



Pengembangan Aplikasi Arsip Digital Berbasis *OCR* Dan *Rule-Based System*

Development of a Digital Archiving Application Based on OCR and Rule-Based System

Eci Wulandari¹, Iwan Setiawan^{*2}, Riza Kartina³

¹²³Program Studi Informatika, Universitas Prabumulih, Prabumulih

*Korespodensi Email: iwanhen2@gmail.com

Abstract

The management of correspondence archives in higher education institutions continues to face various challenges, including dependence on physical document storage, limited archival space, and difficulties in information retrieval. These conditions highlight the need for an archival management system capable of supporting document digitalization while facilitating archive management and retrieval. This study develops a digital archive system based on Optical Character Recognition (OCR) and Rule-Based System for the rectorate unit of Universitas Prabumulih. System development was conducted using the Design Thinking approach, which consists of the empathize, define, ideate, prototype, and test stages to ensure that the proposed solution aligns with user needs. OCR technology was utilized to extract textual information from scanned documents, while the Rule-Based System was implemented to identify archive metadata based on predefined rules. The developed system provides a more structured archival management mechanism through document digitalization, metadata identification, and content-based document retrieval. The integration of these technologies supports the management of incoming and outgoing correspondence and provides an alternative for transforming archival practices from physical media into digital archives that are more accessible, manageable, and maintainable.

Keywords: *Digital Archiving, Optical Character Recognition, Rule-Based System*

Abstrak

Pengelolaan arsip surat di lingkungan perguruan tinggi masih menghadapi berbagai tantangan, mulai dari proses penyimpanan yang bergantung pada dokumen fisik, keterbatasan ruang arsip, hingga kesulitan dalam proses temu kembali informasi. Kondisi tersebut mendorong kebutuhan akan sistem pengelolaan arsip yang mampu mendukung digitalisasi dokumen sekaligus mempermudah proses pengelolaan dan pencarian arsip. Penelitian ini mengembangkan sistem arsip digital berbasis *Optical Character Recognition (OCR)* dan *Rule-Based System* pada unit rektorat Universitas Prabumulih. Pengembangan sistem dilakukan menggunakan pendekatan *Design Thinking* yang terdiri dari tahapan *empathize*, *define*, *ideate*, *prototype*, dan *test* untuk memastikan solusi yang dirancang sesuai dengan kebutuhan pengguna. Teknologi *OCR* dimanfaatkan untuk mengekstraksi informasi tekstual dari dokumen hasil pemindaian, sedangkan *Rule-Based System* diterapkan untuk mengidentifikasi metadata arsip berdasarkan aturan yang telah ditentukan. Implementasi sistem menghasilkan mekanisme pengelolaan arsip yang lebih terstruktur melalui proses digitalisasi, identifikasi metadata, dan pencarian dokumen berbasis isi arsip. Integrasi kedua teknologi tersebut mendukung pengelolaan surat masuk dan surat keluar serta memberikan alternatif transformasi pengarsipan dari media fisik menuju arsip digital yang lebih mudah diakses, dikelola, dan dipelihara.

Kata kunci: *Optical Character Recognition, Pengarsipan Digital, Rule-Based System*

1. Pendahuluan

Pengelolaan arsip merupakan bagian penting dalam mendukung kelancaran administrasi dan penyediaan informasi pada suatu Institusi [1]. Di lingkungan perguruan tinggi, arsip surat masuk dan surat keluar menjadi sumber informasi yang memiliki nilai administratif, hukum, dan historis sehingga memerlukan pengelolaan yang terstruktur. Namun, pengelolaan arsip yang masih bergantung pada dokumen fisik sering menimbulkan berbagai kendala, seperti kebutuhan ruang penyimpanan yang semakin besar, proses pencarian dokumen yang memerlukan waktu lebih lama, serta meningkatnya risiko kehilangan dan duplikasi arsip [2]. Kondisi tersebut juga ditemukan pada unit rektorat Universitas Prabumulih yang masih mengelola arsip administrasi secara konvensional dan terpisah antar unit kerja.

