

Implementation & Integration of a Custom Optical Character Recognition (OCR) Module in an Incentivised Blockchain-based Document Verification Framework

Supervised by

Dr. Nirnay Ghosh

Team Members

Neeraj Pratap Hazarika (2020CSB040)

Vinita Ramdular Yadav (2020CSB026)

Niket Bahety (2020CSB074)

Eliza Debbarma (2020CSB042)

Navya Jain (2020CSB098)

Sudireddy Shravya (2020CSB073)

Acknowledgement

We would like to express our deep gratitude to Professor Nirnay Ghosh, Department of Computer Science and Technology, Indian Institute of Engineering Science and Technology, Shibpur, our project supervisor, for his patient guidance, enthusiastic encouragement and helpful critiques for this project.

We would like to thank Utsa Roy for his guidance. We would also like to thank all the faculty members and staff of the Department, as well as our classmates, for their help and support.

CONTENTS

Abstract.....	4
Introduction.....	4
Overview of the Platform.....	4
Essential Features of the Platform.....	4
Benefits of the Platform.....	5
Motivation.....	6
Proposed Methodology.....	7
Workflow.....	8
Result.....	10
Conclusion.....	11

Abstract

This report outlines the development of a blockchain-based personal document and information-sharing platform tailored for job markets. The platform is a permissioned (private) blockchain where companies and institutions register themselves as peers to exchange data securely. It includes document parsing capabilities for categorizing uploaded documents (e.g., mark sheets, PAN cards, Aadhar cards). When an organization seeks verification from another peer, a reward mechanism is triggered upon successful verification, facilitated by an integrated payment gateway. This project streamlines the verification process in job markets, enhancing trust, efficiency, and security, ultimately fostering smoother hiring processes and reducing fraudulent activities.

Introduction

In response to the evolving needs of modern job markets, the blockchain-based personal document and information-sharing platform introduces innovative features designed to revolutionize the verification process. By seamlessly integrating parsing capabilities and a dynamic reward mechanism into our platform, we aim to enhance efficiency, trust, and security within the employment ecosystem. Through this platform, organizations can securely exchange and verify crucial documents while incentivizing collaboration and participation among peers. This transformative approach not only streamlines hiring processes but also empowers individuals with greater control over their personal data.

Overview of the Platform

The proposed blockchain-based personal document and information-sharing platform for job markets is a permissioned (private) blockchain where companies and institutions must register as peers to exchange data. The platform automates document categorization through advanced parsing, eliminating manual input. Additionally, it incentivizes document verification by rewarding peers via an integrated payment gateway. This streamlined process enhances trust and efficiency in job markets, revolutionizing document verification procedures.

Essential Features of the Platform

1. Registration of Companies and Institutions:

The first essential feature of the platform is the registration of companies and institutions. This feature ensures that only authorized peers can exchange data. The registration process includes a verification process to ensure that the participating companies and institutions are legitimate entities.

2. One-Time Submission of Documents:

Once a document is submitted to the platform, it does not need to be given again to another company while switching jobs, provided the latter is a part of this platform. This feature reduces the need for repeated submissions of the same documents, saving time and effort for job seekers and reducing the verification task for companies.

3. Consent-Based Data Sharing:

The platform ensures that until the owner of the information gives consent, their data cannot be shared by any other participating agency. This feature ensures that individuals have control over their personal information and prevents unauthorized access to data.

4. Validation of Certificates:

The platform includes a mechanism to validate the provided certificate if the institute or company which issues the certificate is a part of the system. This feature ensures the accuracy and authenticity of the certificates provided.

5. Document Parsing:

The platform offers automated Document Parsing, which categorizes uploaded documents based on their type. This feature ensures efficient management and retrieval with utmost accuracy and reliability.

6. Reward System:

The Reward System incentivizes participation and cooperation among peers by offering rewards for successful document verifications. This feature ensures a fair and transparent process that fosters trust and collaboration.

7. Integrated Payment Gateway:

The platform has Integrated Payment Gateway that facilitates seamless transactions for reward distribution, guaranteeing secure and efficient payment processing, thereby enhancing the platform's usability and scalability while ensuring financial integrity and security.

8. Immutability of Data:

The platform ensures that nobody can modify or change the data once verified. The immutability of data ensures the integrity of the information and prevents fraudulent activities.

Benefits of the Platform

1. Enhanced Efficiency:

The platform reduces the time and effort required to verify and exchange personal documents and information, enhancing the efficiency of the recruitment process.

2. Cost-Effective:

The platform reduces the cost of repeated verification tasks for companies, making the recruitment process cost-effective.

