Real Time Mobile Cloud Audio Reading System for Blind Persons

Mohammed Ali Mohammed*
Computer Department, College of Science, Mustansiriyah University, Baghdad, Iraq.
E-mail: h.mahammed.p@gmail.com

Dr. Karim Q. Hussein
Assistant Professor, Computer Department, College of Science, Mustansiriyah University, Baghdad, Iraq.
E-mail: Karim.q.h@uomustansiriyah.edu.iq

Dr. Mustafa Dhiaa Al-Hassani
Assistant Professor, Computer Department, College of Science, Mustansiriyah University, Baghdad, Iraq.
E-mail: dr_mdhiaa77@uomustansiriyah.edu.iq

Received August 07, 2021; Accepted November 24, 2021
ISSN: 1735-188X
DOI: 10.14704/WEB/V19I1/WEB19024

Abstract

According to the World Health Organization there are approximately 285 million blind people around the world. These people are faced challenges when reading a book. This paper aims to design and implementation new real time mobile cloud audio reading system for blind persons. The proposed methodology consists of following steps: In Client Side, firstly, capture image of text by camera. Secondly, check the page localization. Thirdly, send image to server. In Cloud Side (Server Side), firstly, apply the modify EAST algorithm on received image to text detection. Secondly, apply OCR algorithm to extract text from image. Thirdly, apply post-processing step to correct the in corrected text. Finally, return text to client side to speak it using text-to-speech algorithm.

Keywords

Audio Reading System, OCR, EAST, Page Localization, Blind People, Mobile Cloud.

Introduction

In recent years, the information is available on the websites, books, and etc. The blind people are faced challenges when reading this information. Especially, when information and people are used the messages (electronic and paper based) to communication with

each other rather than talking. So, can invest the technology to improving the blind people life by using the audio reading system.

Audio Reading System used to help blind people to read the text based on camera as input device and speaker as output device. The Audio Reading System Consist three general steps: input step, processing step and output step. First step, capture image for text want to read it using a camera, then sent it to the processing step. Second step, image processing where text will be filtered and will be extracted by Optical Character Recognition (OCR) algorithm, and finally, output step used the Text-to-Speech algorithm to speak the text using speaker device (Shirke & Patil, 2018) (Goel et al., 2018).

The main characteristics of our system are: It simulate the real time of the system, using cloud computing to avoid the client hardware limitation and collect the data (data centralization), page localization to help the blind person to capture the complete view of text, and modify the Efficient and Accurate Scene Text (EAST) algorithm to detect the text sequentially. Moreover, Text correction step and index table technology.

The reset of this paper is organized as follows: in section 2 discussed the related work of the audio reading system. In section 3 described our methodology system. In section 4 described the cloud computing. While in Section 5 showed our contribution. In section 6 showed the experimental results and discuss them. Finally, conclusion this paper in section 7.

Related work

In (Gurav et al., 2017) described a smart specification for the blind people. The system can perform text detection then produce a voice output. The system has a camera as an input device to feed the printed text document for digitization and the scanned document is processed by a software module the OCR. Finally, the output of the OCR step will be convert to voice output using Text-to-Speech technology. The flow chart diagram of this research shown in figure 1.

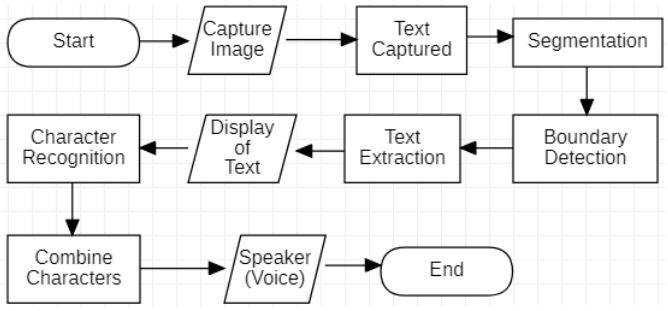

Figure 1 Flow Chart of the (Gurav et al., 2017) system

In (Barathkumar et al., 2019). This paper focus to read name on the products through image capturing. The portable system takes the image by Raspberry Pi camera. The Image is fed to input of Raspberry Pi processor to process the OCR. OCR used to convert image containing written text into machine-readable text data. Then the system use Google Text to Speech Convertor (GTTS) to convert text entered in to the audio. The flow chart diagram of this research shown in figure 2.