Selain permasalahan penyimpanan, pengelolaan arsip fisik juga menghadapi tantangan dalam menjaga keberlanjutan informasi. Berdasarkan kebijakan kearsipan, dokumen administrasi memiliki masa simpan tertentu sebelum dipindahkan ke gudang arsip atau dimusnahkan. Meskipun demikian, informasi yang terkandung di dalam arsip tersebut masih berpotensi diperlukan untuk berbagai keperluan administratif maupun pengambilan keputusan di masa mendatang. Oleh karena itu, transformasi pengelolaan arsip dari media fisik menuju media digital menjadi kebutuhan yang semakin penting untuk mendukung aksesibilitas, efisiensi, dan keberlanjutan pengelolaan informasi [3].

Salah satu teknologi yang banyak digunakan dalam proses digitalisasi arsip adalah *Optical Character Recognition (OCR)*. Teknologi ini memungkinkan sistem mengenali karakter teks dari dokumen hasil pemindaian dan mengubahnya menjadi data digital yang dapat diproses oleh komputer [4]. Pemanfaatan *OCR* pada sistem pengarsipan mampu mempercepat proses ekstraksi informasi serta mendukung pencarian dokumen berdasarkan isi arsip, sehingga pengelolaan dokumen tidak lagi bergantung pada pencatatan manual [5].

Selain kemampuan ekstraksi teks, sistem arsip digital juga memerlukan mekanisme untuk mengidentifikasi metadata dokumen agar proses pengelolaan arsip menjadi lebih efektif. Salah satu pendekatan yang digunakan

adalah *Rule-Based System*, yaitu metode yang menerapkan aturan tertentu dalam proses pengambilan keputusan. Dalam konteks pengarsipan, *Rule-Based System* dapat digunakan untuk mengidentifikasi karakteristik dokumen berdasarkan aturan yang telah ditentukan sebelumnya serta memberikan proses yang transparan dan mudah disesuaikan dengan kebutuhan organisasi [6]. Dalam sistem ini, aturan disusun berdasarkan kebutuhan pengarsipan Unit Rektorat untuk mengekstraksi informasi seperti asal surat, tujuan surat, nomor surat, tanggal surat, dan perihal. Sebagai contoh, jika surat masuk memiliki kata kunci seperti “FAKULTAS”, “UNIVERSITAS”, “DINAS”, atau “LEMBAGA” pada bagian awal dokumen, maka baris tersebut digunakan sebagai asal surat. Pada surat keluar, keberadaan penanda “Kepada”, “Yth.”, “Yang Terhormat”, atau “To” digunakan untuk menentukan tujuan surat. Selain itu, pola “Nomor” atau “No” digunakan untuk mengidentifikasi nomor surat, sedangkan penanda “Perihal” atau “Hal” digunakan untuk memperoleh informasi perihal. Aturan tersebut diterapkan dalam bentuk kondisi *IF-THEN* sehingga proses ekstraksi metadata dapat dilakukan secara transparan dan disesuaikan dengan pola dokumen yang digunakan pada Unit Rektorat.

Beberapa penelitian terdahulu telah membahas pengembangan sistem arsip digital, implementasi *OCR*, maupun penerapan *Rule-Based System* secara terpisah. Penelitian mengenai arsip digital menunjukkan bahwa digitalisasi dokumen mampu meningkatkan efisiensi penyimpanan dan mempercepat proses temu kembali informasi [7]. Penelitian lain menerapkan *OCR* untuk mendukung ekstraksi teks pada dokumen arsip sehingga proses pencarian menjadi lebih efektif [8].

Sementara itu, penelitian mengenai *Rule-Based System* menunjukkan bahwa pendekatan berbasis aturan mampu mendukung proses klasifikasi dokumen secara otomatis dan konsisten [9]. Meskipun demikian, penelitian yang mengintegrasikan *OCR* dan *Rule-Based System* dalam satu sistem arsip digital untuk mendukung pengelolaan surat masuk dan surat keluar di lingkungan perguruan tinggi masih relatif terbatas.

