**“Images to Editable Documents Converter”**

**A**

MINOR PROJECT REPORT

Submitted for fulfilment of the requirement for the award of degree

**Bachelor of Engineering**

**In**

Computer Science & Engineering



**Submitted By:** **Guided By:**

Chitransh S. Vishwakarma (0101CS171029) Prof. Uday Chourasia

Bhupendra S. Chaudhary (0101CS161026)

**Submitted To:**

Department of Computer Science & Engineering,

University Institute of Technology,

Rajiv Gandhi Proudyogiki Vishwavidhyalaya, Bhopal 462033

**May 2020**

**UNIVERSITY INSTITUTE OF TECHNOLOGY**

**RAJIV GANDHI PROUDYOGIKI VISHWAVIDHYALAYA, BHOPAL**



**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**CERTIFICATE**

This is to certify that **Chitransh S. Vishwakarma** and **Bhupendra S. Chaudhary** of BE 3rd year, Computer Science & Engineering have completed their Minor project entitled “**Images to Editable Document Converter**” during the academic year 2019 – 20 under our guidance and supervision.

We approve the project for submission for the fulfilment of the requirement for the award of degree in Computer Science and Engineering.

**Prof. Uday Chaurasia**

Project Guide

DoCSE, UIT RGPV



**DECLARATION BY THE CANDIDATES**

We hereby declare that the work which is being presented in the Minor Project entitled – **Images to Editable Document Converter** is submitted for the fulfilment requirement for the award of the degree in Computer Science & Engineering. The work has been carried out at University Institute of Technology, RGPV, Bhopal in the academic session 2019 – 2020 is an authentication record of our work carried under the guidance of **Prof. Uday Chaurasia,** Department of Computer Science & Engineering, UIT RGPV, Bhopal.

The matter written in this project has not been submitted for award of any other degree.

Chitransh S. Vishwakarma – 0101CS171029

Bhupendra S. Chaudhary – 0101CS161026

**ACKNOWLEDGEMENT**

We are thankful to all the individuals who have lent us support and guidelines without which we could not have completed our project in the stipulated time.

First and foremost we would like to express the deepest gratitude to our project supervisors **Prof. Uday Chaurasia** for the invaluable support, guidance, motivation, and encouragement throughout the period. It was there enthusiastic and progressive outlook toward the project which inspired us throughout the work during this period.

We are grateful for our source of inspiration **Dr. Sanjay Silakari**, Prof. and Head of DoCSE, UIT RGPV, Bhopal, for his unforgettable support and inspiration and staff of the department who were involved in the project either directly or indirectly for their valuable cooperation.

In the end, we would like to forward our gratitude to our teammates, friends, and colleagues for their help and cooperation throughout this work.

Chitransh S. Vishwakarma – 0101CS171029

Bhupendra S. Chaudhary – 0101CS161026

**ABSTRACT**

OCR (Optical Character Recognition) is gaining attention in the field of computer vision due to its promising applications in the areas of artificial intelligence, education, documentation, and pattern recognition. Conversion of Images containing textual matter into Editable Documents has always been a cumbersome task until the past few years. As of now, we have various tools and software available to accomplish this task with ease and get accurate results within a matter of seconds. With the advent of the field of OCR and computer vision techniques, it is possible to identify even the handwritten characters, which is one of the most significant research topics in the field of pattern recognition and computer vision. Existing technologies do not provide a local, easy-to-run web interface for students who are getting to know this field. Thus, the main goal of this study was to create an easy-to-use, getting started guide for students who wish to get started with research at the early stages of their careers in the field of OCR. We are providing a ready-made web-app and the core code along with the assembly of all the required libraries and tools which would be used by any student to get started in the field of optical character recognition. All that the students would have to do is to follow the step-by-step guidelines mentioned in the documentation for the project and they are prepared to explore the world with endless possibilities with OCR. The core functionality of the project begins with uploaded an image with some textual matter in it, to the web app interface which then gets passed on to the Tesseract-OCR Engine which extracts all textual matter from the image and then the 'python-docx' tool appends the extracted text from the image onto a blank MS Word document and the web interface enables the download button which may then be used to download the converted editable document.

**TABLE OF CONTENTS**

**ANNEXURE**

CERTIFICATE…………………………………………………………………………..ii

DECLARATION……………………………………………………………………….iii

ACKNOWLEDGEMENT……………………………………………………………...iv

ABSTRACT……………………………………………………………………………..v

LIST OF FIGURES……………………………………… ………………….……… viii

**CHAPTERS**

Chapter 1**- INTRODUCTION**

* 1. Introduction

Chapter 2 - **LITERATURE SURVEY**

Chapter 3 - **PROBLEM DESCRIPTION**

Chapter 4 - **PROPOSED WORK**

Chapter 5 - **DESIGN & DEVELOPMENT**

Chapter 6 - **SYSTEM ANALYSIS (IMPLEMENTATION)**

Chapter 7 - **RESULT**

Chapter 8 - **CONCLUSION**

Chapter 9 - **REFERENCES**

**LIST OF FIGURES**