

PENERAPAN TEKNIK OCR (*OPTICAL CHARACTER RECOGNITION*) PADA APLIKASI TERJEMAHAN KITAB *SAFINAH AN-NAJA* MENGUNAKAN *READIRIS*

R. Sandhika Galih A.¹⁾, Erik²⁾, Muhammad Lukman Hakim³⁾

^{1,2,3)}Jurusan Teknik Informatika Universitas Pasundan Bandung

Jl. DR. Setiabudhi No. 193, Bandung

sandhikagalih@unpas.ac.id¹⁾, eikic1@unpas.ac.id¹⁾, muhhammad.lukman@mail.unpas.ac.id¹⁾,

Abstrak

Safinah An-Naja adalah kitab fiqih dasar yang banyak dikaji di pesantren-pesantren salafiyah. Di pondok pesantren Hidayatul Islamiyah Karawang, kitab ini sudah mulai di-digitalisasi dengan cara menyalin kitab tersebut ke dalam word-processor dan disimpan kedalam arsip. Kesulitannya adalah proses penginputan yang lama dan penulisan terjemahannya disimpan pada arsip terpisah. Pada penelitian ini akan dibuat sebuah aplikasi berbasis web yang dapat mengelola ayat-ayat beserta terjemahan kitab *safinah an-naja* sehingga pengguna dapat membaca dan mencari ayat dan terjemahan dengan mudah. Untuk mempermudah proses penginputan ayat ke dalam aplikasi digunakan software *ReadIRIS* yang mengimplementasikan metode OCR (*optical character recognition*) sehingga pengelola dapat dengan mudah memindai berkas agar langsung dapat dimasukkan ke dalam aplikasi. Aplikasi *ReadIRIS* ini nantinya juga diuji agar diketahui kondisi ideal pada saat memindai berkas kitab agar hasilnya optimal.

Kata Kunci : *Optical Character Recognition, Aplikasi Web, ReadIRIS*

1. PENDAHULUAN

Perkembangan teknologi saat ini tidak hanya dipakai untuk bidang-bidang tertentu, hampir segala bidang telah memanfaatkan kemajuan teknologi informasi. Salah satu penggunaan teknologi informasi adalah aplikasi terjemahan yang telah banyak diterapkan dalam bidang keagamaan. Contohnya adalah aplikasi terjemahan Al-Quran dan aplikasi tafsir Al-Quran. Namun dari sekian banyak aplikasi terjemahan tersebut, aplikasi yang membahas tentang terjemahan kitab fiqih khususnya *safinah an-naja* jumlahnya masih sangat terbatas. Kitab fiqih *safinah an-naja* merupakan salah satu kitab fiqih madzhab syafi'i yang banyak dipelajari di pesantren – pesantren di Indonesia. Kitab fiqih yang ada sekarang sebagian besar ditulis atau dicetak dalam bahasa arab. Digitalisasi kitab – kitab fiqih berbahasa arab sangat diperlukan, tidak hanya untuk dokumentasi namun juga untuk memudahkan dalam pengembangan aplikasi keagamaan yang menyangkut terjemahan buku – buku yang berbahasa arab.

Pondok Pesantren Hidayatul Islamiyah Karawang merupakan lembaga pendidikan islam yang mencoba mengadaptasi teknologi informasi dalam proses belajar mengajar. Dengan jumlah alumni yang tersebar ke beberapa provinsi di luar jawa barat, maka Pondok Pesantren Hidayatul Islamiyah membutuhkan satu cara agar tetap bisa berhubungan dengan alumninya, salah satu kebutuhan yang terasa saat ini yaitu penyebaran materi ajar tentang kitab – kitab kuning, salah satunya yaitu kitab fiqih *safinah an-naja*. Para alumni terkadang harus kembali ke Pondok Pesantren hanya untuk meminta salinan terjemahan dan penjelasan kitab fiqih *safinah an-naja*. Dengan kemajuan teknologi informasi dan konsep digitalisasi kitab fiqih, diharapkan alumni madrasah akan bisa lebih mudah untuk mendapatkan dokumen kitab fiqih dalam bentuk digital. Di Pondok Pesantren Hidayatul Islamiyah ini, pengelola sudah mulai melakukan digitalisasi kitab dengan cara yang masih manual yaitu dengan menyalin isi kitab ke dalam aplikasi *word processor*, namun cara yang dilakukan tersebut memakan waktu yang lama dan terjemahan kitabnya masih tersimpan di arsip yang terpisah sehingga sulit untuk mencari dan menghubungkan ayat yang satu dengan terjemahannya.