3. Enhanced Security:

The platform ensures the confidentiality and privacy of personal information and prevents unauthorized access, enhancing security for individuals and companies.

4. Improved Transparency:

The platform ensures transparency in the verification and exchange of personal information, improving trust between individuals and companies.

Motivation

The proposed platform, a blockchain-based permissioned network for document sharing and authentication in job markets, addresses significant challenges in the current recruitment process. Motivated by the inefficiencies and risks associated with traditional methods, our platform aims to streamline document exchange while enhancing security and privacy.

Switching jobs often entails repetitive sharing of personal data and documents, leading to unnecessary costs and time consumption for both individuals and companies. This practice not only burdens candidates with manual verification tasks but also raises concerns about data privacy and security breaches.

To tackle these issues, our platform introduces innovative solutions to automate document categorization and streamline verification processes. By incentivizing peer cooperation and integrating secure payment mechanisms, we aim to foster trust, efficiency, and transparency within the recruitment ecosystem.

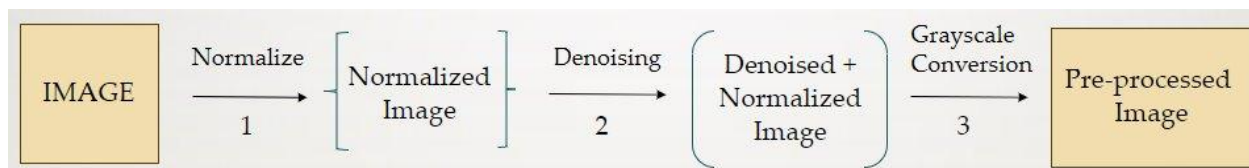
Through these efforts, our platform seeks to provide a secure, efficient, and cost-effective solution to the challenges plaguing the current recruitment process, ultimately benefiting individuals and companies alike.

Proposed Methodology

The proposed methodology for the implementation and integration of a Custom Optical Character Recognition (OCR) Module in a Blockchain-based Document Verification Framework involves several steps:

1. Image Pre-processing : Pre-processing, including normalization, denoising, and grayscale conversion, collectively enhancing OCR accuracy by improving text clarity and contrast, is essential to optimize document images for accurate OCR and parsing.

Following is the flow chart of pre-processing :



1. Color Normalization :

- Color normalization in computer vision adjusts the color distribution of an image to reduce variability caused by lighting conditions, cameras, and other factors, enabling more effective object recognition based on color despite these variations.

2. Non-Local Means Denoising algorithm :

- It is a method used to remove noise from images while retaining fine details in images, including textures and edges unlike other denoising algorithms which might blur these features.

3. RGB to grayscale :

- The conversion from RGB (Red, Green, Blue) to grayscale is a process that transforms a color image into shades of gray.

2. Integration of OCR Technology: OCR converts scanned document images into machine-readable text, enabling computers to interpret and analyze the content. The integration of OCR facilitates accessibility and searchability of document content by converting scanned images into searchable text. This functionality enables users to quickly locate relevant information within documents, enhancing usability and user experience.

We tested the speed and effectiveness of four distinct OCR models : Keras, PyTesseract, Easy OCR, and DocTR in handling diverse document types for optimal OCR selection and conducted extensive testing on GPU servers.