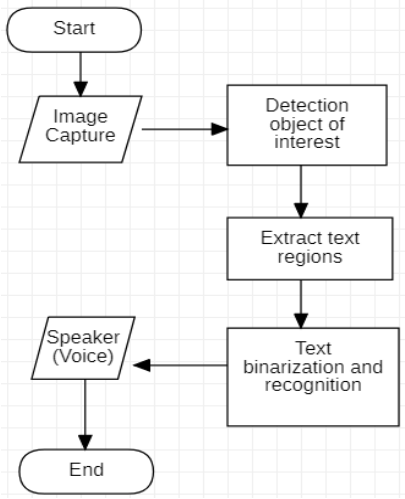

Figure 2 Flow Chart of the (Barathkumar et al., 2019) system

In (binti Harum et al., 2019) focused on development of Smart Book Reader will help the blind people to read the text and using Internet of Things (IoT) technology. The system is design by utilises IoT technology with the use of an IoT device, infrastructure and service. One of the advantages of this system is user can read both softcopy and hardcopy books which base on the online text to voice tools due it is using the IoT technology. The flow chart diagram of this research shown in figure 3.

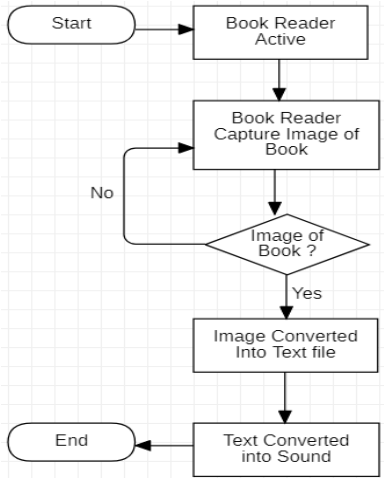

Figure 3: Flow Chart of the (Binti Harum et al., 2019) system

In (Sahu et al., 2020) design and implementation assistive text reading system for visually impaired people to read the text labels from the Prescription as android application. Through camera the application capture the image of Prescription. Using a motion-based method to define Region of Interest (ROI) image to isolate the Prescription Image from background. Then using OCR algorithm to recognize text character. Finally, using text to speech converter to speak the text. The flow chart diagram of this research shown in figure 4.

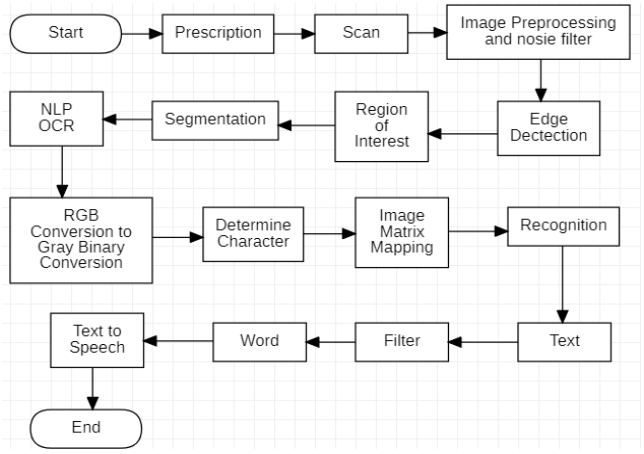

Figure 4 Flow Chart of the (Sahu et al., 2020) system

In (Ravi et al., 2020) presented A smart reader system for Blind people with integration of a complete text read out system with page turning mechanism and dictionary query feature. This project design and implementation a smart reader system when the image book is input to system, the book text is read as sound output using python program and camera device. The flow chart diagram of this research shown in figure 5.

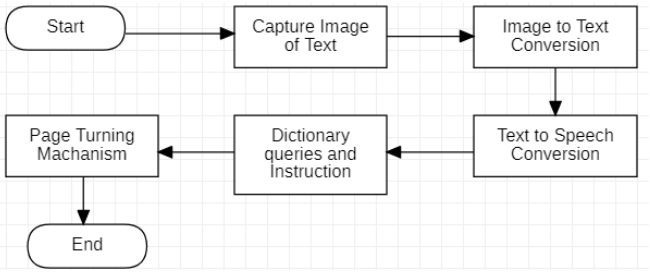

Figure 5 Flow Chart of the (Ravi et al., 2020) system

In (Bhagat-Thakre, 2020) In this research, the book image is converted into the audio output through the use of Text-to-speech and OCR technology which help the blind person to read text. This system is developed using Raspberry Pi which again uses PyTesseract library, Python programming and Google-Text-to-speech (GTTS). The flow chart diagram of this research shown in figure 6.