Berdasarkan permasalahan dan hasil penelitian terdahulu tersebut, penelitian ini

mengembangkan sistem arsip digital berbasis *Optical Character Recognition (OCR)* dan *Rule-Based System* pada unit rektorat Universitas Prabumulih. Pengembangan sistem dilakukan menggunakan pendekatan *Design Thinking* yang terdiri dari tahapan *empathize*, *define*, *ideate*, *prototype*, dan *test* untuk menghasilkan solusi yang sesuai dengan kebutuhan pengguna. Integrasi *OCR* dan *Rule-Based System* diharapkan mampu mendukung proses digitalisasi, klasifikasi, serta pencarian arsip secara lebih efektif sehingga dapat meningkatkan efisiensi pengelolaan dokumen dan mendukung transformasi digital di lingkungan perguruan tinggi.

2. Metode Penelitian

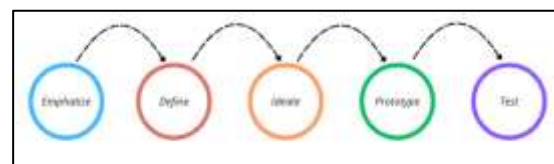
Penelitian ini menggunakan pendekatan deskriptif kualitatif yang berfokus pada pengembangan sistem arsip digital berbasis *Optical Character Recognition (OCR)* dan *Rule-Based System* pada Unit Rektorat Universitas Prabumulih. Pendekatan ini digunakan untuk memahami kebutuhan pengguna, mengidentifikasi permasalahan dalam pengelolaan arsip, serta menghasilkan solusi yang sesuai dengan kondisi nyata di lingkungan kerja [10].

Data dalam penelitian ini bersumber dari data primer dan data sekunder. Data primer diperoleh secara langsung melalui observasi dan wawancara dengan staf administrasi Unit Rektorat Universitas Prabumulih untuk memperoleh informasi mengenai proses pengelolaan dan penyimpanan arsip, alur administrasi, kebutuhan pengguna, proses pencarian dokumen, serta kendala yang dihadapi dalam pengelolaan arsip secara konvensional [11]. Sementara itu, data sekunder diperoleh dari berbagai sumber yang telah tersedia, seperti buku, jurnal, artikel ilmiah, penelitian terdahulu, dan dokumen resmi universitas yang relevan dengan penelitian [12]. Data sekunder digunakan sebagai referensi dan landasan teoritis dalam mendukung pengembangan sistem, khususnya terkait pengelolaan arsip digital, penerapan *OCR*, dan *Rule-Based System*.

Pengumpulan data dilakukan melalui observasi, wawancara, dokumentasi, dan studi

literatur. Observasi dilakukan dengan mengamati secara langsung proses pengelolaan arsip yang sedang berjalan pada Unit Rektorat Universitas Prabumulih [13]. Wawancara dilakukan dengan staf administrasi untuk memperoleh informasi mengenai kebutuhan sistem dan permasalahan yang ditemukan dalam proses pengarsipan. Dokumentasi dilakukan dengan mengumpulkan data dan contoh arsip yang digunakan sebagai objek penelitian. Dokumen yang digunakan berupa surat masuk dan surat keluar dalam format *PDF*, *JPG*, *JPEG*, dan *PNG*, termasuk dokumen hasil pemindaian (scan) dari surat fisik. Dokumen tersebut digunakan sebagai masukan pada proses *OCR* untuk menghasilkan teks yang selanjutnya diproses oleh *Rule-Based System*. Studi literatur dilakukan dengan menelaah buku, jurnal, dan penelitian terdahulu yang relevan dengan topik penelitian [14].