Penelitian ini fokus pada membuat aplikasi berbasis web agar dapat diakses oleh semua pihak yang membutuhkan, dan berfungsi untuk mengelola ayat dan terjemahan dari kitab *safinah an-naja*. Pada penelitian ini juga akan membahas mengenai penggunaan alat bantu berupa *software* yang bernama *ReadIRIS* yang berfungsi untuk memindai berkas sehingga dapat dibaca secara digital. Penggunaan *software* ini dibutuhkan agar proses penginputan data-data digital dari ayat dapat lebih mudah dibandingkan dengan menginputkan secara manual. *Software ReadIRIS* ini kemudian akan diuji terhadap beberapa parameter uji, seperti jenis berkas yang dipindai, tipe *scanner* yang digunakan untuk memindai dan resolusi dari hasil pemindaian. Pengujian ini dilakukan untuk mendapatkan keadaan ideal pada saat melakukan pemindaian berkas, sehingga hasil pemindaian bisa lebih optimal dan waktu yang dibutuhkan untuk proses digitalisasi menjadi lebih singkat.

Metode yang digunakan dalam penelitian ini adalah metode *waterfall* sebagai metode dalam pembangunan perangkat lunak berbasis web dan eksperimen sebagai metode untuk pengujian dari *software ReadIRIS*.

2. TINJAUAN PUSTAKA

2.1 *Safinah An-naja*

Safinah an-Naja (bahtera keselamatan) adalah salah satu kitab fiqh dasar yang banyak dikaji di pesantren-pesantren *salafiyah*. Kajian kitab ini meliputi tauhid dan fiqh. Dalam bidang tauhid mengikuti aqidah islamiyah *ahlus sunnah wal jama'ah*, yaitu aqidah yang akan membawa keselamatan di dunia dan akhirat. Sedangkan dalam bidang fiqh mengikuti madzhab Syafi'i, madzhab mayoritas muslim Indonesia.

Pokok bahasan kitab ini meliputi:

- a. *Bagian Aqidah*, menguraikan rukun iman, rukun islam dan makna kalimat thoyyibah لا اله الا الله
- b. *Bagian Fiqih*, menguraikan tatacara bersuci, sholat, penyelenggaraan jenazah, zakat dan puasa.

Ada banyak faktor yang menjadikan kitab ini banyak dipelajari, diantaranya:

1. Kitab ini mencakup pokok-pokok agama secara terpadu, lengkap dan utuh, dimulai dengan Bab dasar-dasar syari'at, kemudian Bab bersuci, Bab shalat, Bab zakat, Bab puasa dan Bab haji yang ditambahkan oleh para ulama lainnya.
2. Kitab ini disajikan dengan bahasa yang mudah, susunan yang ringan dan redaksi yang mudah untuk dipahami serta dihafal. Seseorang yang serius dan memiliki ke-mauan tinggi akan mampu menghafalkan seluruh isinya hanya dalam masa dua atau tiga bulan atau mungkin lebih cepat.
3. Kitab ini ditulis oleh seorang ulama yang terkemuka dalam berbagai bidang ilmu keagamaan, terutama fiqh dan tasawwuf. Yang sangat menarik, orang lebih menegenal nama kitabnya dari pada nama penulisnya. Hal yang demikian itu mungkin saja berkat keikhlasan dan ketulusan penulis.
4. Kitab ini menjadi acuan para ulama dalam memberikan pengetahuan dasar agama bagi para pemula. Di Hadramaut Yaman, Madinah, Mekkah dan kota lainnya, para ulama menggunakan kitab ini sebagai silibus bagi kelas pengajian mereka.
5. Kitab ini membicarakan hal-hal yang selalu menjadi kebutuhan seorang muslim dalam kehidupan sehari-hari, sehingga semua orang merasa perlu untuk mempelajarinya.

