Electronic Invoicing for the supply Chain Management

An Engineering Project in Community Service

Phase – II Report

Submitted by

1. **18BCE10047-Anupriya Shroti**
2. **18BCE10052-Anushka Shukla**
3. **18BCE10234-Sameer Bhadoriya**
4. **18BCG10080-Saksham Goyal**
5. **18BCG10085-Shivam Shakya**
6. **18BCY10002-Aakash Vats**
7. **18BCY10007-Adil Mustafa Khokhawala**
8. **18BCY10019-Ashutosh Bisawa**
9. **18BCY10031-Devyani Senwar**
10. **18BCY10107-Tanmay D. Shelat**

in partial fulfillment of the requirements for the degree of

Bachlore of Engineering and Technology

VIT Bhopal University
Bhopal
Madhyapradhesh

<< May, 2021>>

VIT® BHOPAL
www.vitbhopal.ac.in

Bonafide Certificate

Certified that this project report titled **"Electronic Invoicing for the Supply Chain Management"** is the bonafide work of 18BCE10047-Anupriya Shroti, 18BCE10052-Anushka Shukla, 18BCE10234-Sameer Bhadoriya, 18BCG10080-Saksham Goyal, 18BCG10085-Shivam Shakya, 18BCY10002-Aakash Vats, 18BCY10007-Adil Mustafa Khokhawala, 18BCY10019-Ashutosh Bisawa, 18BCY10031-Devyani Senwar, 18BCY10107-Tanmay D. Shelat who carried out the project work under my supervision.

This project report (Phase II) is submitted for the Project Viva-Voce examination held on …………..

Supervisor

Comments & Signature (Reviewer 1)

Comments & Signature (Reviewer 2)

1. INTRODUCTION

1.1 Motivation

We are motivated to make this community project in order to help the individuals of the income tax departments , suppliers to generate E bills as they struggle with the traditional methods and often get fooled by the trickery of Legal Consultants etc. Filing out the income tax return is a complicated and tedious process and since most middle class vendors don't know technology much they have to consult a Chartered Accountant which charges heavily for it.

The invoicing technologies have been all over the market and many multi national companies like amazon and flipkart are already using them but the catch is softwares which are already existing are either very complex to operate or their subscription is very costly which forces a small businessman to get help from a CA hence the objective of our project is to help local vendors by assisting them in filling the Income Tax and making the process much easier. Hence we took it a step further and created an automated legal filing system. Notably, we are in the intention of helping these people to make the world a tech driven place where people don't get bullied by 3rd party sources and generate bills , form in a systematic way. Also it reduces the manpower that consumed a lot of time.

1.2 Objective

Our main objective is to provide the local vendors a fully automated and developed platform on which they can file their income tax return with a click of a button. For that we first had to learn how to process the invoices and how to fill the income tax return form. The government website has a pre-structured ITR form available as per the norms, we studied the form and decided which crucial information we have to take from the vendor's invoices.

In the first phase we achieved the pre-processing of the input, which will inevitably help us in extracting a more efficient output data. The OCR helped in the picture to strings conversion. And we had finalized a website prototype and how would it look to the customer.

The objective for the phase 2 is to run post-processing on the output data. This post-processing will help us filter out the crucial and particular information we need to fill out the ITR form. In this we are going to use modules from natural language processing, mainly regex function, which has a pre-build library in Python. We will create a regex format for each unique word and then run the function on the output, it will fetch us the distinct word or number that we are looking for. After fetching those words we need to save them in a secure database. This database is a structured CSV file managed by the back-end. With the help of Python function at last we will fill out the ITR form using these extracted data rom the invoices.

While the backend runs all these functions the fron-end website will be a user friendly guide on how to upload their invoices, we have given them the options to upload jpg, jpeg, png or pdf file as their invoice. Then they simply have to click a button and the backend will do all the work for them.

And to make it possible we will need to create a server that will connect the both front-end and the back-end. The server will link the both so that the file can be uploaded and downloaded from and to the server for the client.

