SCANNED IMAGE AND WEB BASED OMR FOR ETHIOPIAN NATIONAL EXAMINATIONS

A THESIS SUBMITTED TO THE GRADUATE SCHOOL OF APPLIED SCIENCES OF NEAR EAST UNIVERSITY

By
GETINET AMARE MEKONNEN

In Partial Fulfillment of the Requirements for the Degree of Master of Science in Software Engineering

NICOSIA, 2019

GETINET AMARE SCANNED IMAGE AND WEB BASED OMR FOR ETHIOPIAN 2019
MEKONNEN NATIONAL EXAMINATIONS NEU

SCANNED IMAGE AND WEB BASED OMR FOR ETHIOPIAN NATIONAL EXAMINATIONS

A THESIS SUBMITTED TO THE GRADUATE SCHOOL OF APPLIED SCIENCES OF NEAR EAST UNIVERSITY

By
GETINET AMARE MEKONNEN

In Partial Fulfillment of the Requirements for the Degree of Master of Science in Software Engineering

NICOSIA, 2019

Getinet Amare MEKONNEN: SCANNED IMAGE AND WEB BASED OMR FOR ETHIOPIAN NATIONAL EXAMINATIONS

Approval of Director of Graduate School of Applied Sciences

Prof.Dr. Nadire ÇAVUŞ

We certify this thesis is satisfactory for the award of the degree of Master of Sciences in Software Engineering

Examine committee in charge:

Assoc. Prof. Dr. Yöney Kırsal Ever	Department of Software Engineering, NEU
Assoc. Prof. Dr.Kamil Dimililer	Supervisor, Department of Automotive Engineering, NEU
Assist. Prof. Dr. Boran Şekeroğlu	Department of Information System Engineering, NEU

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original of this work.

Name, Last name: Getinet Amare Mekonnen

Signature:

Date: Jun. 12, 19

ACKNOWLEDGEMENT

I would like to thank my thesis supervisor, Assoc. Prof. Dr. Kamil Dimililer for his energetic advice, allocating his valuable time, helpful comments, idea to understand the problem well and motivation all over each phase of the research.

I would also like to express my gratitude to Assist. Prof. Dr. Boran Şekeroğlu and Assoc. Prof. Dr. Yöney Kırsal Ever at Near East University who have been offering helpful comments and suggestions to achieve my goal.

I am grateful to my sibling, sisters, Mastewale, Alganesh, Tiringo, Banchayehu; brothers, Niguse, Zemenu, Markose; mother, Bosena; and father, Amare who have provided me through moral and emotional support in my life. I am also grateful to my other family members and friends who have supported me along the way.

Finally, I would like to acknowledge Ethiopian Ministry of Education especially Debre Tabor University and Betre Science Scholarship program leaders to cover all my expenses and adjust scholarship program here in North Cyprus, Nicosia.
Thanks for all your encouragement.

To my parents …

ABSTRACT

Although optical mark reader system already exists for multiple choice questions, they have still affordability and compatibility issue for low income educational institutions. The proposed system has eight types of scoring sheets and its own unique barcode to classify them. The scoring sheet contains 15, 20, 30, 45, 60, 80, 100 and 125 questions with five choices in each. Ordinary printers, scanners and A4 paper with 70 GSM have been used for the system. Depending on the amount of questions in sheets and system functionality, the speed of this system varies and increases its speed when we go from 125 to 15 questions. Compatibility is not an issue for this system, because it integrates to web and accessible by every web enable devices such as mobiles, tablets, laptops and desktop computers are some of them. This system has two types of users; the one feed sheets and can see OMR result, the second only see OMR result i.e. students. Each user needs to be login by their own username and password. The system enables sheet feeders to upload and evaluate multiple sheets at same time. If one marked scoring sheet needs to be evaluated by this system, the system would check barcode, student id and course code are there in database table before going to the actual student marked answer. This system sends the recognized scoring sheet by bordering the answer area with different colors; these are green border for correctly marked answers, red border for duplicate marked answers, and yellow border for totally unmarked questions. Students can send back to the first user of system if error might happen during the recognition via their own web account.

***Keywords*:** OMR; GSM; barcode; scoring sheet; database; web

ÖZET

Her ne kadar optik işaret okuyucu sistemi çoktan seçmeli sorular için mevcut olsa da, düşük gelirli eğitim kurumları için hala karşılanabilirlik ve uyumluluk sorununa sahiptir. Önerilen sistem, sekiz çeşit puanlama sayfasına ve bunları sınıflandırmak için kendine özgü bir barkoda sahiptir. Skor sayfasında 15, 20, 30, 45, 60, 80, 100 ve her birinde beş seçenek bulunan 125 soru bulunmaktadır. Sistem için sıradan yazıcılar, tarayıcılar ve 70 GSM'li A4 kağıt kullanılmıştır. Sayfalardaki ve sistem işlevselliğindeki soruların miktarına bağlı olarak, bu sistemin hızı 125 ila 15 soruya gittiğimizde hızını değiştirir ve arttırır. Uyumluluk bu sistem için bir sorun değildir, çünkü web ile bütünleşir ve her web tarafından erişilebilen cep telefonları, tabletler, dizüstü bilgisayarlar ve masaüstü bilgisayarlar gibi cihazlardan bazıları bunlardır. Bu sistem iki tip kullanıcıya sahiptir; bir besleme sayfası ve OMR sonucunu görebilir, ikincisi yalnızca OMR sonucunu, yani öğrencileri görebilir. Her kullanıcının kendi kullanıcı adı ve şifresiyle giriş yapması gerekir. Sistem, sayfa besleyicilerin aynı anda birden fazla sayfa yüklemesini ve değerlendirmesini sağlar. Eğer işaretlenmiş bir puanlama kağıdının bu sistem tarafından değerlendirilmesi gerekiyorsa, sistem barkodu kontrol eder, öğrencinin kimliği ve ders kodu, veritabanındaki tabloda cevaplanan asıl öğrenciye gitmeden önce orada bulunur. Bu sistem, kabul edilen puanlama kağıdını, cevap alanını farklı renklerle sınırlayarak gönderir; bunlar doğru işaretlenmiş cevaplar için yeşil sınır, çift işaretli cevaplar için kırmızı sınır ve tamamen işaretlenmemiş sorular için sarı sınırdır. Tanınma sırasında hata kendi web hesapları aracılığıyla gerçekleşirse, öğrenciler sistemin ilk kullanıcısına geri gönderebilirler.

Anahtar Kelimeler: OMR; GSM; barkod; puan tablosu; veritabanı; ağ

TABLE OF CONTENTS

ACKNOWLEDGMENT ii
ABSTRACT iii
ÖZET iv
TABLE OF CONTENTS v
LIST OF FIGURES viii
LIST OF TABLES x
APPENDICES xi
LIST OF ABBREVIATIONS xii

CHAPTER 1: INTRODUCTION
1.1 Statement of The Problem 3
1.2 Motivation of The Study 4
1.3 Objectives 5
1.3.1 General objective 5
1.3.2 Specific objectives 5
1.4 Research Questions 6
1.5 Scope 6
1.6 Significance 7
1.7 Thesis Outlines 7

CHAPTER 2:LITERATURE REVIEW
2.1 Background Theory 8
2.1.1 Image processing 8
2.1.1.1 Pixel 8
2.1.1.2 Image 9
2.1.1.3 Types of images 10
2.1.1.3.1 Binary images 10
2.1.1.3.2 Grayscale images 10
2.1.1.3.3 Color images 10
2.1.2 Web 10

2.1.3 Database 11
2.1.4 Computer graphics 13
2.1.4.1 Transformation 14
2.2 Literature Review 18

CHAPTER 3 : METHODOLOGY

3.1 Research Methodology 25
3.1.1 Data collection method 25
3.1.2 Development method 26
3.1.2.1 Scoring sheet layout 26
3.1.2.2 Scoring sheet registration 27
3.1.2.3 Image pre-processing 30
3.1.2.4 Adjust scoring sheet layout. 30
3.1.2.5 Barcode 33
3.1.2.6 Course code recognition 34
3.1.2.7 Student code recognitions 35
3.1.2.8 Students answer recognition 35
3.1.2.9 Make border for squares 37

CHAPTER 4 : IMPLEMENTATION DETAILS

4.1 Tools Used 44
4.1.1 Software tools 44
4.1.1.1 Python 44
4.1.1.2 Flask framework 45
4.1.2 Hardware tools 46
4.2 Implementations 46
4.2.1 Data preparation 46
4.2.2 OMR without border 48
4.2.3 OMR with border 50
4.2.4 OMR with answer key 51

CHAPTER 5 : RESULTS AND DISCUSION

CHAPTER 6 : CONCLUSION AND RECOMMENDATION

6.1 Conclusion 66
6.2 Recommendation 66

REFERENCES 68

BIBLIOGRAPHY 70

LIST OF FIGURES

Figure 2.1: Original image 8
Figure 2.2: The first row of zoomed image 8
Figure 2.3: Shows the pixel location and its intensity on an image 9
Figure 2.4: Shows shades of grayscale changes from 0 to 255 and vice-versa 10
Figure 2.5: Simplified database system environment 12
Figure 2.6: 2D point translation 14
Figure 2.7: Rotation by angle phi(ϕ) 15
Figure 2.8: Rotation of points both in clockwise and counter clockwise direction 18
Figure 3.1: Proposed scoring sheet 28
Figure 3.2: Proposed 125 questions scoring sheet 29
Figure 3.3: 180 degree rotated scanned scoring sheet 32
Figure 3.4: Template scoring sheet 32
Figure 3.5: Correctly scanned scoring sheet 32
Figure 3.6: Incorrectly scanned scoring sheet 32
Figure 3.7: -90 degree rotated scoring scanned sheet 32
Figure 3.8: +90 degree rotated scoring scanned sheet 32
Figure 3.9: Barcode parts 33
Figure 3.10: Sample barcode example 34
Figure 3.11: General workflow of proposed system 38
Figure 3.12: Flow chart for course code recognition 39
Figure 3.13: Flow chart for student id recognition 40
Figure 3.14: Flow chart for question numbers recognition 41
Figure 3.15: Flow chart for multiple choice mark recognition 42
Figure 3.16: Flow chart for student result computation 43
Figure 4.1: Sample student's data in database table 46
Figure 4.2: student registration page 47
Figure 4.3: list of services when we login as a teacher 47
Figure 4.4: Uploading 15 questions scoring sheet and retrieve its result. 49
Figure 4.5: OMR with border 50

Figure 4.6: Uploading scoring sheet having 100 questions in it 51
Figure 4.7: Result of uploaded sheet on figure 4.6..... 52
Figure 4.8: Login to omr system by 1000000000 student id..... 53
Figure 4.9: Result of student after logged by 1000000000 id 53
Figure 4.10: Message box 53
Figure 4.11: Recognized bordered sheet 54
Figure 5.1: 125 Questions OMR result..... 59
Figure 1. 1: Snapshot of 15 questions result with answer key 71
Figure 1. 2: Snapshot of computed results from Figure 1.1 71
Figure 1. 3: login as 8000000006 student user name 72
Figure 1. 4: Computed results after Figure 1.3..... 72
Figure 1. 5: OMR result of 15 questions with border 73
Figure 1. 6: Snapshot of 20 questions result with answer key 74
Figure 1. 7: Snapshot of students result after logged by 80000000004 id 75
Figure 1. 8: OMR result of 20 questions with border 76
Figure 1. 9: Snapshot of 30 questions result with answer key 77
Figure 1. 10: Snapshot of students result after logged by 80000000003 id 78

LIST OF TABLES

Table 3.1: The content of scoring sheets 25
Table 3.2: The representation of question types with decimal number 33
Table 5.1: Sheets used for testing the system 55
Table 5.2: SRTs before switched to web and connected to database without border. 56
Table 5.3: SRTs before switched to web and connected to database with border..... 57
Table 5.4: SRTs after switched to web and connected to database without border. 57
Table 5.5: SRTs after switched to web and connected to database with border. 58
Table 5.6: Detected student answers from Figure 5.1 62
Table 5.7: Retrieved answer keys from database table 63
Table 5.8: Computed and detected results 64
Table 5.9: Letter and numeric grade values..... 64

APPENDICES

1. More implementation results 71

LIST OF ABBREVIATIONS

RL:	Relational Model
RDMS:	Relational Database Management System
MYSQL:	My Structured Query Language
CRUD:	Create, Retrieve, Update, Delete
func:	Function
ENEs:	Ethiopian National Examinations
IOS:	Internet Over Satellite
SQL:	Structured Query Language
OMR:	Optical Mark Reader
DB:	Database
Std:	Student
RIO:	Region of Interest
RAM:	Random Access Memory
HDD:	Hard Disk Drive
OS:	Operating System
NEC:	Nippon Electric Company
SF:	System Functionality
TNQs:	Total Number of Questions

CHAPTER 1
INTRODUCTION

Optical mark recognition(OMR) is the process of scanning optical paper sheets to decide the appearance and non-appearance of marks in a prearranged location. It has been used by analyst's for almost more than 80 years to convert paper sheet data to computer readable format (Nalan Karunanayake 2015). So, its inputs are well marked optical sheets to have accurate results for both software and hardware based optical mark reader.

Optical mark recognition also known as optical mark reading or optical mark reader. It is a technology that extracts useful data from marked shapes such as fill-in squares, fill-in rectangles, fill-in triangles, fill-in circles and other fill-in shapes as well.
Even though OMR have many applications though out of our life in this world, the most common use of optical mark recognition is in research centers, universities, schools and industries where large number of hand filled documents needs to be processed such as time attendance, elections, surveys, questionnaires, examinations and replay cards with short period of time and painless.

In case of Ethiopian universities, industries, schools and other organizations take employees attendance in manual form, after that count which employee absent or present by concerned body; except grade twelve and ten national examinations; grade eight regional examination does not use any optical mark reader hardware machine for reading students multiple choice questions. But, my concern is not being applying a hardware based optical mark reader machine for those did not use it till, rather I will take images from ordinary scanner device they use for other purpose, with self-simple prepared scoring sheets and make an optical mark recognition algorithm.

If we compare to human beings, OMR can handle many documents with high accuracy and faster time. So, we have no more doubt between the performance of human beings and scanned and web based OMR. The scoring sheet paper quality, the limitation of pencils used for marking squares and weight of scoring sheet papers are some limitations of hardware based optical mark recognition.

This work has database tables that stores the correct answer or answer key of each courses, course code or id and student id kept before. The algorithm that I have develop for recognizing optical marks is that receives scoring sheet images from ordinary scanner, apply pre-image processing techniques, recognize course code from marked sheets, recognize student id from marked sheet, recognize quantity of questions by its attached barcode and recognize students answer from marked squares as well.

In case, students may not mark the correct course code and or student id, at that time the algorithm has recognize the marked squares and check whether the recognized course code is existed on database table or not. If the recognized course code will have not found on database table, the optical mark recognition algorithm does not continue to recognize the remaining scoring sheet data's i.e. students answer.

On the other hand, students may mark the course code on scoring sheet correctly, the algorithm tries to check the existence of recognized course code that are kept on database table by concerned body before. After that the optical mark recognition algorithm pass to check the correctness of students marked id as well. By taking students recognized id the optical mark recognition algorithm tries to check its existence on database table. If its exist, the optical mark recognition algorithm will go to the main body of scoring sheet.

Scoring sheets have divided into four main parts such as student's id, course code, barcode and main body that contains multiple choice squares. The number of squares both in student's id and course code remain constant for any eight pre-prepared types of scoring sheet design. The student's id part on scoring sheet section has ten rows that contains the first ten Arabic numbers and columns too. The course code section has ten rows that contains the first ten Arabic numbers and three columns. So, there will have 100, 30 squares for student's id and course code respectively.

The main body part of scoring sheet will have five alternatives such as A, B, C, D and E with corresponding square spaces to fill students for their own correct answers. The scoring sheet have design in eight ways based on the number of questions they contain in it.

The first design contains 125 questions with twenty-five rows, five alternatives and columns as well. The second design contains 100 questions with twenty-five rows, five alternatives and four columns. The third scoring sheet design contains 80 questions with twenty rows, five alternatives and four columns. The fourth scoring sheets design contains 60 questions with twenty rows, five alternatives and three columns. The fifth scoring sheet design contains 45 questions with fifty rows, five alternatives and three columns. The sixth scoring sheet design contains 30 questions with fifty rows, five alternatives and two columns. The seventh scoring sheet design contains 20 questions with ten rows, five alternatives and two columns. The last scoring sheet design contains fifty questions with five alternatives and one column. Therefore, the scoring sheet design have 755, 630, 530, 430, 355, 280, 230, 205 squares for the first, second, third, fourth, fifth, sixth, seventh and eighth respectively.

After checking the validity of marked squares for student's id and course code, the optical mark recognition algorithm compares recognized students answer with the answer key of selected course retrieved from database table by the help of marked course code and recognized barcode. Then the optical mark recognition algorithm counts how many marks are matched with the answer key, how many questions are not marked by students, count how many questions marked but not correct answer. After counting total correct students answer, the algorithm changes number to letter grade if its needed.

Finally, the details of optical mark recognition data will send to html forms by post method for displaying its result to student's own web account and other concerned bodies.

1.1 Statement of The Problem

The problem is price of hardware based optical mark recognition machine and its scoring sheets costs up to 125 dollars per 1000 pages (Nalan Karunanayake, 2015), this leads us to find an alternative solution for solving mark recognition problem. Rejection of scoring sheets for more than one answers were filled per question (Omrhome, 2016) is also another problem of hardware based optical mark recognition machine.

In case students might fill up two choices per question, some machines reject scoring sheets totally and others put zero result. As long as marked choices are there on same question, the

optical mark reader machine or software should recognize one square as a correct student answer. When we start to solve one's world problem, we have to think worst case rather than that of the best and average cases of the selected problem so, students also might fill answers on scoring sheets outside of the given fill-in squares. During this time, the existing optical mark recognition machine and software recognize that fill-in squares are as students' answer. If this happens on scoring sheets, we will get either unexpected result and or rejection of scoring sheets by OMR machine and or software.

Another problem is that, after collecting recognizing fill-in students mark from scoring sheets, they put on spreadsheet file and it waits for another task i.e. comparing with answer key and way of forwarding comparison results to students. There is mobile based optical mark recognition software's but it needs another second software to use on windows and other platforms to integrate. So, why do not we integrate it to web and accessible by web enabled devices remotely without local installation or worried about what ever platforms we will have to use. As long as web enabled devices we have to use, just type the URL of the file and follow the steps asked by the system. Therefore, like uploading profile pictures on Facebook, Twitter, Instagram and other social websites, needs to upload scoring sheets and get result on their own students' web account rather than that of classifying tasks into recognize students answer from scoring sheets, save it to external file and compare to answer keys on different time gap

1.2 Motivation of The Study

My motivation for deciding to conduct this work is that, when I was study my bachelor, I did checking of grade eight student's multiple choice examination with correct answer manually during summer time. After cross checking students marked sheets with correct answer key, I was count how many questions are correct by and answer by whom and put its sum on paper. There were a lot of student's sheet needed to convert to grade letter by human being's effort so, as we know such task is too traditional, takes too much time and boring as well to do. We did that much boring task without the help of any technology but, there is a scanner and printer on that office they accomplish another office tasks. If we know how to do solve that problem by using the existing scanners and printers, we were so lucky

and painless. Another reason that pushes to study this work is that the cost of hardware based optical mark reader is expensive, not only the machine but also the scoring sheets are 125 dollars per 1000 pages (Nalan Karunanayake, 2015).

If one's small organization, school and industry have a scanner and computer that are used for solving their common problems, why they cost their budget for buying the hardware based optical mark reader. So, I decide to solve such problem by using the existing scanners, printers and computers.All related conducted researches as I read on internet, they did recognition and retrieve scanned image data, but not web features were added and most of them were their future work.

1.2 Objectives

Objectives of the study are classified into two parts such as general and specific objectives.

1.3.1 General objective

The general objective of this study is to solve an ENEs problem by using ordinary papers, scanners and printer too.

1.3.2 Specific objectives

To achieve the general objective of the study that are written on above, the following specific objectives should be accomplished.

1. Develop an algorithm: -
    - That recognizes scanned score sheets data i.e. barcode, student's id, course code and multiple choice square filled data.
    - That checks the uploaded scoring sheet is valid or not
    - That computes total number of students answer by comparing the correct answer key were retrieved from database table.
    - That shows totally unmarked question number by student's
    - That shows total number of wrong student's answer

- That converts total students numerical result to equivalent letter grade and its average grade point.
- That minimizes to zero error rate as much as I can.

2. Retrieve information stored on database table by recognized data from scanned scoring sheets.
3. Design eight types of scoring sheets that contains 125, 100, 80, 60, 45, 30, 20 and 15 questions to test the proposed system.
4. Display the recognized marked data on web by flask pythons web framework to students account.

1.3 Research Questions

The following research questions have answered for this study

- Why do we need to develop scanned image and web based OMR? Over the existing hardware based optical mark reader machine?
- What is the limitation and advantages of hardware based optical mark reader machine?
- How to solve, when more than one choice was filled per question?
- Which choices were filled in which sheet type?
- Why do we need to integrate the recognition result to web and can students see their results remotely?

1.5 Scope

The boundary of this study is that, scanned image and web based optical mark recognition for ENEs. Even though I wrote ENEs, it can be used by other schools by making a little modification on the scoring sheets. After getting scoring sheet images from scanner, the optical mark recognition algorithm applies image processing techniques and then going to mark recognition for the specific pattern. The result of recognition has an integration and display on web for remote accessing. I acquire scoring sheet images from ordinary scanners and prepared by printers with A4 paper too. The study conducted for multiple choice exam question mark recognition specifically.

1.6 Significance

Primarily, the significance of the study is switching the use of hardware based optical mark recognition to software based, by using scanners and printers for scoring sheet preparation. It plays an important role in cost and the use of usual scanners and printer for additional task. Secondly, for schools' examination starting from scoring sheet preparation up to see students grade letter result on their own web account remotely.

