DIGITIZATION OF INDIAN DOCUMENTS USING

OPTICAL CHARACTER RECOGNITION

A PROJECT REPORT

Submitted by

HARI KRISHNAN S G

In partial fulfillment for the award of the
degree Of

MASTER OF COMPUTER APPLICATIONS

HAJI C.H.M.M. COLLEGE FOR ADVANCED STUDIES
CHAVARCODE, PALAYAMKUNNU P O – 695146
THIRUVANANTHAPURAM DIST
KERALA

UNIVERSITY OF KERALA, THIRUVANANTHAPURAM

DECEMBER 2021

HAJI C.H.M.M. COLLEGE FOR ADVANCED STUDIES
CHAVARCODE, PALAYAMKUNNU P O – 695146
THIRUVANANTHAPURAM DIST
KERALA

MASTER OF COMPUTER APPLICATIONS

BONAFIDE CERTIFICATE

Certified that this project report **"DIGITIZATION OF INDIAN DOCUMENTS USING OPTICAL CHARACTER RECOGNITION"** is the bonafide work of **HARI KRISHNAN S G** who carried out the project work under my supervision.

HARI KRISHNAN S G: 95518801003

Mr. RAJESH S
Associate Professor
HEAD OF THE DEPARTMENT & INTERNAL GUIDE

EXTERNAL EXAMINER

TABLE OF CONTENTS

PAGES

ACKNOWLEDGEMENT……. iii

LIST OF FIGURES…… vi

ABSTRACT…… vii

CHAPTER

1. INTRODUCTION……. 1
   1.1 Statement of the problem…… 2
2. SYSTEM ANALYSIS……... 3
   2.1 Present System……. 3
   2.2 Limitations of present system……. 3
   2.3 Proposed system……. 4
   2.4 Advantages of Proposed system…….. 4
   2.5 Feasibility Study……. 6
3. SYSTEM SPECIFICATION……. 8
   3.1 Hardware Requirements…….. 8
   3.2 Software Requirements…… 8
4. SYSTEM DESIGN…….. 9
   4.1 Context Level Diagram…… 9
   4.2 Data Flow Diagram……. … 10
   4.3 Data Base Diagram…… 11
   4.4 Normalization……. 13
   4.5 Design of Each Subsystem……. 15
   4.6 UML Diagram…….. 16
   4.7 Flow charts……. 18
5. CODING…… 22
   5.1 Features of Language…… 22
   5.2 Functional Description…….. 25
6. TESTING…… 52
7. IMPLEMENTATION……. 55
   6.1 Implementation of Proposed System……. … 55

8.CONCLUSION…… 56
9.FUTURE ENHANCEMENT…… 57
APPENDIX…… 58
BIBLIOGRAPHY…… 62

ACKNOWLEDGEMENT

I would like to express our gratitude to God for giving us good health and better courage to accomplish this project successfully.

I express my sincere gratitude to the Director of MCA Prof.(Dr).SIRAJUDEEN.M, for providing us an opportunity for doing this project work.

Special thanks to Mr. RAJESH S., Associate Professor, Head of Department for his expert and valuable advice, inspiration and facilities rendered throughout for successful completion of the project.

I express my sincere thanks to our internal guide Mr. RAJESH S, Associate Professor, for her expert and valuable advice, inspiration and facilities rendered throughout for successful completion of the project.

With great pleasure we may record our deep gratitude to our parents, friends and to all staff members of MCA Department for the immensurable help rendered to us during the course of the project.

HARI KRISHNAN S G

LIST OF TABLES

Table Page

4.3.1 Pancard…… 11
4.3.2 Aadhar…….. 11
4.3.3 Driving license……. 12
4.3.4 Voter Id……. 12

LIST OF FIGURES

Figure		Page
4.1	Context Level Diagram…….……..	9
4.2	Data Flow Diagram…….	10
	4.2.1 Level 1 DFD……	10
4.5	Design of Each Subsystem……...……...….	15
4.6	UML Diagram……...……	16
	4.6.1 Use Case Diagram…….	16
	4.6.2 Sequence Diagram…….	17
4.7	Flow charts…….	18
	4.7.1 Pancard flowchart…….	18
	4.7.2 Aadhar card flowchart……..	19
	4.7.3 Driving Licence……	20
	4.7.4 Voter Id……	21

ABSTRACT

DID (Digitization of Indian Documents using optical character Recognition) system offers a unique identity verification service platform that allows identity document verification at high accuracy and high speed with use of optical character recognition. In DID system document of a person scanned through a camera or upload from system and data is extracted from this document after extraction the data is cross checked to ensure that applicants are who they claim to be.