



Department of Digital Business

Journal of Artificial Intelligence and Digital Business (RIGGS)

Homepage: <https://journal.ilmudata.co.id/index.php/RIGGS>

Vol. 5 No. 2 (2026) pp: 11298-11305

P-ISSN: 2963-9298, e-ISSN: 2963-914X

Perancangan Sistem Arsip Surat Digital dengan Deteksi Duplikasi Berbasis Nomor Surat Menggunakan Regular Expression dan Exact Matching

Dian Sari Amalia

Sistem Informasi, Fakultas Sains dan Teknologi, UIN Raden Fatah Palembang

dianne.amalia@gmail.com

Abstrak

Kemajuan teknologi informasi telah mendorong lembaga pemerintah untuk melakukan transformasi digital dalam berbagai aspek administrasi, termasuk pengelolaan arsip dokumen. Arsip dokumen merupakan sumber informasi yang sangat penting dan memainkan peran krusial dalam mendukung proses administrasi, pengambilan keputusan, serta menyediakan bukti kegiatan organisasi. Namun, proses pengelolaan arsip yang masih bergantung pada dokumen fisik dan pencatatan manual seringkali menimbulkan berbagai masalah, seperti kesulitan dalam menemukan dokumen, ketidakkonsistenan data, dan duplikasi catatan akibat entri data yang berulang. Masalah-masalah ini berpotensi mengganggu kualitas data dan menghambat efektivitas pengelolaan arsip di lembaga pemerintah. Penelitian ini bertujuan untuk mengembangkan Smartchive, sebuah sistem pengarsipan dokumen digital yang mampu mendeteksi dokumen duplikat secara otomatis berdasarkan nomor dokumen. Sistem ini dikembangkan untuk memenuhi kebutuhan pengelolaan dokumen Badan Pengelola Keuangan dan Aset Daerah (BPKAD) Provinsi Sumatera Selatan. Proses deteksi duplikasi dilakukan dengan mengekstrak nomor dokumen dari dokumen yang diarsipkan menggunakan metode Regular Expression (Regex), dilanjutkan dengan perbandingan menggunakan metode Exact Matching untuk mengidentifikasi catatan dengan nomor dokumen yang identik. Sistem ini juga dirancang untuk membaca data arsip dari berkas Microsoft Excel, mengekstrak metadata arsip, menyimpan data dalam repositori arsip digital, serta menyediakan fasilitas pencarian data terpusat. Hasil implementasi menunjukkan bahwa Smartchive mampu secara otomatis mengekstrak nomor dokumen, mendeteksi data arsip yang duplikat, dan menyajikan informasi arsip dalam format yang lebih terstruktur. Sistem ini membantu mengurangi risiko redundansi data, meningkatkan kualitas pengelolaan arsip, dan mempercepat proses pencarian informasi. Dengan demikian, Smartchive dapat berfungsi sebagai solusi sederhana yang mendukung digitalisasi pengelolaan dokumen di lingkungan pemerintahan.

Kata Kunci: Arsip Surat, Deteksi Duplikasi, Regular Expression, Exact Matching, Sistem Arsip Digital, Smartchive.

1. Latar Belakang

Transformasi digital telah menjadi salah satu fokus utama tata kelola modern untuk meningkatkan efektivitas layanan publik dan kualitas pengelolaan informasi. Penerapan teknologi informasi dalam administrasi pemerintahan mendorong berbagai lembaga untuk beralih dari sistem konvensional ke sistem elektronik yang lebih terintegrasi, transparan, dan mudah diakses [1], [2], [3]. Salah satu bentuk transformasi digital ini adalah digitalisasi arsip, yang bertujuan untuk meningkatkan efisiensi penyimpanan, pencarian, dan pengelolaan dokumen di dalam organisasi.

Arsip surat merupakan komponen kunci dalam kegiatan administratif suatu organisasi, karena berfungsi sebagai bukti, sumber informasi, dan memori kelembagaan yang mendokumentasikan berbagai kegiatan organisasi [4], [5]. Di sektor pemerintahan, arsip korespondensi digunakan sebagai landasan untuk melaksanakan kegiatan, melakukan koordinasi antarlembaga, dan mendokumentasikan berbagai kebijakan yang telah ditetapkan. Oleh karena itu, pengelolaan arsip yang efektif merupakan faktor krusial dalam mendukung tata kelola yang efektif dan akuntabel.