1.7 Thesis Outlines

The rest of this document is organized as follows. Chapter 2 talks about literature review and background theories, it contains related works for this area, terms and concepts used for conducting the research. Chapter 3 states about methodology of the research. Methodology contains design, amount of sample data used for testing the proposed system. Chapter 4 is about implementation details; I have used python's programming language for backend, MySQL database management for manipulating retrieved sheets data, Hypertext markup language(HTML), cascading stylesheet(css) and JavaScript(js) used for front end development. Chapter 5 contains result and discussion and the last chapter 6 contains conclusion and recommendation of the research.

CHAPTER 2

LITERATURE REVIEW

2.1 Background Theory

In this section, terms and concepts specifically in the area of optical mark reader are discussed. This section also plays an important role for those who wants to read this work and get better understanding too. So, readers can know easily the basic concepts and terms used for throughout the thesis.

2.1.1 Image processing

Like a collection of pixels makes an image, let's start by defining commonly used terms in image processing and will do more on it later.

2.1.1.1 Pixel

PEL stands for Picture Element and it's another alternative name of pixel. Many sources use these two words interchangeably. The last smallest element we use to measure image is pixel even though, larger units are there like mega pixel.

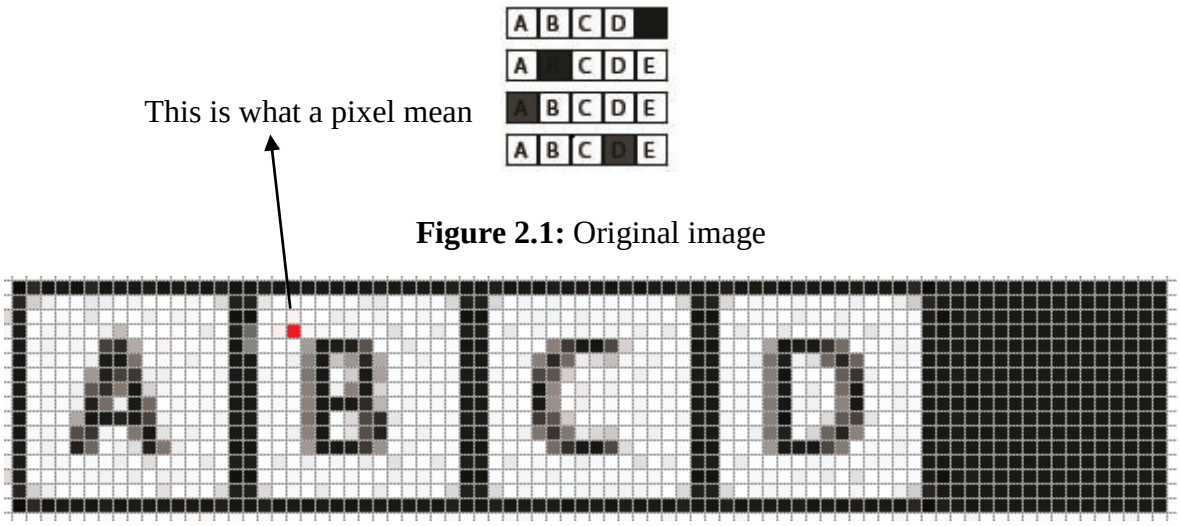


Figure 2.1: Original image

Figure 2.2: The first row of zoomed image

As we saw in the above Figure 2.2, there are thousands of pixels to build an image. More on zoomed image shows clearly the sub-division of pixels. In order to calculate the size of an image, we need to count how many number of pixels are there in an image. This concept leads to resolution; it is the total number of pixels per rows multiplied by columns in an image. Then, image resolution can be defined as M X N where 'M' is the width and 'N' is height of an image.

2.1.1.2 Image

An image is one real world object created by image signals. Signals are mathematical and or statistical approach that interacts us to the world and conveys image information from visual scene by taker devices to its storage, in case of image. but, for sound it conveys sound signals, for light it conveys light signals from one location to another and the like. So, an image can have expressed in two dimensional function func (h, v) where 'h' and 'v' are special coordinates of pixel for horizontal and vertical directions respectively and 'func' is an amplitude at any pairs of spatial coordinates also known as intensity value of small portion of an image i.e. pixel.

The process of converting pair of coordinate values to digital is called sampling whereas amplitude known as quantization. Therefore, the result of quantization and sampling is a set of real numbers and expressed in matrix form shown below on Figure 2.3.

$$
\text{func}(h,v) = \begin{bmatrix}
\text{func}(0,0) & \text{func}(0,1) & \text{func}(0,2) & \text{func}(0,3)\ldots\ldots\ldots\ldots\text{func}(0,N-1) \\
\text{func}(1,0) & \text{func}(1,1) & \text{func}(1,2) & \text{func}(1,3)\ldots\ldots\ldots\ldots\text{func}(1,N-1) \\
\text{func}(2,0) & \text{func}(2,1) & \text{func}(2,2) & \text{func}(2,3)\ldots\ldots\ldots\ldots\text{func}(2,N-1) \\
\vdots & & & \vdots \\
\text{func}(M-1,0) & \text{func}(M-1,1) & \text{func}(M-1,2) & \text{func}(M-1,3)\ldots\ldots\ldots\text{func}(M-1,N-1)
\end{bmatrix}
$$

Figure 2.3: Shows the pixel location and its intensity on an image

2.1.1.3 Types of images

Based on the number of colors they contain, images classified into different types, the most common types are defined below

2.1.1.3.1 Binary images

As the name indicates that binary image contains only two colors i.e. zero(black) and one(white). Binary images also called Black and white images, there is no gray level on it.

2.1.1.3.2 Grayscale images

Grayscale images are the combinations of binary and gray-level shades. If the brightness of grayscale image becomes 255 and its darkness 0 that image is same as black and white(binary) image. But, in between darkness and whiteness, there are $2^8 - 2$ gray levels, this is why grayscale images are the combinations of both gray-levels and binary images. When we read pixel values of grayscale images, we get 2D array of size H X V. on Figure2.3, each array values measure one byte and can have values in between 0 and 255, by convention 0 represents black and 255 represents white as well. In order to clear this idea, have a look the following image label on Figure 2.4 that shows pure black, pure white and in-between 0 and 255 shades changes from 0 to 255 and vice-versa.

Figure 2.4: Shows shades of grayscale changes from 0 to 255 and vice-versa

2.1.1.3.3 Color images

Color images commonly known by the name of RGB images. It's the combination of three channels i.e. Red, Green and Blue respectively. When we read RGB images with a size H X V, we get H X V X 3 images i.e. we get a three dimensional images. Each plane in the array corresponds to Red, Green and Blue channels. Each element in an array of color image have 8 bits (1 byte). So any element on an array of color images have in between 0 and 255 including the starting and its ending values i.e. 0 and 255 respectively. When we write (R=0, G=0, B=0) it means pure black RGB image, where as white (R=255, G=255, B=255). To

get pure red, green and blue color from RGB images we should write (R=255, G=0, B=0), (R=0, G=255, B=0), (R=0, G=0, B=255) respectively.

2.1.2 Web

A computer communicates to another computer that links via either telephone lines, IOS and other linkers called an internet. It can be also defined as a network of networks that combined together for common task. In the early 1960, 1970 and 1980, it was rare to exchange information at all, because there is not the concept of internet and web at all. Even, the same brand personal computers could not run same programs. Instead, each type of same brand personal computers had to have computer programs written to it specifically, for instance applications like Facebook, email, WhatsApp and functionalities like chatting in text, audio and video all are impossible things. Chris, W. (2018). But, now no more such type of complications are there and we are so lucky, live in uncomplicated world. If such much type of change is there with in twenty years and I expect more inventions still, there in this real world that are hidden by human beings. So, the task of unhide this world with full of hidden secret in technology is obviously human begins divided into different levels like researchers, students, teachers, scientists and so on.

Web is the collection of texts, videos, audios, images, animations and we can access those things over the technology that are invented since 1983 i.e. internet. The web is more special that all those things are there in a connected way together. Web pages are the basic building blocks of web that contains texts, videos, images, audios and animations on it. The term website also taken from the collection of web pages together. So, as web makes this world much easier and flexible and it plays an important role for accessing its results and feed sheets too.

2.1.3 Database

Database technology has a major importance on the growing computer software whether in web, standalone and mobile too. It is rational to say that database plays a vital role in almost every area where computers are used, including education, business, engineering and e-commerce, Elmasri, R. (2011). Database is a collection of known facts having its own

implicit meaning. It can be designed, populated and built for the purpose we want. It can be populated or maintained both in computerized and in manual format. As I wrote previously, I have populated and maintained only computerized database for this work. A Database Management System(DBMS) is a group of computer programs that enable users to maintain and construct databases. It is universal purpose computer software system that facilitates us to define, construct, manipulate and sharing databases to other software systems and users as well. defining a database contains selecting attribute datatypes, structure of database, constraints of data to be stored on database. Constructing the database is the method of storing the data on some storage medium and controlled by relational database management system. Manipulating a database contains functions that enables us to do CRUD data from the database. Sharing a database permits several programs and users to use the database at same time. Sharing databases to users and programs may varies its privilege. Software systems access the database by sending query for data to database management system, Elmasri, R. (2011). The following Figure 2.5 shows a simplified database system environment.

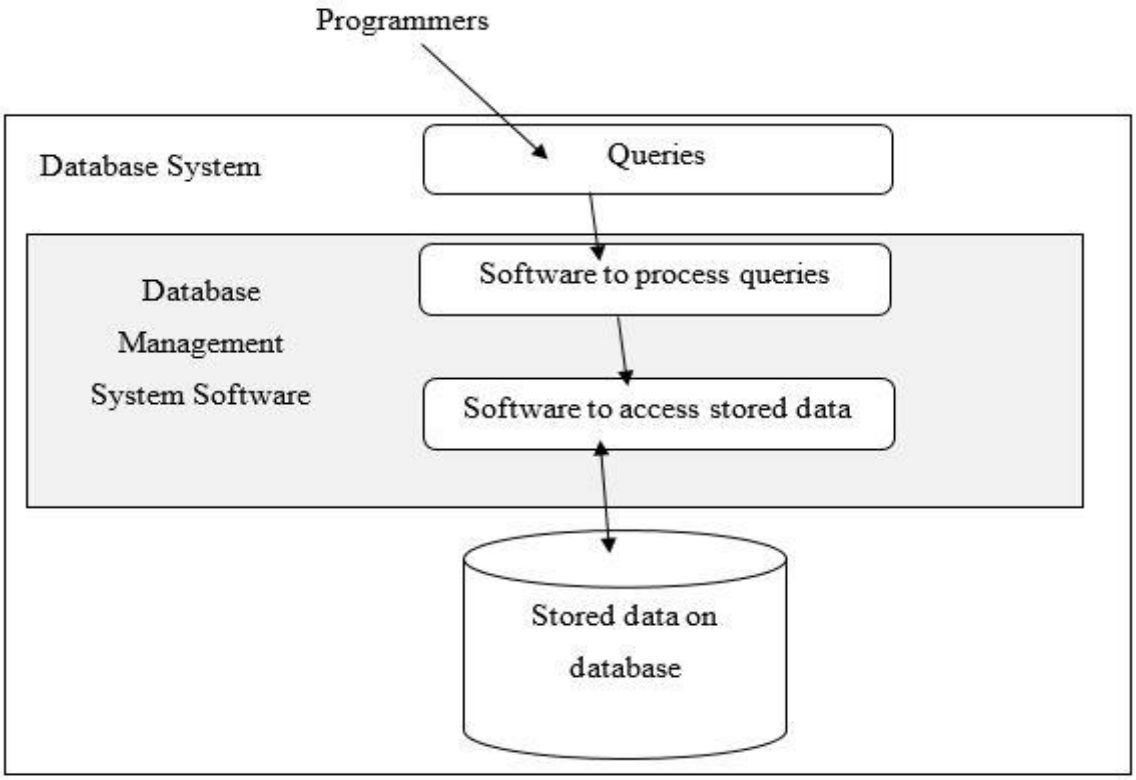


Figure 2.5: Simplified database system environment

The need of database for this work is that, to store scoring sheet recognition results such as student's id, course code, students answer and barcodes as well. after storing such data to database management system, there are some CRUD queries applied to previously stored data. Relational Database Management System(RDMS) is a DBMS based on RM. Relational model is a model used by many database management systems currently available, (Pratt, Philip J.; Last, Mary Z. (2014-09-08)). The purpose of RM is to offer the declarative function for specifying data and queries. Some commonly used RDMS are Oracle, MySQL and MS SQL. I have used MySQL database management software for this work. The reason behind to use this database management software is that, its free, used by popular web applications like Drupal, WordPress, Joomla, Facebook and Twitter and then easy to learn are some of them. And now it is time to see that the concept of image processing, image processing is the process of applying computer algorithms to extract useful information from an image to make an image more useful after acquiring an images from image taker devices such as digital cameras or scanners.

Why we need to process images? Have a look the following two main reasons why we process images.

- To improve image information for human interpretation.
- To improve image information for machine perception.

So, this work focuses on the second reason that is why we process an image, extract scanned scoring sheets image data.

2.1.4 Computer graphics

Displaying an image on a screen for the size we want is a hard method. Computer graphics are used to lighten this difficulty. It is a skill of plotting an image on a computer display area with the help of computer programming. It comprises the creation and operation of images. Transformation is one main concept in computer graphics and have a look the following subtopic.

2.1.4.1 Transformation

As the name indicates that transformation means changing some graphics to something else by applying procedures. There are numerous forms of transformations for instance translation and rotation are some of them. Transformations that are takes place on 2D plane we call them 2D transformation. It plays an imperative role that relocate graphics on the display area of computers and modification of their size and orientations.

2.1.4.1.1 Translation

A translation transfers an object to another location on a computer display area i.e. screen. We can get a new pair of coordinate (X_1, Y_1) by adding the translation pair of coordinates (X_t, Y_t) to the original pair of coordinates (X_0, Y_0).

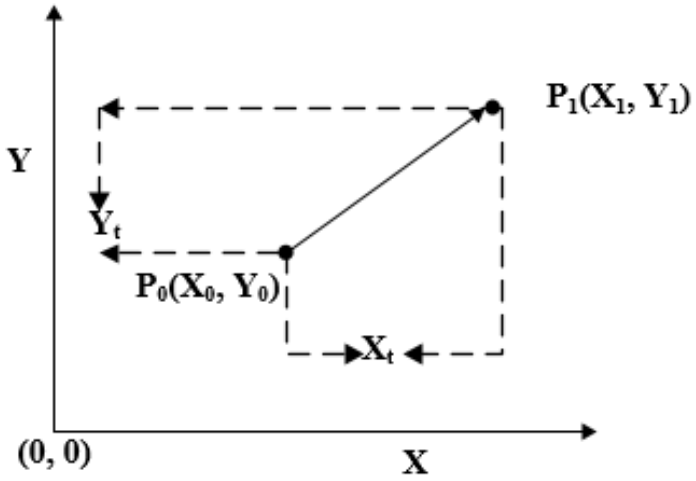


Figure 2.6: 2D point translation

From the above Figure 2.6 we can write in the following ways: -

$$X_1 = X_0 + X_t \quad (2.1)$$

$$Y_1 = Y_0 + Y_t \quad (2.2)$$

The coordinate pairs P_t (X_t, Y_t) is named as the translation coordinate and we can write also in the following ways: -

$$P_1 = P_0 + P_t \quad (2.3)$$

2.1.4.1.2 Rotation

In rotation, we rotate graphics object with a certain angle phi(ϕ) from its starting pair of coordinates. Have a look the following figure 2.7: -

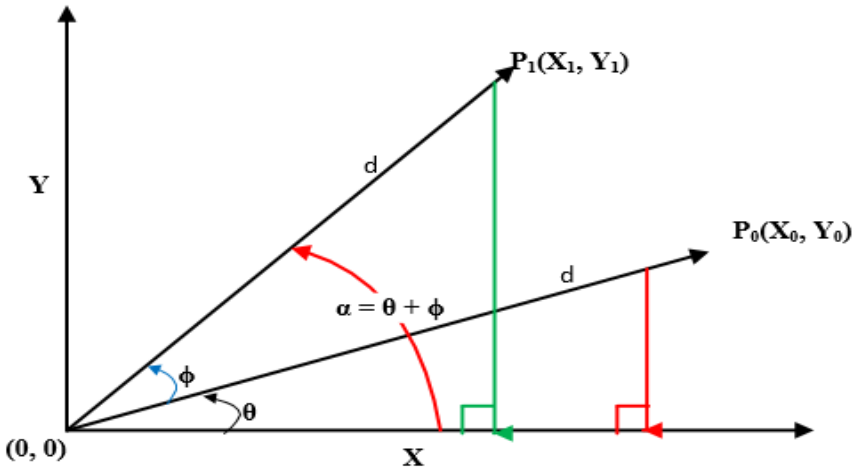


Figure 2.7: Rotation by angle phi(ϕ)

From the above Figure 2.7, we can see that the point P_0 (X_0, Y_0) is located at angle θ from the horizontal axis X with distance d from the origin (0, 0). Let's suppose we rotate the point P_0 (X_0, Y_0) by angle phi(ϕ) with the existing angle θ and we will get the point P_1 (X_1, Y_1) at new location. By using Pythagoras's theorem, we can represent X and Y as follows: -

$$X_0 = d * \cos \theta \quad (2.4)$$

$$Y_0 = d * \sin \theta \quad (2.5)$$

On the same fashion, we can represent the point P_1 (X_1, Y_1) from the fundamental formulas of angle addition in trigonometry as follows: -

$$X1 = d * \cos (\theta + \phi) = d \cos \theta \cos \phi – d \sin \theta \sin \phi) \quad (2.6)$$

$$Y1 = d * \sin(\theta + \phi) = d \cos\theta \sin\phi + d \sin\theta \cos\phi) \quad (2.7)$$

And then, substitute equation 2.6 and 2.7 to 2.4 and 2.5 respectively, we will get the following result.

$$X1 = X_0 \cos\phi - Y_0 \sin\phi$$

$$Y1 = X_0 \sin\phi + Y_0 \cos\phi$$

Represent the above equation in matrix form,

$$\begin{bmatrix} X1 \\ Y1 \end{bmatrix} = \begin{bmatrix} X_0 \\ Y_0 \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin \\ \sin\phi & \cos\phi \end{bmatrix}$$

OR, $P_1 = P_0 * R$

Where R is the rotation matrix,

$$R = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$$

But for negative angle rotation, the matrix changes in the following ways,

$$-R = \begin{bmatrix} \cos(-\phi) & -\sin(-\phi) \\ \sin(-\phi) & \cos(-\phi) \end{bmatrix} = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix}$$

From mathematical trigonometry negative angle identity where, $\sin(-\phi)$ is equivalent to minus of $\sin\phi$ and $\cos(-\phi)$ is equivalent to positive of $\cos\phi$.

For instance, the point P_0 (2, 3) rotate with counter clockwise rotation angle 90 ($\pi/2$), we can calculate the rotation and will get the new point in the following way,

$$X1 = 2 \cos(\pi/2) - 3 \sin(\pi/2) = 2 * 0 - 3 * 1 = -3$$

$$Y1 = 2 \sin(\pi/2) + 3 \cos(\pi/2) = 2 * 1 + 3 * 0 = 2$$

Therefore, the point P_0 (2, 3) rotated by ($\pi/2$) with counter clockwise direction becomes P_1(-3, 2), whereas negative rotation becomes,

$$X1 = 2 \cos(\pi/2) + 3 \sin(\pi/2) = 2 * 0 + 3 * 1 = 3$$

$$Y1 = 2 \text{ -sin}(\pi/2) + 3 \cos(\pi/2) = -2 * 1 + 3 * 0 = -2$$

On the other hand, the point P_0 (2, 3) rotated by ($\pi/2$) with clockwise direction becomes P_1(3, -2).

The following Figure 2.8 shows, the clockwise and counter clockwise rotation of point P_0 (2, 3) by rotation angle ($\pi/2$).

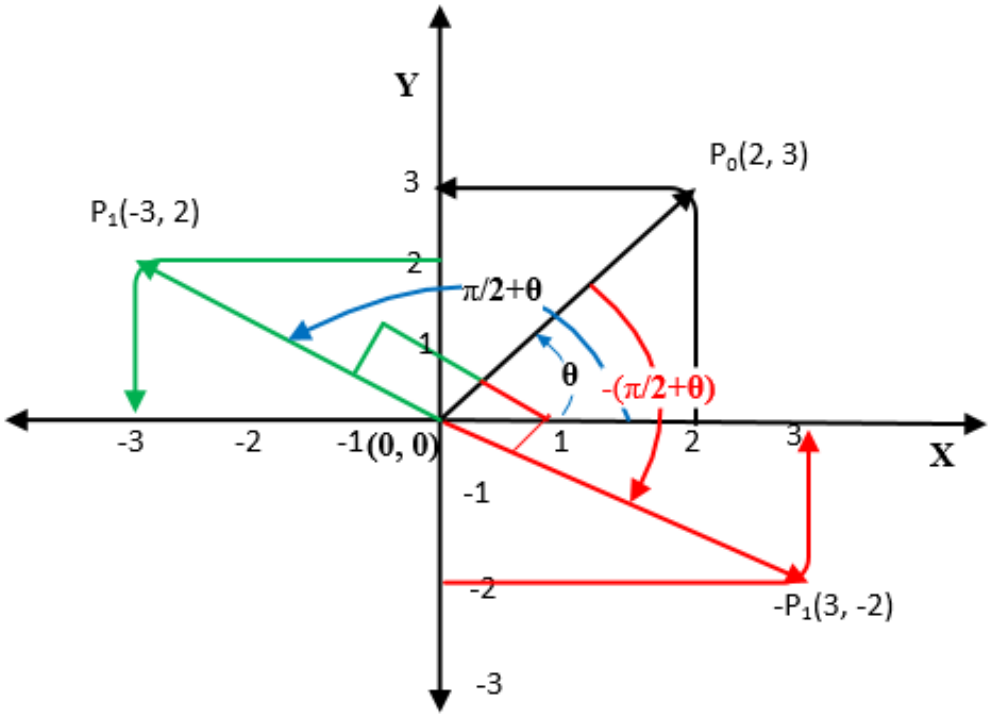


Figure 2.8: Rotation of points both in clockwise and counter clockwise direction

2.2 Literature Review

Rakesh S., Kailash Atal, & Ashish Arora. (2013) were conducted a research on an optical mark reader for Kolkata institute entrance multiple question examinations. The study uses normal printer and scanner for printing and scanning scoring sheets respectively. 39 scanned scoring sheets out of 5000 were ambiguously erased so, this ambiguity leads to manual verification of those scoring sheets for fair evaluation, otherwise the system couldn't generate exact result as they were expected and error rate is around 0.78%.

