Code Extraction from Image with Language Detector and Compiler using Optical Character Recognizer

Ms.Roqia Tabassum, Asst. Professor*, Rienie Niharika Ganji**, Supraja Choukula***, Mulli Deepthi Reddy****

*(CSE Department, Sphoorthy Engineering College, Nadergul
Email: roqia041@gmail.com
** (CSE Department, Sphoorthy Engineering College, Nadergul
Email: ganjirienieniharika007@gmail.com)
*** (CSE Department, Sphoorthy Engineering College, Nadergul
Email: choukulasupraja2001@gmail.com)
**** (CSE Department, Sphoorthy Engineering College, Nadergul
Email: mulli.deepthireddy@gmail.com)

1. ABSTRACT-

With the aid of online compilers, you can compile source code for use in a variety of computer languages and run it online. It is an online compiler and debugging utility that enables online code compilation and execution in a variety of computer languages. To simultaneously build programs in various languages, a number of compilers must be installed on the same computer. The issues here are that the user must compose all the code, that it is unable to determine the computer language used, and that it is unable to extract the code from the scanned pictures. A method to conduct numerous pages of text recognition and make use of the cloud server's processing power is the answer to the as for mentioned issue working hard. Based on processing time and character recognition precision, the findings will be contrasted with OCR output from a few commercial OCR apps. A person can use an "Optical Character Recognizer (OCR)" in this program to extract code from a picture. The name of the computer language is displayed after it instantly recognizes it.

Keywords: *Optical Character Recognizer (OCR)*

2. INTRODUCTION

There is a great deal of interest these days in "saving the material present in these paper papers in to a computer storage disc and then later editing or reusing this information by searching process".[1] The goal of this application is to enable users to derive code from images using "Optical Character Recognizers (OCR)" that can identify computer languages automatically and present the user with a few choices.[2] You have the choice between running the code or compiling it and by removing the code from the picture, this program reduces the amount of time spent on coding.[3] Users can access the application in guest mode, but they must register in to save their earlier scans and work.[4]

A. Research Problem

Although the user cannot view or change these images, it is quite challenging to read the individual contents and search the contents of these papers line-by-line and word-by-word to reuse this information. There is a keen interest these days in "storing the information present in this paper. documents into a computer storage device and then subsequently modifying or reusing.

The issue here is for software systems to recognize characters in computer systems when information is scanned through paper documents. When we scan papers with a scanner, the documents are saved in the computer as images in formats like JPEG and others.

B. Research Justification

The goal of this application is to allow users to extract code from images using "Optical Character Recognizers (OCR)" that can identify computer languages and present the user with a few choices.

3. LITERATURE SURVEY

Most often associated with academic-oriented literature, such reviews are found in academic journals and are not to be confused with book reviews. C. A narrow-scope literature review may be included as part of a peer-reviewed journal article presenting new research, serving to situate the current study within the body of the relevant literature and to provide context for the reader. In such a case, the review usually precedes the methodology and results sections of the work. Year: 2013 Description: There are several optical character recognition (OCR) mobile applications on the market running on mobile devices, both Android and iOS (iPhone,

iPad, iPod) platforms. This paper proposes a framework for optical character using standalone and server-based OCR on mobile devices and comparing the results of the accuracy and time required for the entire OCR processing. Server-based mobile OCR obtains 5% higher character recognition accuracy than standalone OCR, and its format recognition accuracy is 99.8%. Year: 2016 Description: This paper presents an algorithm for character extraction of paper-printed characters using MATLAB computation software with an image processing toolbox.

4. EXISTING SYSTEM

Online compilers are tools which allows you to compile source code and execute it online in many programming languages. It's an online compiler and debugging tool which allows to compile and run code online in more than 40 programming languages. There is a need to have many compilers in the same machine to compile programs in different languages at the same time. The controller will then decide which compiler server the program should be assigned to compile, depending on the load of backend compilers. The compiler server will compile and run the program. The output is then given back to the user.

As we all know, there are numerous printed newspapers and books on a variety of themes. The issue here is for software systems to recognise characters in computer systems when information is scanned through paper documents. When we scan papers with a scanner, the documents are saved in the computer as images in formats like JPEG and others. Although the user cannot view or change these images, it is quite challenging to read the individual contents and search the contents of these papers line-by-line and word-by-word in order to reuse this information. There is a keen interest these days in "storing the information present in these paper documents in to a computer storage device and then subsequently modifying or reusing.

Disadvantages

- User should write the entire code.
- It cannot detect the programming language in which the code is written.
- It cannot extract the code from the scanned images.

5. PROPOSED SYSTEM

A web application where a user can extract code from image using a "Optical Character Recognizer (OCR)", which automatically detects the programming language. This system will use web-based application. It can be viewed using web browsers. This application saves the typing time by extracting code from the image This application also uses code language detection technology which enables a unique usage of OCR. Code extracted from the Tesseract OCR can be executed using Compile Run. The format and character accuracy is very useful for example in the book reprints. In terms of cost, using OCR with good accuracy would also be more efficient. For example, the reprint of 100 pages of printed book, the typing cost is around USD 20.00 and can be done in about 4 days. Using this method, it will only take 90 minutes for recognition (OCR) on mobile devices using server-based processing.