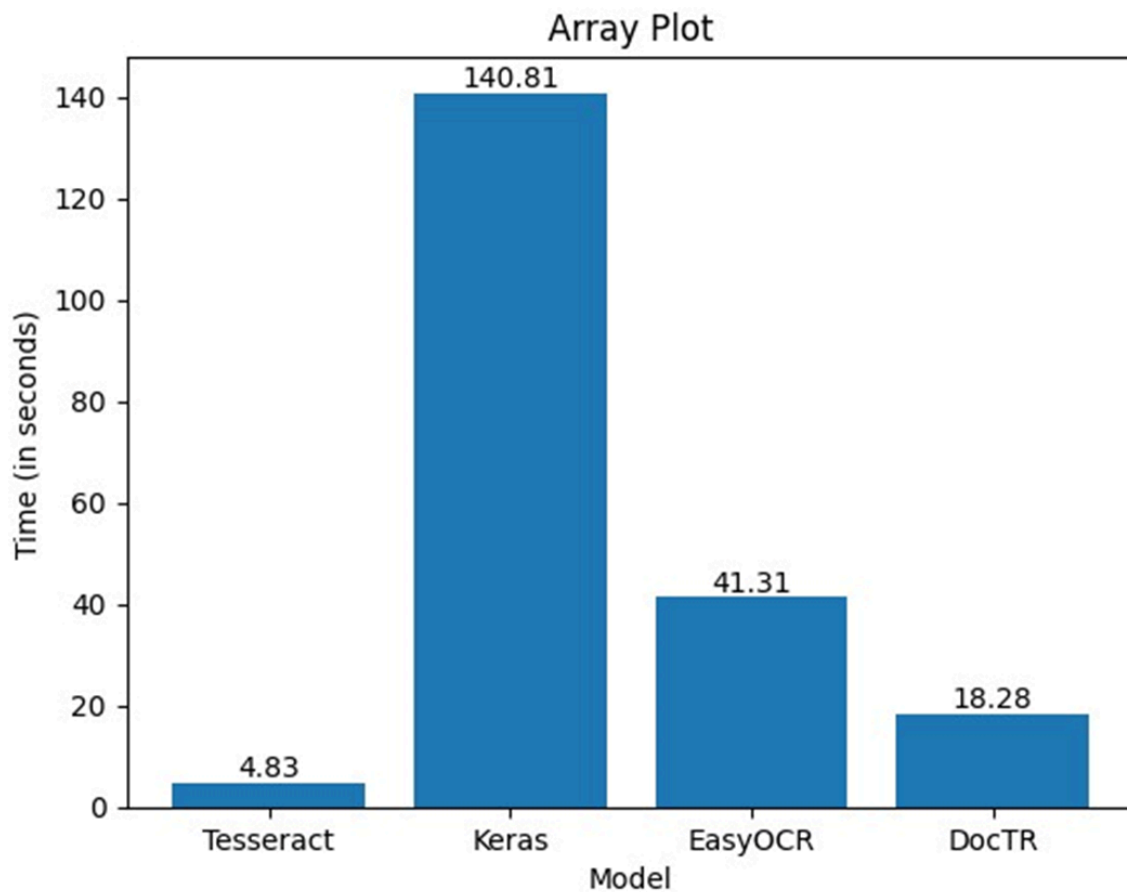


Fig: A comparative graph showcasing the speed performance of different OCR models (Keras, Tesseract, EASY OCR, DocTR) on the same image.

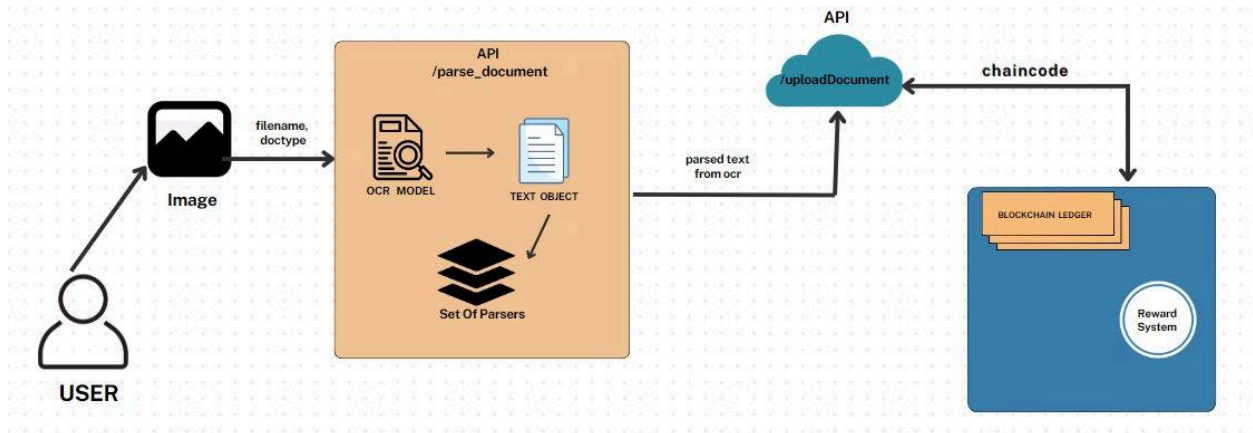
Selected DocTR as the best model.

3. Designing Specialized Parsers for Document Classification: We developed specialized parsers designed to extract relevant information from the OCR text based on the specific requirements of the document type.

4. Implementing Verification and Reward System for blockchain: Organizations can request document verification from other organizations within the blockchain framework.

- When an organization requests verification, it earns credits as part of a reward system based on the number of requests made.
- The verification process involves validating the document against the original in the Issued Document Database (IDD) by the endorser peer of the verifying organization.
- This system incentivizes organizations to participate in document verification and contributes to the overall integrity and security of the blockchain network.

Workflow



The workflow begins with a user uploading an image of a document (such as a marksheet, Aadhar card, or PAN card) through the command line interface. The uploaded image is then sent to the OCR model using an API endpoint that also serves for parser selection.