Meskipun penerapan arsip elektronik terus berkembang, masih ada lembaga yang mengelola arsip menggunakan berkas *spreadsheet* yang disimpan secara terpisah dan diperbarui secara manual. Situasi ini berpotensi menimbulkan kesulitan dalam proses pencarian dokumen, ketidakkonsistenan data, serta peningkatan risiko kehilangan informasi akibat kurangnya sistem manajemen terintegrasi [2], [4], [6]. Selain itu, proses entri data yang dilakukan oleh banyak pengguna sering kali menimbulkan masalah seperti data ganda atau catatan ganda.

Kualitas data merupakan aspek krusial dalam pengelolaan informasi digital. Data yang tidak akurat, tidak konsisten, atau mengandung duplikat dapat menurunkan kualitas informasi yang dihasilkan dan memengaruhi proses pengambilan keputusan suatu organisasi [7]. Menurut Fatima dkk. [8], salah satu masalah utama dalam pengelolaan data adalah adanya data duplikat, yang menyebabkan redundansi penyimpanan dan mengurangi efisiensi sistem informasi. Martins dkk. [9] juga menjelaskan bahwa proses pembersihan data merupakan langkah krusial dalam meningkatkan kualitas data dan mendukung efektivitas sistem informasi.

Masalah data duplikat juga ditemui dalam pengelolaan arsip korespondensi. Duplikasi dapat terjadi ketika surat yang sama dimasukkan lebih dari sekali ke dalam sistem penyimpanan arsip, yang mengakibatkan redundansi data dan mempersulit proses pengelolaan dokumen [10], [11]. Berbagai penelitian telah mengembangkan metode deteksi duplikasi dengan menggunakan pendekatan pencocokan data (*record linkage*) dan deduplikasi data [12]. Namun, untuk data dengan atribut unik seperti nomor dokumen, pendekatan Pencocokan Tepat (*Exact Matching*) dianggap lebih sederhana dan lebih efektif karena proses pencocokannya didasarkan pada nilai-nilai yang identik [10].

Selain deteksi duplikat, pengelolaan arsip digital juga memerlukan proses ekstraksi informasi untuk memperoleh data penting dari dokumen yang disimpan. Teknik ekstraksi informasi banyak digunakan untuk secara otomatis mengekstrak data tertentu dari teks menggunakan pola yang telah ditentukan sebelumnya [13], [10]. Salah satu metode yang umum digunakan adalah *Regular Expressions* (Regex), yaitu teknik pencocokan pola teks yang mampu mengidentifikasi informasi dalam format tertentu, termasuk nomor dokumen.

Berdasarkan pengamatan yang dilakukan selama magang di Badan Pengelolaan Keuangan dan Aset Daerah (BPKAD) Provinsi Sumatera Selatan, ditemukan bahwa pengelolaan arsip surat masih menghadapi tantangan dalam proses pencarian dokumen dan identifikasi data duplikat. Arsip yang disimpan dalam berbagai berkas menyebabkan proses verifikasi data memakan waktu yang relatif lama, terutama karena jumlah arsip terus bertambah seiring berjalannya waktu.

Penelitian ini bertujuan untuk merancang sistem pengarsipan dokumen digital bernama Smartchive yang dapat membantu dalam penyimpanan, pencarian, dan deteksi dokumen duplikat secara otomatis. Sistem ini dikembangkan menggunakan kerangka kerja *Streamlit* dengan berkas *Excel* sebagai media penyimpanan. Deteksi duplikat dilakukan dengan mengekstrak nomor dokumen menggunakan ekspresi reguler dan mencocokkannya menggunakan metode pencocokan eksak. Diharapkan sistem yang dikembangkan ini dapat membantu meningkatkan kualitas pengelolaan dokumen digital di BPKAD Provinsi Sumatera Selatan serta mendukung implementasi transformasi digital di bidang pengelolaan dokumen.

2. Metode Penelitian

2.1. Objek Penelitian

Penelitian ini dilakukan di Badan Pengelolaan Keuangan dan Aset Daerah (BPKAD) Provinsi Sumatera Selatan. Objek penelitian terdiri dari data arsip surat yang digunakan dalam kegiatan administrasi dan pengelolaan dokumen di instansi tersebut. Data arsip yang digunakan disimpan dalam format *Microsoft Excel* dan bersumber dari dokumen-dokumen yang dikelola selama kegiatan operasional instansi. Data ini dimanfaatkan sebagai sumber data dalam proses perancangan, implementasi, dan pengujian sistem arsip surat digital yang dikembangkan dalam penelitian ini.

