Recognition Pengertian Menurut Priyatma, dkk (2004). [Online] Available at: <http://docplayer.info/47360592-Bab-ii-tinjauan-pustaka-optical-character-recognition-pengertian-menurut-priyatma-dkk-2004-pengenalan-karakter-dengan.html>. Di Akses: 4 Agustus 2019.
- [4] S Kusumaningratri, 2018. Tinjauan Pustaka Dan Dasar Teori. [Online] Available at: http://eprints.akakom.ac.id/6526/3/3_145410101_BAB_II.pdf. Di Akses: 4 Agustus 2019.
- [5] Jamal, 2015. Pengertian Emgu CV. [Online] Available at: <https://makalah28.blogspot.com/2015/11/pengertian-emgu-cv.html>. Di Akses: 4 Agustus 2019.
- [6] Arni, Ulti Desi, 2018. Pengertian Visual Studio Dan Visual Studio Toolbox. [Online] Available at: <https://garudacyber.co.id/artikel/1202-pengertian-visual-studio-dan-visual-studio-toolbox>. Di Akses: 4 Agustus 2019.