2. Existing Work / Literature Review

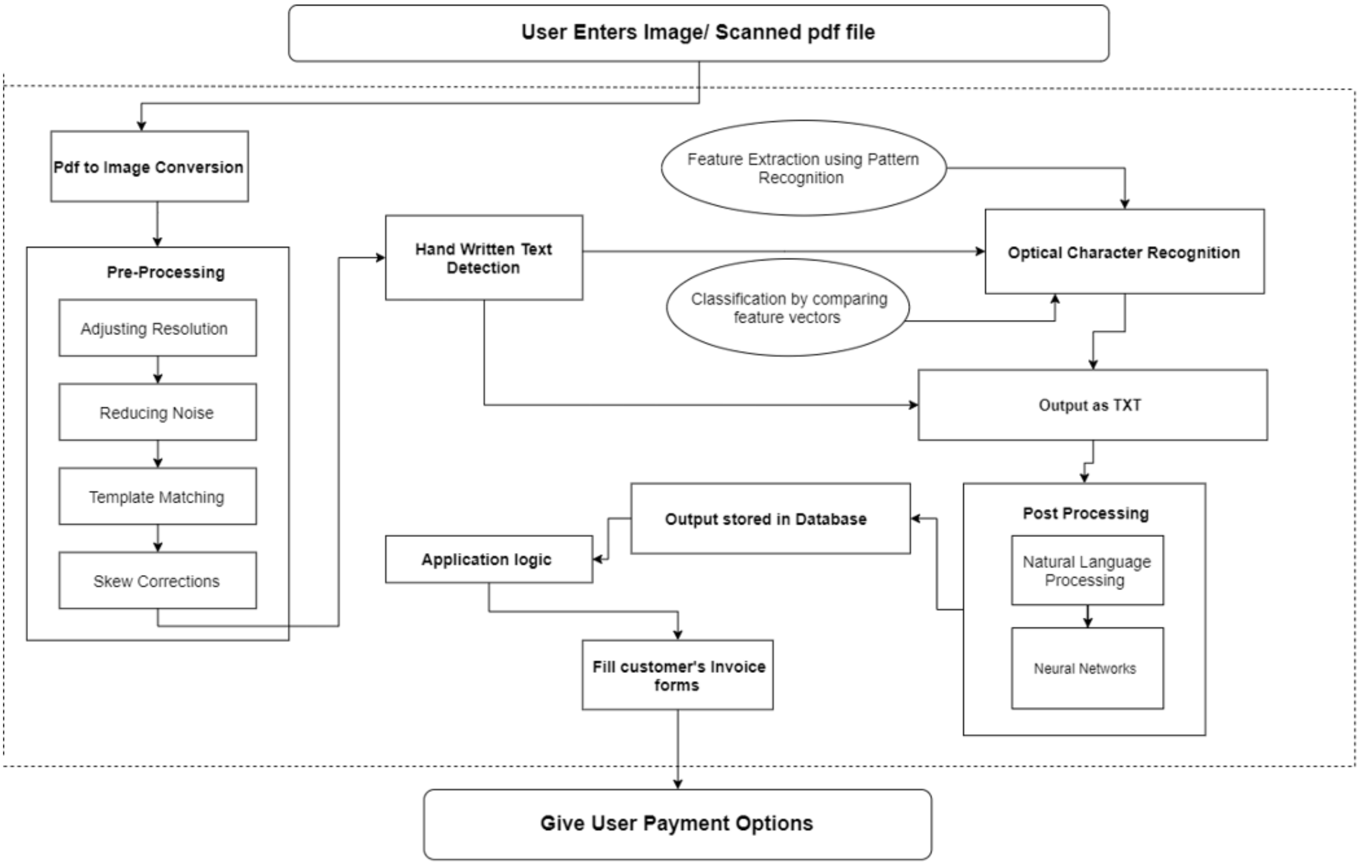
This application has two major concepts. First is using the OCR (Optical Character Reader) for reading the handwritten or machine generated invoices into a digital text.
The project mainly targets the local vendors which are present in the community.
The former projects provide all the facilities of OCR and storing the data on a server.
This successor will provide the OCR functionalities as well as you can directly store the data in a CSV file if you want, which can help the vendors to keep track of what products their customers need or what vendors need for themselves. It can shorten the long chain of multiple confirmations boosting the time efficiency.

Lexmark's Oracle Invoice Processing: Lexmark, a company specializing in high-end IT-solutions for established companies. The oracle invoice processing solution offers automatic scanning, data extraction, filing and payment of invoices. It also offers support for other business suites and ERPs, boasting a network of partners in FPA, Financial Process Automation, from over 70 countries. **One major drawback with Lexmark's solution is that it does require integration with the end-user business suite.** If that is not possible then Lexmark cannot offer automatic invoice processing for said end-user. Our application differs in this aspect by being standalone, while it is supposed to be able to communicate with a given business suite it does not attempt to integrate with the business software but rather attempts to tag and sort the information on the invoice and present it in such a way that the business suite can read it right away. Preferably by passing a file with a specific convention for tagging and sorting.

Fortnox: Fortnox provides a service named 'Fakturatolkning', which translates roughly to 'invoice interpretation'. The end-users of this service send invoices they want to have digitized to Fortnox. Once the invoice has arrived, physically or electronically, Fortnox applies their software solution onto the invoice to extract the vital information needed for bookkeeping from the invoice. The extracted information is validated by an employee at Fortnox after which it is inserted into the bookkeeping service offered by Fortnox. Fortnox chose to store an image of the original invoice paired with the digitized copy. As of today the current price for this service is 4.90 or 9.90 SEK, Swedish Krona, per invoice where the lower price is assuming the invoice is delivered electronically to Fortnox. **The major drawback of Fortnox's is that the actual processing is done at a third party company**, in this case Fortnox. Which leaves the possibility of a company renting this service having little to no idea of what products and services they've actually paid for. If there is a service that offers a one-time payment solution, then a service that is priced proportionally to the amount of invoices received will be undesirable for a load big enough.