Pemilihan BPKAD Provinsi Sumatera Selatan sebagai objek penelitian didasarkan pada pengamatan yang dilakukan selama magang, yang menunjukkan adanya kebutuhan akan sistem yang mampu memfasilitasi proses pengelolaan arsip yang lebih terstruktur terutama dalam mengidentifikasi arsip dengan nomor referensi yang sama. Kebutuhan ini menjadi dasar pengembangan sistem *Smartchive* sebagai solusi untuk mendukung proses pengarsipan digital dan memastikan kualitas arsip yang disimpan.

2.2 Tahapan Penelitian

Penelitian ini menggunakan pendekatan perancangan dan pengembangan sistem yang terdiri dari beberapa tahap: identifikasi masalah, pengumpulan data, analisis persyaratan sistem, perancangan sistem, implementasi sistem, pengujian sistem, dan analisis hasil pengujian. Tahapan-tahapan penelitian tersebut dilaksanakan secara berurutan untuk memastikan bahwa sistem yang dikembangkan memenuhi kebutuhan pengguna dan mampu mengatasi masalah-masalah yang teridentifikasi dalam proses pengelolaan dokumen.

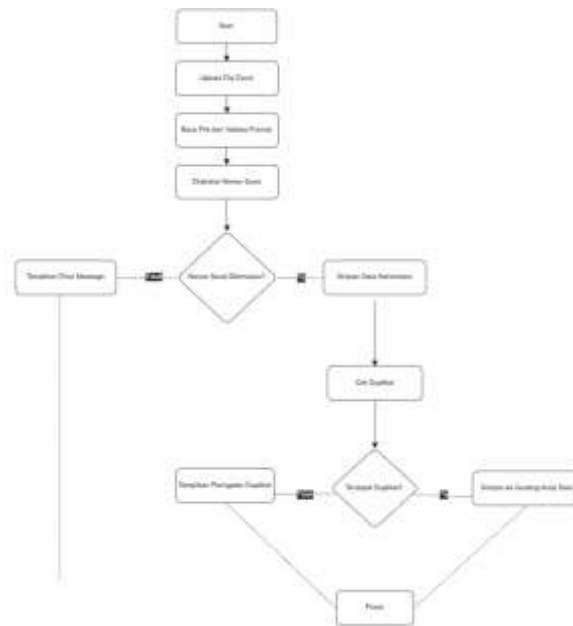


Gambar 1. Flowchart Tahapan Penelitian

Berdasarkan Gambar 1, penelitian ini dimulai dengan tahap identifikasi masalah melalui pengamatan terhadap proses pengelolaan dokumen di BPKAD Provinsi Sumatera Selatan. Fase berikutnya melibatkan pengumpulan data arsip yang akan digunakan sebagai dasar untuk menganalisis persyaratan sistem. Setelah persyaratan sistem ditetapkan, proses perancangan sistem dilaksanakan, yang mencakup perancangan alur kerja, struktur data, dan antarmuka pengguna. Sistem tersebut kemudian diimplementasikan menggunakan kerangka kerja Streamlit dan diuji untuk memastikan bahwa semua fungsi beroperasi sesuai yang diharapkan. Hasil pengujian kemudian dianalisis untuk mengevaluasi kemampuan sistem dalam mendeteksi catatan duplikat dan mendukung proses pengelolaan arsip digital.

2.3 Perancangan Sistem

Sistem yang dikembangkan dalam penelitian ini dinamakan *Smartchive* dan dirancang sebagai sistem pengarsipan dokumen digital berbasis web yang menggunakan kerangka kerja *Streamlit*. Sistem ini dikembangkan untuk menyederhanakan proses penyimpanan, pencarian, dan identifikasi data yang diarsipkan dengan nomor dokumen yang sama. *Smartchive* terdiri dari tiga halaman utama: Dashboard, Unggah Data, dan Repositori Data Arsip. Halaman Dashboard menampilkan ringkasan informasi mengenai arsip yang tersimpan; halaman Unggah Data digunakan untuk mengunggah arsip baru dan memeriksa duplikat; sedangkan halaman Repositori Data Arsip menampilkan seluruh data arsip yang tersimpan dalam sistem.