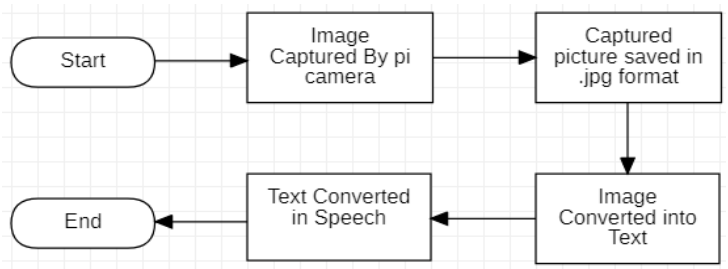

Figure 6 Flow Chart of the (Bhagat-Thakre, 2020) system

Proposed Methodology

New design and implementation of proposed audio reading system which is used to help blind people to read a text. Our system is characterized by many algorithms and steps to make it efficient and easy to use. Our system uses the cloud computing as server side and new design glasses as client side.

Our methodology consists of following steps: In Client Side, firstly, capture image of text by Raspberry pi camera. Secondly, use page localization algorithm to check the position of the page. Thirdly, send image to server. In Cloud Side (Server Side), firstly, apply the modify EAST algorithm on received image to text detection. Secondly, text recognition by applying OCR algorithm to extract the text. Thirdly, apply text correction step to correct the in corrected text. Finally, return text to client side to speak it using text-to-speech algorithm. Figure 7 shows the flow chat form each step.

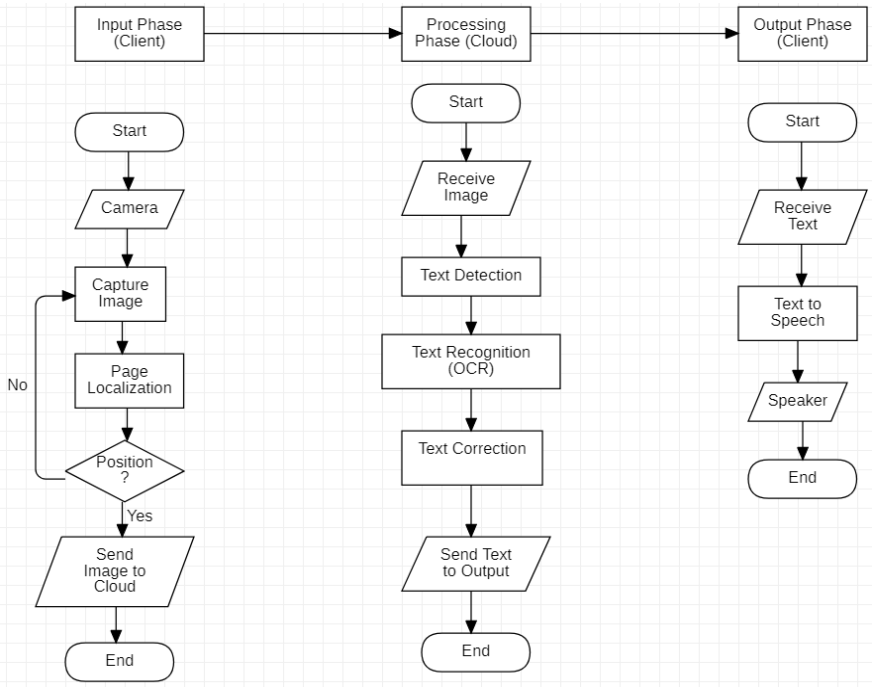

Figure 7 Flow Chart of the proposed system

The main steps of the proposed system are:

Client Side

Step 1: Capture Image, the camera is a main input device in our system, capture image of text by Raspberry pi camera which is placed in the glasses. The image should be high resolution and containing the text that wanted to read.