Upon receiving the image, the OCR model processes the document image to extract text from it, converting the textual content within the image into machine-readable text.

Following the OCR text extraction, the extracted text is subsequently sent to the same API endpoint where the user specifies the type of document they have uploaded (e.g., marksheet, Aadhar card, PAN card). This API dynamically determines and selects the appropriate parser designed specifically for the identified document type.

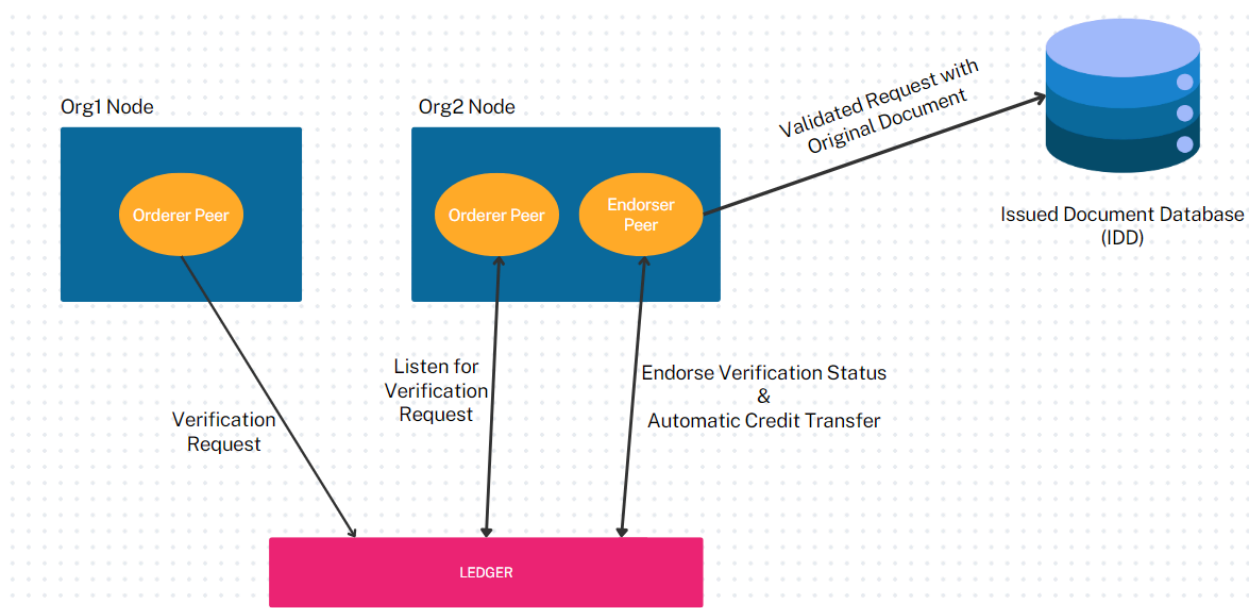
The selected parser then processes the extracted text, analyze and interpret it to retrieve specific information relevant to that document type. For example, the parser might extract fields like name, date of birth, address, identification numbers, or academic qualifications depending on the selected document.

After parsing the text and extracting the required information, the parsed data is formatted and sent to another API endpoint. This API acts as a gateway to a verification service responsible for validating the accuracy and authenticity of the parsed information against predefined rules or databases.

It is our assumption that the structure of the data in the organisation's database is similar to the parsed result we get from parser(key value pair) as we do not have access to peer's database.

Through chaincode which is an interface to interact with the blockchain , the parsed result will be uploaded on blockchain and will be verified by the issuing organization. upon validation and verification, after some time it will be updated on the node js server again if its verified or pending. And finally the organization will get credits by the added reward system .

Upon successful validation, the verified information is then updated or recorded in a blockchain network. The blockchain serves as an immutable ledger, securely storing the verified data, ensuring transparency, traceability, and data integrity.



Then after the validation, the verification will be done by the blockchain and in the blockchain we have a different number of nodes which represent different organizations and the blockchain ledger will try to listen for verification requests from the organization nodes.

In an organization node, the orderer peer will send a verification request and it is the one which listens for verification requests from other nodes and the endorser peer will be

responsible for sending the validated request of the given document with the original document to the issued document database.

The endorser peer after getting to know whether it is validated or not from the request that was sent to it will send the verification status that was obtained to that respective organization and also automatically there will be credit transfer from the requested organization to the blockchain. This is called the reward system.