Kitab *Safinah* telah tersebar secara luas di kalangan pesantren yang mempelajari ilmu fiqh terutama yang menganut Madzhab Imam Syafi'i ra. Kitab ini dikenali di berbagai Negara, baik Arab maupun Ajam seperti Yaman, Mekkah, Madinah, Jeddah, Somalia, Ethiopia, Tanzania, Kenya, Zanjibar, dan di berbagai belahan negara-negara Afrika. Namun demikian perhatian yang paling besar terhadap kitab ini telah diberikan oleh para ulama dan pecinta ilmu, yang hidup di Alam Melayu termasuk Indonesia, Malaysia, Singapura, dan negara-negara lainnya.

2.2 *OCR (Optical Character Recognition)*

Optical Character Recognition (OCR) dalam arti luas merupakan cabang dari *artificial intelligence* dan *computer vision*. *OCR* adalah sebuah aplikasi komputer yang digunakan untuk mengidentifikasi citra huruf maupun angka untuk dikonversi ke dalam bentuk file tulisan. Sistem pengenalan huruf ini dapat meningkatkan fleksibilitas atau kemampuan dan kecerdasan komputer. Sistem pengenalan huruf yang cerdas sangat membantu usaha digitalisasi informasi dan pengetahuan, misalnya dalam pembuatan koleksi pustaka digital, koleksi sastra kuno, dan lain – lain.

Secara umum proses OCR dapat dilihat pada gambar berikut :



Gambar 1. Skema proses OCR.

- a. File Input
File input berupa file citra digital dengan format *.bmp atau *.jpg atau *.TIFF
- b. Preprocessing
Preprocessing merupakan suatu proses untuk menghilangkan bagian – bagian yang tidak diperlukan pada gambar input untuk proses selanjutnya.
- c. Segmentasi
Segmentasi adalah proses memisahkan area pengamatan (*region*) pada tiap karakter yang dideteksi.
- d. Normalisasi
Normalisasi adalah proses merubah dimensi region tiap karakter dan ketebalan karakter.
- e. Ekstraksi ciri
Ekstraksi ciri adalah proses untuk mengambil ciri – ciri tertentu dari karakter yang diamati.
- f. Recognition
Recognition merupakan proses untuk mengenali karakter yang diamati dengan cara membandingkan ciri – ciri karakter yang ada di dalam *database*.

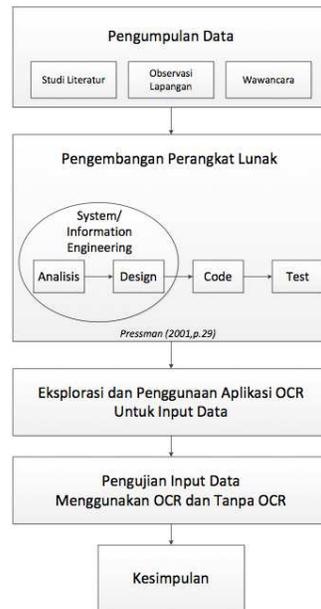
2.3 ReadIRIS

ReadIRIS adalah sebuah *software OCR (Optical Character Recognition)* yang dirancang untuk meningkatkan kemampuan piranti *scanner* agar dapat mengenali dan mengkonversi berkas, gambar, atau dokumen PDF agar dapat diubah dan dikelola kedalam file digital (*word, excel, PDF, HTML, dll*). *ReadIRIS* memungkinkan kita untuk membuat file PDF mudah dicari dan diindex agar kompatibel untuk semua komputer. Berikut adalah beberapa fitur yang dimiliki oleh *ReadIRIS* berdasarkan situsny di <http://www.irislink.com> :

- a. *Convert paper, images or PDF files into editable digital text*
- b. *Upload your documents to the Cloud*
- c. *Import any paper document, image or PDF file*
- d. *Import documents from any scanner, digital camera or mobile phone*
- e. *Export to Word, Excel, PDF and many more*
- f. *Convert your documents into searchable PDF files*
- g. *Easy file viewing on the Web*
- h. *Recognize more than 130 languages (including Arabic Text)*
- i. *New user-friendly interface*
- j. *Powerful picture enhancement tools*
- k. *Multi-page document recognition*

3. METODE PENELITIAN

Metode yang digunakan pada penelitian ini dapat dilihat pada **Gambar 2** di bawah ini:



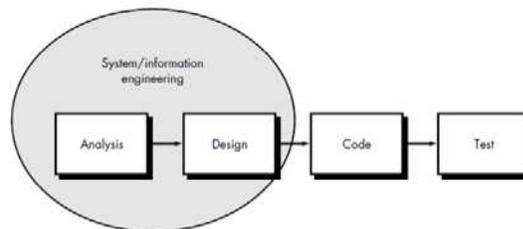
Gambar 2. Metode Penelitian

3.1 Pengumpulan Data

Langkah-langkah yang dilakukan dalam tahap pengumpulan data adalah sebagai berikut:

1. Studi Literatur
Tahap ini dilakukan untuk mempelajari dan memahami teori yang berkaitan dengan topic penelitian. Literatur yang digunakan adalah buku, paper juga artikel – artikel yang tersedia di internet yang menjelaskan mengenai konsep dan software OCR.
2. Observasi Lapangan
Tahap ini dilakukan untuk mempelajari kebutuhan user mengenai terjemahan kitab fiqh safinah an-naja. Observasi dilakukan di Pondok Pesantren Hidayatul Islamiyah Karawang.
3. Wawancara
Tahapan ini dilakukan untuk mendapatkan kebutuhan – kebutuhan mengenai kitab fiqh safinah an-naja juga dan terjemahannya. Wawancara dilakukan kepada pengurus Pondok Perantren Hidayatul Islamiyah Karawang.

3.2 Pengembangan Perangkat Lunak



Gambar 3. Diagram metode pengembangan perangkat lunak menurut Pressman

Berdasarkan Gambar 3 di atas, langkah-langkah yang dilakukan dalam pengembangan perangkat lunak dideskripsikan sebagai berikut:

1. *Analysis*

Langkah ini merupakan analisis terhadap kebutuhan *software*, dan tahap untuk mengadakan pengumpulan data yang berkaitan dengan proses digitalisasi kitab fiqih safinah an - naja. Pada tahapan ini akan dihasilkan *user requirement* atau bisa dikatakan sebagai data yang berhubungan dengan keinginan user dalam pembuatan aplikasi terjemahan kitab fiqih *safinah an-naja*.

2. *Design*

Proses *design* ini akan menterjemahkan syarat kebutuhan ke sebuah perancangan *software* yang dapat diperkirakan sebelum dibuat *coding*. Proses ini fokus pada rancangan struktur data, representasi antarmuka, dan detail (algoritma) yang dibutuhkan dalam pembuatan aplikasi terjemahan kitab fiqih *safinah an-naja*.

3. *Code*

Code merupakan proses membuat kode. Tahapan ini merupakan tahapan secara nyata dalam pengerjaan suatu aplikasi. Pada tahap ini akan diterjemahkan transaksi yang sudah dirancang sebelumnya dengan menggunakan bahasa pemrograman. Pada penelitian ini, bahasa pemrograman yang digunakan adalah php dan database yang digunakan adalah mysql.

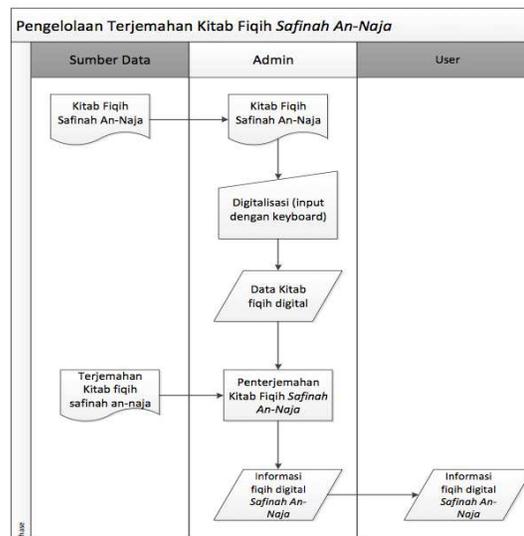
4. *Test*

Setelah tahapan *coding* selesai maka akan dilaklkan *testing* terhadap aplikasi pengelolaan terjemahan kitab fiqih *safinah an-naja* yang telah dibuat. Tujuan *testing* adalah menemukan kesalahan – kesalahan pada aplikasi tersebut dan kemudian bisa diperbaiki.