Comparison methods are proposed in this paper by conducting a series of tests.

ADVANTAGES OF PROPOSED SYSTEM:

1.The benefit of proposed system that overcomes the drawback of the existing system is that it supports multiple functionalities such as editing and searching. It also adds benefit by providing heterogeneous characters recognition.

2.Presently this application accepts a single image of code to convert and compile/run, in future it can be made to accept multiple images of code and combine them all together and be able to compile/run at a time. At present this application allows a single user to work on a code, in future we can implement collaboration on the same project so that multiple users can work on a single piece of code and fix their problem.

3.Users cannot share the code with other people in the platform in the present application, but in future we can implement a feature to share their work the platform with other users. We can also implement forums in this application which allows users to share his/her problem with the code publically and get solutions from the developers all over the platform.

4.A mobile application where a user can extract code from image using a "Optical Character Recognizer (OCR)", which automatically detects the programming language and provides a couple of options to the user. One of the option is to Compile the code and the other is to Run the code. User can use the application in guest mode, but to save their previous scans and work one should login. Optical character recognition (OCR) is the electronic or mechanical change of images of typed, handwritten or printed text into

machine-encoded text, whether from a scanned document, a image of a document.
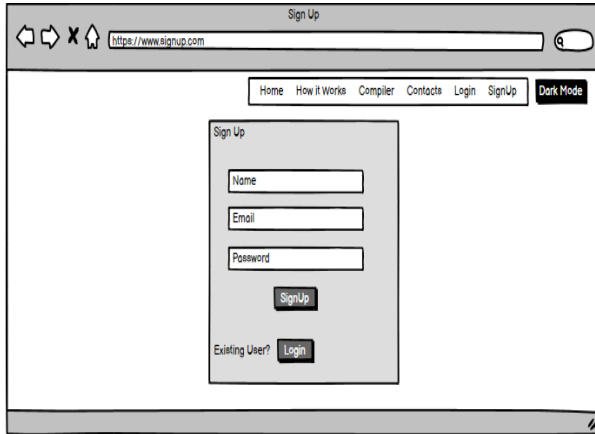
6. Results

6.1 Sign Up page.

Fig.6.1. The registration page where the user provides the

7. CONCLUSION

The most important details in this text are the use of an optical character recognition (OCR) web tool to extract code from an image. This is made possible by the use of code language detection technology, Compile and Run, and how to conduct multipage OCR on mobile devices. The topic of discussion was how to use OCR applications to output formats and characters with precision while also minimizing processing time. OCR that is used accurately would be more effective in terms of cost as well. Our technique produces better results than standalone OCR tools but requires longer upload time.

8. ACKNOWLEDGEMENT

details to get registered in the website.

6.2 Login Page
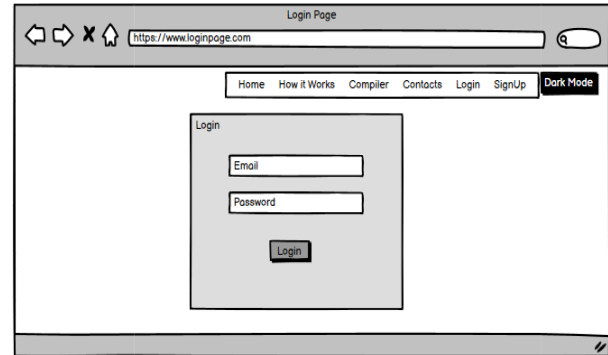
Fig.6.2.Here after registration the user directly can log into the web app by entering the credentials.

9. REFERENCES

[1] Arjun Datta ; Arnab Kumar Paul. "Online compiler as a cloud service" Year 2014

[2] Teddy Mantoro ; Abdul Muis Sobri ; Wendi Usino "Optical Character Recognition (OCR) Performance in Server-Based Mobile Environment" Year 2013.

[3] Uday chauhan ; "Online Cloud Based Compilers System" Year 2016.

[4] Sakchai Tangwannawit; Wanida Saetang; "Recognition of Lottery Digits Using OCR Technology", Signal-Image Technology & Internet-Based Systems (SITIS) Year 2016.

[5] Nishal Ancelette Pereira; Prajwal Rao; Akshay K Kallianpur; "Discrete Artificial Bee Colony Algorithm based Optical Character Recognition" Year 2017.

[6] Smart Mobile Software – Mobile OCR. Internet:http://www.smartmobilesoftware.com/mobile-

ocr.html [Accessed: April 13, 2013].

[7] Wojciech Bieniecki, Szymon Grabowski and Wojciech Rozenberg, Image Preprocessing for Improving OCR Accuracy, MEMSTECH'2007, May 23-26, 2007, Lviv-Polyana, UKRAINE.

[8] Bunke, Horst, and Patrick S.P. Wang (1997) Handbook of Character