



Getafix
TECHNOLOGIES

Whitepaper on

Invoice Validation using OCR based automated system

● Executive Summary:

This whitepaper is an analysis of the challenges in developing an invoice validation product for a multinational software company and how technology can significantly reduce time and improve efficiency.

The Client is a North American based leading cloud platform and cognitive solutions company. The company planned to automate a few of its processes and under them invoice validation was identified as one which was proving to be error prone. All invoices were received via email by the invoice validation team comprising of over 400 members. The workflow for how invoices enter the system is as described below:

Procurement department floats tenders for requirements and the same is fulfilled by registered vendors who then place a bids on the same. The winning bid would be approved and the vendor would be issued a Purchase Order (PO) from the procurement division. All purchases made by the company were to be accompanied by an invoice, following which the invoice management team had to mirror the purchase order (PO) with the invoice and manually check for below details

- 1. If the invoice had the items as per PO**
- 2. The delivery address was correct**
- 3. The pricing particulars were matching**

The Client hence wanted to have a tool built that could reduce redundancies and introduce automation to make this process faster and error free.

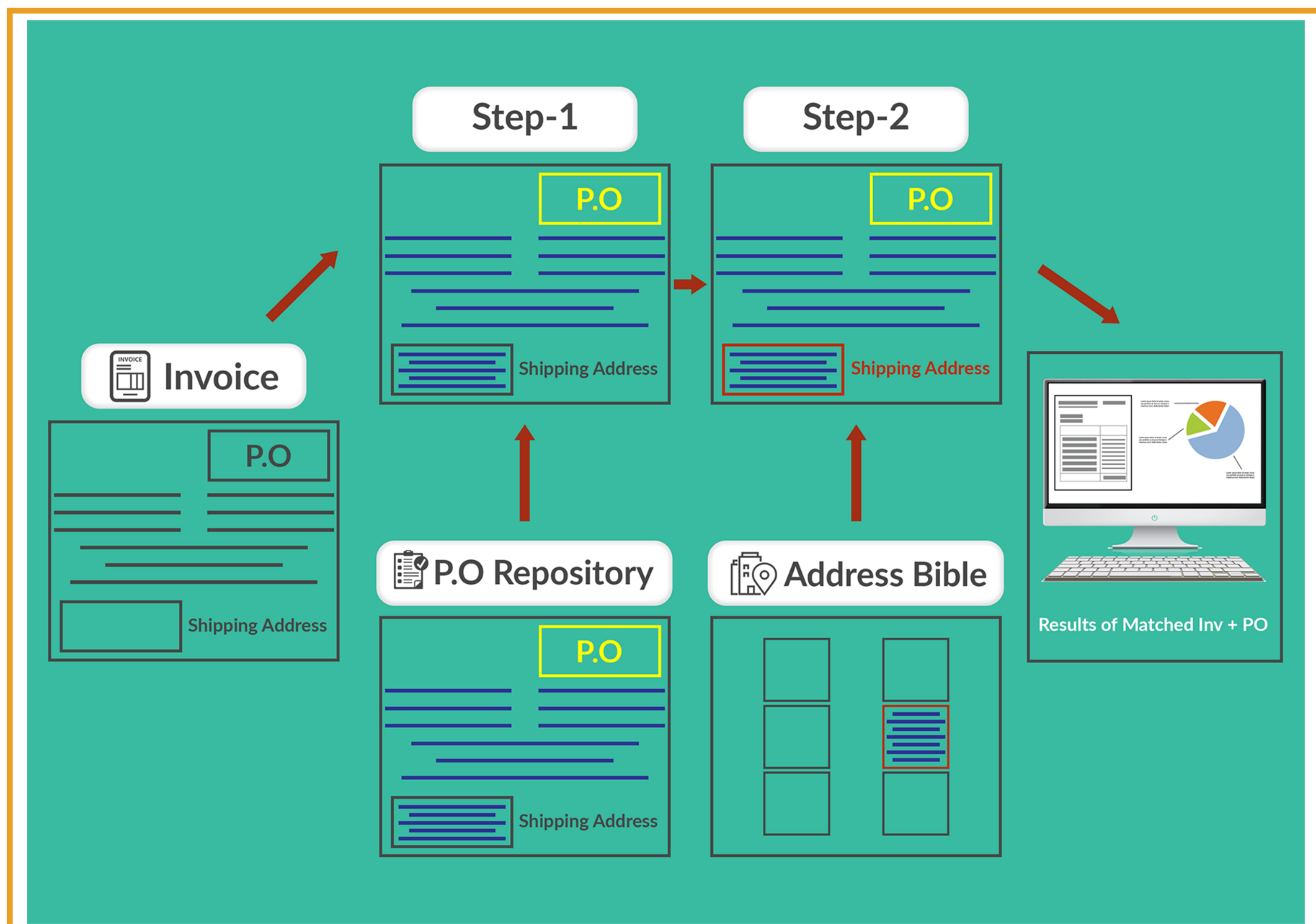
● Task Details:

- ✓ Create single file format for all invoices
- ✓ Building an automated tool which could match the PO with the Invoice
- ✓ Define rules within the tool to fetch the PO number

● Proposed Solution:

Most document verification products under the RPA banner are slow and very error prone. In this scenario, an Optical character recognition (OCR) logic alone would not be able to perform all tasks as desired.

For example, the shipping address has to be rightly matched with the Client's different office locations. The pincode in the address has to be a perfect match.



Owing to these challenges, a python-based document verification and authentication system was built which trumps up all RPA's. The invoice file formats of each vendor were different and had their own unique layout. The first task would be to convert all invoices into a single format, when an invoice came into the system, our python application picked it up and converted it into PDF if it was not in the format on a real time basis. The pdf mining was performed across 3 layers. A highly qualified optical character recognition engine called Tesseract was used which performs data extraction with the help of tags and provided filtered data. Followed by which a heatmap like hOCR was deployed that would scan the document, this application picked up the PO number and did an exhaustive search for the PO in the PO repo, if a PO was found, data points were extracted from the PO and matched with the filtered data from the invoice, once all the data matches were done, we needed to validate the shipping address. The client had an address bible where they stored all their office location address, the tool would extract the shipping address from the invoice and find a 100% match from the address bible. Once a match was found, the invoice was considered valid and an auto generated mail was sent to invoice management team for further action. In situations where any of these checks failed or the proper data points were not found, we provided an intuitive dashboard where all such invoices were listed and the closest matching PO and shipping address were shown. With the help of manual intervention, these invoices were either sent back to the supplier for correction and were put back into the system.

The product we have designed and developed performs an activity that giant killing RPA's fail to perform in a cost-effective manner. If a document needs to be compared and verified for particulars corresponding to a parent document this is done with super ease using our parent-child logic for document authentication. This tool would read, verify, ascertain and report back on a dashboard the authentication levels of any document within seconds.



| P.O Number | Invoice Number | Matching Rate | Remarks |
|------------|----------------|---------------|-------------------------------|
| 10 | 68 | 86% | Partial Match |
| 35 | 46 | 62% | Shipping Address not Matching |
| 60 | 52 | 100% | Sent for Clearance |
| 72 | 88 | 100% | Sent for Clearance |
| 55 | 23 | 68% | Shipping Address not Matching |
| 19 | 57 | 92% | Partial Match |
| 84 | 44 | 100% | Sent for Clearance |
| 97 | 74 | 81% | Partial Match |

● Conclusion:

- ✓ The company now has a robust and easy to use product
- ✓ Our System is capable of processing 100's of invoices per hour
- ✓ The tool offers more efficiency to the organisation