4. HASIL DAN PEMBAHASAN

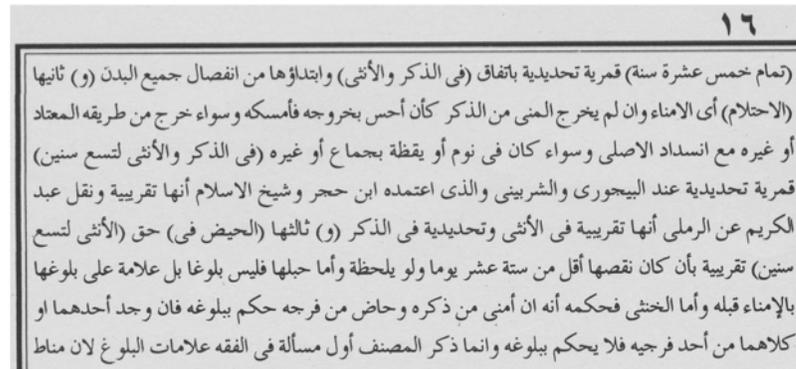
4.1 Analisis sistem yang sedang berjalan

Analisis sistem yang sedang berjalan memberikan gambaran mengenai sistem yang sedang berjalan saat ini. Melalui analisis terhadap sistem yang sedang berjalan memungkinkan untuk menemukan kelebihan ataupun kekurangan dalam sistem tersebut sehingga akan mempermudah dalam pengembangan sistem baru yang lebih baik, yang diharapkan mampu mengatasi segala kelemahan ataupun kekurangan dalam sistem yang lama. Aliran kegiatan dari sistem yang sekarang sedang berjalan di Pondok Pesantren Hidayatul Islamiyah Karawang dapat dilihat pada *flowmap* di bawah ini:



Gambar 4. Flowmap sistem yang sedang berjalan

Gambar berikut ini adalah contoh dokumen kitab fiqih *safinah an-naja* yang akan diinputkan pada aplikasi *word processor* secara manual.



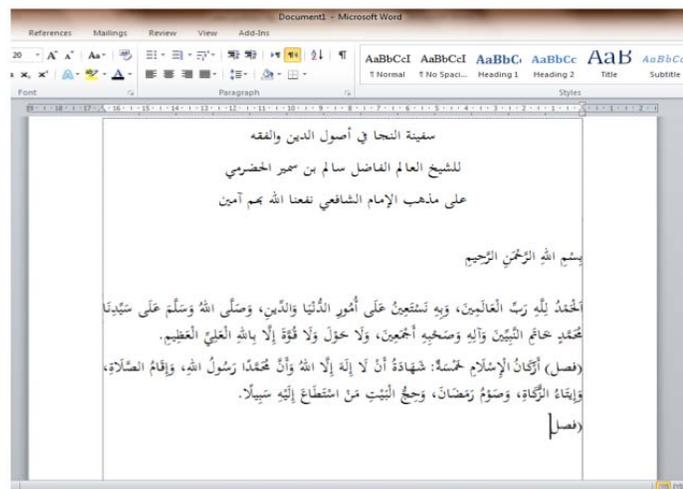
Gambar 5. Contoh potongan kitab fiqih *safinah an-naja* yang akan digunakan

Pada Flowmap diatas terdapat kekurangan dari sistem yang sedang berjalan saat ini, yaitu pada tahap pengetikan kitab fiqih *safinah an-naja* yang dilakukan secara manual. Pengetikan dilakukan dengan menggunakan bantuan *software* pengolah kata Microsoft word, dan menggunakan *keyboard* khusus yang memiliki huruf arab agar lebih memudahkan pengelola untuk menginputkan ayat. **Gambar 6** berikut ini menunjukkan keyboard dengan font arab untuk proses pengetikan manual.