3. Topic of the work

a) System Design / Architecture


User Enters Image/ Scanned pdf file
Pdf to Image Conversion
Pre-Processing
Adjusting Resolution
Reducing Noise
Template Matching
Skew Corrections
Hand Written Text Detection
Feature Extraction using Pattern Recognition
Optical Character Recognition
Classification by comparing feature vectors
Output as TXT
Post Processing
Natural Language Processing
Neural Networks
Output stored in Database
Application logic
Fill customer's Invoice forms
Give User Payment Options

b) Working Principle

1. PDF to Image conversion :
    a. Using python modules like PIL and ImageMagick we will convert the user PDF/JPG/PNG images into a standard JPEG format.

2. Pre-processing :
    a. On the JPEG format we will run a series of functions using Open-CV and numpy modules to make OCR efficient
        i. Grey Scaling
        ii. Resolution Adjustment

iii. Skew correction
iv. Template Matching
v. Noise Reduction

3. Optical Character Recognition
   a. JPEG to string conversion using pytesseract
   b. Output in txt format

4. Post-procssing
   a. Latural language processing for efficient feature extraction.
      i. RE : Regular Expressions to help find the unique key words we are looking for
5. Data extraction to a secure database
   a. Using CVE module in python extracting the User data to our Standardised Database format

6. Server Creation
   a. Using Flask module to build a server and host a website
   b. Using GET and POST methods to download and upload the desire files

c) Results and Discussion :

We have completed the frontend and backend prototypes , i.e entire working of the project modules that include the working of OCR , generation of output text files in .txt format. Along with this all data is been extracted through regex and later the files are converted into **.csv** format and then put into a ITR pdf.

Notable , we have completed the frontend website too where input files are easily browsed and later these files are uploaded to the backend server for final downloadable pdf file. The server is also been generated in order to integrate both front-end and back-end work . Hosting of website and sending requests through GET and POST are been implemented after which the final file is been sent to the website to get desirable pdf output.

This can be considered the basic version of the project, there are many scope of improvemnets and further research work in the field of computer vision and automated filing systems. This project was just a small effort to give back to the community, these hard times have affected them the most.

d) Individual Contribution by members :

1. Anushka Shukla
    a. Worked on the project objective
    b. Project modules recognition
    c. Built and assembled Architecture diagram
    d. Set up the server using various Python modules
    e. Deployed the website on the server
    f. Linked the front-end and back-end
    g. Made python functions for uploading and downloading files
    h. Documentation and presentation making

2. Tanmay D. Shelat
    a. Worked on project objective
    b. Project modules recognition
    c. Working methodology of the project
    d. Implementation of each module in back-end
    e. Put together a software prototype for Phase 1
    f. Coding each module function in Python
    g. Completing and finalizing the back-end code
    h. Documentation and presentation making

3. Saksham Goyal
    a. Worked on the aim and purpose
    b. Front-end ideation
    c. Researched on the problems faced by current methodologies
    d. Put together a website protptype for Phase 1
    e. Worked with App UI Design
    f. Documentation and presentation making

4. Sameer Bhadoriya
    a. Worked on the front-end ideation
    b. Implemented various web designing modules
    c. Put together HTML and CSS code for the website
    d. Worked on the website UI Designs

5. Ashutosh Bisawa
    a. Researched about the existing technologies
    b. Researched on the limitation faces by the current trends
    c. Worked on integrating the back-end modules
    d. Back-end modules testing

6. Anupriya Shroti
    a. Worked on the front-end implementation
    b. Recognition of Hardware and Software requirements
    c. Put together various web designing modules
    d. Documentation and editing
    e. Presentation making

7. Shivam Shakya
    a. Researched on the history of OCR
    b. Comprehension of existing projects in the field
    c. Worked on the fron-end implementation
    d. Helped on the front-end modules integration
    e. Documentation

8. Adil Mustafa Khokhawala
    a. Documentation
    b. Phase 1 and phase 2 report generation
    c. Literature Review

9. Devyani Senwar
    a. Documentation
    b. Phase 1 and phase 2 report generation
    c. Abstarct from the research paper on OCR

10. Aakash Vats
    a. Documentation
    b. Phase 1 and phase 2 report generation
    c. Abstract from the Post-processing applications

4. CONCLUSION

From the phase 1 till now we have concluded the following modules.

- Full working of backend modules
    - Pre processing
    - OCR
    - Output text file generation (.txt)
    - Succesful data extraction through regex
    - Filling data in csv format
    - Turning csv file into desirable file format output

- Front-end website completion
    - Browsing the input file.
    - Uploading file to server

  - Sending the uploaded file to backend
  - Operation execution on input files
  - Final output file is ready to download

- Server generation
  - Backend and frontend integration
  - Hosting the website
  - Using GET and POST method to transfer files from frontend to backend
  - Sending final file to website
  - Downloadable file is ready.

All the project modules are successfully completed and executed fulfilling the purpose of this project.

5. Reference: (Minimum 1 Page)

1.https://ieeexplore.ieee.org/document/8576202
2.https://ieeexplore.ieee.org/document/9151144
3.https://link.springer.com/chapter/10.1007/978-3-642-23777-5_22
4.Automatic reading and interpretation of paper invoices (UPPSALA UNIVERSITY)