Gambar 2. Flowchart Sistem Smartchive

Berdasarkan Gambar 2, proses dimulai ketika pengguna mengunggah berkas arsip dalam format *Microsoft Excel* ke sistem. Data yang berhasil diunggah kemudian diproses untuk mengekstrak informasi kunci, khususnya nomor referensi yang terdapat pada kolom “uraian masalah”. Nomor referensi yang diperoleh selanjutnya digunakan dalam proses pemeriksaan duplikasi menggunakan metode Pencocokan Tepat. Jika ditemukan nomor referensi yang identik dengan data yang telah disimpan sebelumnya, sistem akan menandai data tersebut sebagai duplikat. Sebaliknya, jika tidak ditemukan nomor referensi yang cocok, data tersebut akan disimpan dalam arsip dan dapat diakses kembali melalui sistem.

2.4 Implementasi Regular Expression

Proses identifikasi angka huruf dalam penelitian ini dilakukan dengan menggunakan metode *Regular Expression* (Regex). Metode ini digunakan untuk mengekstrak informasi tertentu dari data teks berdasarkan pola karakter yang telah ditentukan sebelumnya. Dalam penelitian ini, Regex digunakan untuk secara otomatis mengambil angka huruf dari kolom “uraian masalah” pada arsip yang diunggah ke sistem. Penggunaan teknik ekstraksi informasi berbasis pola memungkinkan proses identifikasi nomor dokumen dilakukan dengan cepat dan konsisten tanpa memerlukan input manual dari pengguna [13].

Penerapan Regex dalam sistem dilakukan dengan menggunakan pola pencarian yang dirancang untuk mengenali format nomor dokumen yang terdapat dalam dokumen-dokumen yang diarsipkan. Contoh penerapan yang digunakan dalam sistem ditunjukkan pada Kode 1.

```
re.search( r'(?i)Nomor:\s*([A-Za-z0-9\.\-]+)', str(x) )
```

Pola ini digunakan untuk mencari teks yang diawali dengan kata “Nomor:” dan mengekstrak karakter-karakter yang membentuk nomor dokumen yang mengikuti kata tersebut. Data yang diekstrak kemudian disimpan dalam atribut `NOMOR_SISTEM`, yang berfungsi sebagai pengenal unik untuk setiap dokumen dan digunakan dalam proses pendeteksian duplikat.

2.5 Implementasi *Exact Matching*

Deteksi duplikat dalam penelitian ini dilakukan dengan menggunakan metode *Exact Matching*. Metode ini bekerja dengan membandingkan nomor dokumen yang diekstraksi dari setiap catatan secara satu per satu. Jika ditemukan dua atau lebih catatan yang memiliki nomor dokumen yang sama, sistem akan mengidentifikasinya sebagai data duplikat [14], [13].

Proses pemeriksaan duplikasi dilakukan menggunakan fungsi deteksi duplikat dalam pustaka `Pandas`. Implementasi metode *Exact Matching* yang digunakan dalam sistem ditunjukkan pada Kode 2.

Duplicated (subset=['NOMOR_SISTEM'], keep=False)

Fungsi ini digunakan untuk menelusuri seluruh data arsip yang tersimpan dan mengidentifikasi nomor dokumen yang muncul lebih dari satu kali. Data yang terdeteksi sebagai duplikat kemudian ditampilkan kepada pengguna di halaman pemeriksaan duplikat untuk verifikasi lebih lanjut. Metode Pencocokan Tepat dipilih karena nomor dokumen merupakan atribut unik yang digunakan sebagai pengenal arsip; oleh karena itu, pendekatan pencocokan tepat dianggap cukup efektif dan sesuai dengan kebutuhan sistem yang sedang dikembangkan.