Metode pengembangan sistem yang digunakan adalah *Design Thinking* karena berorientasi pada kebutuhan pengguna dan penyelesaian masalah secara iteratif [15]. Tahapan *Design Thinking* terdiri atas *empathize*, *define*, *ideate*, *prototype*, dan *test* sebagaimana ditunjukkan pada Gambar 1.



Gambar 1. Metode *Design Thinking*

1. *Empathize*, dilakukan dengan mengidentifikasi kebutuhan pengguna melalui observasi dan wawancara terhadap staf administrasi rektorat.
2. *Define*, dilakukan dengan menganalisis dan merumuskan permasalahan yang ditemukan pada proses pengelolaan arsip.
3. *Ideate*, dilakukan dengan menyusun alternatif solusi berupa sistem arsip digital yang mengintegrasikan *OCR* dan *Rule-Based System*.

4. *Prototype*, dilakukan dengan merancang antarmuka dan alur kerja sistem sesuai kebutuhan pengguna.
5. *Test*, dilakukan melalui pengujian sistem menggunakan *White Box Testing* untuk memastikan logika dan fungsi program berjalan sesuai rancangan, serta demonstrasi aplikasi kepada staf administrasi rektorat untuk memperoleh umpan balik mengenai kesesuaian fitur dengan kebutuhan pengguna. Demonstrasi dilakukan dengan memperlihatkan proses upload surat masuk, ekstraksi *OCR*, hasil ekstraksi metadata, pengelolaan arsip, dan disposisi surat. Berdasarkan demonstrasi tersebut, staf administrasi memberikan umpan balik terutama terkait kebutuhan fitur pembuatan surat keluar. Umpan balik tersebut digunakan sebagai bahan evaluasi dan ditindaklanjuti dengan pengembangan halaman Editor Surat Keluar pada sistem.

Pengujian sistem dilakukan untuk memastikan aplikasi arsip digital berbasis *Optical Character Recognition (OCR)* dan *Rule-Based System* berjalan sesuai dengan rancangan. Metode yang digunakan adalah *White Box Testing*, yaitu pengujian yang berfokus pada struktur internal, alur kontrol, dan logika program dengan menelusuri jalur eksekusi serta keluaran dari setiap modul [16]. Metode ini digunakan untuk mengevaluasi fungsi utama sistem, meliputi proses ekstraksi teks menggunakan *OCR*, ekstraksi dan identifikasi metadata berdasarkan aturan pada *Rule-Based System*, serta penyimpanan data arsip. Setiap modul diuji berdasarkan kondisi masukan dan keluaran yang dihasilkan untuk mengidentifikasi kesalahan logika atau *error* pada program. Hasil pengujian digunakan sebagai dasar perbaikan sistem sehingga fungsi pengelolaan arsip dapat berjalan sesuai kebutuhan dan rancangan yang telah ditetapkan.

3. Hasil dan Pembahasan

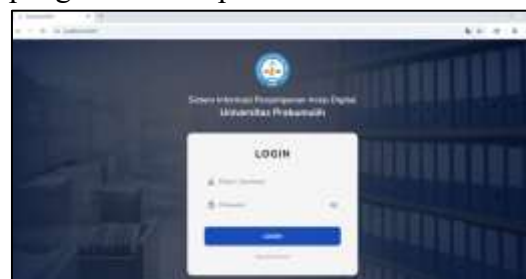
Pada bagian ini disajikan hasil implementasi dan pengujian sistem arsip

digital berbasis *Optical Character Recognition (OCR)* dan *Rule-Based System* yang telah dikembangkan. Hasil penelitian mencakup implementasi sistem pada lingkungan pengguna serta pengujian untuk memastikan seluruh fungsi sistem berjalan sesuai dengan kebutuhan yang telah ditetapkan.