The study had a minimum of 250 bubbles per one scoring sheet but, what if at least two bubbles per question will darken by mistake? It should be done in manual way. If we think like two bubbles gets darken for one question per scoring sheet and continue this mistake up to 5000 scoring sheets, this work will haven't more advantage and switches to manual matching to answer keys. Because, as they wrote two bubbles were seems marked per question needs manual assistant for correct result either erasing one marked bubble. The

type of programming language they had used for implementation side is that java with standalone type of software. As long as, they were developed standalone software, it needs local installation for those machines wants to do optical mark recognition task. Most developers said that java is not pure platform independent programming languages and they agree with only scripting languages are, Rai.Skumar. (2013). For instance, System.getProperty("os.name") is clearly shows operating system dependent. The problem gets bigger when new operating system will develop.

Ms. Sumitra B. Gaikwad. (2015). has presented Image Processing Based OMR Sheet Scanning using ordinary scanner. She had also suggested that with small modification of the OMR system will works by taking sheet mages from digital camera. C# as backend and OpenCV as image processing library used for OMR system implementation. The system enables us to design answer sheets that contains 50 multiple choice questions. Its implementation is based on region of interest built in OpenCV algorithm. The algorithm used by the system works by taking RGB images from scanner and or digital camera with a little modification of source code, convert RGB image to binary, crop region of every questions so that visible only four alternatives and finally decide the answer is based on the location of X and Y co-ordinate. The result of the study is not defined on the paper for deciding its work

When two or more bubbles gets marked per question and another bubbles there outside the target question, the paper do nothing. But, there might happens a problem with optical mark recognition system and generate unexpected result.

Ishu Madan, Kashif, Narendra Sahu and Garima Krishna, Hemant Ram Rana. (2013). They did a research paper on the Implementation of optical mark reader technology with the Help of Ordinary Scanner. They have used java programming language with swing components for graphical user interface design and its implementation as well for software tool and then printer and scanner as hardware tool. They divide the system into three usual simple phases named as input, process and output. The first phase contains searching and selecting current working folder and answer key to be assign to recognition result. the second phase contains two modules i.e. recognition and analysis. Recognition phase is the key phase in this work

that converts fill-in image data i.e. marked bubbles data to digital data(ASCII). The last phase provides to users the result of recognition.

Its objective is to reduce cost of implementation of OMR, cost of scanning sheet is around 0.25USD, to make the application available to wide range of scanners, to make the application that works on a certain level of damaged sheets, the damage includes stain of scoring sheets. They have developed an algorithm with three main steps namely finding corner points of sheet, check orientation and finally reading the fields. The first step solves the problems that might happen by human error and its problem is that the sheets not being keeps its straight during scanning. In order to find the corner points, we need to know that the sheet was rotated toward to left side or towards to right side or not rotated at all. The second step that after founding the corner points of sheet, the precise angle of rotation needs to be calculated. The angle of rotation found by Pythagoras theorem i.e. angle = tan inverse of perpendicular over base of sheet. Finally, rotate the sheet by the angle we found earlier to make it straight. While the optical mark reader sheets were designed by users previously, the attributes of the sheets were being stored like horizontal and vertical space between bubbles, radius of bubbles and the like and finally a rough estimate of sheets attribute will be taken. Its result is nearly accurate even though not defined by number. However, the optical mark reader system will not be working while the sheet is knowingly injured.

Astha Gupta and Sandhya Avasthi. (2016) were conducted a research on the area of optical mark reader for surveys. They have written a method of developing an optical mark reader software for processing surveys and researches but, the body their text talks about multiple choice questions optical mark reader. They have focused on errors that will happen on during scanning and photocopying of questionnaire forms. They have recommended that the standard template questionnaire sheet would photocopied after getting one printed form for cost reduction purpose. Because the cost we pay for print and photocopy have a little difference per page even though have its own risk on OMR software they had developed.
The errors that will happen during scanning and photocopied questionnaire forms are due to translation, scaling and rotation of both scanned and photocopied questionnaire forms and also proposed the appropriate solution as well. To solve such detected problems, they proposed their solutions with four phases. These phases are listed below

- design a template
- take questionaries' image
- applying 2D transformation over questionaries' image
- mark finding

The first phase of the work is that preparing questionaries' image by using Microsoft office word. All of the questionaries' image corner coordinates have to be keep in database, these coordinates are top left corner coordinate, top right corner coordinate, bottom left corner coordinate and bottom right coordinates as well. for accurate result, all saved coordinate values should be consistent.

The second phase is that capturing an image, to capture an image they were used flatbed scanners because it has consisted two functionalities in combined way i.e. scanning and printing. Another main reason that they were used these scanner is that its cost lies between 40USD to 75USD, while OMR scanners costs between 570USD to 1290USD. The third phase is applying 2D transformation over scanned images, during scanning and printing the sheets will not set to scanners and printers correctly because of human beings problem. If there is tilted scanned image, would needs rotation with a certain angle. Its efficiency is comparable to an old OMR scanner(reader) even though not written in numerical way and error rate too. Mark finding is the last step of their work, to find the marks the previous steps should process correctly for accurate result. all of questionaries' coordinate that are discussed earlier have to be matched exactly with template image layout. All of the marks on questionaries' would be marked or black. If 50% and more pixel coordinates are counted per one alternative, then the mark is expected to be filled.

Hui Deng, Feng Wang and Bo Lian. (2008) have presented a price effective optical mark reader application on an international conference. Their main aim is to solve the drawback of traditional OMR and the developed system works on both windows vista and XP. Their system named by low cost optical mark reader(LCOMR) and has the following advantages over the traditional OMR.

- To design the questionaries' sheet they use Microsoft office word.
- Image of questionaries' sheet can obtain from any kind of scanners.

Their system consists of four main steps: optical mark reader answer sheet preparation, answer sheet acquisition, layout correction and mark recognition. Most instructors in school are experts in Microsoft office word so, the system enables them to design it easily. The layout of the sheet can adjustable by sheet designers easily. Depend on the requirement, the sheet design can vary from designer to designer. Their sheet has two types of marks: darken and hollow marks. The solid mark used by the system that helps to read solid marks whereas hollow marks are marks filled by students i.e. actual marks of the sheet is located in this area. Each hollow marks have its own unique coordinates values. In order to take sheet images, they use high speed scanners: both TWIN and ISIS scanners are supported. Their average processing speed are around 50 sheets per minute. For questionnaire sheets less than or equal to 70gsm, will tilt while pass thorough scanners, so such incline will extremely affect the accuracy of optical mark recognition system. To correct the tilt image, they use the angle calculated from by what angle the image is tilt using Pythagoras theorem i.e. angle of tilt equal to tan inverse of quotient of opposite over adjacent.

If the mark filled 40%, it will have considered as filled alternative. In case more than two alternatives will have filled more than 40%, the proposed system selects the maximum filled alternative by comparing them. While the recognition of multi filled choice is difficult for hardware based OMR, the LCOMR could propose by selecting the maximum filled choice becomes the one that is correctly filled. But, the problem is that equal percentage recognition per multiple choice in one question. Their proposed system needs more manual works to achieve 100% accurate recognition result. The use thin questionnaire sheets increase the error rate of mark recognition.

Douglas Chai (2016) has presented an optical mark reader research paper for multiple choice questions. The work tested by 50 multiple questions with 5 choices. Each question has 5 choices to fill students answer. On the top left corner side of the sheet contains students' id with ten columns by 10 rows size as well. So, the answer sheet contains 350 total squares if it will have 50 questions in it. His answer sheet design has six main parts namely: on the first top center area of the sheet contains the text" multiple choice answer sheet", next center area contains the text" title of an exam", and then it comes the text "name of student", the fourth part of the sheet contains two texts with one row and two columns i.e. student's id with 10

by 10 matrix and exam instruction on the other side. The last section is that the area of answer box. Based on the number of questions are there in the answer sheet, the answer sheet contains different amount of questions. His marking system takes scanned sheets from document scanners in the form of PDF (portable document format). But to feed the proposed system, its file format should convert to common image file name extensions such as PNG (portable network graphics) JPEG (joint photographic experts group) are some of them.
According to his study, we cannot know how to identify sheets course code. His system starts the recognition process starting from students answer and continues to their id. After accomplishing the recognition process, the system sends it to students' email address for feedback.

The average proposed speed of his system is 1.4 second per sheet including saving images to disk and excluding annotation. With annotation the system takes 3.5 seconds per sheet. The type of tools used by the system is that: intel core i5 processor laptop, the programming language were developed by MATLAB R2014b, scanner, A4 paper for preparing answer sheet and total 100 answer sheets with 50 questions in it. The significance of the work is presented earlier, but now it's time to see the drawback of their work as they conducted and I reviewed as well. Have a look below,

- The application cannot process multiple sheets simultaneously, means that if we have 1000 scanned sheets needs to be process, we have to click an upload button 1000 times. This process almost comparable to manual way of checking fill-in marks with correct answers, it is time consuming process and boring task when we have many scanned images.
- They couldn't identify the sheets identity and will add barcode technology for their future work.
- They cannot read students id from scanned sheet image.
- We don't know what will happen if all of the marks on questionaries' not marked.
- They were developed a desktop and mobile based OMR, and its web featurette is their future work.
- If more than one choices are filled per a single question, then the higher percentage of colored alternative is considered as a filled alternative but, what if they were

filled them in equal percentage? Will it evaluate as filled mark or considered as non-mark question? in addition to this limitation, how about none of alternatives were filled per question? will it has a problem in related to the order of the question?

Their proposed system needs manual assistant like inserting student's id and sheets identity. Therefore, the above lists of limitations are detected and I will solve them on my paper.

CHAPTER 3
METHODOLOGY

3.1 Research Methodology

Is scientific or systematic means of answering the research problem. Researchers not only know its methods and techniques but also its methodology. So they need to know which of these methods or techniques are appropriate or not appropriate, what would they mean? The scope of methodology is wider than that of methods or techniques. Methods or techniques are also included in methodology.

3.1.1 Data collection method

It is a process of collecting scoring sheets that helps to test the proposed system. After printing 30 scoring sheets for each type, I have invited friends to fill their own answer and then test my work with those collected scoring sheets. The collected scoring sheets are same as following.

Table 3.1: The content of scoring sheets

NUMBER OF QUESTIONS	TOTAL NUMBER OF SQUARES	REQUESTED SHEETS	COLLECTED SHEETS
125	30*((125*5)+(10*10)+(3*10)) = 22650	30	30
100	30*((100*5)+(10*10)+(3*10)) = 18900	30	30
80	30*((80*5)+(10*10)+(3*10)) = 15900	30	30
60	30*((60*5)+(10*10)+(3*10)) = 12900	30	30
45	30*((45*5)+(10*10)+(3*10)) = 10650	30	30
30	30*((30*5)+(10*10)+(3*10)) = 8400	30	30
20	30*((20*5)+(10*10)+(3*10)) = 6900	30	30
15	30*((15*5)+(10*10)+(3*10)) = 6150	30	30

3.1.2 Development method

From the review section of this work, most of the implementation algorithms are the aspect of mark, student id and course id detection. However, it is fair to say that bar code plays an important role for this work. The need to add bar code to scoring sheet is that, to differentiate which scoring sheet is currently getting recognize among 8 types.

3.1.2.1 Scoring sheet layout

An overview of proposed answer sheet is illustrated in figure 3.1. Which shows the different parts of scoring sheet. The scoring sheet as shown in figure 3.1 has eleven parts namely:

- Barcode
- Student id
- Course code
- Answer area
- Date
- Logo
- Name of ministry
- Student signature
- Invigilator signature
- Instructions
- Corner Points:
    1. Top left corner point
    2. Top right corner point
    3. Bottom left corner point
    4. Bottom right corner point

Even though the scoring sheet contains the above 11 parts in it, only five parts are needed for this proposed system. The remaining six part are not needed by this proposed system rather they are needed by organizations. For instance, logo is needed to indicate the scoring sheet belongs to the specific organization. As long as I have used an image processing technique, the above five parts of scoring sheet is mandatory. These parts are listed below:

- Corner Points:
    1. Top left corner point
    2. Top right corner point
    3. Bottom left corner point
    4. Bottom right corner point
- Barcode
- Student id
- Course code
- Answer area

3.1.2.2 Scoring sheet registration

This process should have done after scoring sheet is designed and before filled by students. As I mentioned earlier, there are eight types of scoring sheets were designed. To identify those sheets, its layout information with barcode is registered to database. On the other hand, this task helps us unregistered scoring sheets cannot process with this OMR system. The following steps shows that how to register scoring sheets to database table:

1) Upload scoring sheet
2) Fill text boxes that how many number of questions are there in scoring sheet
3) Fill the starting coordinate location values of column alternatives
4) Fill the corner coordinate location values of scoring sheet
5) Fill the starting coordinate location values of student code
6) Fill the starting coordinate location values of course code
7) Fill the starting coordinate location values of barcode

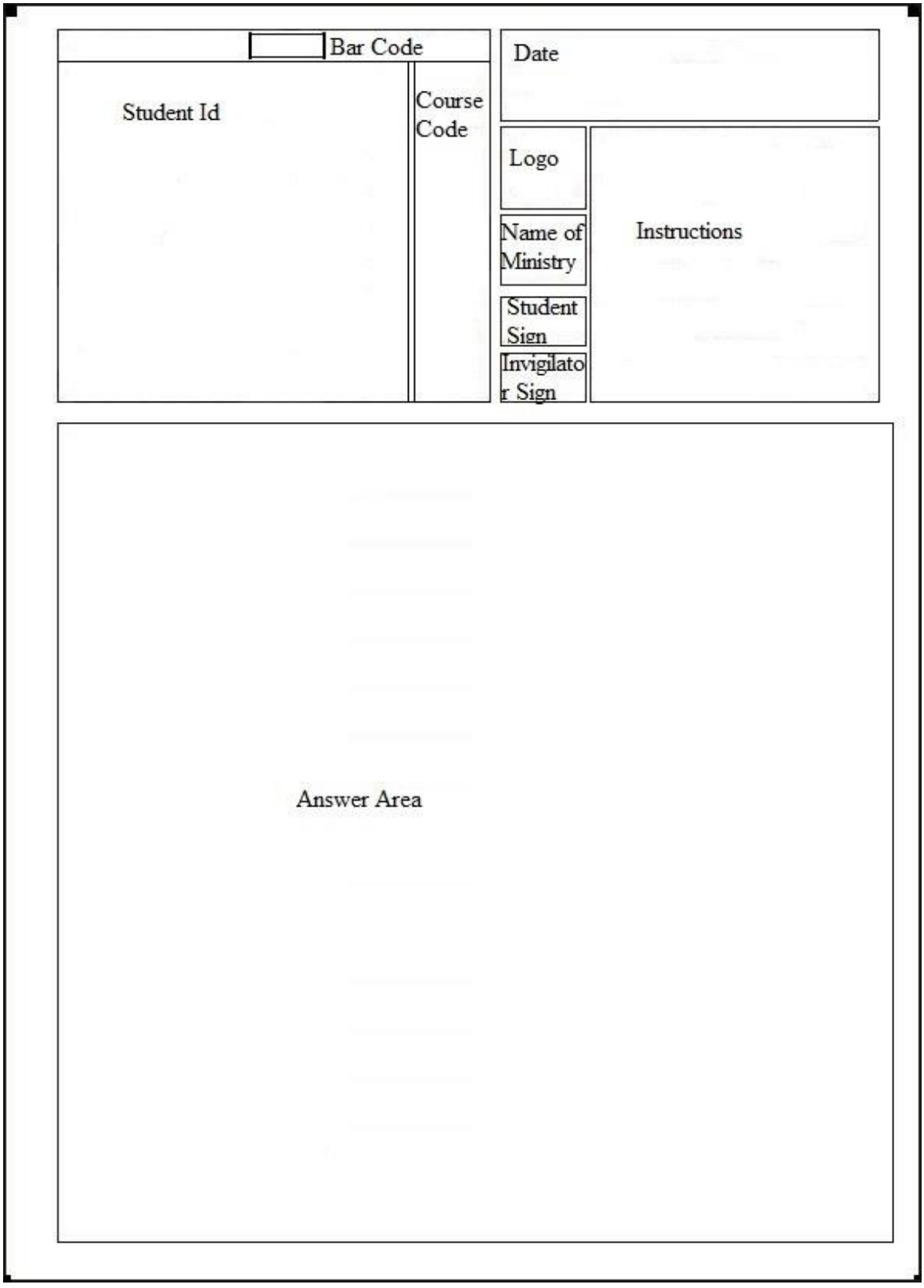


Figure 3.1: Proposed scoring sheet

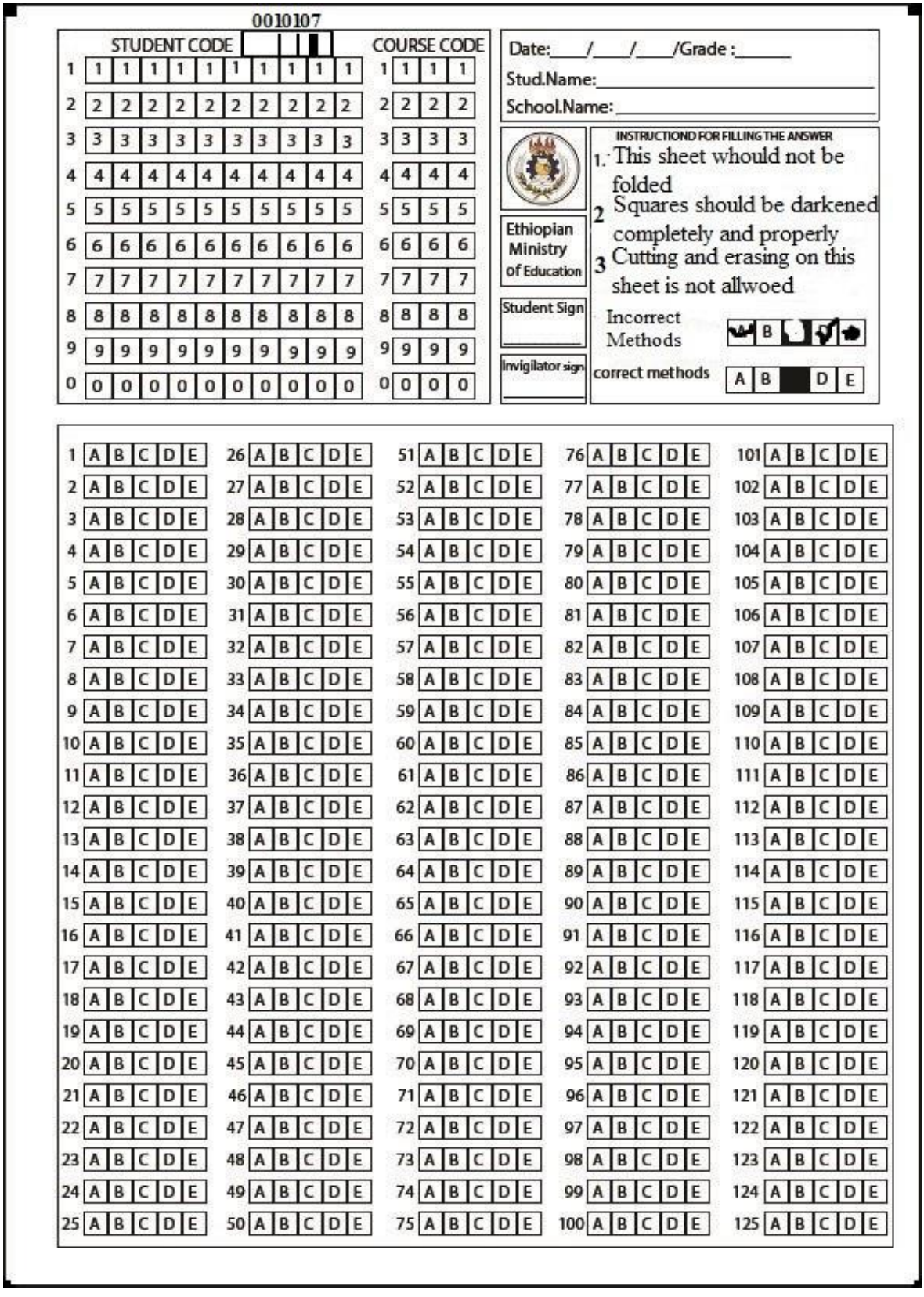


Figure 3.2: Proposed 125 questions scoring sheet

8) Fill decimal equivalent of that sheet barcode
9) Finally, insert number fields and text fields to database table that OMR system gets from step 2 up to 8

3.1.2.3 Image pre-processing

Converting color image to grayscale is key for minimizing CPU time, K. Dimililer & A. ilhanb, (2016). Before going to do the actual or target aim of the work i.e. mark recognition, an image should pass through the basic pre image processing techniques. The reason that an image should pass through pre image processing technique is that, the optical mark reader system returns correct result. Converting an image to grayscale is, the first pre image processing technique. As I stated earlier on literature review section, color images should convert to grayscale because of its size. Its size has a big effect on CPU time. The size of a single pixel in grayscale is equal to one third multiplication of one pixel of color image.

3.1.2.4 Adjust scoring sheet layout.

Before go to adjusting scoring sheet layout, we should know the scoring sheet is in either correct layout or incorrect. The following algorithm does the scoring sheet layout is in correct or incorrect manner.