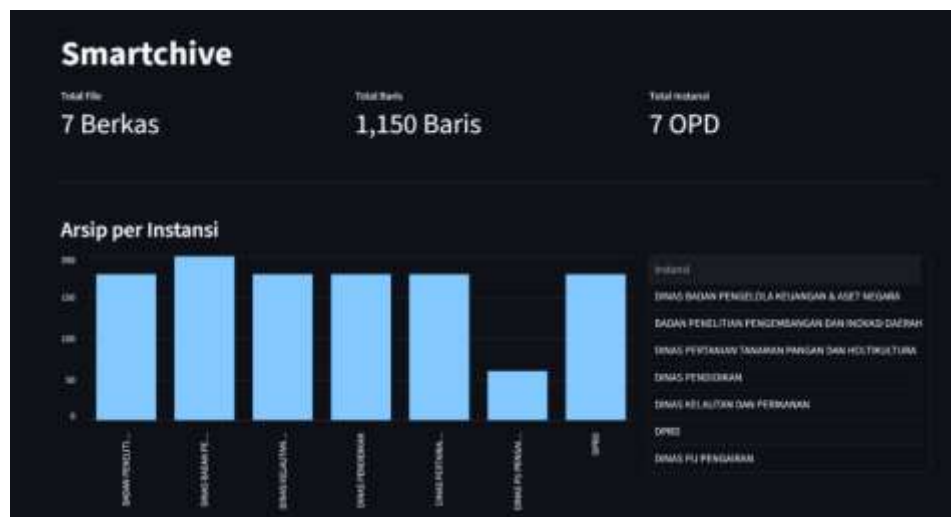
3. Hasil dan Diskusi

3.1 Hasil

Penelitian ini menghasilkan sebuah sistem pengarsipan dokumen digital berbasis web yang disebut *Smartchive*. Sistem ini dikembangkan menggunakan kerangka kerja Streamlit dan dirancang untuk memudahkan proses pengarsipan serta mendeteksi data duplikat berdasarkan nomor dokumen. Sistem ini terdiri dari tiga menu utama yaitu Dashboard, Unggah Data, dan Repositori Data Arsip yang terintegrasi ke dalam satu platform.

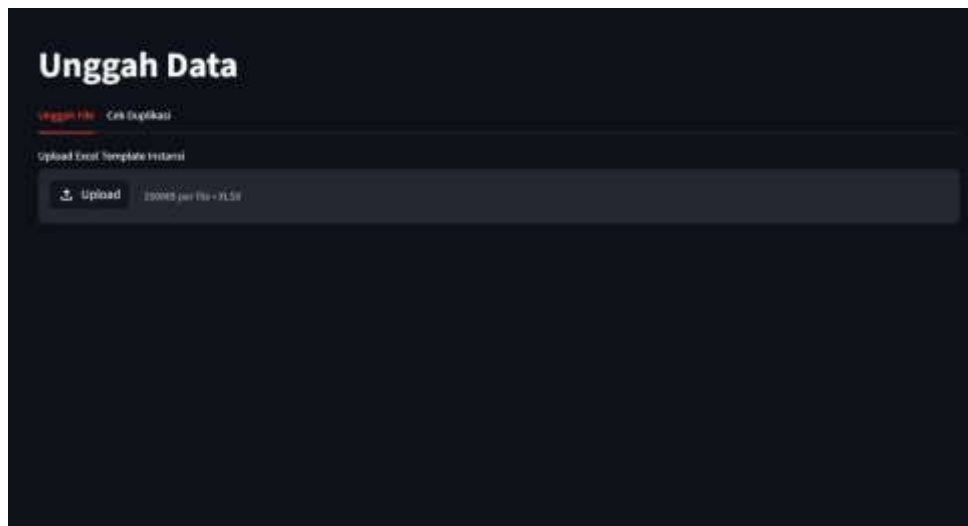
Hasil Implementasi Sistem

Sistem ini berhasil mengelola data arsip dari berkas *Microsoft Excel* dan menyimpannya dalam basis data arsip digital. Dasbor menampilkan informasi mengenai jumlah berkas arsip, volume data arsip, dan jumlah lembaga yang tersimpan dalam sistem.



Gambar 3. Tampilan Dashboard Smartchive

Menu Unggah Data memungkinkan pengguna untuk mengunggah berkas baru ke dalam sistem. Selama proses ini, sistem secara otomatis mengekstrak nomor dokumen dan memeriksa adanya duplikat sebelum data disimpan.



Gambar 4. Proses Unggah Data dan Pemeriksaan Duplikasi

Data yang berhasil diproses kemudian disimpan pada Gudang Data Arsip yang berfungsi sebagai repositori seluruh arsip yang telah diunggah.