Result

This is the result of the DocTR result of the CBSE marksheet:

DocTR RESULT	Output				
Input	COUNCIL FOR THE INDIAN SCHOOL CERTIFICATE EXAMINATIONS, NEW DELHI INDIAN CERTIFICATE OF SECONDARY EDUCATION (CLASS-X)-YEAR2018 No.TF 20161065 1958 2018 1185686/072 STATEMENT OF MARKS Name NIKET BAHETY UNIQUE ID 6812642 Son of Sunita	SUNITA BAHETY Shri RAJ KUMAR BAHETY of AGRASAIN BOYS' SCHOOL, HOWRAH CHOO SUBJECTS External Examination ENGLISH HINDI MATHEMATICS Percentage Marks 91 NINE ONE 95 NINE FIVE 92 NINE TWO 93 NINE THREE 96	NINE SIX 99 NINE NINE HISTORY, CIVICS & GEOGRAPHY (HCS-A,GEO- B). SCIENCE (PHY-A,CHE-A, BIO-A) COMPUTER APPLICATION S Internal Assessment SUPV & COMMUNITYS ERVICE Date of birth as Head of the School registration RESULT-	Grade A certified by the (in words) Ninth September Two Thousand Two at the time of (in figures) 09.09.2002 PASS & CERTIFICATE AWARDED Date of declaration of Result 14.05.2018 Note: 1.The pass mark for	each subject is 33% 2.No divisions are awarded. Jaites Chief Executive & Secretary (See Overleaf)

The below given code is the parsing algorithm for school certificate:

```
def extract_information(text):
    # Define regular expressions for different pieces of information
    roll_number_pattern = re.compile(r'Roll No[.-:~]? (\d+)')
    unique_id_pattern = re.compile(r'UNIQUE ID[.-:~]? (\d+)')
    name_pattern = re.compile(r'Name(?: of Candidate)?[.-:~]? (\w+ \w+)')
    school_pattern = re.compile(r'\nSchool[.-:~]?[\'"]? [a-zA-Z ]*[\'"]?\d*([A-Za-z,:\' ]+)')
    school_pattern2 = re.compile(r'of ([A-Za-z,\' ]+)')
    marks_pattern = re.compile(r'(\d{3}) ([A-Z &]+)\n(\d{3}) (\d{2,3})')
    issue_date = re.compile(r'date[a-zA-Z: ]* (([0-9]{12}|[0-9]{13}[01])[-.](0[1-9]|1[0,1,2])[-.](19|20)\d{2})', re.IGNORECASE)

    # Extract information using regular expressions
    roll_number_match = roll_number_pattern.search(text)
    unique_id_match = unique_id_pattern.search(text)
    name_match = name_pattern.search(text)
    school_match = school_pattern.search(text)
    school_match2 = school_pattern2.search(text)
    marks_matches = marks_pattern.findall(text)
    issue_date_match = issue_date.search(text)

    # Create a dictionary to store extracted information
    extracted_info = {
        'Roll Number / Unique ID': roll_number_match.group(1) if roll_number_match else unique_id_match.group(1) if unique_id_match else None,
        'Name': name_match.group(1) if name_match else None,
        'School': school_match.group(1).strip() if school_match else school_match2.group(1).strip() if school_match2 else None,
        'Marks': {subject: (theory, practical) for theory, subject, practical in marks_matches} if marks_matches else None,
        'Issue date': issue_date_match.group(1) if issue_date_match else None
    }

    return extracted_info
```

```
0071715
10986/00109
futtran
ta,,
0071715
Regn.No. A220/16492/0117
SRA
chgfa Hreafuche ARTEIT ais
CENTRALBOARD OFSECONDARY EDUCATION
31ch faarfucht
MARKS STATEMENT
efTR Foprt efefonchz uitenr, 2020
NIKETBAHETY
03.08.2010
SENIOR SCHOOLCERTIFICATE EXAMINATION, 2020
vedf 21 HTH
Name of Candidate NIKET BAHETY
STSHIA
Roll No. 11607640
HIAT 21 HTY
fat/erers 21 TTH
faaima
Mother's Name SUNITA BAHETY
Father's/Guardian's Name RAJ KUMAR BAHETY
School 10986 LAL BAHADUR SHASTRI SR SEC SCHOOL RANPUR KOTA RJ
HTHICS MARKS OBTAINED
Taga 2E
SUB.
CODE
fagy
SUBJECT
fearia 43
POSITIONAL
B1
C1
B1
B2
A2
A1
A2
B1
fefera ari. ATT
2T1
(sreat #)
EIGHTY SIX
SIXTY NINE
```

```
SIXTY NINE
EIGHTY TWO
SEVENTY SEVEN
NINETY ONE
THEORY HT.PR. IA/ TOTAL TOTAL (IN WORDS) GRADE
301 ENGLISH CORE
041 MATHEMATICS
042 PHYSICS
043 CHEMISTRY
048 PHYSICAL EDUCATION
500 WORK EXPERIENCE
502 HEALTH & PHYSICAL EDUCATION
503 GENERAL STUDIES
066
049
053
047
062
020 086
020 069
029 082
030 077
029 091
eferferat 21 3rf Abbreviations
AB: : argufara Absent
RP: VRTIr4A H grgfr Repeat in Practical
RT: fefera # yrrigfr Repeat in Theory
ER: : afrari yrigfr zrefat faget #
Essential Repeat in all subjects
URUITH Result PASS
rOwniy
uiterr frisc
Controller of Examinations
foreft Delhi
feach Dated: 13-07-2020
```