Step 2: Page Localization, the biggest challenge in audio reading system is the input image. The image captured should be covered all the text that wanted to read. So, our methodology focus in this step to check the position of the page in image. Page Localization consist of the following steps: convert the image to grayscale, find edges in the image using Canny Edge Detection algorithm, find the contours in the edged image, sort contours and get the largest one. Check if the size of page in image is greater than the threshold value and if the contour has four points, then our system assume that have found the page, else go to step 1. Algorithm 1 show the page localization algorithm.

Algorithm 1: Page Localization	
Input:	Image, thrshould_area, number_points=4
Output:	True of False
Steps:	
1.	Convert the image to grayscale.
2.	Find edges in the image using Canny algorithm.
3.	Find the contours in the edged image.
4.	Sort contours and get the largest one.
5.	If contour_area > thrshould_area then
6.	If contour_point == number_points then
7.	Return True
8.	Return False

Step 3: send image to the Cloud side.

Cloud Side

Step 1: received image from the client side.

Step 2: Text Detection, because he is blind person, the image captured doesn't contain only the text. So we should apply text detection algorithm on image to determine the region of interest (ROI). This step is used to decrease the execution time of the text recognition step because we apply it only on the region of interest (text only).

EAST (Efficient and Accurate Scene Text) text detector is a deep learning model, based on a novel architecture and training pattern. It is capable of (1) running at near real-time at 13 FPS on 720p images and (2) obtains state-of-the-art text detection accuracy (Zhou et al., 2017).

The drawback of the EAST algorithm in this system is it doesn't detection the text in the correct sequence of text as in image. So, the EAST algorithm should modify to detection the text sequence correctly. Algorithm 2 shows the new steps of the EAST algorithm.

Algorithm 2: Modify EAST algorithm	
Input:	Image
Output:	Region of Interest (ROI)
Steps:	
1.	Apply Radon Transform on image to correct text skew.
2.	Resize the Image.
3.	Apply predictions step.
4.	Generating Boxes using non_max_suppression algorithm.
5.	Threshold: Calculate the average height of the boxes.
6.	Sort the boxes by x0 then y0 coordinates (if a box is defined as (x0, y0) for top left corner and (x1, y1) for bottom right corner).
6.1	For i=2 to Boxes:
6.2	If the difference between y0 coordinates of box i and i - 1 are greater than half the threshold, then box i belongs to the next line of text. So we increase its sort index by 1 from the sort index of the previous line.
6.3	Else: box i is in the same line as box i - 1, so we keep the same sort index for it.
7.	Sort the boxes by x0 coordinate then by the sort index.
8.	For i=1 to Boxes
9.	Compute start_x, end_x, start_y, end_y as ROI

Step 3: Text Recognition, in this step the Tesseract OCR algorithm will be apply to extracted text from ROI of image. In another word, translation of the character image into character codes, such as ASCII.

One state-of-the-art solution is the Tesseract OCR engine, considered to be one of the best OCR engines available (Luscombe et al., 2020). Tesseract (Tesseract OCR, 2021) is arguably one of the most popular open source OCR engine. It features page layout analysis and a flexible text recognition module. It was originally created at Hewlett-Packard between 1985 and 1994 and was open sourced in 2005. Tesseract OCR Engine makes use of Long Short Term Memory (LSTM) which is a part of Recurrent Neural Networks (Clausner et al., 2020).

Step 4: Text Correction, text correction is a task for detecting and fixing any spelling word errors. We are to substitution one word into another using a set of prescribed edit operations. The accuracy of the output text depends on the input image. Due he is blind people, the input image may be unclear, distortions or etc. So, the output of OCR

algorithm should be checked if there is incorrect word in a text (this step called post-processing of OCR).

This paper proposed the following processes to correct the text: English-word dataset is a dataset which contain all words (sorted) in the English language. Using Damerau–Levenshtein distance algorithm to get the predicate correct word instead of incorrect word from the set of correct words from the dataset.

Damerau–Levenshtein distance algorithm is a matrix measurement obtained from the calculation of two words. In Damerau–Levenshtein distance the two words represent the number of minimum changes required to substitute a word with another word. The operations used in this algorithm are insert, delete, substitution, and transposition (Christanti & Naga, 2018) (Santosoa et al., 2019) (Zhao & Sahni, 2019).