3.1 Implementasi Sistem

a. Halaman Login

Halaman login berfungsi sebagai mekanisme autentikasi pengguna sebelum mengakses sistem arsip digital. Pengguna diwajibkan memasukkan username dan password yang valid agar dapat menggunakan fitur yang tersedia sesuai hak akses masing-masing. Proses autentikasi ini bertujuan untuk menjaga keamanan data arsip dan membatasi akses hanya kepada pengguna yang berwenang. Sistem juga menerapkan *Role-Based Access Control (RBAC)* untuk membatasi akses pengguna berdasarkan peran dan unit yang diberikan, sehingga pengguna hanya dapat melihat dan mengelola arsip sesuai dengan lingkup hak aksesnya. Selain itu, setiap arsip menyimpan *id_user_input* sebagai identitas pengguna yang melakukan input arsip. Namun, sistem belum menerapkan audit trail secara menyeluruh untuk mencatat riwayat perubahan atau penghapusan arsip. Pengembangan audit trail dapat dilakukan pada penelitian selanjutnya untuk meningkatkan keterlacakan aktivitas pengelolaan arsip.



Gambar 2. Halaman Login

b. Halaman Dashboard Operator

Dashboard merupakan halaman utama yang ditampilkan setelah pengguna berhasil masuk ke dalam sistem. Halaman ini menyediakan informasi ringkas mengenai jumlah arsip yang tersimpan serta menu

navigasi untuk mengakses berbagai fitur pengelolaan arsip.



Gambar 3. Halaman Dashboard Operator

c. Halaman Upload Surat Masuk

Halaman *upload* surat masuk digunakan untuk memasukkan dokumen arsip ke dalam sistem. Pengguna dapat mengunggah dokumen surat masuk dalam format yang telah ditentukan untuk selanjutnya diproses menggunakan *OCR*. Tahap ini menjadi langkah awal dalam proses digitalisasi arsip sehingga informasi yang terdapat pada dokumen fisik dapat diolah secara otomatis oleh sistem.



Gambar 4. Halaman Upload Surat Masuk

d. Halaman Form Upload Surat Masuk

Halaman ini menampilkan hasil ekstraksi dokumen surat masuk menggunakan *OCR* dan identifikasi informasi menggunakan *Rule-Based System*. *OCR* digunakan untuk mengenali teks dari dokumen yang diunggah, sedangkan *Rule-Based System* mengidentifikasi informasi seperti asal surat, nomor surat, tanggal surat, dan perihal secara otomatis sebelum arsip disimpan ke dalam sistem.



Gambar 5. Halaman Form Upload Surat Masuk

e. Halaman Upload Surat Keluar

Halaman Upload Surat Keluar digunakan untuk mengunggah dokumen surat keluar yang akan diarsipkan. Selain fitur unggah dokumen, tersedia opsi untuk membuat surat keluar baru secara langsung melalui sistem. Antarmuka dirancang konsisten dengan halaman surat masuk untuk menjaga kemudahan penggunaan.



Gambar 6. Halaman Upload Surat Keluar

f. Halaman Form Upload Surat Keluar

Halaman ini menampilkan hasil ekstraksi teks beserta metadata surat keluar yang telah diidentifikasi sistem. Pengguna dapat memeriksa dan memperbaiki data sebelum proses penyimpanan dilakukan. Fitur ini memastikan informasi surat tersimpan secara akurat dalam basis data.



Gambar 7. Halaman Form Upload Surat Keluar

g. Halaman Editor Surat Keluar

Berdasarkan hasil demonstrasi aplikasi dan umpan balik dari staf administrasi rektorat, dikembangkan halaman Editor Surat Keluar untuk mendukung proses pembuatan surat secara digital. Halaman ini memungkinkan pengguna mengisi data surat melalui formulir yang tersedia serta melihat hasil surat secara langsung melalui fitur pratinjau. Setelah surat selesai dibuat, pengguna dapat menyimpannya ke dalam arsip atau mencetaknya sesuai kebutuhan.



Gambar 8. Halaman Editor Surat Keluar

lanjut. Tampilan dirancang untuk mendukung proses pengambilan keputusan secara cepat.