This is the Input (OCR extracted Text) for CBSE certificate.

Classified Text for CBSE certificate is:

```
{  
  
    "Issue date": "13-07-2020",  
  
    "Marks": null,  
  
    "Name": "NIKET BAHETY",  
  
    "Roll Number / Unique ID": "11667640",  
  
    "School": "LAL BAHADUR SHASTRI SR SEC SCHOOL RANPUR KOTA RJ"  
  
}
```

This is the parsing algorithm for identity documents (aadhar):

```
# Define regular expressions for different pieces of information  
name_pattern = re.compile(r'Name[\n\s]+(?:[\n\s]+(?:DOB|Year of Birth|DOB)', re.DOTALL | re.IGNORECASE)  
dob_pattern = re.compile(r'(DOB[:/\ ](\d{2}/\d{2}/\d{4})|\d{8}|\d{4})|Year of Birth*[:/\ ]*(\d{4})|DOB[:/\ ]*(\d{2}/\d{2}/\d{4})|\d{8}|\d{4}'))', re.DOTALL | re.IGNORECASE)  
gender_pattern = re.compile(r'(MALE|FEMALE|TRANSGENDER|OTHERS?)[\n\s]+(?:[\n\s]+(\d{4} \d{4} \d{4}))?', re.DOTALL | re.IGNORECASE)  
aadhar_number_pattern = re.compile(r'(\d{4} \d{4} \d{4})', re.DOTALL)
```

```
def extract_information(text):  
    # Define regular expressions for different pieces of information  
    name_pattern = re.compile(r'HTA Name[\n\s]+(?:[\n\s]+(?:DOB|Year of Birth|DOB)', re.DOTALL | re.IGNORECASE)  
    dob_pattern = re.compile(r'(DOB[:/\ ](\d{2}/\d{2}/\d{4})|\d{8}|\d{4})|Year of Birth*[:/\ ]*(\d{4})|DOB[:/\ ]*(\d{2}/\d{2}/\d{4})|\d{8}|\d{4}'))', re.DOTALL | re.IGNORECASE)  
    gender_pattern = re.compile(r'(MALE|FEMALE|TRANSGENDER|OTHERS?)[\n\s]+(?:[\n\s]+(\d{4} \d{4} \d{4}))?', re.DOTALL | re.IGNORECASE)  
    aadhar_number_pattern = re.compile(r'(\d{4} \d{4} \d{4})', re.DOTALL)  
  
    # Extract information using regular expressions  
    name_match = name_pattern.search(text)  
    dob_match = dob_pattern.search(text)  
    gender_match = gender_pattern.search(text)  
    aadhar_number_match = aadhar_number_pattern.search(text)  
  
    lines = text.split('\n')  
  
    prev_line = ""  
    name = ""  
  
    for i in range(1, len(lines)):  
  
        dob_match1 = dob_pattern.search(lines[i])  
        if dob_match1:  
            name = lines[i-1]  
  
    # Extracted information  
    extracted_info = {  
        'Name': name,  
        'Date of Birth': dob_match.group(2) or dob_match.group(3) or dob_match.group(4) if dob_match else None,  
        'Gender': gender_match.group(1).strip() if gender_match else None,  
        'Aadhar Number': aadhar_number_match.group(1).replace(" ", "") if aadhar_number_match else None,  
    }  
}
```

Input (OCR extracted Text)
for Aadhar card

-
TTava
9819 HSSPS
Government of India
wetw aet
1s5 39TT 251801
Neeraj Pratap Hazarika
SWSFAN/DO8:24012002
TPA/N Male
3967 9731 1111
aTais 9IfSb3
afais Gtais

Classified Text for
Aadhar card

Information from Aadhar:
{'Name': 'Neeraj Pratap Hazarika ',
'Date of Birth': '24012002',
'Gender': 'Male',
'Aadhar Number': '39679731111'}