NAMA INSTANSI (OPD)	NO	KODE KLASIFIKASI	JENIS ARSIP	URAIAN MASALAH
BADAN PENELITIAN PENGEMBANGAN DAN INOVASI DAERAH	1	931	SP20	Nomor: 0940/SP20/1.20.07.01/2014 Tanggal: 25 Nopember 2014 Pembayaran gaji
BADAN PENELITIAN PENGEMBANGAN DAN INOVASI DAERAH	2	931	SP20	Nomor: 0133/SP20/1.20.07.01/2014 Tanggal: 25 September 2014 Pembayaran gaji
BADAN PENELITIAN PENGEMBANGAN DAN INOVASI DAERAH	3	931	SP20	Nomor: 0077/SP20/1.20.07.01/2014 Tanggal: 25 Februari 2014 Pembayaran gaji d
BADAN PENELITIAN PENGEMBANGAN DAN INOVASI DAERAH	4	931	SP20	Nomor: 07570/SP20/1.20.07.01/2014 Tanggal: 25 September 2014 Pembayaran Tji
BADAN PENELITIAN PENGEMBANGAN DAN INOVASI DAERAH	5	931	SP20	Nomor: 0030/SP20/1.00.01.01/2014 Tanggal: 5 Desember 2014 Pembayaran Lang
BADAN PENELITIAN PENGEMBANGAN DAN INOVASI DAERAH	6	931	SP20	Nomor: 06820/SP20/1.20.07.01/2014 Tanggal: 4 September 2014 Pembayaran Tji
BADAN PENELITIAN PENGEMBANGAN DAN INOVASI DAERAH	7	931	SP20	Nomor: 04353/SP20/1.20.07.01/2014 Tanggal: 27 Juni 2014 Pembayaran gaji d
BADAN PENELITIAN PENGEMBANGAN DAN INOVASI DAERAH	8	931	SP20	Nomor: 01089/SP20/1.20.07.01/2014 Tanggal: 23 April 2014 Pembayaran gaji d
BADAN PENELITIAN PENGEMBANGAN DAN INOVASI DAERAH	9	931	SP20	Nomor: 08530/SP20/1.20.07.01/2014 Tanggal: 24 Oktober 2014 Pembayaran gaji d
BADAN PENELITIAN PENGEMBANGAN DAN INOVASI DAERAH	10	931	SP20	Nomor: 11581/SP20/1.20.07.01/2014 Tanggal: 12 Desember 2014 Balanja Langrai

Gambar 5. Gudang Data Arsip Smartchive

Hasil Ekstraksi Nomor Surat

Pengujian dilakukan pada data arsip yang berisi angka huruf di kolom uraian masalah. Sistem berhasil mengekstrak angka huruf tersebut secara otomatis menggunakan pola ekspresi reguler yang telah ditentukan sebelumnya. Tabel 1 menunjukkan contoh hasil ekstraksi angka huruf yang dihasilkan oleh sistem.

Tabel 1. Contoh Hasil Ekstraksi Nomor Surat

Uraian Masalah	Nomor Surat Hasil Ekstraksi
Nomor: 800/123/BPKAD/2025 Permohonan Data	800/123/BPKAD/2025
Nomor: 090/045/BPKAD/2025 Laporan Kegiatan	090/045/BPKAD/2025
Nomor: 050/011/BPKAD/2025 Permintaan Informasi	050/011/BPKAD/2025

Hasil pengujian menunjukkan bahwa sistem mampu mengenali pola huruf dan angka yang terdapat dalam data yang diarsipkan dan menyimpannya dalam atribut NOMOR_SISTEM, yang digunakan dalam proses deteksi duplikasi.

Hasil Deteksi Duplikasi

Deteksi duplikat dilakukan dengan menggunakan metode Pencocokan Tepat (*Exact Matching*) melalui perbandingan nilai `SYSTEM_NUMBER` di seluruh catatan yang tersimpan. Tabel 2 menunjukkan contoh data yang terdeteksi sebagai duplikat.

Tabel 2. Contoh Hasil Deteksi Duplikasi

Instansi	Nomor Surat
BPKAD	800/123/BPKAD/2025
BPKAD	800/123/BPKAD/2025

Hasil pengujian menunjukkan bahwa sistem mampu mengidentifikasi catatan dengan nomor referensi yang sama dan menampilkannya sebagai entri duplikat pada halaman pemeriksaan duplikat.