1) Upload scoring sheet
2) Convert it to grayscale image type.
3) Get the height and width of scoring sheet
4) Compare results that we obtained on step 3 to the height and width of template scoring sheet.
5) If we got approached same result, the scoring sheet becomes either in correct layout or already rotated by 180^0 i.e. scanned in opposite direction.
6) Compare four layouts of scoring sheet that we got on step 5 with template image, if they are same, the sheet is in correct layout otherwise rotate it by 180 degrees.
7) If height and width of template and uploaded scoring sheet are not same, we need to find four corner points. We can find four corner points in the following way:

- Create a loop equal to half of height and width for its row and column respectively.
- Get intensity value of a pixel. The loop iterates until it gets more than 40% of grayscale pixel
- If the loop gets more than 40% of grayscale pixel; it will return the location of that pixel and this is top left corner point.
- Create a loop with half of height for row side and column side goes from total width size to half of width by one decrement.
- Get intensity value of a pixel. The loop iterates until it gets more than 40% of that of grayscale pixel.
- If the loop gets more than 40% of grayscale pixel; it will return the location of that pixel and this is top right corner point.
- Create a loop with half of width for column side and row side goes from total height size to half of height by one decrement.
- Get intensity value of a pixel. The loop iterates until it gets more than 40% of that of grayscale pixel.
- If the loop gets more than 40% of grayscale pixel; it will return the location of that pixel and this is bottom left corner point.
- Create a loop that starts from total height and width size to half of height and width size respectively by one decrement.
- Get intensity value of a pixel. The loop iterates until it gets more than 40% of that of grayscale pixel.
- If the loop gets more than 40% of grayscale pixel; it will return the location of that pixel and this is bottom right corner point.

8) After collecting four corner points on step 7, make transformation to template layout.
9) If height and width of template scoring sheet image is same as that of width and height of uploaded scoring sheet image respectively, that image is rotated either clockwise and anticlockwise direction by 90 degrees. So, rotate an image by 90 degrees and got to step 6.

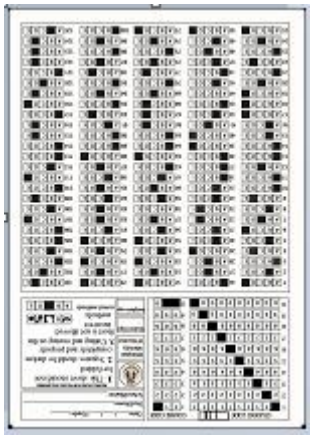


Figure 3.3: 180 degree rotated scanned scoring sheet

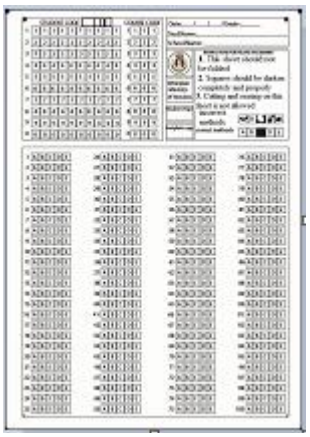


Figure 3.4: Template scoring sheet

Figure 3.5: Correctly scanned scoring sheet

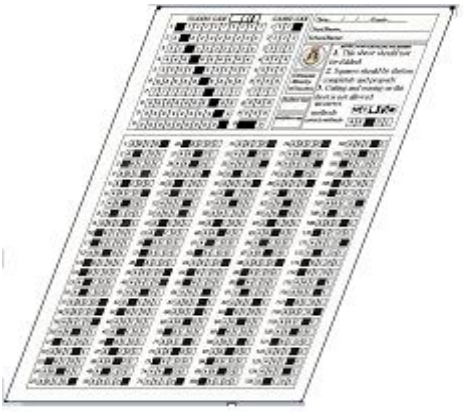


Figure 3.6: Incorrectly scanned scoring sheet

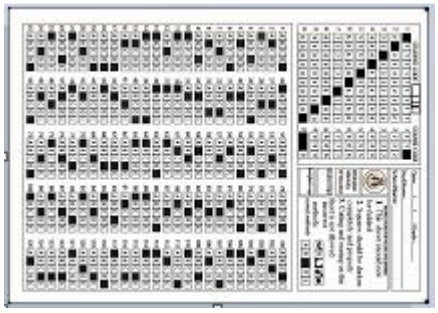


Figure 3.7: -90 degree rotated scoring scanned sheet

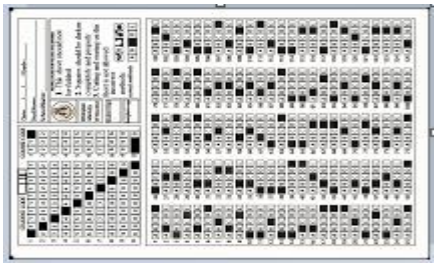


Figure 3.8: +90 degree rotated scoring scanned sheet

3.1.2.5 Barcode

In this research I have used bar code for categorizing sheets. When we translate this bar code to decimal number it has 7-digits. The first three digits shows an organization code, the middle two digit shows course code and the last two digit shows question type. The representation of question type with decimal number shown below:

Table 3.2: The representation of question types with decimal number

TYPE	TOTAL NUMBER OF QUESTIONS	EQUIVALENT DECIMAL NUMBER
1	125	1
2	100	2
3	80	3
4	60	4
5	45	5
6	30	6
7	20	7
8	15	8

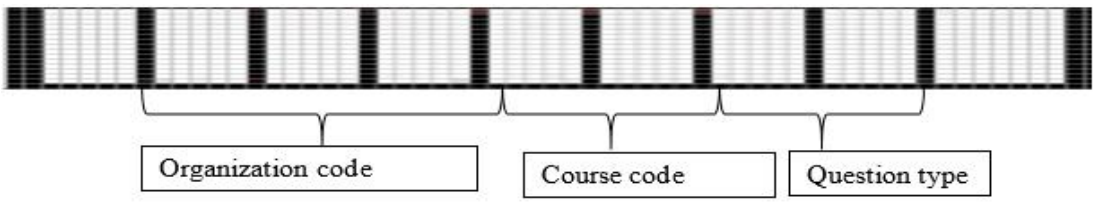


Figure 3.9: Barcode parts

On the above Figure 3.9, I use vertical black lines for separating one digit to next. Each seven parts contains a number from zero to seven in decimal. These concepts lead us to octal number system. 7 is the largest number when we convert binary to decimal number system for each 7 parts. But, for decimal number system there are ten letters to represent numeral system. So, we need to two extra digits i.e. 8 and 9 in it. To do such tasks we have to add extra bars both on left and right parts of the bar code. For example, the following bar code shows 0010103

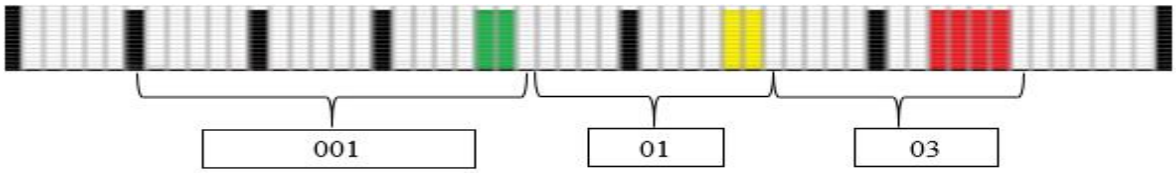


Figure 3.10: Sample barcode example

3.1.2.6 Course code recognition

After scoring sheet is registered successfully, the next step is reading course code. An algorithm to read course code is same as following:

1) Upload scoring sheet
2) Convert scoring sheet to grayscale image type
3) Retrieve barcode coordinate location from database table and read barcode
4) Check if the recognized barcode result will exist on previously inserted data to database table about barcode information.
5) If step 4 is correct (barcode information is existing on database table), the OMR system will retrieve course code from database table.
6) Retrieve the starting coordinate location of course code from database table that are kept on it during scoring sheet registration.
7) Get the intensity values of that region of image and count its gray shade intensity value.
8) If shades of gray will greater than 40%, that region of image considered as marked(filled). This process continues up to 10 rows and 3 columns. The result of this step will be 30 columns with one row array i.e. the result gets from multiplication of ten rows by three columns.
9) Reshape an array that are returned from step 8 to 10 rows by 3 columns.
10) Find an array location that intensity values satisfy the gray level shade rule and check the recognized course code digit length exactly equal to the length of course code retrieved from database table.
11) If the returned digit number is greater than that of three, the recognized course code will not valid else return recognized course code. In case, error may happen by students during filling more or less than three squares per sheet.

3.1.2.7 Student code recognitions

After recognizing valid course code, the next step is recognizing students code. An algorithm that reads students code is almost same as that of course code and listed below: Assume we have valid recognized course code

1) Retrieve the starting coordinate values of student code square location starts from database table that are inserted during scoring sheet registration.
2) Then get intensity values of that region of image and count its gray shade intensity values.
3) If shades of gray will greater than 40%, that region of image considered as marked(filled). This process continues up to 10 rows and 10 columns. The result of this step will be 100 columns with one row array i.e. the result gets from multiplication of ten rows by ten columns.
4) Reshape an array that are returned from step 3 to 10 rows by 10 columns.
5) Find an array index in which intensity values satisfy the gray level greater than 40%.
6) Accumulate results returned on step 5.
7) Check the result returned on step 6(recognized student code) is exist on previously kept in database table.
8) If the result exists on database table, the OMR system would return recognized student code.
9) Otherwise the process ends that means the uploaded sheet is invalid either incorrect student code is filled by students or some other reason.

From the above algorithms we have student code, course code, total number of questions and other information are known and needed by next recognition i.e. students answer.

3.1.2.8 Students answer recognition

This task will have done after successful completion of recognition of barcode, student and course code as well. otherwise an optical mark reader system considered as an uploaded image is invalid and will reject it. An algorithm used to recognize students answer on scoring sheet is shown below: -

1) Retrieve the starting coordinate values of columns by using its barcode from database table.
2) Then get intensity values of that region of image and count its gray shade intensity values.
3) If shades of gray will greater than 40%, that region of image considered as marked(filled). This process will continue up to end of questions. The result of this step will not be same, it depends on the number of questions on sheet. As long as we have different types of questions, we can calculate its squares by multiplication of alternatives with total number of questions.
4) Accumulate results returned on step3 to numeric array.
5) Reshape an array that are returned on step 4 to total number of questions and choices per column. The optical mark reader system gets total number of questions and alternatives from database table.
6) Because of student's mistake questions might not get marked so, we need to find which number of question gets marked or not marked. But in this step lets return marked question number. Create two dimensional array that we got from step 5 with length of first dimension equal to total number of questions and length of second dimension equal to its choices(alternatives).
7) If elements in step 6 array greater than 40 percent of total gray shade square, an OMR would append marked question numbers by instantiating a list until end of questions. As long as they obey an OMR system rules, whether all alternatives or two alternatives gets marked(filled) equally per question, this system recognizes the first shade as student answer.
8) If elements in step 6 array greater than 40% of gray shade, an OMR system would append column or choices(alternatives) values.
9) Translate column values that are returned from step 8 to multiple choice letters. if column values 0, 1, 2, 3 and 4 set letter "A", "B", "C", "D", and "E" respectively. if we think that every question number gets marked exactly one choice, we would stop here. But, we should think in worst case rather than best: it may happen totally unmarked alternatives per question and more than one alternatives gets marked per question.
10) Create a numeric array with length equal to number of questions.

11) If results returned by step 7 does not exist in results returned on step 4, create a dictionary with inexisted keys and set ‘0’values. This loop iterates until length of results returned on step 4 and finally returns a dictionary with key question number and values its alternatives.
12) Sort dictionary values returned on step 11 in ascending order with keys.
13) Fetch answer keys from database table that satisfies a specific SQL query conditions. Attributes used for SQL query comparison are course code, current academic year and then total number of question.
14) Check length of arrays returned on step 12 and 13 are equal.
15) If step 14 returns correct, create a loop with length equal to either results returned on step 12 or 13 and check its similarity. This step returns total correct and incorrect result of students’ mark.
16) Convert numeric result returned on step 15 to letter grade, numeric to letter grade conversion is varies based on the rule and regulation of an organization.
17) Write results returned on steps 15 and 16 to database table.
18) OMR system notify to students for their own web account on notification icon and can see their results.
19) If errors might happen, students can send message back to an OMR system about that error.

From the above students answer recognition algorithm, we can get the following things: -

- Question number
- Marked alternative(choice)letter even if it has more than one equal marked alternatives are there on same question.
- Unmarked question number

3.1.2.9 Make border for squares

The above work is used to convert marked image data to machine understandable format. But the following algorithm is used to border marked squares in different color for visual aid. Colors used for bordering are listed below: -

- Green color: - used for squares that are marked(filled) correctly.

- Red color: - used for squares that are marked(filled) incorrectly i.e. multiple squares filled per question.
- Yellow color: - used for squares that are total unmarked(unfilled) correctly per question.

1) Retrieve the starting horizontal and vertical location of squares from database table.
2) Count the number of gray shades that are greater than that of 40%
3) Sum up the results returned on step 2
4) If results, we got from step 3 are greater than 50% of total shade of squares, accumulate the starting vertices of that square.
5) If an algorithm did not get gray shades that satisfies this OMR system rule per question, an OMR system takes the starting point of that question and accumulate its results.
6) In this step we do have vertices returned on step 4 and 5. Then convert a gray scale image to color image.
7) Create a loop in which its length is equal to results we got from step 4 and draw a border around a square.