The below given is the parsing algorithm for identity documents(PAN):

```
def extract_information(text):  
    # Define regular expressions for different pieces of information  
    name_pattern = re.compile(r'Name[\n\s]+(.*?)?:\n|$', re.DOTALL)  
    father_name_pattern = re.compile(r'Father\s Name[\n\s]+(.*?)?:\n|$', re.DOTALL)  
    pan_number_pattern = re.compile(r'Permanent Account Number Card[\n\s]+(.*?)?:\n|$', re.DOTALL)  
    dob_pattern = re.compile(r'Date of Birth[\n\s]+(.*?)?:\n|$', re.DOTALL)  
  
    # Extract information using regular expressions  
    name_match = name_pattern.search(text)  
    father_name_match = father_name_pattern.search(text)  
    pan_number_match = pan_number_pattern.search(text)  
    dob_match = dob_pattern.search(text)  
  
    # Extracted information  
    extracted_info = {  
        'Name': ' '.join(name_match.group(1).strip().split()) if name_match else None,  
        'Father\s Name': ' '.join(father_name_match.group(1).strip().split()) if father_name_match else None,  
        'PAN Number': pan_number_match.group(1) if pan_number_match else None,  
        'DOB': dob_match.group(1) if dob_match else None,  
    }  
  
    return extracted_info
```


Input (OCR extracted Text)
for PAN card

314a
fAHTST
HRA
RCOR
INCOME TAX DEPARTMENT
GOVT. OF INDIA
TeITAT aar TSAT 2618
Permanent Account Number Card
BHFPY3100C
HTA Name
VINITA RAMDULAR YADAV
furnr ET "TH/Father's Name
RAMDULAR YADAV
FE aht ata
Date of Birth
17/01/2003
Nieis
EeiTat/Signatu

Classified Text for
PAN card

Information from Pancard:

{'Name': 'VINITA RAMDULAR YADAV',
"Father's Name": 'RAMDULAR YADAV',
'PAN Number': 'BHFPY3100C',
'DOB': '17/01/2003' }

The given below is the output of the ICSE certificate:

COUNCIL FOR THE INDIAN SCHOOL CERTIFICATE EXAMINATIONS, NEW DELHI
INDIAN CERTIFICATE OF SECONDARY EDUCATION (CLASS-X)-YEAR2018
No. TF 20161065
1958
2018
Education
1185686/072
Pioneering
STATEMENT OF MARKS
Name NIKET BAHETY
UNIQUE ID 6812642
Son of
Smt. SUNITA BAHETY
Shri RAJ KUMAR BAHETY
of AGRASAIN BOYS' SCHOOL, HOWRAH
SUBJECTS
External Examination
ENGLISH
HINDI
MATHEMATICS
Percentage Marks
01 NINE ONE
05 NINE FIVE
02 NINE TWO
03 NINE THREE
06 NINE SIX
09 NINE NINE
HISTORY, CIVICS & GEOGRAPHY (HCS-A, GEO-B)
SCIENCE (PHY-A, CHE-A, BIO-A)
COMPUTER APPLICATIONS
Internal Assessment
SUPW & COMMUNITYSERVICE
Date of birth as
Head of the School
registration
RESULT-
Grade
A
certified by the (in words) Ninth September Two Thousand Two
at the time of (in figures) 09.09.2002
PASS CERTIFICATE AWARDED
Date of declaration of Result 14.05.2018
Note: 1. The pass mark for each subject is 33%
2.1 No divisions are awarded.
Jhaless
Chief Executive & Secretary
(See Overleaf)

Input (OCR extracted Text)
for ICSE certificate

Classified Text for
ICSE certificate

{
"Issue date": "14.05.2018",
"Marks": null,
"Name": "NIKET BAHETY",
"Roll Number / Unique ID": "6812642",
"School": "AGRASAIN BOYS' SCHOOL,
HOWRAH"
}