Gambar 11. Halaman Dashboard Pimpinan

h. **Halaman Arsip Surat Masuk Operator**

Halaman arsip surat masuk di sisi operator menampilkan data dokumen yang telah berhasil disimpan ke dalam sistem arsip digital. Informasi arsip disajikan secara terstruktur sehingga memudahkan proses pencarian, pengelolaan, dan temu kembali dokumen sesuai kebutuhan pengguna.



Gambar 9. Halaman Arsip Surat Masuk Operator

k. **Halaman Arsip Surat Masuk Pimpinan**

Halaman ini menampilkan daftar surat masuk yang dapat dipantau oleh pimpinan. Selain melihat detail arsip, pimpinan dapat memberikan disposisi kepada unit terkait melalui fitur yang tersedia. Fitur ini mendukung pengelolaan surat secara lebih terstruktur.



Gambar 12. Halaman Arsip Surat Masuk Pimpinan

i. **Halaman Arsip Surat Keluar Operator**

Halaman Arsip Surat Keluar digunakan untuk menampilkan seluruh data surat keluar yang telah tersimpan dalam sistem. Data disajikan dalam bentuk tabel yang dilengkapi fitur pencarian dan pengelolaan arsip. Halaman ini memudahkan proses penelusuran kembali dokumen yang dibutuhkan.



Gambar 10. Halaman Arsip Surat Keluar Operator

l. **Tampilan Form Disposisi**

Form Disposisi digunakan untuk memberikan instruksi tindak lanjut terhadap surat masuk. Pada halaman ini pimpinan dapat menentukan tujuan disposisi, menambahkan keterangan, serta menetapkan tingkat prioritas surat. Data disposisi kemudian disimpan dan diteruskan kepada unit yang dituju.



Gambar 13. Tampilan Form Disposisi

j. **Halaman Dashboard Pimpinan**

Dashboard Pimpinan berfungsi sebagai pusat pemantauan surat masuk dan surat keluar. Halaman ini menyediakan informasi statistik arsip, fitur pencarian, serta akses menuju daftar arsip yang memerlukan tindak

m. **Tampilan Lihat Disposisi**

Halaman Lihat Disposisi digunakan untuk menampilkan informasi disposisi yang telah dibuat sebelumnya. Pengguna dapat melihat tujuan, instruksi, dan status disposisi sebagai bagian dari proses pengelolaan arsip

digital. Fitur ini mendukung transparansi alur administrasi surat.

unit kerja. Tampilan dibuat sederhana untuk mendukung efisiensi pengelolaan sistem.



Gambar 14. Tampilan Lihat Disposisi



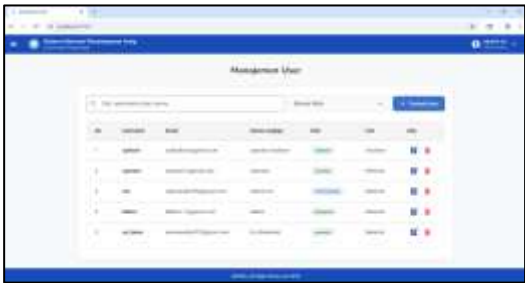
Gambar 17. Halaman Dashboard Administrator

n. **Halaman *Inbox Disposisi***
Halaman *Inbox Disposisi* digunakan untuk mengelola disposisi masuk dan disposisi terkirim. Pengguna dapat melihat informasi surat, status disposisi, serta melakukan pencarian data dengan cepat. Halaman ini membantu proses distribusi dan pemantauan disposisi antar unit kerja.

q. **Halaman *Manajemen User***
Halaman *Manajemen User* digunakan untuk mengelola akun pengguna beserta hak aksesnya. Administrator dapat menambah, mengubah, maupun menghapus data pengguna sesuai kebutuhan. Fitur ini mendukung penerapan mekanisme *Role-Based Access Control (RBAC)* pada sistem.



