

ABSTRAK

PT Nusa Raya Cipta Tbk masih menggunakan sistem pengarsipan dokumen secara manual, yang menghambat efisiensi serta aksesibilitas dalam pengelolaan administrasi perusahaan. Untuk mengatasi permasalahan tersebut, penelitian ini bertujuan untuk mengembangkan sistem arsip digital guna meningkatkan efisiensi serta kemudahan akses terhadap dokumen. Sistem yang dikembangkan berbasis Machine Learning dan mengintegrasikan Optical Character Recognition (OCR) untuk mengekstraksi teks dari dokumen fisik, serta Natural Language Processing (NLP) untuk menerjemahkan teks dari Bahasa Indonesia ke Bahasa Inggris secara otomatis. Implementasi OCR dilakukan dengan menyesuaikan konfigurasi serta memanfaatkan regex untuk mendeteksi pola teks tertentu, guna memastikan ekstraksi data yang lebih akurat dan sesuai dengan kebutuhan sistem. Pengembangan sistem menggunakan metode Kanban secara iteratif, memungkinkan fleksibilitas tinggi dalam menyesuaikan perubahan serta penyempurnaan fitur secara berkelanjutan. Proses pengembangan mencakup integrasi OCR dengan regex untuk otomatisasi pengarsipan dokumen dan pemrosesan NLP untuk penerjemahan bahasa. Hasil penelitian ini menunjukkan bahwa sistem arsip digital yang dikembangkan mampu mengubah dokumen fisik menjadi format digital yang dapat diakses, diindeks, dan dianalisis secara optimal. Sistem ini meningkatkan efisiensi pencarian serta pengelolaan dokumen, sehingga mendukung peningkatan produktivitas dan efektivitas dalam administrasi perusahaan.

Kata Kunci: Digitalisasi Arsip, Optical Character Recognition (OCR), Natural Language Processing (NLP), Machine learning, Penerjemahan Otomatis, Regex, Pengelolaan Dokumen, Kanban