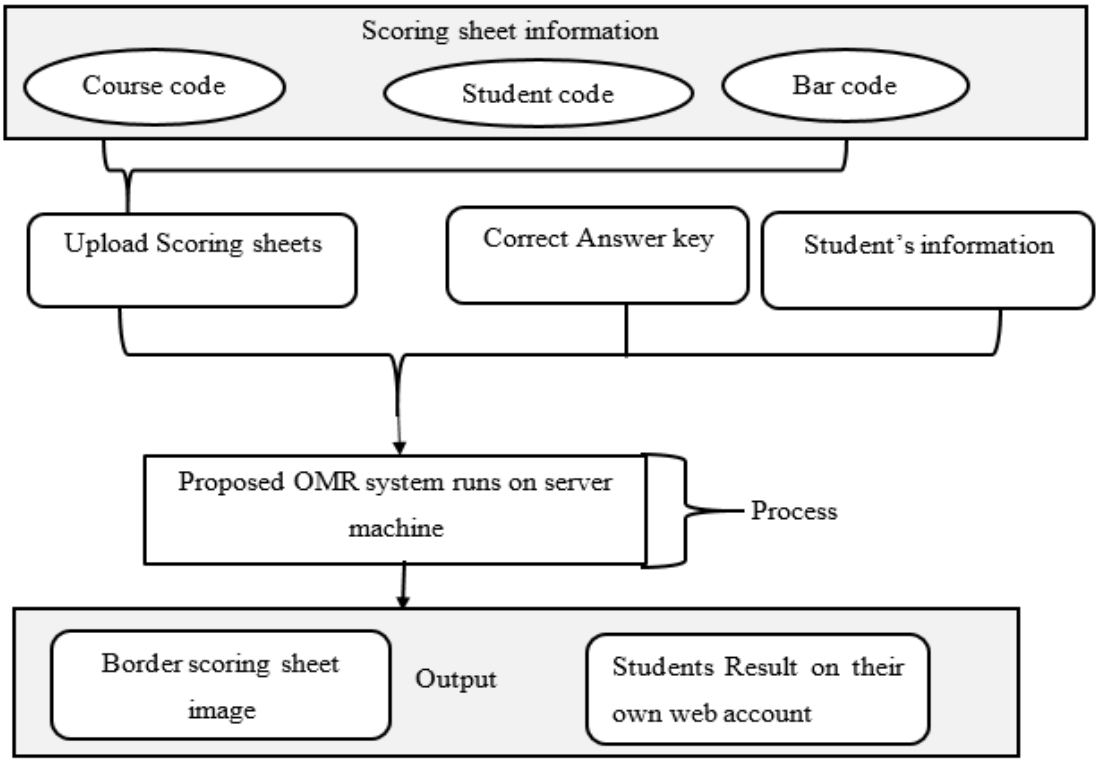


Figure 3.11: General workflow of proposed system

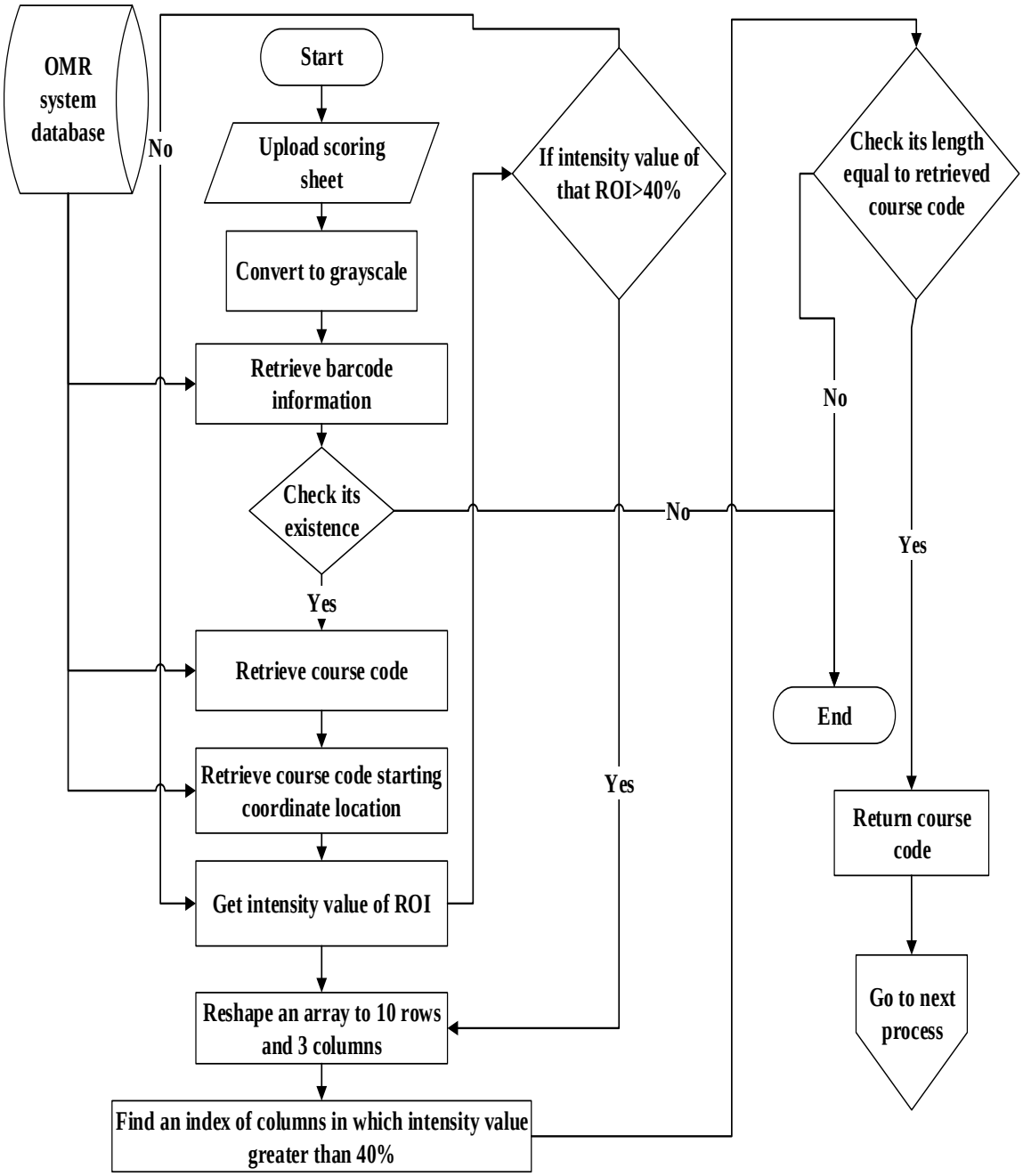


Figure 3.12: Flow chart for course code recognition

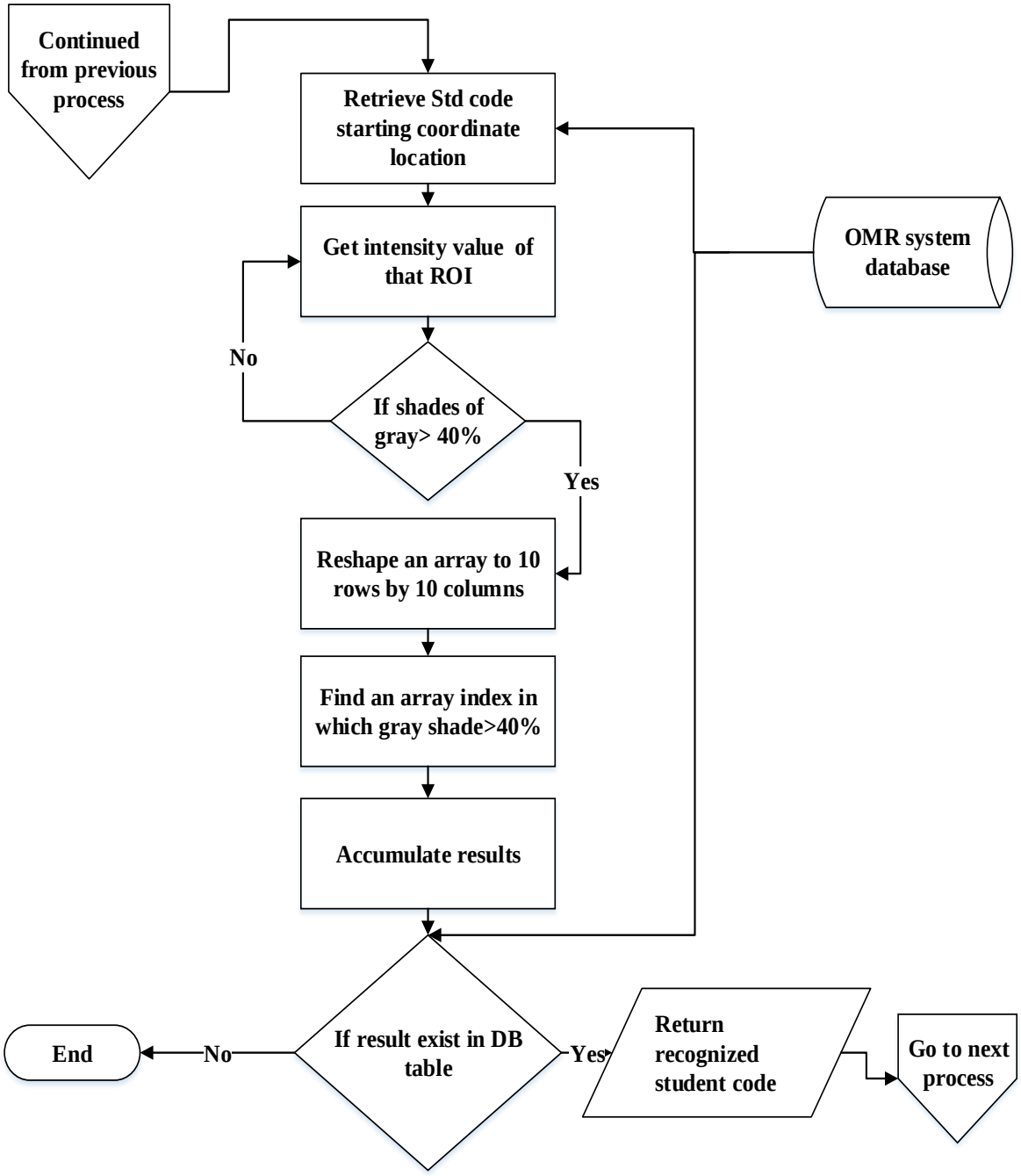


Figure 3.13: Flow chart for student id recognition

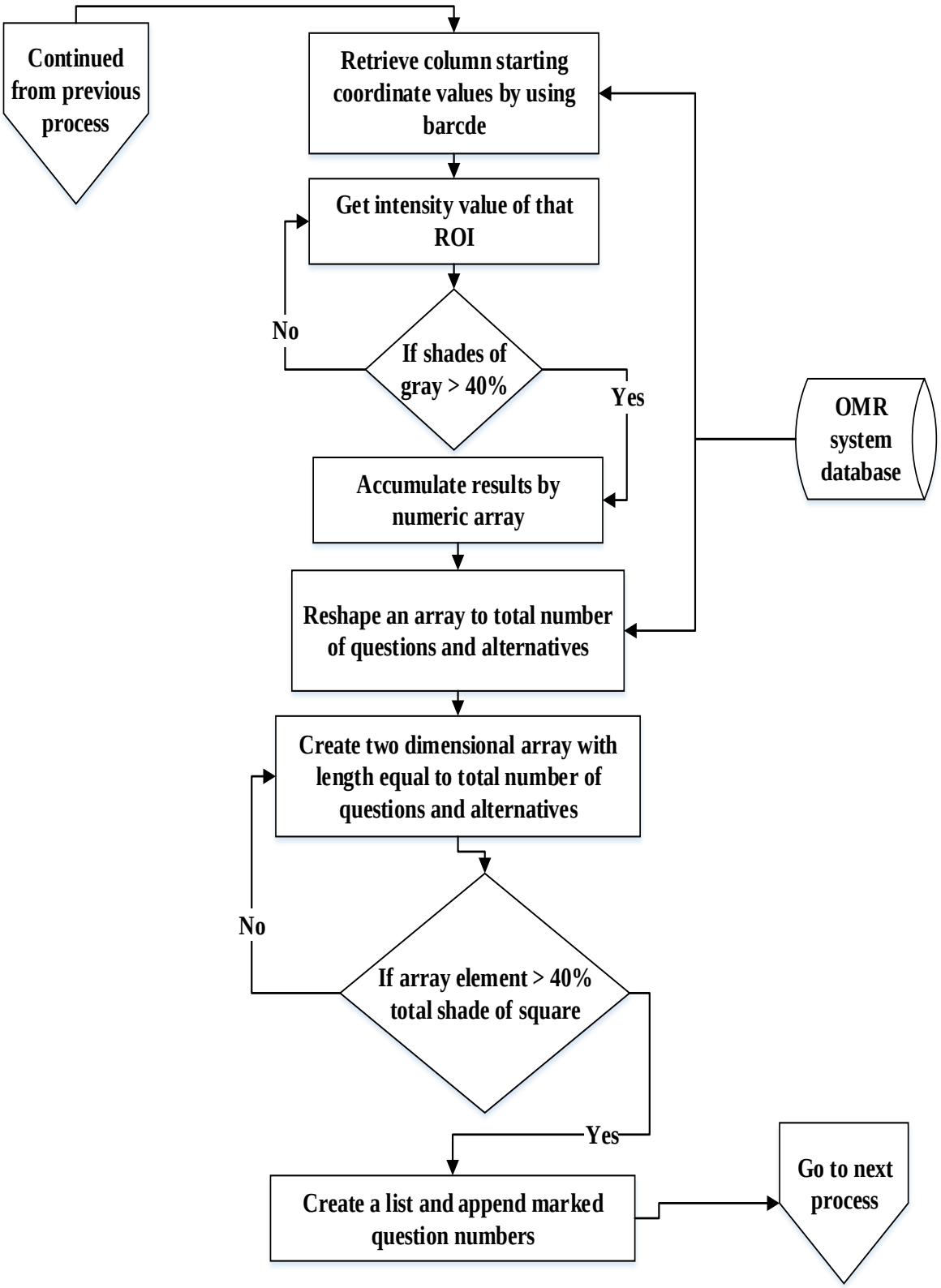


Figure 3.14: Flow chart for question numbers recognition

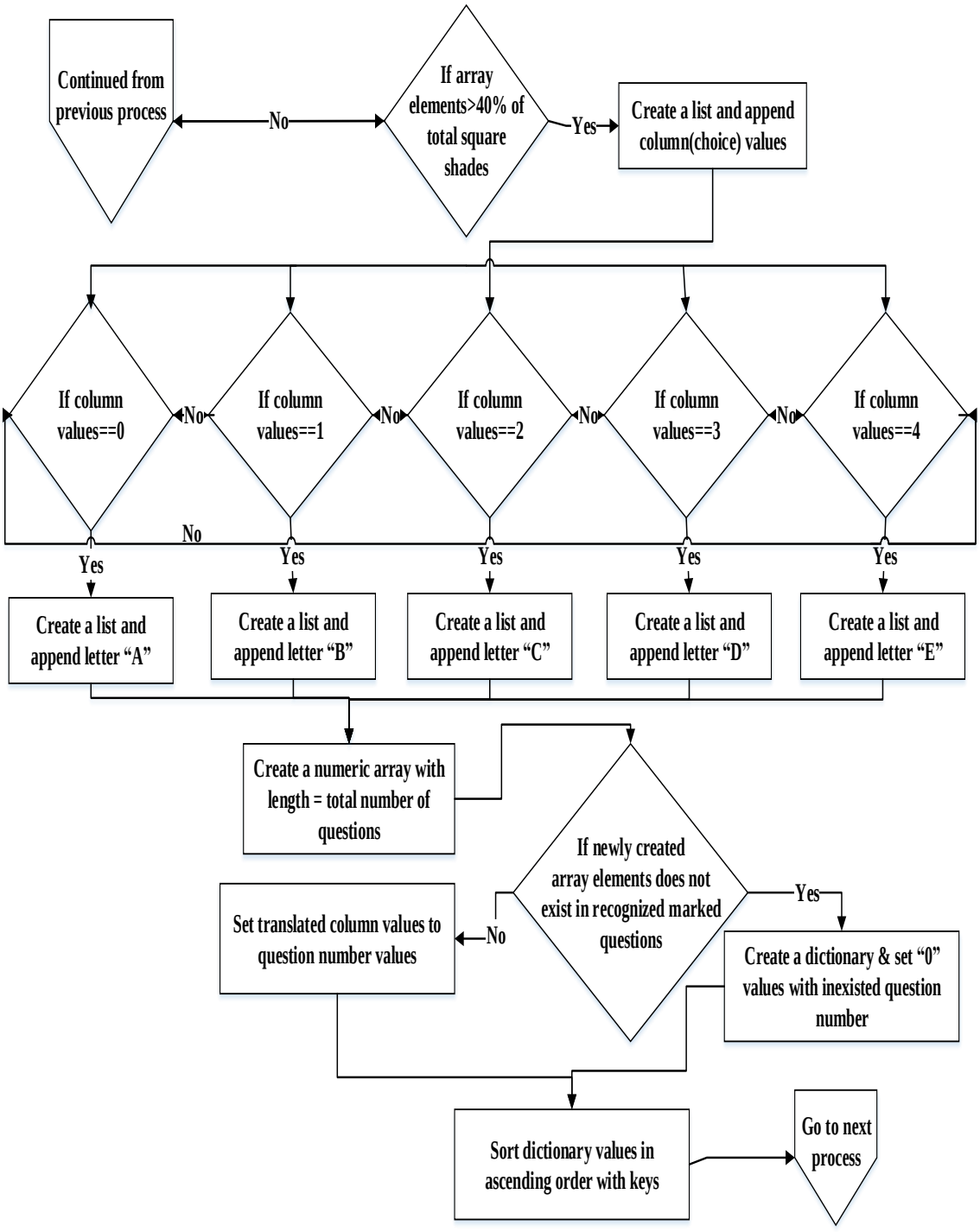


Figure 3.15: Flow chart for multiple choice mark recognition

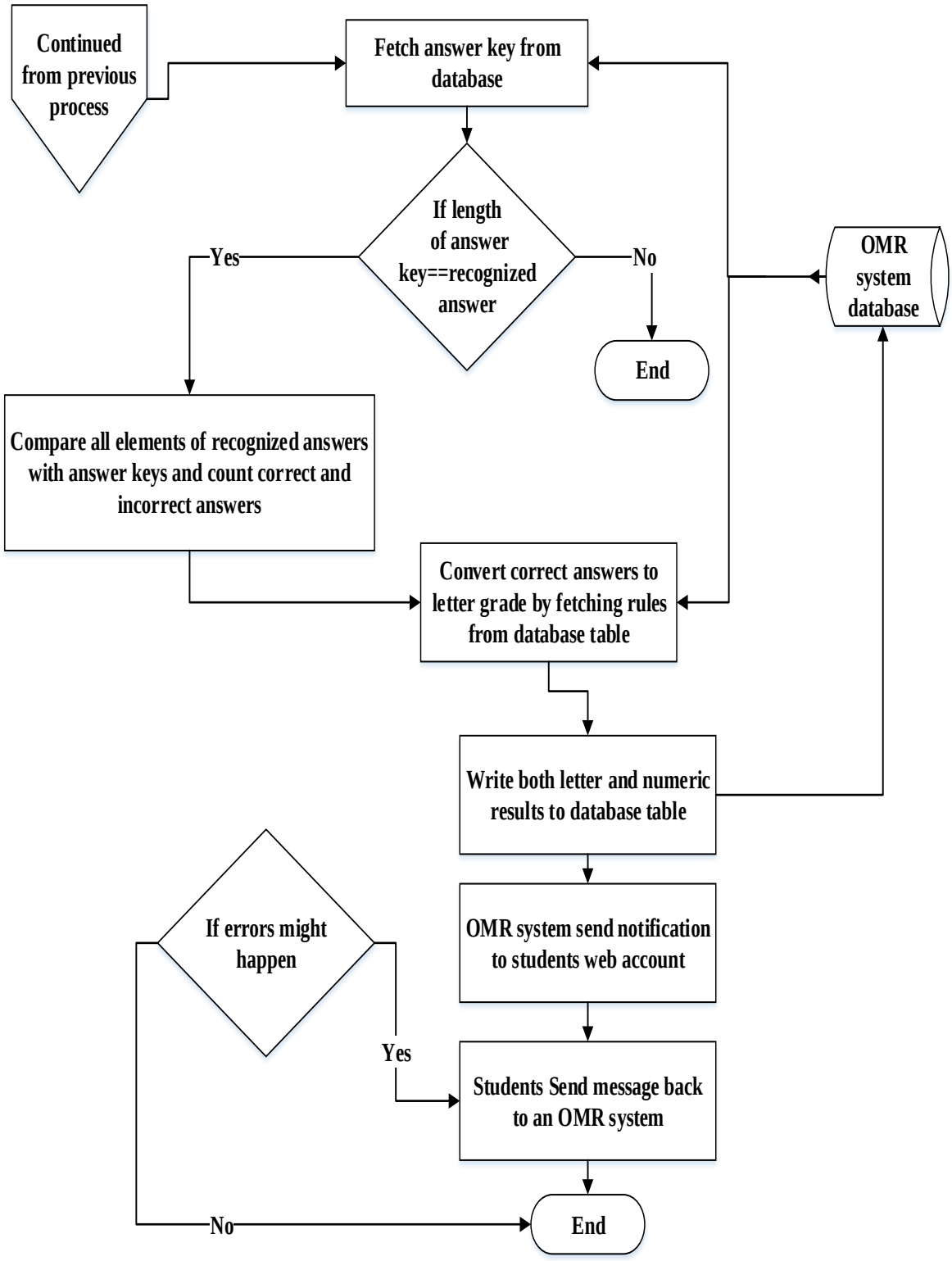


Figure 3.16: Flow chart for student result computation

CHAPTER 4
IMPLEMENTATION DETAILS

This chapter contains tools and implementation details of image and web based optical mark reader system. Systems that have both front and back end part needs a full stack developer or a separate back end and front end developers that should work in a smoothly manner. Python 3.6.5 language with flask 1.0.2 framework, MySQL database as back end and HTML5, css, js with bootstrap framework as front end is used for this system. As discussed in literature review, many papers future work is included in this work. These future works that are presented in literature review section is that adding course code to differentiate courses, barcode to differentiate sheets and registered or not to database, integrate the recognition result to web and enables it to view students result in their own registered id as user name and mothers name as password and then finally have a feedback to concerned body if errors will happen. This section offers a precise introduction of which tools are used for the accomplishment of the system and its implementations.

4.1 Tools Used

In this system both software and hardware tools are used for data preparation and implementation as well, so those used tools are listed below.

4.1.1 Software tools

4.1.1.1 Python

Python is a high level, popular, easy to use and learn and available on Unix, mac and windows operating systems. Until past year, python has 10,000 open source libraries. This leads us to say, python is rich in built-in libraries and many developers attracted by it. Because of its easiness and simplicity, python known as beginner's level programming language. The following are the list and usage of built-in libraries for this system.

1. OpenCV (stands for Open Source Computer Vision Library) is a huge open source computer vision library. But for this system, I have used image processing (imgproc module) for the following tasks: -

- To read an image from file by using cv2.imread() function
- To show an image by using cv2.imshow() function
- To write an image by using cv2.imwrite() function
- To convert color to grayscale image by using cv2.IMREAD_GRAYSCALE function
- To convert grayscale images back to color image by using cv2.IMREAD_COLOR function
- To draw a rectangle around marked and unmarked squares by using cv2.rectangle() function
- To access and modify pixel values for required region of interest in an image.
- To transform an image same as to template image by using cv2.getPerspectiveTransfrom() and cv2.warpPerspective().

2. Numpy is a short form of numerical python. Even though it has many functionalities, I have used Numpy for array manipulation. As we know an image is a collection of rows and columns that are expressed in number so, these tasks that are done by Numpy like reshape the size of an array that are generated from an image and to know the shape of generated array (contains rows and columns).

4.1.1.2 Flask framework

In order to integrate this system to web, I have used flask web framework. This system is not that much broad and enough to use flask framework for web integration.

Paint for scoring sheet preparation

HTML5, CSS and JS used for front end development

Bootstrap is an open source front end framework. To have a responsive web application that are accessible by web enable devices for example mobile phones, computers and tablets without losing the content of web pages from device to device bootstrap is best choice.

MySQL database is an open source database management system established by oracle corporation. For this system all sheets and student's information are kept in MySQL database table either for later usage or for further processing.

4.1.2 Hardware tools

Most hardware tools listed below are used for data preparation
Printer is used for printing the prepared scoring sheets
Pen and Pencil is used to fill students answer, student's id, and course code in sheet
A4 paper having 210x297mm size, 70gsm weight and white color paper are used for scoring sheet preparation. Scanner to convert the printed paper after filled by students to softcopy format. Laptop computer is used for preparing scoring sheets, made implementation by using NEC, 8GB ram, core i5, 500GB hard drive, third generation processor, 2.50 GHz processor speed with windows 10 operating system. But, for testing the system tablets, different brand mobile phones and laptop computers are used. From laptop Lenovo, Hp and NEC; from mobile phone Techno, Samsung and Huawei; from tablet Samsung are used for testing the system.

4.2 Implementations

4.2.1 Data preparation

For the very first time to start implementation, we should have students, scoring sheet, course, login and other needed information in database table.

std_id	first_name	middle_namae	last_name	grade	mother_first_name	photo	registered_date
1000000000	Getinet	Amare	Mekonnen	12	5436e673595ac055701acf2ac5026dbbf656	getinet	2019-04-09 10:22:51
2000000000	Fithalew	Ashagrie	Mengiste	12	1c42f005ae796bdbe75ad044c5d6ef679d1a	fithalew	2019-04-09 10:22:51
3000000000	helen	bedilu	Amare	12	41f5f4ba9b611b122ded6cd28718949ab60c	helen	2019-04-09 10:25:26
4000000000	Ermiayse	Tefera	Mekonnen	12	9f2e63034115f7ae60a869139a7c75dae084	Ermiase	2019-04-09 10:25:26
5000000000	Addisu	Fentahun	Alem	12	40a10fc04e302d0ac5fdd9759a389fbec7087	addisu	2019-04-09 10:25:26
6000000000	Woriku	Mekonnen	Yohannes	12	15c2fab7c23f557ab9177ac3c81288ca7f7c8	woriku	2019-04-09 10:27:47
7000000000	Einsten	Alem	Debretsion	12	5b21b2706b6acde349542caa51221b8b7eb	Einsten	2019-04-09 10:27:47
8000000000	Jemal	Endriese	Mohammed	12	c960cdbc9225856055ebad232463b028d03	jemal	2019-04-09 10:27:47

Figure 4.1: Sample student's data in database table

On the above figure that are shown on column named by mother_first_name is encrypted, because it has used for student's password when they login in the very first time. We do have such tables for course, scoring sheet and other information. The system has two users i.e.

student and teacher. We have seen teachers task. We obtain the above data label on Figure 4.1 from html forms as shown on Figure 4.2: -

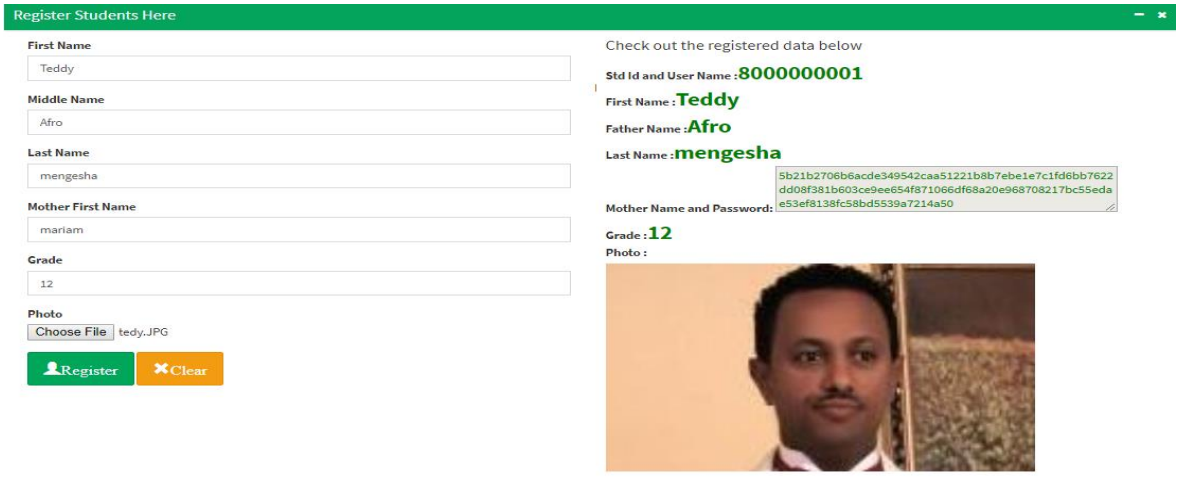


Figure 4.2: student registration page

After registering students, they can get user name and password and will login to their own account. When we login to the system as teacher, we got the above and following services on Figure 4.3

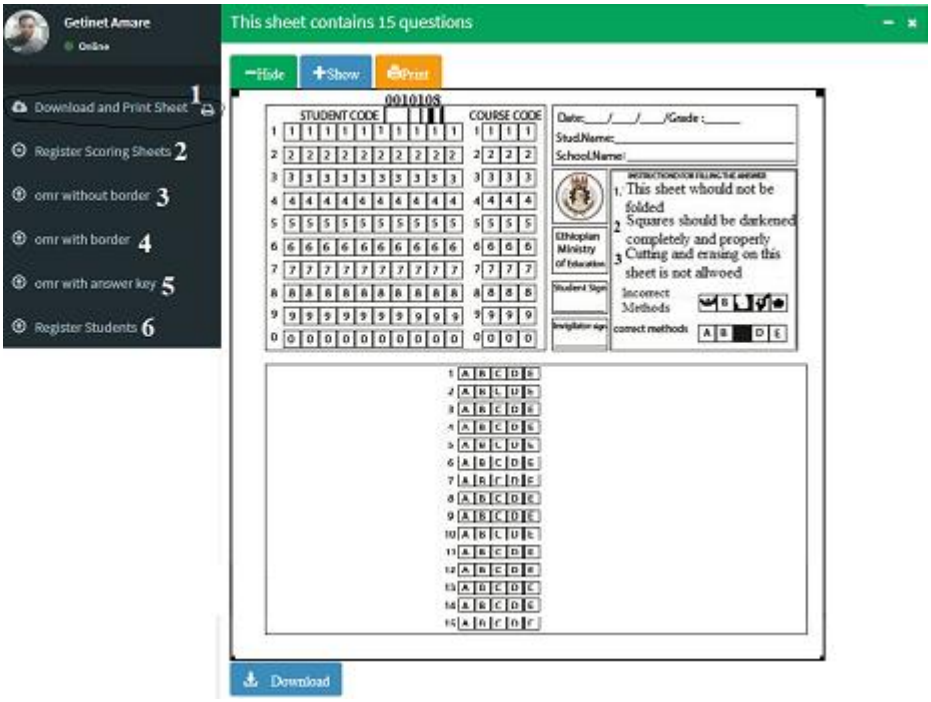


Figure 4.3: list of services when we login as a teacher

On the above figure 4.3 left navigator of the page labeled by number 1 contains download and print eight types of unmarked scoring sheets, meaning those sheets contain 15, 20, 30, 45, 60, 80, 100 and 125 number of questions. On register scoring sheet left navigator menu labeled by number 2 enables us to register new scoring sheets to database table for further processing and later usage.