Gambar 15. Tampilan *Inbox Disposisi*



Gambar 18. Halaman *Manajemen User*

o. **Halaman *Arsip Surat Keluar Pimpinan***
Halaman ini digunakan pimpinan untuk memantau arsip surat keluar yang telah dibuat oleh operator. Data ditampilkan dalam bentuk tabel yang dilengkapi fitur pencarian dan pratinjau dokumen. Hak akses pimpinan dibatasi pada fungsi monitoring tanpa mengubah data arsip.

r. **Halaman *Manajemen Unit***
Halaman *Manajemen Unit* digunakan untuk mengelola data unit kerja yang terdaftar dalam sistem. Administrator dapat menambah, mengubah, mengaktifkan, atau menonaktifkan unit sesuai struktur organisasi yang berlaku. Data unit ini menjadi referensi penting dalam proses pengarsipan dan disposisi surat.



Gambar 16. Halaman *Arsip Surat Keluar Pimpinan*



Gambar 19. Halaman *Manajemen Unit*

p. **Halaman *Dashboard Administrator***
Dashboard Administrator merupakan pusat pengelolaan data master sistem. Halaman ini menyediakan akses cepat menuju fitur manajemen pengguna dan manajemen

3.2 **Pengujian *White Box Testing***
Pengujian *White Box Testing* dilakukan untuk mengevaluasi logika program pada sistem arsip digital berbasis *OCR* dan *Rule-Based System*. Pengujian difokuskan pada modul-modul utama yang berperan

dalam proses ekstraksi data, pengelolaan arsip, dan penyimpanan informasi ke dalam basis data. Hasil pengujian menunjukkan bahwa seluruh jalur independen yang diidentifikasi dapat dieksekusi dengan baik dan tidak ditemukan kesalahan logika yang menyebabkan kegagalan proses. Selain itu, nilai *statement coverage* yang diperoleh mencapai 100%, sehingga seluruh pernyataan program pada modul yang diuji telah berhasil dieksekusi oleh skenario pengujian yang dirancang.

Pengujian pada penelitian ini difokuskan pada verifikasi logika dan ketercakupan eksekusi kode. Evaluasi kuantitatif terhadap kualitas hasil ekstraksi OCR dan ketepatan hasil Rule-Based System belum dilakukan dalam penelitian ini. Dengan demikian, *statement coverage* digunakan untuk menunjukkan ketercakupan pengujian kode, sedangkan evaluasi terhadap kualitas hasil ekstraksi dan ketepatan Rule-Based System dapat dilakukan sebagai pengembangan penelitian selanjutnya.

4. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penelitian ini berhasil mengembangkan sistem arsip digital berbasis *Optical Character Recognition (OCR)* dan *Rule-Based System* untuk mendukung pengelolaan surat di Universitas Prabumulih. Sistem yang dikembangkan mendukung proses pengarsipan secara lebih terstruktur dan mudah diakses dibandingkan dengan metode pengarsipan secara manual. Berdasarkan hasil pengujian menggunakan metode White Box Testing, seluruh fungsi utama sistem telah berjalan sesuai dengan rancangan yang ditetapkan, dengan nilai *statement coverage* sebesar 100%. Namun, hasil tersebut menunjukkan ketercakupan pengujian kode dan belum menggambarkan kualitas hasil ekstraksi OCR maupun ketepatan Rule-Based System. Oleh karena itu, pengembangan lebih lanjut diperlukan untuk mengevaluasi kualitas hasil ekstraksi OCR dan ketepatan Rule-Based System, serta menguji sistem menggunakan dokumen dengan karakteristik dan format yang lebih beragam agar penerapannya dapat dikembangkan pada berbagai kondisi pengarsipan.