3.2 Diskusi

Hasil penelitian menunjukkan bahwa penerapan *Smartchive* dapat membantu pengelolaan arsip digital dengan mengintegrasikan fungsi penyimpanan arsip dan deteksi duplikat ke dalam satu sistem. Keberadaan sistem arsip digital mendukung proses pengelolaan dokumen yang lebih terstruktur dan memudahkan pencarian informasi dibandingkan dengan pengelolaan arsip secara manual [1], [2], [4], [6].

Penggunaan ekspresi reguler dalam penelitian ini berhasil memungkinkan ekstraksi otomatis nomor dokumen dari data arsip yang diunggah oleh pengguna. Temuan ini menunjukkan bahwa pendekatan berbasis pola teks dapat dimanfaatkan untuk mengekstraksi informasi penting dari dokumen tanpa memerlukan proses pelatihan model yang rumit. Hasil ini sejalan dengan penelitian tentang ekstraksi informasi, yang menjelaskan bahwa pola teks dapat digunakan untuk secara otomatis mengidentifikasi dan mengekstraksi informasi tertentu dari dokumen [13], [15].

Selain itu, penggunaan metode Pencocokan Tepat (*Exact Matching*) telah terbukti efektif dalam mendeteksi catatan dengan nomor dokumen yang identik. Karena nomor dokumen berfungsi sebagai pengenalan unik bagi dokumen, pendekatan pencocokan tepat ini dianggap cukup efektif untuk mengidentifikasi data duplikat dalam sistem arsip. Temuan ini mendukung penelitian sebelumnya mengenai deteksi duplikat, yang menjelaskan bahwa atribut unik dapat digunakan untuk mengidentifikasi data yang mewakili objek yang sama dalam basis data [14], [10], [7], [11].

Meskipun demikian, sistem yang dikembangkan masih memiliki keterbatasan. Proses deteksi duplikasi hanya didasarkan pada nomor dokumen yang berhasil diekstraksi dari dokumen-dokumen tersebut. Akibatnya, sistem ini belum dapat mendeteksi dokumen dengan isi yang serupa namun menggunakan format nomor dokumen yang berbeda. Keterbatasan ini menunjukkan bahwa metode Pencocokan Tepat sangat bergantung pada keseragaman format data yang digunakan. Oleh karena itu, pengembangan di masa mendatang dapat mempertimbangkan penerapan metode deduplikasi yang lebih adaptif, seperti pencocokan *fuzzy* atau pendekatan berbasis pembelajaran mesin, sebagaimana digunakan dalam penelitian sebelumnya [14], [7], [10].

4. Kesimpulan

Penelitian ini berhasil merancang dan mengimplementasikan sistem pengarsipan dokumen digital berbasis web bernama *Smartchive* untuk mendukung pengelolaan arsip di Badan Pengelolaan Keuangan dan Aset Daerah (BPKAD) Provinsi Sumatera Selatan. Sistem yang dikembangkan ini mampu mengelola data arsip secara terintegrasi melalui fitur-fitur untuk menyimpan, mencari, dan memantau data arsip dalam satu platform. Metode Regular Expression berhasil digunakan untuk mengekstrak nomor dokumen secara otomatis dari data arsip yang diunggah, sedangkan metode *Exact Matching* mampu mendeteksi duplikat berdasarkan kecocokan nomor dokumen. Hasil pengujian menunjukkan bahwa semua fungsi utama sistem beroperasi sesuai dengan persyaratan yang ditetapkan, sehingga membantu meningkatkan pengorganisasian penyimpanan arsip dan memfasilitasi proses pencarian informasi. Pengembangan *Smartchive* menunjukkan bahwa metode sederhana seperti *Regular Expression* dan *Exact Matching* dapat digunakan secara efektif untuk mendukung pengelolaan arsip digital di lembaga pemerintah yang masih mengandalkan data berbasis *spreadsheet*. Sistem yang dihasilkan berpotensi berfungsi sebagai solusi awal untuk mendukung transformasi pengelolaan arsip menuju lingkungan kerja yang lebih terdigitalisasi tanpa memerlukan infrastruktur yang rumit. Untuk penelitian di masa mendatang, sistem ini dapat ditingkatkan dengan mengintegrasikan metode deteksi kesamaan data yang lebih adaptif serta kemampuan untuk memproses dokumen yang dipindai guna menangani karakteristik arsip yang lebih beragam.