4.2.2 OMR without border

OMR without border that are labeled on number 3 for the above Figure 4.3 contains the following: -

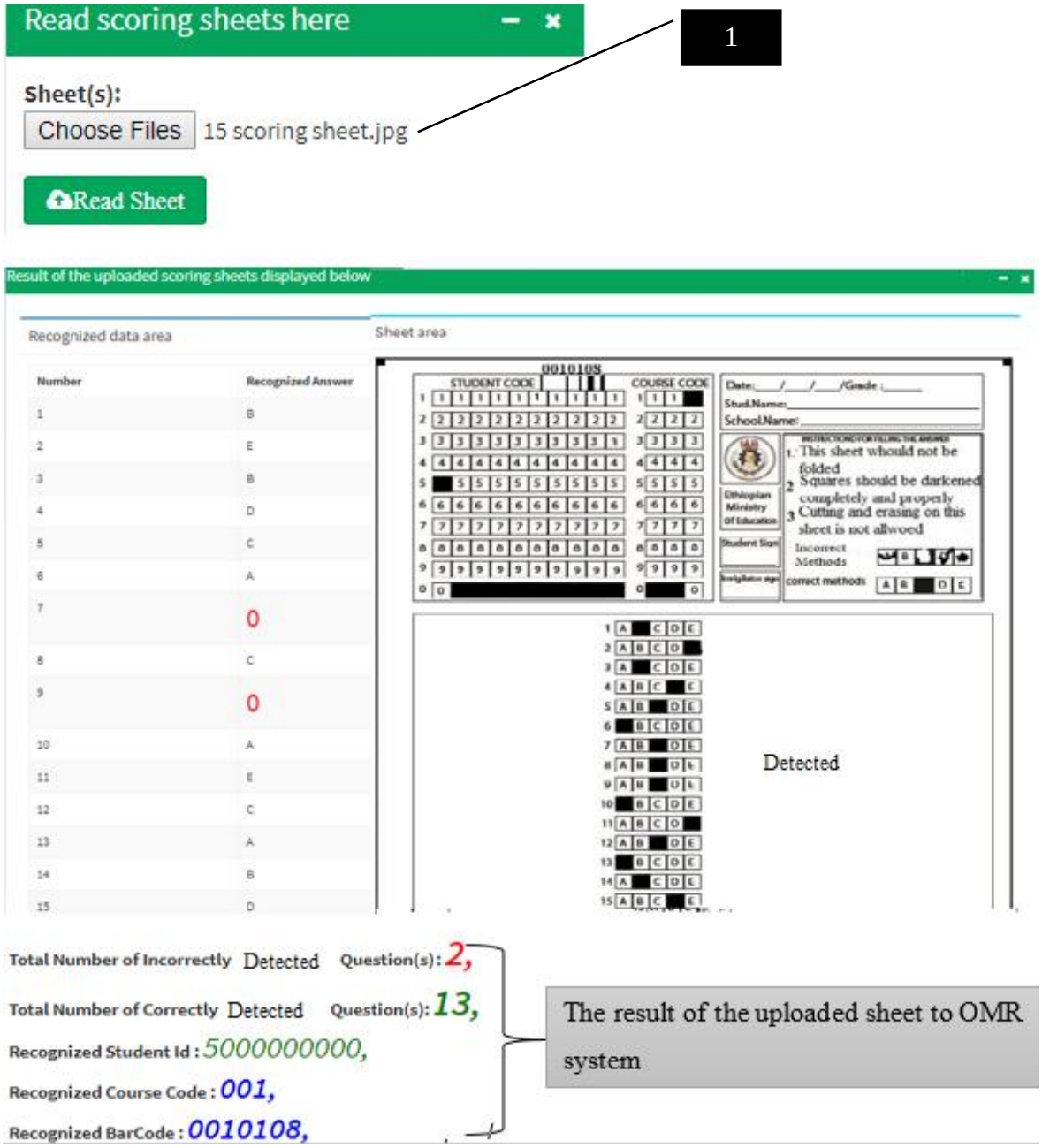


Figure 4.4: Uploading 15 questions scoring sheet and retrieve its result.

As shown on the above Figure 4.4, label 1 shows that select the scoring sheet we want to recognize by OMR system and click

button. After click read sheet button, the system displays the uploaded sheet and recognized data. Uploaded sheet it the sheet we got from scanner after filled by students. For this figure we uploaded 15 question sheet and we got two questions are not filled(marked) by 5000000000 student id, 13 questions are

filled correctly, 001 is the recognized course code and 0010108 is the barcode of the sheet. The system can read multiple scoring sheets at the same time by selecting multiple sheets to OMR system.

4.2.3 OMR with border

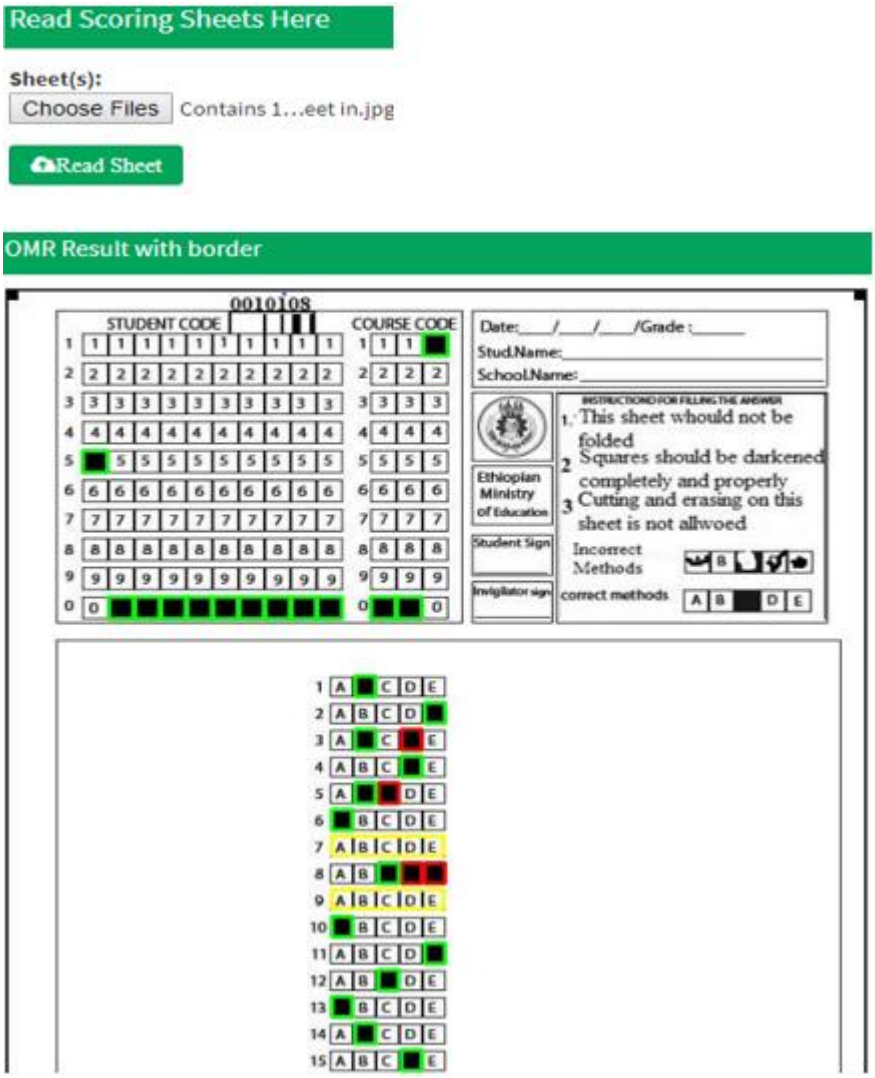


Figure 4.5: OMR with border

On the above Figure 4.5, those marked(filled) squares that are bordered by green color is recognized as correctly marked for example student, course id and question number 1 choice B are some of them. So those questions that are bordered by green color, counted by the system as correct answer. Question number 3 choice D, 5 choice C, 8 choice D and E are

bordered by red color, because there are more than one filled(marked) answer(s) is(are) detected per question. as long as the system detects more than one marked(filled) choices per question, the system takes the first choice is the correct answer and counted as correct answer. The last case is that totally unmarked(unfilled) questions might happen, this situation is happened on question number 7 and 9, that are bordered by yellow color as we see on figure 4.5. For this case OMR system mark counter considered as zero (no mark) for those questions.

4.2.4 OMR with answer key

In this case, an OMR system compares the recognized student answers that are done on previous step to appropriate answer key retrieved from database table. The system checks that the recognized scoring sheet data is there the appropriate answer key in database table or not? If it is there, the system starts to compare the students recognized answer with answer key kept in database table and generate the result as shown on Figure 4.7.

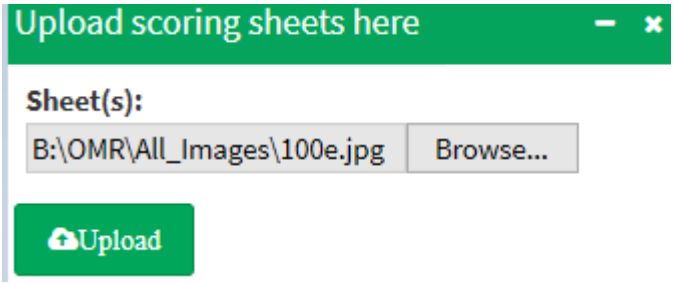


Figure 4.6: Uploading scoring sheet having 100 questions in it

When we click an upload button on Figure 4.6, the system read marks, student id, course code and barcode in it. After that the system computes the following things for each sheet.

- Total number of incorrectly marked question(s)
- Total number of correctly marked question(s)
- Total student result
- Letter grade
- Numerical value

For more sheets implementation result, have a look at appendix 1

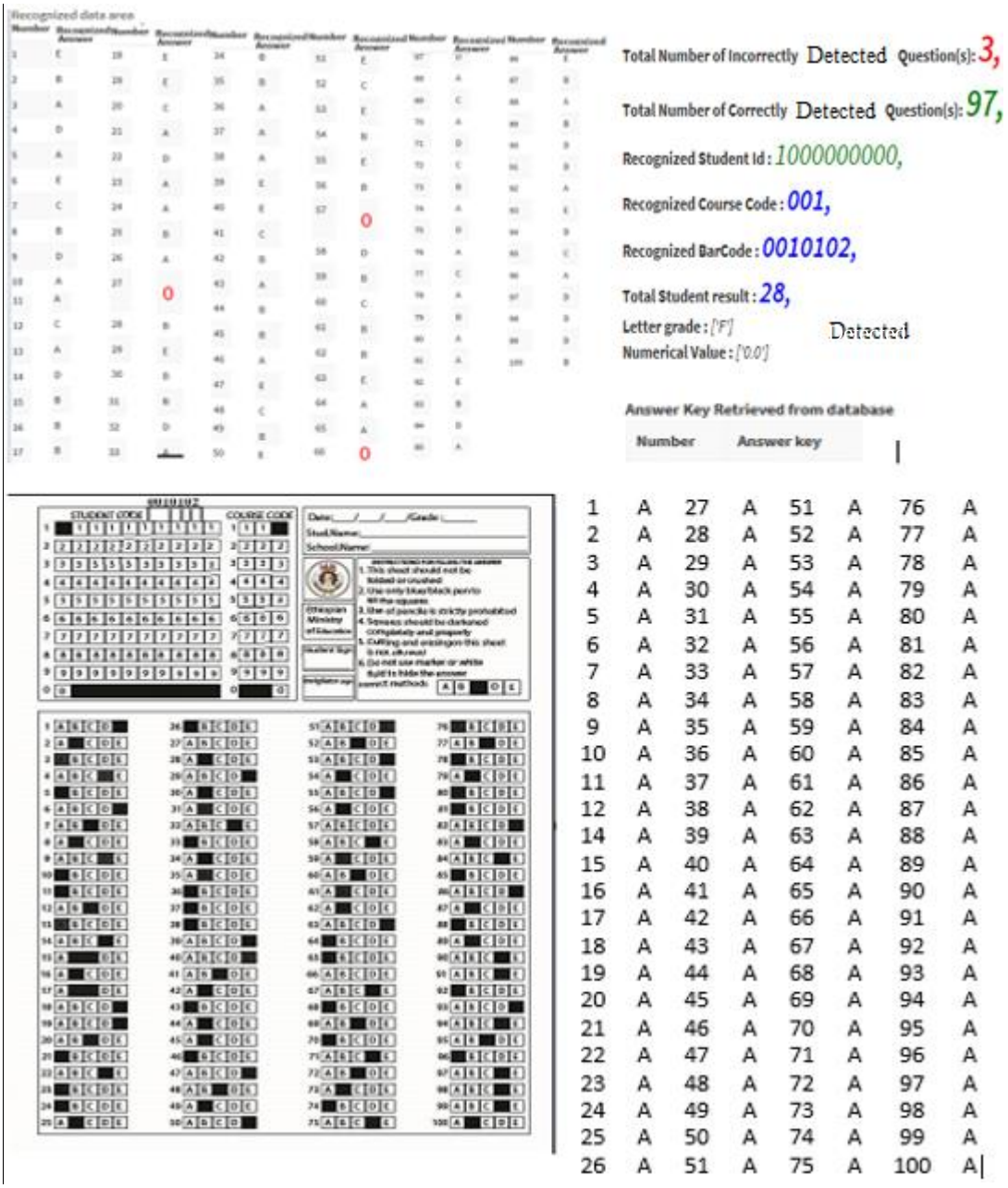


Figure 4.7: Result of uploaded sheet on figure 4.6

After the sheet processed by OMR system as shown on figure 4.7, that data is ready to be shown by student with id number shown on figure 4.7 i.e. 1000000000 and password is same as mothers first name as we saw on data preparation figure 4.1. Students have to login first to see their results on their own web account, have a look this process below: -

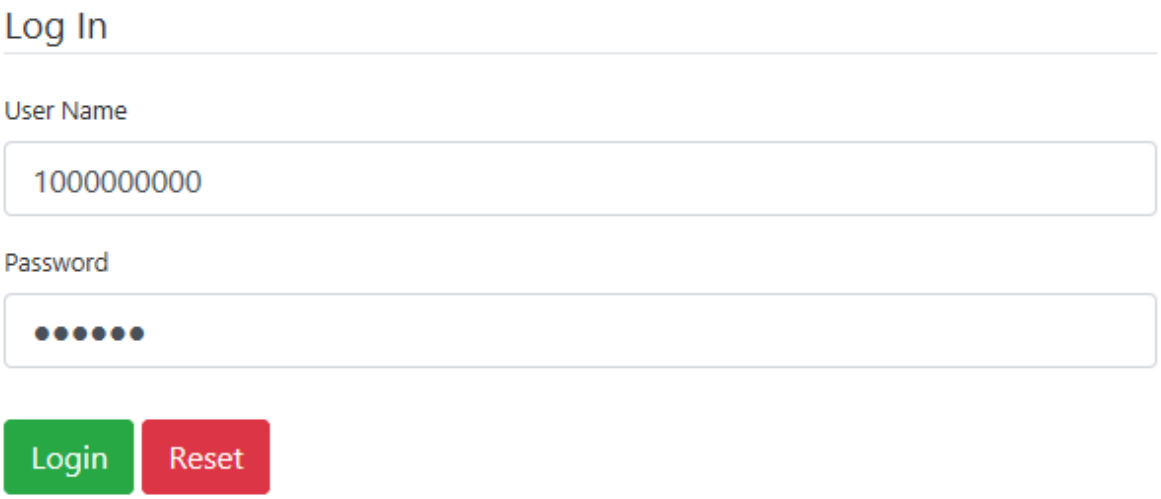


Figure 4.8: Login to omr system by 1000000000 student id

After login to system as shown on Figure 4.8, omr system displayed the following result as shown on Figure 4.9.

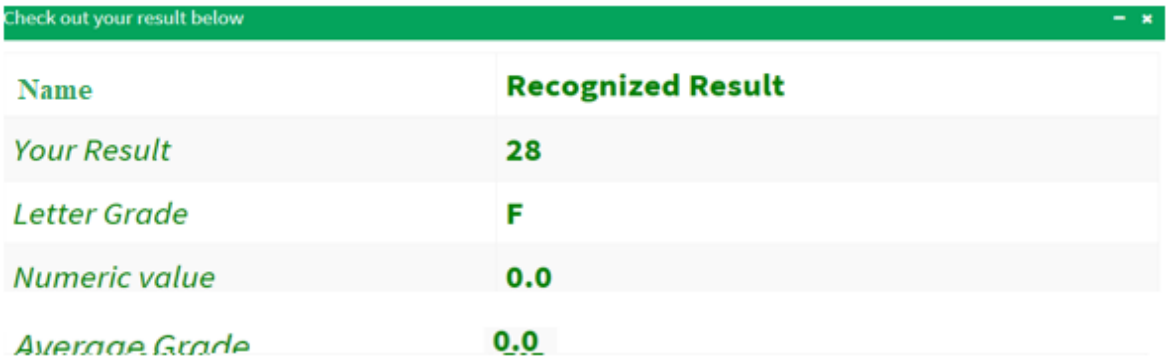


Figure 4.9: Result of student after logged by 1000000000 id

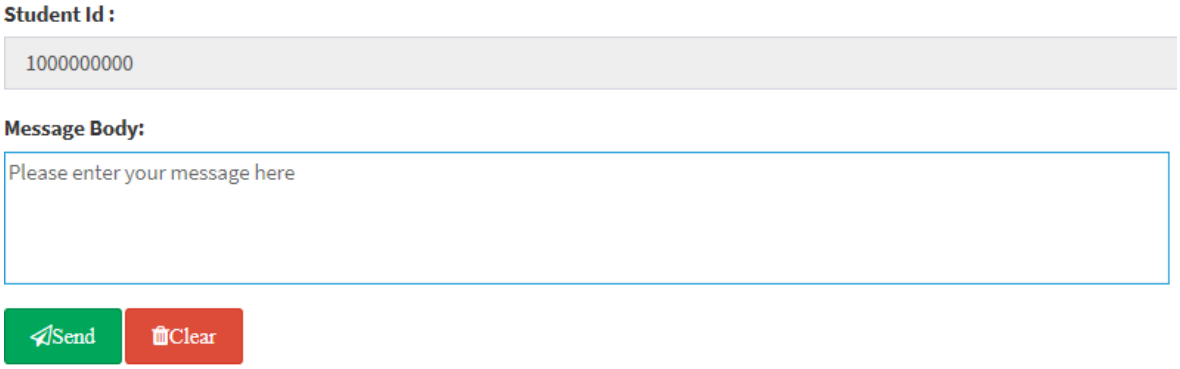


Figure 4.10: Message box

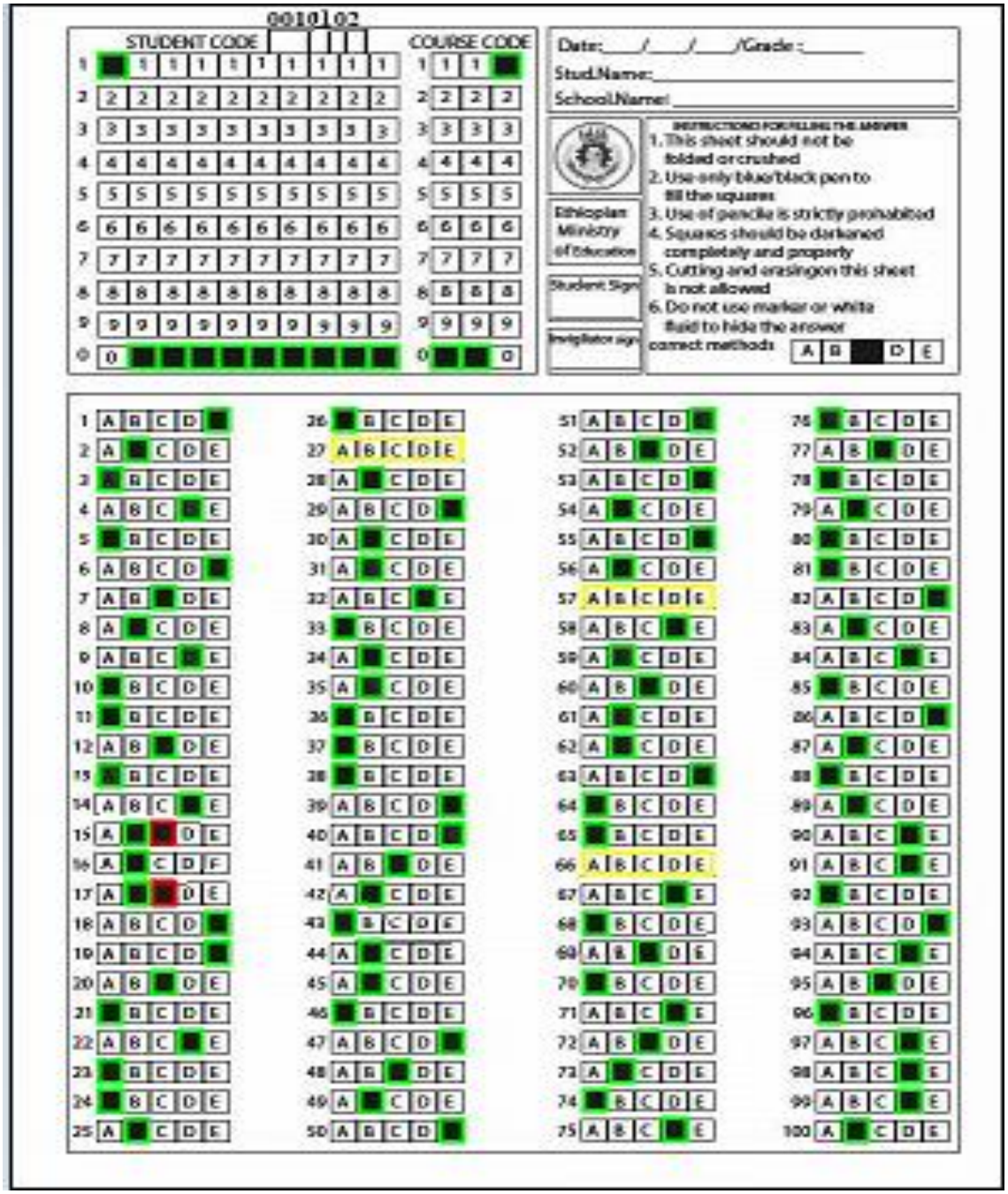


Figure 4.11: Recognized bordered sheet

As we saw on the above Figure 4.10, there is message box, on Figure 4.11 there is recognized bordered sheet seen by student and its importance is that to enable students can send their message to concerned body if he/she can found an error during recognition. But for this sheet, we don't have recognized error.