Gambar 6. Keyboard Arabic

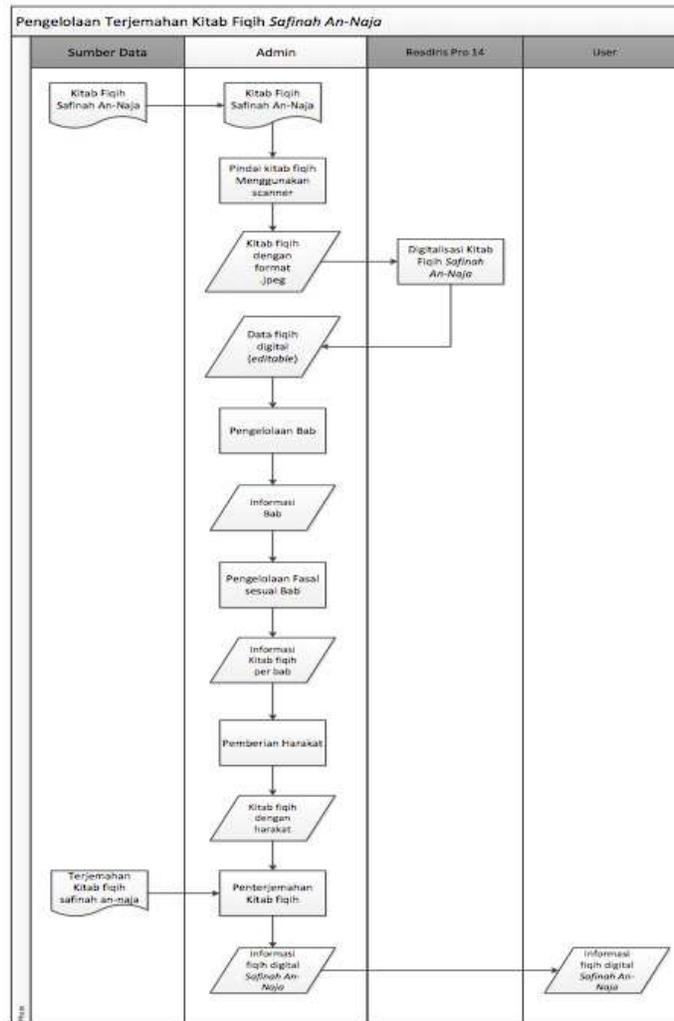
Berikut ini adalah hasil penginputan / proses penyalinan huruf arab ke dalam aplikasi *word processing*, aplikasi yang digunakan adalah Microsoft Word.



Gambar 7. Pengetikan menggunakan Microsoft Word

4.2 Analisis sistem yang diusulkan

Kitab *saḥīḥ an-najā* yang pada awalnya merupakan lembaran kertas, kemudian dilakukan proses digitalisasi menggunakan aplikasi OCR (*Optical Character Recognition*) *ReadIRIS* sehingga menjadi teks digital. Hasil dari digitalisasi kitab fiqh tersebut kemudian diolah didalam aplikasi pengelolaan terjemahan kitab fiqh *saḥīḥ an-najā* yang akan dibangun. Berikut adalah *flowmap* dari sistem yang diusulkan.



Gambar 8. Flowmap sistem yang diusulkan

Pada flow map diatas dapat dilihat perbedaan proses digitalisasi dengan yang ada pada sistem yang sedang berjalan. Pada sistem yang diusulkan, proses digitalisasi dikerjakan dengan bantuan *software Optical Character Recognition (OCR) ReadIRIS*. Dokumen terjemahan yang didapatkan sudah berupa file digital.

4.3 Antarmuka Aplikasi

Antarmuka aplikasi merupakan tampilan dari aplikasi pengelolaan terjemahan kitab fiqh *saḥīḥ an-najā*. Aplikasi ini memiliki halaman untuk *user* (pengguna) dan halaman untuk *admin* (*pengelola*).

1. Halaman user

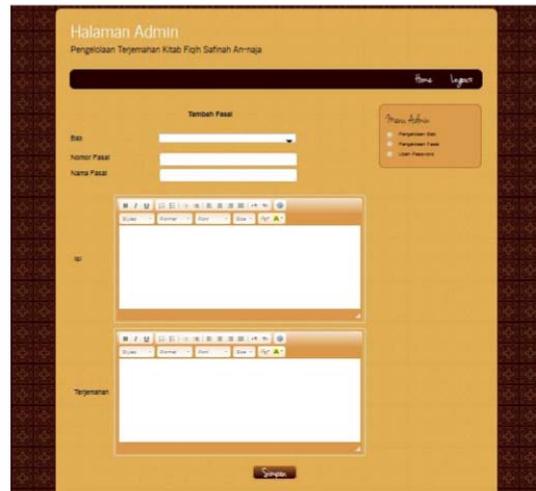
Halaman *user* adalah tampilan depan dari aplikasi pengelolaan terjemahan kitab fiqh *saḥīḥ an-najā* untuk pengguna umum. Beberapa gambar dibawah ini menunjukkan tampilan dari halaman *user*.