The result of the whole project:

```
niket@niket-HP-Pavilion-x360-Convertible-14-dw1xxx:~/Documents/college/FabDVS/Major-Project$ node main_script.js
Choose a document type (PAN/ICSE/CBSE/AADHAR): PAN
Enter the file name: pan2.jpg
Parser result: {
  DOB: '24/01/2002 ',
  father_name: 'PRANABJYOTI HAZARIKA',
  name: 'NEERAJ PRATAP HAZARIKA',
  pan_number: 'ANYPH8727Q '
}
Data: {
  identifier: 'PAN',
  id: 6444,
  data: {
    DOB: '24/01/2002 ',
    father_name: 'PRANABJYOTI HAZARIKA',
    name: 'NEERAJ PRATAP HAZARIKA',
    pan_number: 'ANYPH8727Q '
  }
}
Result: {
  message: 'Document is uploaded to be verified',
  verification: 'Pending'
}
Waiting....
Verdict: {
  message: 'Document is verified',
  verification: 'Verified',
  doc: {
    data: 'fIZTfi7oF0fb0NLWG9E1NW+F2ZTvt+Uo4qrUp+LuLMY+7jLpRroUobdmfWj7VbrWlhcoZaEdw02PJ4i/puA5M5yTb477ABmdvxTf24etm8Yh
ABg4Cfku10njgffF9wAVCLFnYc2xUAhleL5h/I7nJ75sA97N320gj29BY/Triy8oJtf8hiU74m1JGvUODGmyOKCFIaDVGF/ZoYD03tmXuyZXdkgon+oelBrj
c2BI28W0h6cl/djc05yGFM1vmjgtlpI/RGlofXSjmp3SZ/mxcjdoGsGXPCDeioxI5HffnUc2l77LLwjCASu04WBKIFkqJopNMeAPaaAvdNAGN/ZLmSw==',
    id: 6444,
    identifier: 'PAN',
    requestBy: 'CBSE',
    status: 'verified'
  }
}
Document verified!
niket@niket-HP-Pavilion-x360-Convertible-14-dw1xxx:~/Documents/college/FabDVS/Major-Project$
```

The above given is an example of the verified document result.

```
niket@niket-HP-Pavilion-x360-Convertible-14-dw1xxx:~/Documents/college/FabDVS/Major-Project$ node main_script.js
Choose a document type (PAN/ICSE/CBSE/AADHAR): PAN
Enter the file name: pan1.jpg
Parser result: {
  DOB: '17/01/2003 ',
  father_name: 'RAMDULAR YADAV',
  name: 'VINITA RAMDULAR YADAV',
  pan_number: 'BHFPY3100C '
}
Data: {
  identifier: 'PAN',
  id: 4253,
  data: {
    DOB: '17/01/2003 ',
    father_name: 'RAMDULAR YADAV',
    name: 'VINITA RAMDULAR YADAV',
    pan_number: 'BHFPY3100C '
  }
}
Result: {
  message: 'Document is uploaded to be verified',
  verification: 'Pending'
}
Waiting....
Verdict: {
  message: 'Document is uploaded to be verified',
  verification: 'Pending'
}
Document rejected!
niket@niket-HP-Pavilion-x360-Convertible-14-dw1xxx:~/Documents/college/FabDVS/Major-Project$
```

The above given is an example of the rejected document result.

The outputs of reward system are:

```
{
  Key: 'ENTITY1',
  Record: { credits: '10', docType: 'entity', id: 1, org_name: 'PAN' }
},
{
  Key: 'ENTITY2',
  Record: { credits: '10', docType: 'entity', id: 2, org_name: 'CBSE' }
},
```

This above output represents the credits before transaction.

```
{
  Key: 'ENTITY1',
  Record: { credits: 11, docType: 'entity', id: 1, org_name: 'PAN' }
},
{
  Key: 'ENTITY2',
  Record: { credits: 9, docType: 'entity', id: 2, org_name: 'CBSE' }
},
```

This above shows the output which represents the credits of the respective organization after the transaction.

System Benchmarks :

Avg OCR API time : 5.354 sec

Avg Total System Turnaround time: 9.637 sec

Conclusion and Future Work

This work envisions a distributed framework for verifying documents and certificates, offering enhanced security and efficiency. Utilizing the Hyperledger Fabric blockchain technology ensures that the documents remain tamper-proof and protected from unauthorized alterations. The user interface or wallet application facilitates users to check their verification status and sharing status, and they can quickly authorize sharing requests. The proposed framework not only eases the burden on organizations by eliminating the need for manual document verification but also accelerates the process, providing a swift

and effective solution. As the importance of digital documentation continues to grow, this system represents a significant step toward establishing a more reliable and trustworthy document verification process by combining security, privacy, efficiency, and a user-friendly interface into one framework.

In our future work, we aim to enhance the network's reward model for more equitable distribution among participants like endorsers and orderers based on their contributions. This refinement seeks to ensure fair compensation proportional to participants' roles and efforts, fostering an incentive-driven ecosystem that encourages active participation and reliability in transaction processing.

Furthermore, we plan to implement a feature enabling organizations to initiate direct verification requests to each other, bypassing the traditional issuer organization. This streamlines verification by allowing organizations to validate information internally and exchange requests autonomously, promoting collaboration, reducing dependencies, and expediting transaction validation and authorization. These developments aim to optimize network performance, enhance autonomy, and streamline transaction workflows.