CHAPTER 5
RESULTS AND DISCUSION

As we saw in implementation detail chapter, eight type of scoring sheets were designed and tested by the proposed web and image based OMR system. 30 total number of questions were used for testing the system, from this total number of questions, we do have the following squares: -

Table 5.1: Sheets used for testing the system

S.R.NO	Total number of questions per sheet	Sheets used for testing	Total Number of squares
1	15	30	30*((15*5)+(10*10)+(3*10)) = 6150
2	20	30	30*((20*5)+(10*10)+(3*10)) = 6900
3	30	30	30*((30*5)+(10*10)+(3*10)) = 8400
4	45	30	30*((45*5)+(10*10)+(3*10)) = 10650
5	60	30	30*((60*5)+(10*10)+(3*10)) = 12900
6	80	30	30*((80*5)+(10*10)+(3*10)) = 15900
7	100	30	30*((100*5)+(10*10)+(3*10)) = 18900
8	125	30	30*((125*5)+(10*10)+(3*10)) = 22650
Total	475	240	102,450

As we saw on Table 5.1 column 3, we have used 240 total number of sheets for testing an OMR system. The system has worked with zero error rate, that means all sheets are recognized correctly and forwarded to the recognized student id to enable students see their results over their own web account. Depending on the machine(computer) resources used for, an OMR system response time varies. The system is implemented and tested based on the following laptop resources: -

- Brand Name: NEC (Nippon Electric Company)
- Processor: Intel® core i5-3210M CPU @2.50GHz
- System type: 64 based processor

- RAM: 8.00 GB
- HDD: 500GB and windows 10 home edition

Python has time module to handle time related tasks, so the function used for measuring time in second it took to response an OMR system is that time.time(). Before connected to database and switched to web based system without border around answer area, the system took the following response time in second on Table 5.2.

Table 5.2: SRTs before switched to web and connected to database without border.

Total Number of questions per sheet	Average Time they took per sheet in second
15	0.1
20	0.14
30	0.2
45	0.30
60	0.4
80	0.5
100	0.6
125	0.8

Before connected to database and switched to web based system with border around squares, the system took the following response time in second on Table 5.3.

Table 5.3: SRTs before switched to web and connected to database with border.

Total Number of questions per sheet	Average Time they took per sheet in second
15	0.7
20	1.05
30	1.4
45	2.0
60	2.4
80	3.1
100	3.9
125	5.1

After connected to database and switched to web based system without border around squares, the system took the following response time in second on Table 5.4.

Table 5.4: SRTs after switched to web and connected to database without border.

Total Number of questions per sheet	Average Time they took per sheet in second
15	1.6
20	2.0
30	2.5
45	3.4
60	3.6
80	4.5
100	4.8
125	6.0

After connected to database and switched to web based system with border around squares as we see in Figure 5.1, the system took the following responses time in second on Table 5.5.

Table 5.5: SRTs after switched to web and connected to database with border.

Total Number of questions per sheet	Average Time they took per sheet in second
15	1.65
20	2.15
30	2.5
45	3.6
60	4.4
80	5.0
100	6.0
125	7.0

As we saw on the above tables, system response time, OMR system functionality and total number of questions have the following mathematical relationship expressed on Equation 5.1 and 5.2 as well.

$$\text{SRTs} = \frac{1}{SF} \tag{5.1}$$

$$\text{SRTs} = \frac{1}{TNQs} \tag{5.2}$$

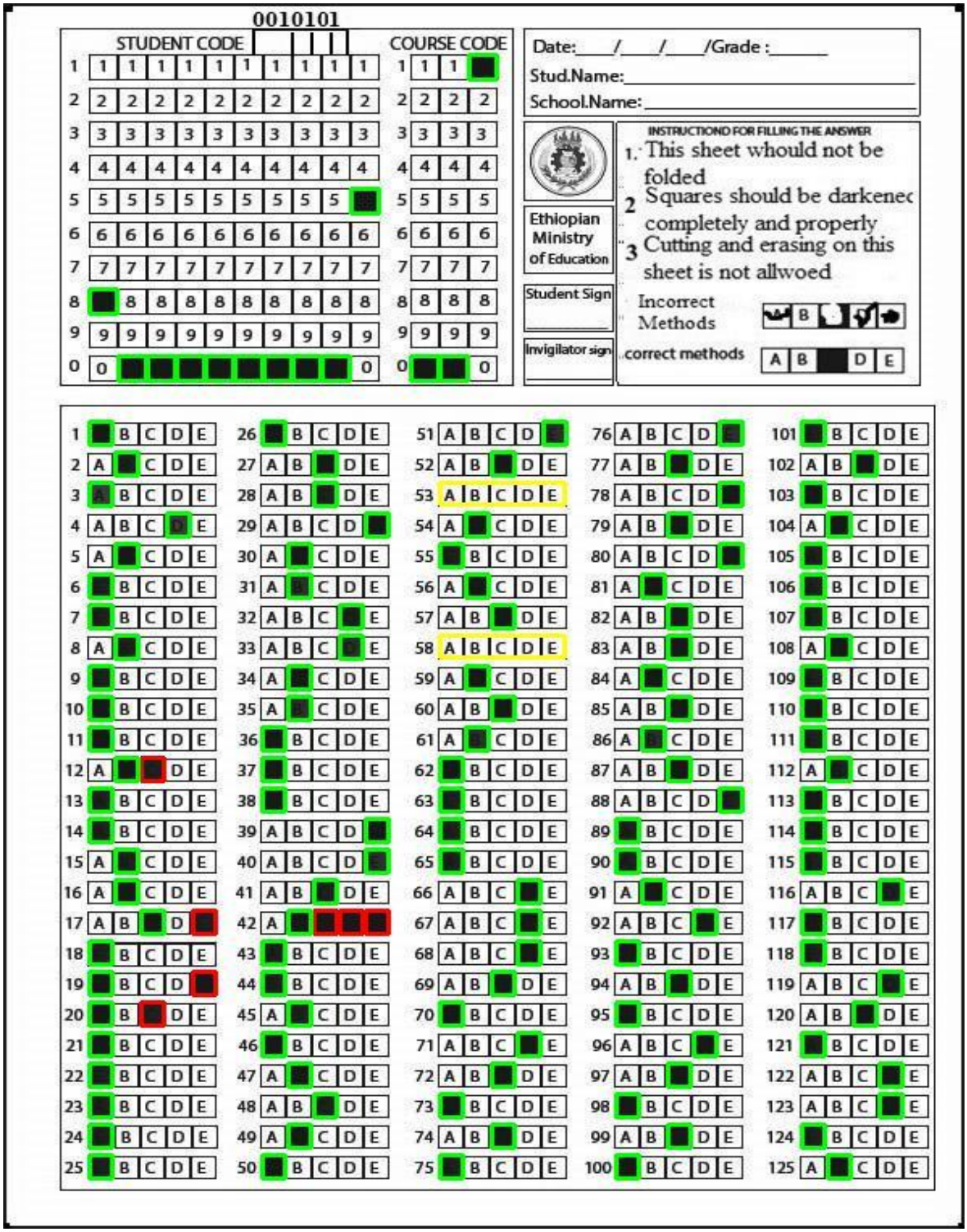


Figure 5.1: 125 Questions OMR result

Where SRTs, SF, TNQs is System Response time, System Functionality and Total Number of Questions respectively. When we upload 15 total number of questions, SRTs increases while 125 is decreases than that of 15. When we come to system functionality(SF), system functionality increases, its response time will decrease as we saw on Equation 5.2. on Table 5.2 an OMR system does not connect to database and border around answer areas, in this case we do have less system functionality, because we did only black mark recognition, no comparison to answer key in database table and border around the recognized answer area. Whereas on Table 5.5, we got better functionality than that of previous one, some of them are; we know which choice is marked correctly, which questions not marked totally, on which question have more than one choice gets marked, compute total student result by comparing to answer key that are stored in database table previously, send the sheet recognition result to student automatically by using the recognized student id in the sheet. Generally, SRTs is inversely proportional to both SF and TNQs as we saw on Equation 5.1 and 5.2.

In the Figure 5.1 shows that the sheet contains 125 questions. From these questions we have seen three border colors that shows correctly, incorrectly and totally unmarked alternatives(choices). At the very beginning in Figure 5.1, there is a bar code located on the top sheet i.e. 0010101. This code is really important for sheet categorization in which sheet is currently recognized by OMR system; to select the targeted answer key on database table and finally make comparison with detected student answers. After that there is a students and courses id area having 100 and 30 squares respectively. For this sheet 8000000005 and 001 student and course id respectively is detected as well. After we got the correct detected sheet, barcode, student code and course code an OMR system starts to detect the main aim of the work i.e. multiple question number with answer detection. From 125 questions, 123 are filled by student, only question number 53 and 58 were left by mistake with 8000000005 student ids. For those question number 53 and 58, an OMR system counts zero for student's result. On question number 12, 17, 19, 20 and 42 a student with 8000000005 id fills more than one choice. For those questions an OMR system borders by red color neither added nor subtracted from total student's result rather they leave as incorrectly filled by mistake instead. For the remaining 123 questions including 12, 17, 19, 20 and 42 that are bordered

by green color around them considers as student's correct answers. Date, student name, student signature and invigilators signature is not needed by this work. So, the detected student answer is shown in Table 5.6 with 20 rows and 8 columns.

Table 5.6: Detected student answers from Figure 5.1

Number	Answer	Answer	Answer	Answer	Answer	Answer	Answer
1	A	A	C	B	B	A	A
2	B	A	B	A	C	C	D
3	A	A	A	A	C	A	D
4	D	A	A	A	B	B	A
5	B	A	B	A	C	A	B
6	A	A	A	D	B	A	
7	A	C	B	D	C	A	
8	B	C	C	D	E	B	
9	A	E	B	C	A	A	
10	A	B	A	A	A	A	
11	A	B	E	D	B	A	
12	B	D	C	C	D	B	
13	A	D	0	A	A	A	
14	A	B	B	C	C	A	
15	B	B	A	A	A	A	
16	B	A	B	E	D	D	
17	C	A	C	C	C	A	
18	A	A	0	E	A	A	
19	A	E	B	C	C	D	
20	A	E	C	E	A	C	

Table 5.7: Retrieved answer keys from database table

Number	Key	Key	Key	Key	Key	Key	Key
1	E	C	B	E	B	B	C
2	C	D	C	C	C	C	C
3	C	B	A	A	E	C	B
4	A	C	C	C	C	A	C
5	D	A	C	B	E	B	C
6	B	C	B	B	C	C	
7	C	A	C	C	C	C	
8	A	B	C	C	E	E	
9	A	C	A	A	E	A	
10	B	C	C	C	A	C	
11	C	E	B	C	C	C	
12	C	D	C	A	A	C	
13	D	C	E	C	C	D	
14	E	C	C	B	B	C	
15	A	A	A	C	B	C	
16	A	A	C	C	C	A	
17	C	C	B	B	C	C	
18	C	C	C	C	A	B	
19	D	C	E	B	C	C	
20	C	D	C	C	C	A	

After detecting students answer and retrieving appropriate answer key, an OMR system compares results on Table 5.6 and Table 5.7 and its computed results are written on Table 5.8.

Table 5.8: Computed and detected results

Name	Result
Your Result	21
Letter Grade	F
Numeric Value	0.0
Average Grade	0.0
Detected barcode	0010101
Incorrectly detected questions	2
Correctly detected questions	123
Detected student id	8000000005
Detected course code	001

Table 5.9: Letter and numeric grade rules

Rules	Letter Grade	Numeric Values
If 90 <= result <= 100	Set “AA”	4.0
If 85 <= result < 90	Set “A”	3.8
If 80 <= result < 85	Set “BA”	3.5
If 70 <= result < 80	Set “BB”	3.0
If 60 <= result < 70	Set “BC”	2.5
If 50 <= result < 60	Set “C”	2.0
If 40 <= result < 50	Set “D”	1.0
If 0<= result < 40	Set “F”	0.0

As we saw on Table 5.9, letter grade and its numeric value have done based on 100 total number of questions but, there are 7 remaining question types. So we have to convert them

to 100. Multiply 125, 80, 60, 45, 30, 20 and 15 by 0.8, 1.25, 1.667, 2.23, 3.334, 5, 6.667 respectively.

CHAPTER 6
CONCLUSION AND RECOMMENDATION

6.1 Conclusion

We have proposed an easy and profitable web and scanned image based optical mark reader system with fair speed and high precision. An implementation result shows that the speed of proposed system depends on, system functionality and total number of questions are in scoring sheets. The proposed system uses ordinary scanners and printers that are available in offices and used by them to run office tasks in order to prepare scoring sheets. Because of availability and cheapness, A4 paper with 70 GSM is used for preparing scoring sheets. This system has two users; the one who feed the scoring sheet and can see the recognition result and the other only see the result of its recognition after compared to answer key. The system offers subsequent functionalities; convert black marks to computer understandable format, indicate which choices are marked by drawing border around the answer area, compare the recognized black marks with correct answer key that are stored in database table and calculate the result, compute the result into letter grade and letter grade equivalent, enables students can see their own result on web account, finally, students can send feedback to concerned body if the answer recognized by proposed system and students marked will not same. in another words, this system takes scoring sheets and finally produce its result with letter grade and numerical letter equivalency. As long as the proposed system is web based, any web enabled devices such as mobile phones, tablets, laptops, desktops computer can have used for accessing it.

6.2 Recommendation

For schools that have affordability concern, I recommend to use the services that can offered by this proposed system, but for schools that haven't affordability issue, I recommend to buy hardware based scanners that can scan up to 120 scoring sheets per minute. Even though C++ needs developer's effort, most developers said that C++ is 400 times faster than python. As part of future work, I will re- implement this system by C++ programming language and

recommend the coming researchers to use C++ for better performance. Detection of student and invigilator signatures are my additional future work.

REFERENCES

Astha Gupta,Sandhya Avasthi. (2016) Image based low cost method to the OMR process for surveys and research. *International Journal of Scientific Engineering and Applied Science (IJSEAS) – Volume-2, Issue-7, July 2016 ISSN: 2395-3470 www.ijseas.com*

Chris, W. (2018, December 30*).* How the World Wide Web (WWW) works? [Blog post]. Retrieved from https://www.explainthatstuff.com/howthewebworks

Douglas Chai (2016). Automated Marking of Printed Multiple Choice Answer Sheets. *2016 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE).*

Elmasri, R. (2011). Fundamentals of Database Systems, Arlington University of Texas: Addison-Wesley.

Garima Krishna., Hemant Ram Rana, Ishu Madan, Kashif., & Narendra Sahu. (2013). Implementation of OMR Technology with the Help of Ordinary Scanner. *International Journal of Advanced Research in Computer Science and Software Engineering*

Hui Deng, Feng Wang, Bo Lian. (2008). A Low-Cost OMR Solution for Educational Applications. *International Symposium on Parallel and Distributed Processing with Applications*

JingTao Yao. (2008). An Introduction to Web-based Support Systems. *J. T. Yao, An Introduction to Web-based Support Systems, Journal of Intelligent Systems, Vol. 17, No. 1-3, pp267-281, 2008.*

K. Dimililer & A. ilhanb. (2016). Effect of image enhancement on MRI brain images with neural networks. *12th International Conference on Application of Fuzzy Systems and Soft Computing, ICAFS 2016, 29-30 August 2016, Vienna, Austria.*

Ms. Sumitra B. Gaikwad. (2015). Image Processing Based OMR Sheet Scanning. *International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE) Volume 4, Issue 3, March 2015.*

Nalan Karunanayake. (2015). OMR Sheet Evaluation by Web Camera Using Template Matching Approach. *International Journal for Research in Emerging Science and Technology*, vol. 2s, 8.

Omrhome. (2016). 5 Common Mistakes While Filling an OMR Sheet. Retrieved February 14, 2019, from https://create.piktochart.com/output/18411492-5-common-mistakes-while- filling-an-omr-sheet.

Patel, R., Sanghavi, S., Gupta, D., & Raval, M. S. (2015). CheckIt - A low cost mobile OMR system. *TENCON 2015 - 2015 IEEE Region 10 Conference*. *doi:10.1109/tencon.2015.7372983*

Pratt, Philip J.; Last, Mary Z. (2014-09-08). Concept of Database Management (8 ed.). Course Technology.

Rakesh S., Kailash Atal, & Ashish Arora. (2013). Cost Effective Optical Mark Reader. *International Journal of Computer Science and Artificial Intelligence*. Jun. 2013, Vol. 3 Iss. 2, PP. 44-49.

BIBLIOGRAPHY

C.R. Kothari (1985). Research Methodology: Methods and Techniques (Second revised Edition). *New Age International (P) Ltd., Publishers.*

David Postill. (2016). What is java virtual machine and what does it have to do with Java? Retrieved February 16,2019, from https://superuser.com/questions/1048425/what-is-java-virtual-machine-and-what-does-it-have-to-do-with-java

Dr.Anil Kumar Maini (2018). Image Processing using MATLAB. Retrieved February 27,2019, from https://electronicsforu.com/electronics-projects/image-processing-using-matlab-part-1

Mohan. (2012). Image Processing- Types of images. Retrieved March 01,2019, from http://dspforyou.blogspot.com/2012/08/image-processing-types-of-images.html

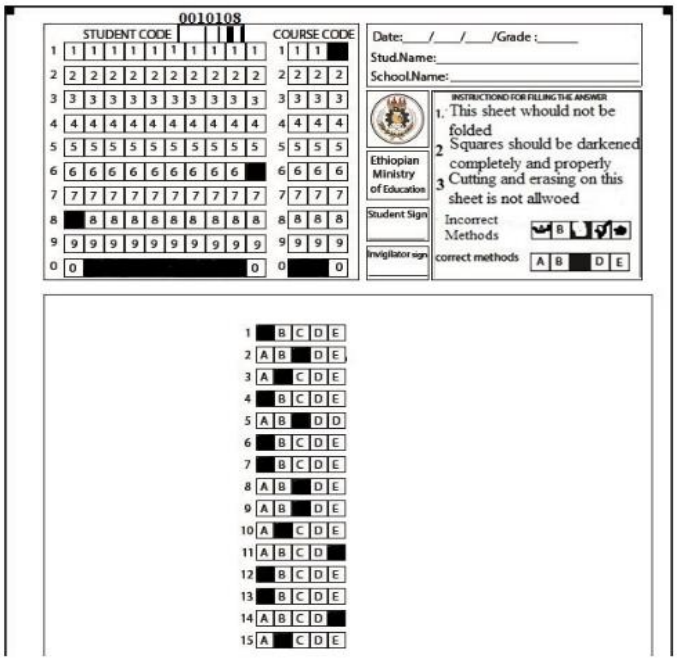
Rai.Skumar. (2013). Java is not very strictly platform independent. Retrieved February 15, 2019, from https://stackoverflow.com/questions/12337427/is-java-really-platform-independent.

Surbhi Gupta, Geetila Singla, and Parvinder Singh Sandhu. (2012). A Generalized Approach to Optical Mark Recognition. *International Conference on Computer and Communication Technologies (ICCCT'2012) May 26-27, 2012 Phuket.*

APPNDIX 1

1. More implementation results

Sheet area



Recognized data area

Number	Recognized Answer	Answer Key Retrieved from database Number	Answer key
1	A	1	A
2	C	2	C
3	B	3	B
4	A	4	A
5	C	5	E
6	A	6	A
7	A	7	A
8	C	8	B
9	C	9	B
10	B	10	B
11	E	11	E
12	A	12	A
13	A	13	A
14	E	14	E
15	B	15	D

Figure 1. 1: Snapshot of 15 questions result with answer key

Total Number of Incorrectly Detected Question(s): 0,

Total Number of Correctly Detected Question(s): 15,

Recognized Student Id : 8000000006,

Recognized Course Code : 001,

Recognized BarCode : 0010108,

Total Student result : 11,

Letter grade : ['BB']

Numerical Value : ['3.0']

Figure 1. 2: Snapshot of computed results from Figure 1.1

In the above snapped Figure 1.2 shows that the result of OMR when we upload scoring sheet having 15 total number of questions in it. We got theses functionalities when we login as a teacher. Figure 1.2 shows the recognition and computation result of Figure 1.2. When we login to the system with recognized student id that are located in Figure 1.2 with registered password by mothers first name we will get the following view.