Referensi

1. H. P. N. Soulthoni and M. Itasari, "The Implementation of Electronic-Based Archiving to Accelerate Government Digitalization in Indonesia," no. 1, pp. 49–57, 2025, doi: 10.47540/ijias.v5i1.1735.
<https://doi.org/10.47540/ijias.v5i1.1735>
2. K. Kurniadi, S. Ibrahim, and B. Badruzzaman, "Evaluation of the efficiency of electronic filing systems in handling records : a case study of SRIKANDI applications in government institutions," *Cogent Educ.*, vol. 12, no. 1, p., 2025, doi: 10.1080/2331186X.2025.2562347.
<https://doi.org/10.1080/2331186X.2025.2562347>
3. D. R. Shabila, J. Pratikno, A. Wahid, and T. Hidayati, "THE INFLUENCE OF AI IMPLEMENTATION IN LOCAL GOVERNMENT INFORMATION SYSTEMS ON THE EFFECTIVENESS OF PUBLIC SERVICE DELIVERY IN INDONESIA," vol. 1, no. 1, pp. 44–55, 2025.
<https://doi.org/10.54443/ijerlas.v4i3.1719>
4. C. Queen, R. Oktavia, and J. D. Pribadi, "Digital Archives Management System for Archives Security at the Balai Pemberdayaan Industri Persepatuan Indonesia (BPIPI)," vol. 4, no. 3, pp. 505–509, 2025.
<https://doi.org/10.58258/bisnis.v4i3.9277>
5. N. S. Sahlimar and E. Dimisyqiyan, "Smart archive management : The interplay of organizational culture and knowledge governance," vol. 11, no. 2, 2025, doi: 10.20473/rlj.V11-I2.2025.424-440.
<https://doi.org/10.20473/rlj.V11-I2.2025.424-440>
6. P. Data, M. Body, and O. F. Knowledge, "MODEL MANAJEMEN DATA PADA PENGELOLAAN ARSIP ELEKTRONIK : PENERAPAN DATA MANAGEMENT BODY OF KNOWLEDGE DATA MANAGEMENT MODEL IN ELECTRONIC ARCHIVES MANAGEMENT ;," pp. 55–76.
<https://doi.org/10.52362/jmijayakarta.v2i1.685>
7. O. Azerual, G. Saake, and M. Abuosba, "Data Quality Measures and Data Cleansing for Research Information Sys- tems," vol. 16, no. 1, 2018.
<http://hdl.handle.net/11366/632>
8. A. Fatima, N. Nazir, and M. G. Khan, "Data Cleaning In Data Warehouse : A Survey of Data Pre-processing Techniques and Tools," no. March, pp. 50–61, 2017, doi: 10.5815/ijitcs.2017.03.06.
<https://doi.org/10.5815/ijitcs.2017.03.06>
9. A. K. Elmagarmid and S. Member, "Duplicate Record Detection : A Survey," vol. 19, no. 1, pp. 1–16, 2007.
<https://doi.org/10.1109/TKDE.2007.250581>
11. S. Aleshin-guendel, M. Sadinle, and M. E. Oct, "Multifile Partitioning for Record Linkage and Duplicate Detection," vol. 98195, pp. 1–63, 2021.
<https://doi.org/10.1080/01621459.2021.1987251>
12. P. Christen, *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*. Berlin, Germany: Springer, 2012, doi: 10.1007/978-3-642-31164-2.
<https://doi.org/10.1007/978-3-642-31164-2>
14. S. Sindhu and R. Suganya, "Decision Tree Formation and Fuzzy Similarity Matching for Duplicates Detection," *International Journal of Engineering Research & Technology (IJERT)*, vol. 2, no. 9, pp. 1187–1199, Sep. 2013, doi: 10.17577/IJERTV2IS90467.
<https://doi.org/10.17577/IJERTV2IS90467>
15. R. Liu, L. Gao, D. An, Z. Jiang, and Z. Tang, "Automatic Document Metadata Extraction Based on Deep Networks," in *Natural Language Processing and Chinese Computing (NLPCC 2017)*, Lecture Notes in Computer Science (LNCS 10619), X. Huang et al., Eds. Cham, Switzerland: Springer, 2018, pp. 305–317, doi: 10.1007/978-3-319-73618-1_26.
https://doi.org/10.1007/978-3-319-73618-1_26