Gambar 9. Halaman depan user

2. Halaman admin

Halaman *admin* adalah tampilan depan dari aplikasi pengelolaan terjemahan kitab fiqih *safinah an-naja* untuk *administrator*. Berikut potongan gambar halaman *admin*.



Gambar 10. Halaman administrator

4.4 Pengujian

Pada penelitian ini, *software OCR ReadIRIS* dipilih setelah dilakukan perbandingan dengan dua *software OCR* yang lain yaitu *Omnipage* dan *ABBYY FineReader*. Pengujian terhadap *ReadIRIS* dilakukan untuk mengetahui tingkat akurasi dari konversi huruf arab yang dilakukan oleh *ReadIRIS*, juga untuk mendapatkan teknik pemindaian yang optimal. Pengujian dilakukan dengan beberapa parameter, yaitu:

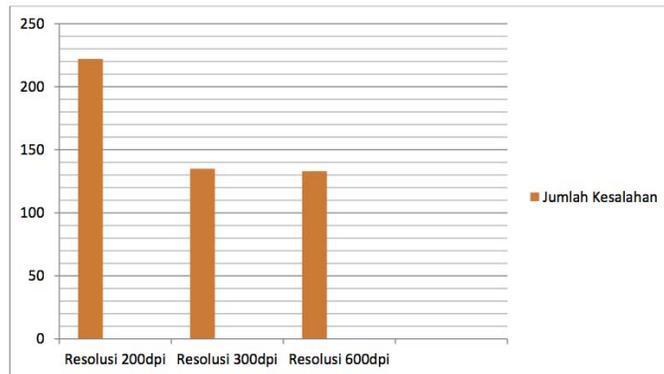
1. Pengujian terhadap jenis huruf arab

Jenis huruf arab yang dimaksud adalah huruf arab yang berharakat dan duruf arab yang tidak berharakat.

2. Pengujian hasil file pemindaian

Ada tiga macam pengujian yang akan dilakukan terhadap file hasil pemindaian. Pertama pengujian dengan membandingkan resolusi, yaitu resolusi dengan 200dpi, 300dpi dan 600dpi. Kedua, membandingkan dua jenis scanner yaitu tipe scanner Scanjet dengan tipe scanner all in one. Ketiga, akan dilakukan pengujian akurasi terhadap file yang dipotret dengan kamer SLR.

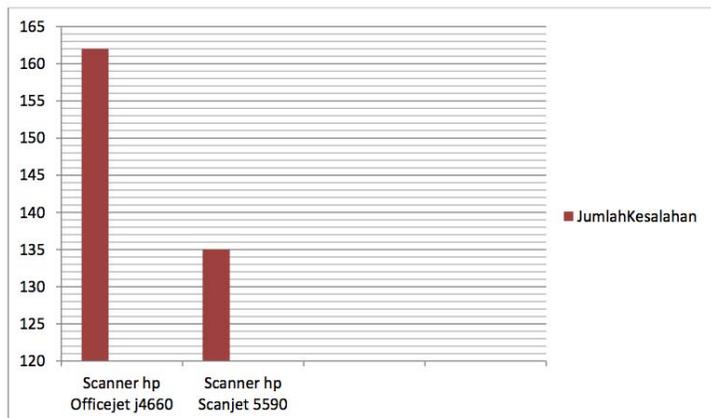
Pengujian *ReadIRIS* dilakukan dengan menggunakan dokumen kitab fiqih *safinah an-naja* tanpa harakat pada halaman ke-16 yang mengandung \pm 3000 huruf. Gambar 11 berikut ini menunjukkan grafik hasil pengujian akurasi *ReadIRIS* berdasarkan resolusi yang digunakan.



Gambar 11. Grafik hasil pengujian berdasarkan resolusi pemindaian

Berdasarkan hasil pengujian resolusi, maka resolusi yang tepat untuk digunakan dalam proses digitalisasi adalah resolusi 300dpi. Resolusi 600dpi tidak digunakan karena pemindaian dengan resolusi 600dpi akan menghasilkan ukuran file yang besar, sehingga akan membutuhkan tempat penyimpanan yang besar.

Selanjutnya dilakukan pengujian untuk menentukan tipe *scanner* yang paling tepat untuk digunakan pada saat memindai dokumen. Tipe *scanner* dibagi menjadi dua kategori yaitu *scanner* multifungsi (biasanya menyatu bersama *printer*) dan *scanner* biasa. Gambar 12 berikut ini menunjukkan hasil pengujiannya.