Daftar Rujukan

- [1] L. Sulkowski, *Managing the digital university : paradigms, leadership, and organization*. Routledge, 2023.
- [2] G. A. Fad'li, Marsofiyati, and Suherdi, "Implementasi Arsip Digital Untuk Penyimpanan Dokumen Digital," *Jurnal Manuhara : Pusat Penelitian Ilmu Manajemen dan Bisnis*, vol. 1, no. 4, pp. 01–10, Oct. 2023, doi: 10.61132/manuhara.v1i4.115.
- [3] Setiorini and D. R. Aulianto, "Digital Archive Management in the Organization Section of the Regional Secretariat of Tuban Regency," 2024. [Online]. Available: <http://ejournal.unp.ac.id/index.php/iipk/index>
- [4] A. T. P. D. Akhsa, M. Agus, Rosmiati, and A. M. B. Ulum, "Designing E-Office Services And Digital Archiving Using Web-Based OCR Methods," *Journal of Information Technology and Computer Science (INTECOMS)*, vol. 7, no. 1, 2024.
- [5] F. A. Gosal and P. T. D. Rompas, "Penerapan Teknologi Optical Character Recognition Pada Pengarsipan Dokumen (Studi Kasus: PT Pertamina Geothermal Energy Area Lahendong)," 2023.
- [6] R. R. Runtu, G. D. P. Maramis, and A. Hasibuan, "Implementasi Algoritma Rule-Based Classification pada Aplikasi Arsip Web Menggunakan Metode RAD dan Evaluasi ISO/IEC 25010," *Jurnal Minfo Polgan*, vol. 14, no. 2, pp. 2792–2798, Nov. 2025, doi: 10.33395/jmp.v14i2.15505.
- [7] F. M. H. Tjiptabudi and R. Ndaumanu, "Implementasi Aplikasi Arsip Dokumen Untuk Mendukung Proses Bisnis Pada PT.XXX," May 2024.
- [8] J. Mulyadi and N. D. M. Veronika, "Penerapan Optical Character Recognition (OCR) dalam Pengambilan Teks Judul Buku untuk Digitalisasi Arsip Perpustakaan," *JSAI : Journal Scientific and Applied Informatics*, vol. 08, no. 2, pp. 447–454, Jun. 2025, doi: 10.36085.
- [9] A. M. Aubaid, A. Mishra, and A. Mishra, "Machine learning and rule-based embedding techniques for classifying text documents,"

- International Journal of System Assurance Engineering and Management*, vol. 15, no. 12, pp. 5637–5652, Dec. 2024, doi: 10.1007/s13198-024-02555-w.
- [10] T. Evi, A. M. Gai, M. Fauza, and I. Hatchi, “Buku Referensi Metode Penelitian Kualitatif dan Kuantitatif Teknik dan Strategi Dalam Mengolah Data,” Jun. 2025.
- [11] R. F. Ulwani, M. Mailin, and Z. Zainun, “Strategi Komunikasi Al-Ittihadiyah Sumatera Utara dalam Mengembangkan Dakwah di Sumatera Utara,” *JURNAL LENSEA MUTIARA KOMUNIKASI*, vol. 5, no. 1, pp. 10–20, Jun. 2021, doi: 10.51544/jlmk.v5i1.1496.
- [12] R. Nuraeni, A. P. A. Santoso, and Rezi, “Perlindungan Konsumen Terhadap Produk Overclaim Skincare Di Indonesia,” 2025. [Online]. Available: <https://ejournals.com/ojs/index.php/jih>
- [13] I. Setiawan, Ariansyah, and Y. Purbasari, *Buku Ajar Kuliah Metode Penelitian Kualitatif*. 2020.
- [14] T. Priatna *et al.*, “Teknik Pengumpulan Data Dalam Penelitian Kualitatif: Kajian Konseptual Dan Berbasis Literatur,” *Social Studies & Humanities Journal (SOSHUM)*, vol. 2, no. 4, pp. 465–470, 2024, doi: 10.62207/40p54795.
- [15] D. A. Lantana and Fauziah, *Design Thinking dan Inovasi*. Lembaga Penerbitan Universitas Nasional (LPU UNAS), 2025.
- [16] A. Fauzi and A. H. Yunial, *Testing dan QA Perangkat Lunak: Memahami Konsep, Metode, Teknik, dan Pendekatan dalam Pengujian*. Eureka Media Aksara, 2025.