

ABSTRAK

Ridvy Avyodri (01679210007)

OTOMASI PEMERIKSAAN *INVOICE* DALAM TRANSAKSI *LETTER OF CREDIT* YANG TUNDUK PADA UCP 600 MENGGUNAKAN OCR DAN *RULE-BASED TEXT MATCHING* (nomor romawi + 89 halaman; 25 gambar; 1 tabel; 0 diagram; 4 lampiran)

Letter of Credit (L/C) adalah produk *trade finance* yang cukup populer digunakan di seluruh dunia. L/C adalah jaminan pembayaran bersyarat dari bank penerbit kepada *beneficiary* atau penjual. Syarat dari jaminan ini adalah presentasi dokumen yang sesuai dengan ketentuan L/C dan peraturan-peraturan terkait seperti UCP 600. Faktur penjualan atau *invoice* adalah suatu jenis dokumen yang paling umum disyaratkan dalam transaksi L/C. Dalam praktiknya, pemeriksaan *invoice* terhadap L/C dan UCP 600 masih dilakukan secara manual. Hal ini mempengaruhi keakuratan, biaya dan waktu yang digunakan untuk proses pemeriksaan tersebut. Oleh karena itu, proses otomasi pemeriksaan *invoice* menjadi penting untuk dilakukan.

Penelitian ini mengajukan suatu alat dan proses kerja yang dapat mengotomasi proses pemeriksaan *invoice* terhadap L/C. Proses kerja ini terdiri dari 3 tahap utama yaitu, proses ekstraksi data dari L/C yang berbentuk teks, proses ekstraksi data dari *invoice* yang berbentuk gambar, dan proses pencocokan data L/C dan *invoice* menggunakan *Rule-based Matching*. Beberapa Teknik yang digunakan dalam penelitian ini adalah *Rule-based Named Entity Recognition* (NER) untuk ekstraksi data dari L/C dengan format teks, *Connected Component Analysis* untuk mendeteksi posisi tanda tangan pada *invoice*, *Hough Transform* untuk meluruskan gambar *invoice*, *Gaussian Blur* untuk mengurangi *noise* pada gambar dan Tesseract OCR 5.0 sebagai modul OCR yang digunakan.

Untuk menguji efektivitas dari proses kerja yang diajukan, eksperimen dilakukan dalam 3 tahap. Pertama, ekstraksi data dari teks L/C. Data yang digunakan dalam tahap ini adalah 47 L/C dalam format .txt. Kedua, ekstraksi data dari gambar *invoice*. Data yang digunakan dalam tahap ini adalah 16 *invoice* dengan 13 format yang berbeda. Ketiga, pencocokan data L/C dan *invoice* menggunakan *Rule-based Matching*. Data yang digunakan pada tahap ini adalah data yang telah diperoleh dari tahap-tahap sebelumnya.

Hasil dari eksperimen dalam penelitian ini adalah 100% keakuratan untuk proses ekstraksi data dari L/C menggunakan daftar perusahaan, rata-rata 90.62% keakuratan untuk proses ekstraksi data dari gambar *invoice*, dan 83.75% keakuratan untuk proses pencocokan data L/C dan *invoice* menggunakan *Rule-based matching*.

Kata kunci: *Rule-based Matching*, *Named Entity Recognition*, *Optical Character Recognition*, Otomasi

46 Referensi (1992 – 2021)

ABSTRACT

Ridvy Avyodri (01679210007)

AN AUTOMATION OF INVOICE EXAMINATION UNDER LETTER OF CREDIT BASED ON UCP 600 USING OCR AND RULE-BASED TEXT MATCHING (nomor romawi + 89 pages; 25 figures; 1 table; 0 diagram; 4 attachments)

Letter of Credit (L/C) is a trade finance product that is popularly used worldwide. L/C is a conditional guarantee of payment from issuing bank to the seller. The condition that has to be met is presentation of documents which complies with the L/C and every applicable rule such as UCP 600. Invoice is one of the most common documents that mostly required in L/C conditions. In current practices, compliance of the invoices is examined manually. These practices affect the accuracy, cost, and time efficiency of the examination considering the high L/C transaction volume. The needs to automate this examination process appear due to this background.

This research proposed a tools pipeline that automate the invoice examination process. Three main processes proposed are; Text data extraction from L/C, Data extraction from invoice, and rule-based matching between pre-extracted data. Several techniques are used for the proposed processes namely, rule-based named entity for text data extraction from L/C, connected component analysis for determination of signature location, Hough transform for de-skewing the image, and Gaussian blur for minimization noise. Tesseract OCR 5.0 is used for OCR engines.

To test the effectiveness of proposed method, experiment is done in 3 processes. First, text data extraction from L/C. Dataset used in this step are 47 L/Cs in text format. Second, data extraction from invoice images. Dataset used in this step are 16 invoices with 13 different formats. Third, rule-based matching between previously extracted data. Dataset used in this step are extracted data from previous step.

Experiments results are 100% accuracy for text data extraction from L/C using gazetteers, average of 90.62% accuracy for data extraction from invoice images and average of 83.75% accuracy for rule-based matching between pre-extracted data.

Keywords: Rule-based Matching, Named Entity Recognition, Optical Character Recognition, Automation

46 References (1992 – 2021)