Gambar 12. Grafik hasil pengujian berdasarkan jenis *scanner*

Berdasarkan hasil pengujian tipe scanner, maka scanner yang tepat untuk digunakan dalam proses digitalisasi adalah scanner dengan tipe scanjet yang khusus untuk scanner.

Dari tahap pengujian *ReadIRIS* dapat disimpulkan beberapa hal berikut:

1. *ReadIRIS* hanya bisa melakukan konversi dokumen arab yang belum menggunakan harakat.
2. Pemindaian dokumen dengan menggunakan scanner akan lebih optimal jika resolusinya diatur menjadi 300 dpi.
3. *ReadIRIS* memiliki keterbatasan akurasi. Beberapa huruf yang sering salah konversi yaitu: ع ('ain), م (mim), س (sin), ش (sin), ال (lam alif), ك (kaf).
4. Penggunaan *scanner* berpengaruh terhadap akurasi, sebaiknya pemindaian dilakukan menggunakan scanner tipe *scanjet* yang khusus untuk pemindaian saja. Jangan menggunakan scanner *officejet all-in-one* (*scanner* multifungsi).

Penggunaan kamera SLR dalam media pemindaian tidak optimal, karena akan tergantung dari pencahayaan pada saat pemotretan dokumen juga teknik dari fotografer yang mengambil gambar.

5. KESIMPULAN

Dari hasil penelitian ini dapat disimpulkan beberapa hal, yaitu:

1. Proses digitalisasi kitab fiqh bisa dilakukan dengan menggunakan *software Optical Character Recognition (OCR)*.
2. Proses digitalisasi dengan menggunakan *software OCR* lebih cepat jika dibandingkan dengan proses digitalisasi dengan pengetikan manual.
3. *ReadIRIS* sebagai *software OCR* yang dipakai masih memiliki keterbatasan akurasi.
4. Terdapat beberapa huruf yang sering salah konversi oleh *ReadIRIS*, yaitu: ع ('ain), م (mim), س (sin), ش (syin), ل (lam alif), ك (kaf), ه (hamzah).
5. Untuk mendapatkan hasil yang optimal dalam proses digitalisasi, ada beberapa hal yang harus diperhatikan, yaitu:
 - a. Dokumen kitab fiqh yang digunakan harus dokumen yang belum menggunakan harakat (arab gundul).
 - b. Pada proses pemindaian menggunakan scanner, akan lebih optimal jika resolusinya menggunakan 300 dpi.
6. Scanner yang digunakan dianjurkan adalah scanner dengan tipe scanjet yang khusus untuk scan saja untuk mendapatkan hasil yang optimal.
7. Aplikasi pengelolaan terjemahan kitab fiqh *safinah an-naja* yang dibangun akan dapat membantu user (pengguna awam) untuk belajar kitab fiqh dasar.

6. UCAPAN TERIMAKASIH

Ucapan terimakasih disampaikan kepada Jurusan Teknik Informatika dan Fakultas Teknik Universitas Pasundan Bandung baik dalam bentuk dana, fasilitas dan peralatan yang telah banyak membantu bagi keberhasilan dan kelancaran kegiatan penelitian.

DAFTAR PUSTAKA

- Cheriet M., Kharma N., Liu C, Suen C.Y., "Character Recognition System A Guide for Student and Practioners", John Willey & Sons. Inc. 2006
- Image Recognition Integrated System (I.R.I.S.), 2012, "Introducing The New ReadIRIS 14", <http://www.irislink.com/c2-2115-189/Readiris-14--OCR-Software--Scan--Convert---Manage-your-Documents-.aspx> , diakses pada tanggal 23 Maret 2014.
- Jannah., Erliyah Nurul, "Rekonstruksi Dokumen Teks Arab yang Terdistorsi Gemetris Pada Saat Akuisisi Citra", Surabaya : Institut Teknologi Sepuluh November., 2010
- Mori S., Nishida H., Yamada H., "Optical Character Recognition", Willey Interscience. 1999
- Munir., Rinaldi, "Pengolahan Citra Digital", Informatika., 2007
- Solichin., Achmad, "Pemrograman web dengan PHP dan MYSQL", Informatika., 2009
- Syakur, Ahmad Bisyr, "The Pocket Fiqih", Salamadina, 2011.