Bahria University
Discovering Knowledge

FINAL YAER PROJECT REPORT

OPTICAL CHARACTER RECOGNITION AND PLAGIARISM DETECTOR

By

SAIF UDDIN (43811)

UZAIR AHMED (43804)

YASHAL ZAFAR SIDDIQUI (43742)

SUPERVISED BY

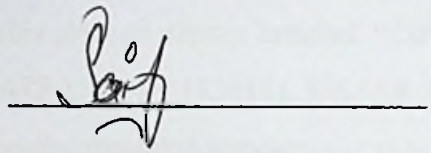
(MA'AM ANIQA NAEEM)

BAHRIA UNIVERSITY (KARACHI CAMPUS)

2019

DECLARATION

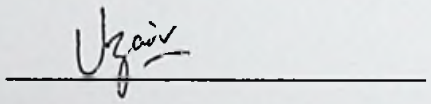
We hereby declare that this project report is based on our original work except for citations and quotations which have been duly acknowledged. We also declare that it has not been previously and concurrently submitted for any other degree or award at Bahria University or other institutions.

Signature : 

Name : SAIF UDDIN

Reg No. : 43811

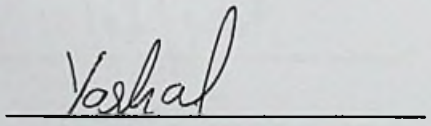
Date : 10-2-2020

Signature : 

Name : UZAIR AHMED

Reg No. : 43804

Date : 10-2-2020

Signature : 

Name : YASHAL ZAFAR SIDDIQUI

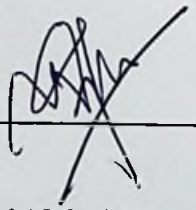
Reg No. : 43742

Date : 10-2-2020

APPROVAL FOR SUBMISSION

I certify that this project report entitled **"OPTICAL CHARACTER RECOGNITION"** was prepared by **SAIF UDDIN (43811), UZAIR AHMED (43804), YASHAL ZAFAR SIDDIQUI (43742)** has met the required standard for submission in partial fulfilment of the requirements for the award of Bachelor of Computer Science at Bahria University.

Approved by,

Signature : 

Supervisor: MA'AM ANIQA NAEEM

Date : 10/02/2020

OPTICAL CHARACTER RECOGNITION AND PLAGIARISM DETECTOR

ABSTRACT

There are many different forms of plagiarism, Plagiarism at schools can be a highly de-motivating factor for teachers and also for students. If plagiarism is not addressed sufficiently, plagiarists could gain undeserved advantage, e.g. more marks for their assignments with less effort. Optical Character Recognition (OCR) technology is being used worldwide today, it extracts the printed text that is present in a picture and converts it into editable text. Accuracy of an OCR can fluctuate over what algorithm and techniques are used in it. This software uses the processes namely pre-processing, segmentation, feature extraction, and classification. So, plagiarism is a global problem, which occurs in many different areas of our life. This report discusses many important aspects in the whole plagiarism detection program that includes detection Tasks, detection process and some of the current detection techniques used all around the world by most students. Also, the tips the tricks that are used most commonly to trick a plagiarism checker. We conclude this report by a comparative study of this technology in many research papers and its application are done by many computer scientists.

TABLE OF CONTENTS

DECLARATION		ii
APPROVAL FOR SUBMISSION		iii
COPYRIGHTS		iv
ABSTRACT		v
TABLE OF CONTENTS		vi
LIST OF FIGURES		ix
LIST OF SYMBOLS / ABBREVIATIONS		xi
LIST OF APPENDICES		xii

CHAPTER

1	INTRODUCTION		
	1.1	Background	1-2
	1.2	Problem Statement	2-3
	1.3	Objectives	3
	1.4	Scope of Projects	3
2	LITERATURE REVIEW		
	2.1	Real – time font edge focus measurement for optical character recognition (OCR)	4-5
	2.2	A proposed approach for character recognition using document analysis with OCR	5-6
	2.3	Optical character recognition systems for the English Language	6-7
	2.4	Optical character recognition from printed text images	7-8
	2.5	PYTESSEARCT: simple python optical character	

	2.5	PYTESSEARCT: simple python optical character Recognition	8-9
3		DESIGN AND METHODOLOGY	
	3.1	Introduction of Agile Methodology	10
	3.1.1	Systems Requirement	10
	3.2	Design	11
		3.2.1 Input of File	11
		3.2.2 Image Detection	12
		3.2.3 Text Extraction	12
		3.2.4 Reassemble Document	12
		3.2.5 Plagiarism Checking	13
		3.2.6 Gantt Chart	13-14
4		IMPLEMENTATION AND TESTING	
	4.1	Module: Development of GUI	15-20
	4.2	Module: Image Detection	21-22
	4.3	Module: Extraction of Text	22-24
	4.4	Module: Deployment of OCR	24-27
	4.5	Module: Plagiarism Checking	27-29
	4.7	Testing	29
		4.7.1 Unit Testing	29-32
		4.7.2 Integration Testing	32-33
5		RESULTS AND DISCUSSIONS	

	5.1	Outcome	34
	5.2	Future of the Project	34-35
6		CONCLUSION AND RECOMMENDATION	
	6.1	Conclusion	36
	6.2	Recommendations	37
REFERENCES			38
APPENDICES			39-44