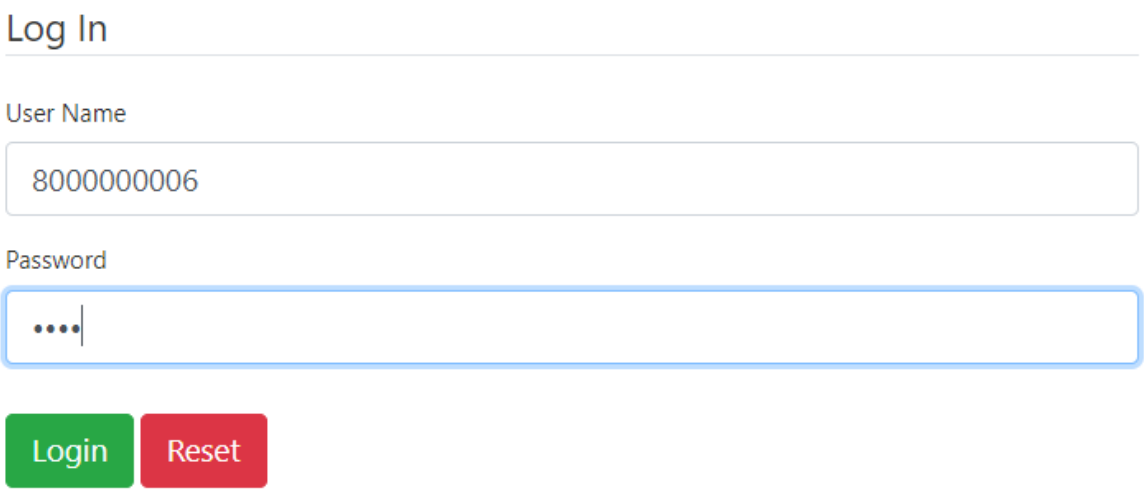


Figure 1. 3: login as 8000000006 student user name

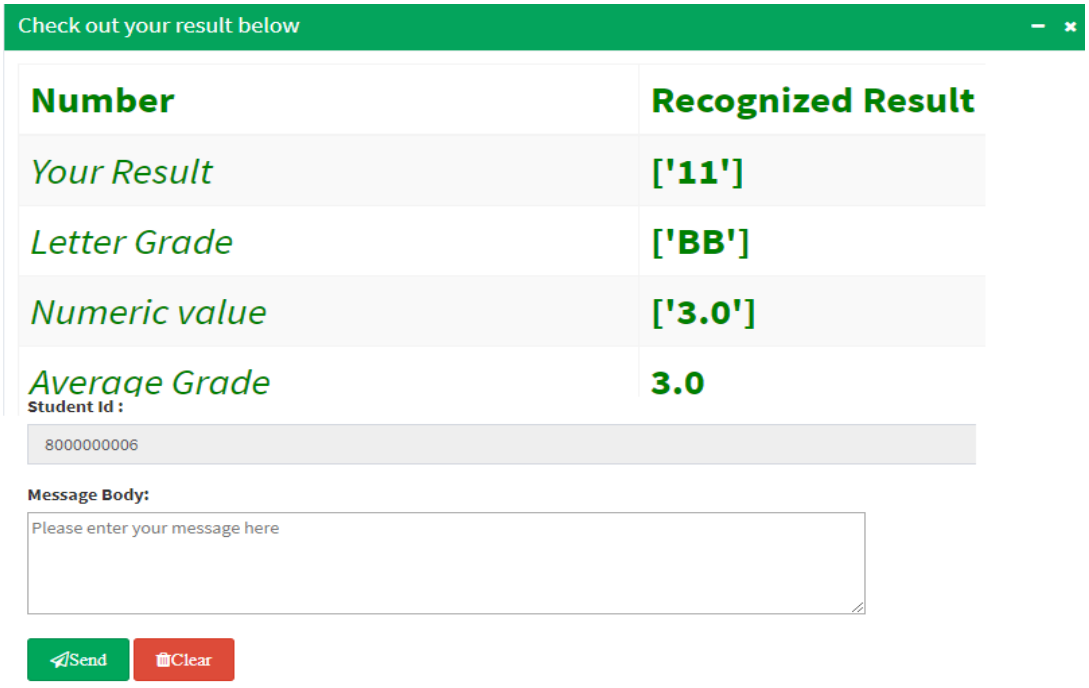


Figure 1. 4: Computed results after Figure 1.3

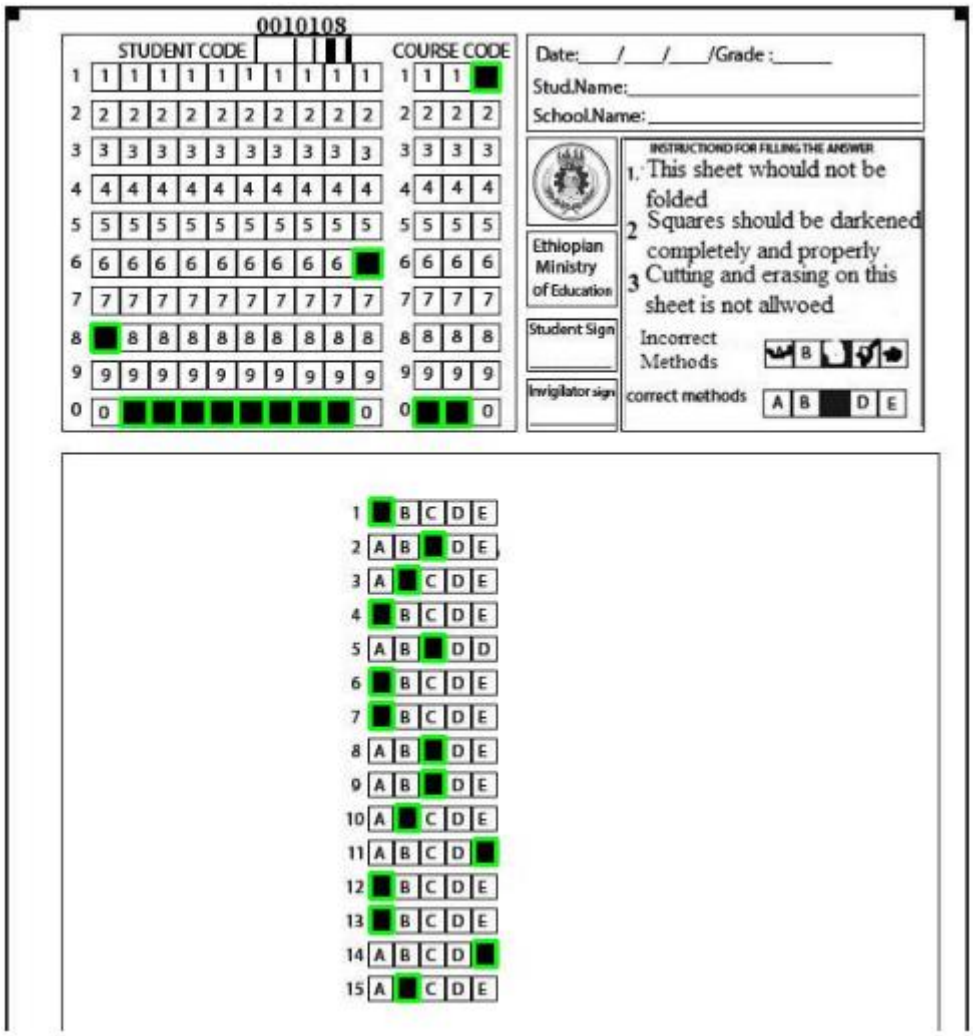


Figure 1. 5: OMR result of 15 questions with border

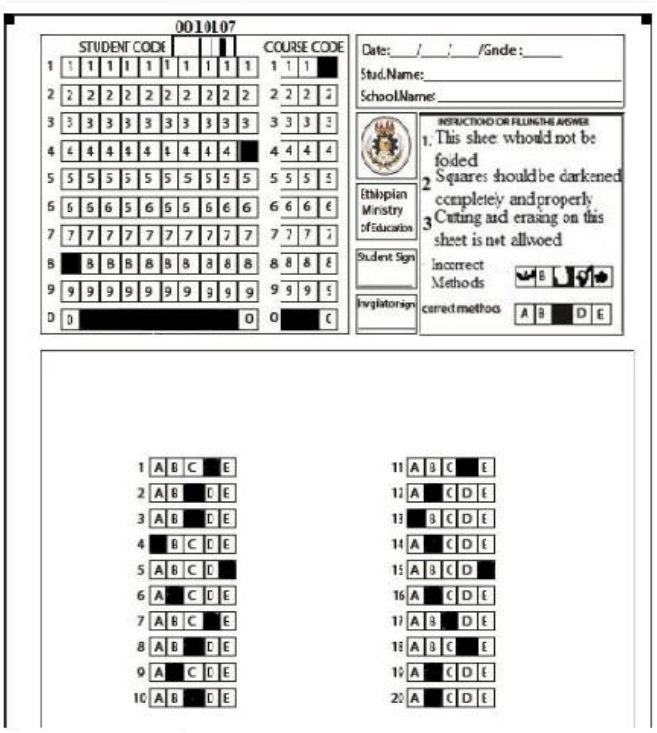
Recognized data area

Number	Recognized Answer
1	D
2	C
3	C
4	A
5	E
6	B
7	D
8	C
9	B
10	C
11	D
12	B
13	A
14	B
15	E
16	B
17	C
18	D
19	B
20	B

Answer Key Retrieved from database

Number	Answer key
1	B
2	B
3	B
4	B
5	B
6	B
7	B
8	B
9	B
10	B
11	B
12	B
13	B
14	B
15	B
16	B
17	B
18	B
19	B
20	B

Sheet area



Total Number of Incorrectly Detected Question(s): 0,

Total Number of Correctly Detected Question(s): 20,

Recognized Student Id : 8000000004,

Recognized Course Code : 001,

Recognized BarCode : 0010107,

Total Student result : 7,

Letter grade : ['F']

Numerical Value : ['0.0']

Figure 1. 6: Snapshot of 20 questions result with answer key

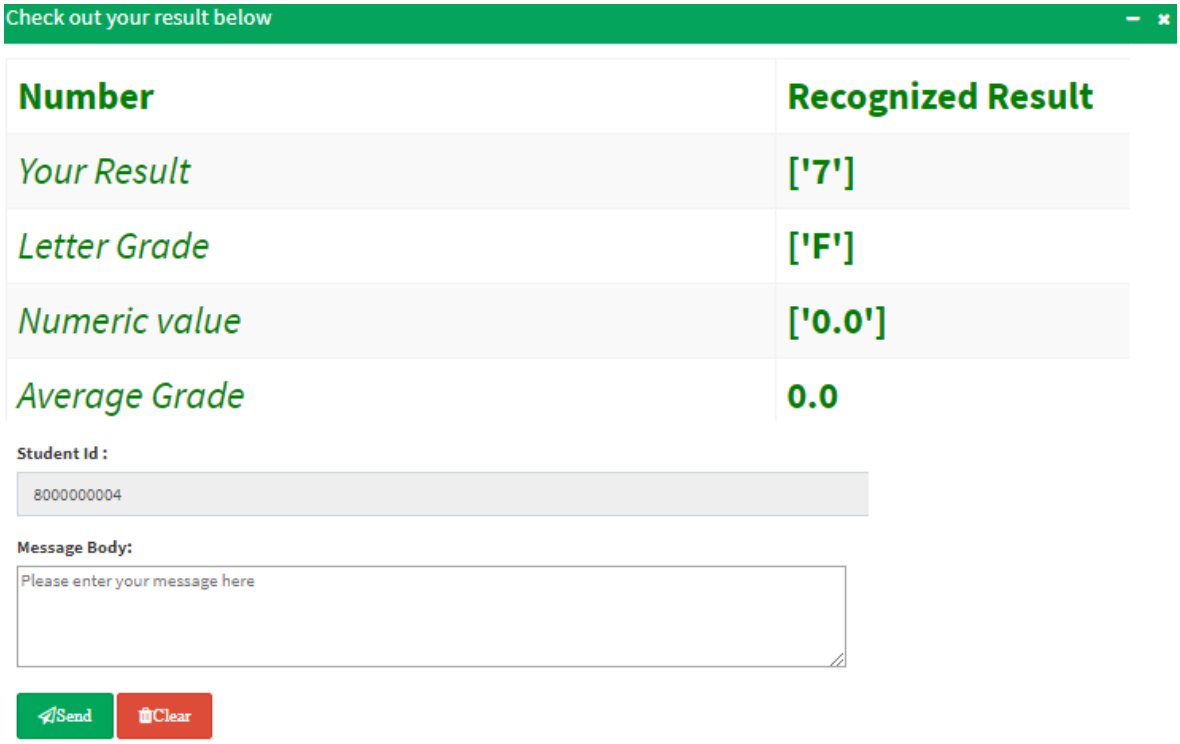


Figure 1. 7: Snapshot of students result after logged by 80000000004 id

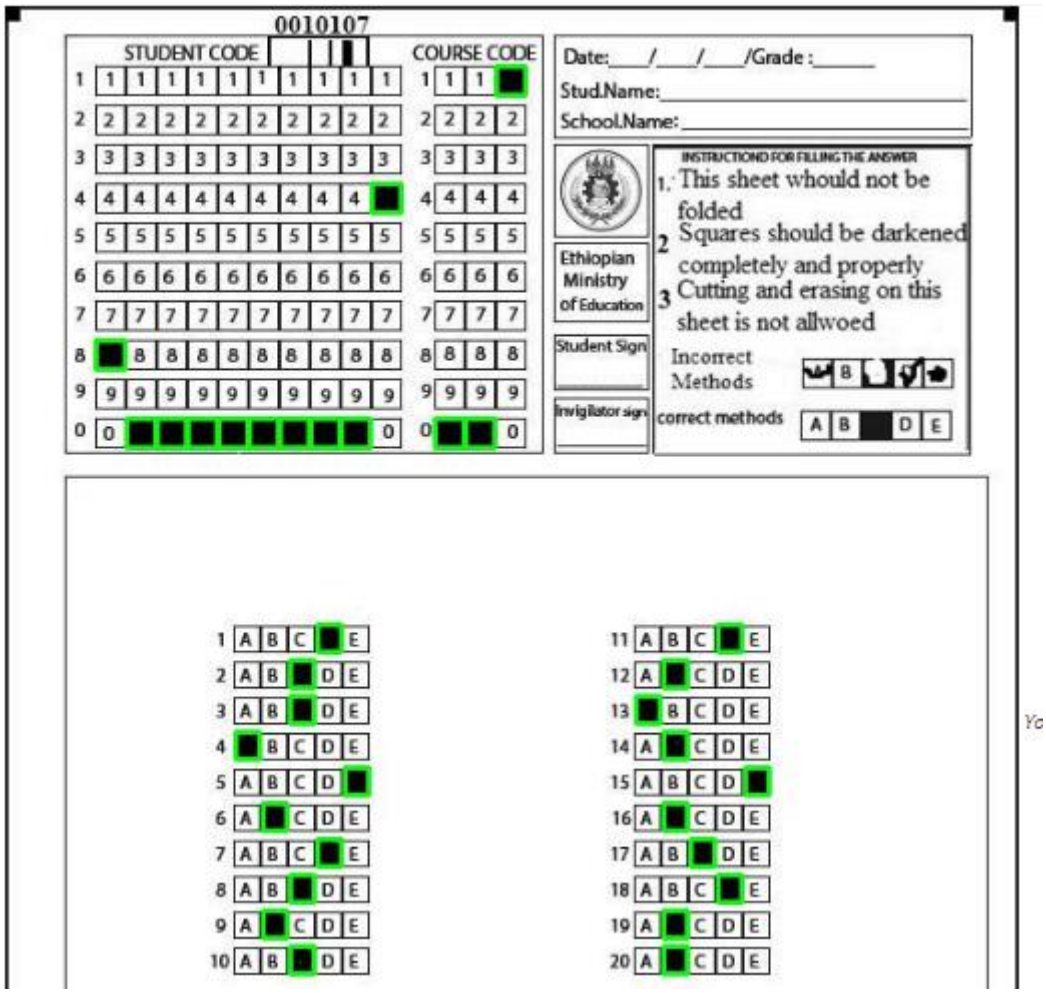
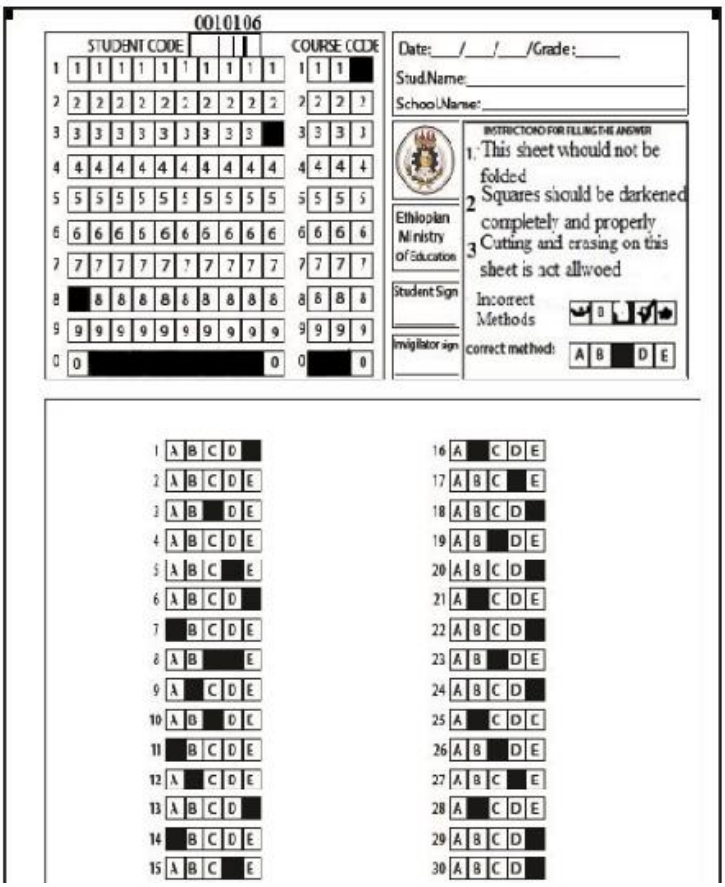


Figure 1. 8: OMR result of 20 questions with border

Number	Recognized Answer	Number	Recognized Answer
1	E	18	D
2	0	19	E
3	C	20	C
4	0	21	E
5	D	22	D
6	E	23	E
7	A	24	C
8	C	25	E
9	B	26	B
10	C	27	C
11	A	28	D
12	B	29	B
13	E	30	E
14	A		
15	D		
16	B		
17	D		



Total Number of Incorrectly Detected Question(s): 2,

Total Number of Correctly Detected Question(s): 28,

Recognized Student Id : 8000000003,

Recognized Course Code : 001,

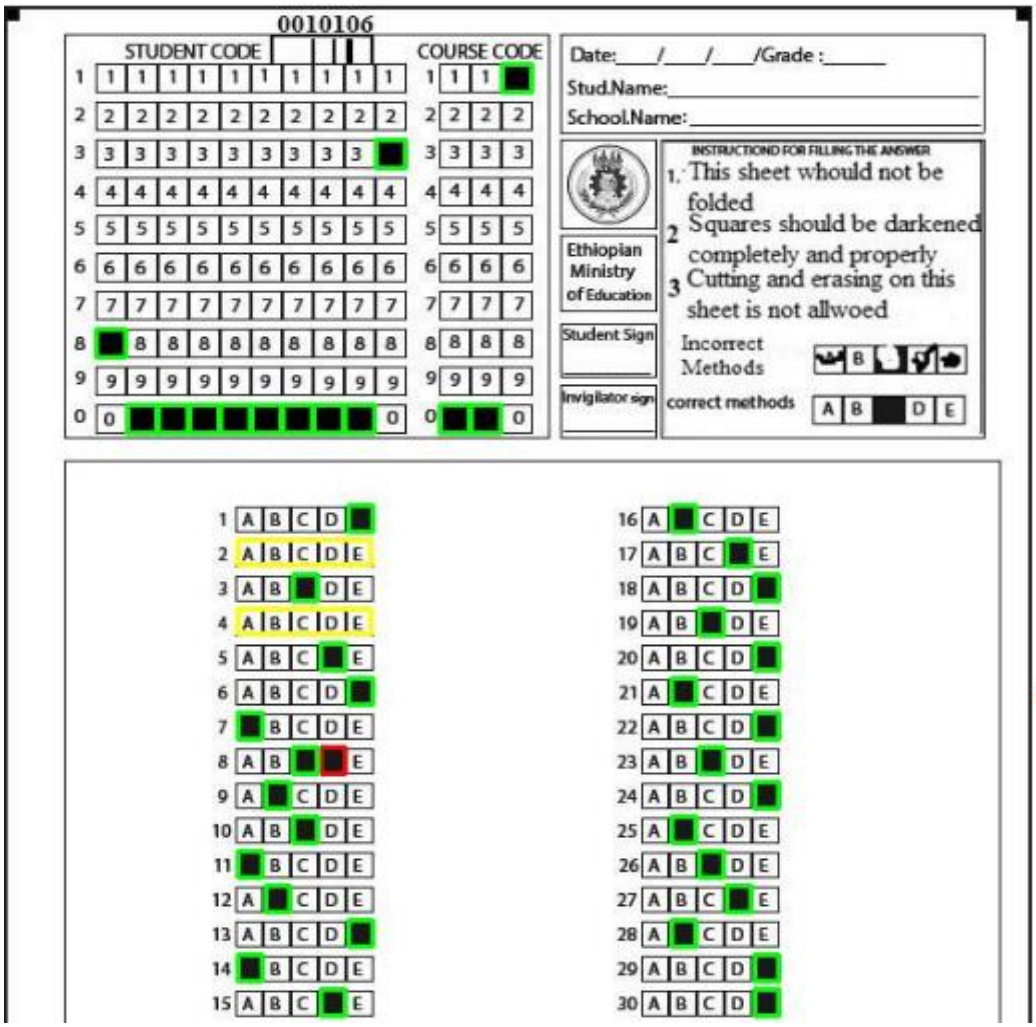
Recognized BarCode : 0010106,

Total Student result : 6,

Letter grade : ['F']

Numerical Value : ['0.0']

Figure 1. 9: Snapshot of 30 questions result with answer key



Check out your result below

Number	Recognized Result
Your Result	['6']
Letter Grade	['F']
Numeric value	['0.0']
Average Grade	0.0

Figure 1. 10: Snapshot of students result after logged by 80000000003 id