The big challenge in Damerau–Levenshtein distance algorithm is a large execution time, due to the large dataset. So our system proposed the index table to store the start and end index for each character in the dataset. By using the index table, our system takes less execution time by get the candidate subset of the correct words from the dataset based on the first character. The candidate subset and incorrect word are input to the Damerau–Levenshtein distance algorithm to predicate the correct word.

Step 5: send the corrected text to the client side.

Return to Client Side

Step 1: Receive the text from the Cloud side.
Step 2: using Text-to-Speech tools to speak the text using the speaker device.

The block diagram of the proposed algorithm shown in the figure 8.

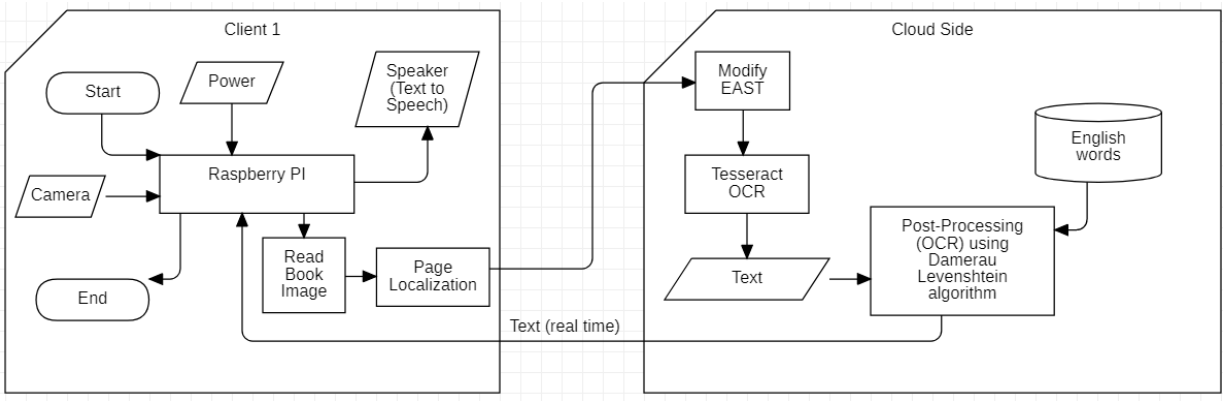

Figure 8 Block Diagram of the proposed system

Finally, the hardware controller block shown in the figure 9.

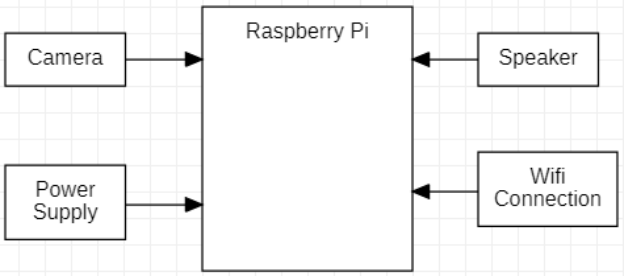

Figure 9 Hardware block diagram of the proposed system

Cloud Computing

Cloud Computing is a standout amongst the most mainstream developing innovations in the recent years. Cloud Computing is the combine of an innovation, service, platform that gives hosting and capacity administration on the Internet. Cloud Computing provide many service can be access by user from the internet (Pooja, 2019).

To provide the cheap glasses which can buys by any blind person. Our system designed the glasses with hardware limitation. The proposed system uses the cloud computing to make all execution (process) needed and to centralization data which collected from many blind persons. The openstack (Openstack, 2021) cloud computing solution was installed to create the cloud which is used in our system.

Our Contribution

There are many contributions in our proposed system:

1. First Real time system with cloud computing.
2. Page localization step to check the position of text page in image.
3. Modify EAST algorithm to detection the text sequence correctly.
4. Text Correction step using Damerau–Levenshtein distance algorithm to correct the incorrect word after applying the OCR algorithm.
5. Index table to decreasing the execution time.

Result

After build the glasses and implementation using Raspberry pi and python programming language. Figure 10 shown the glasses device.

Figure 10 Glasses Device

Example 1: table 1 shows the execution time for each step for the input image (720x720) shown in the figure 11.

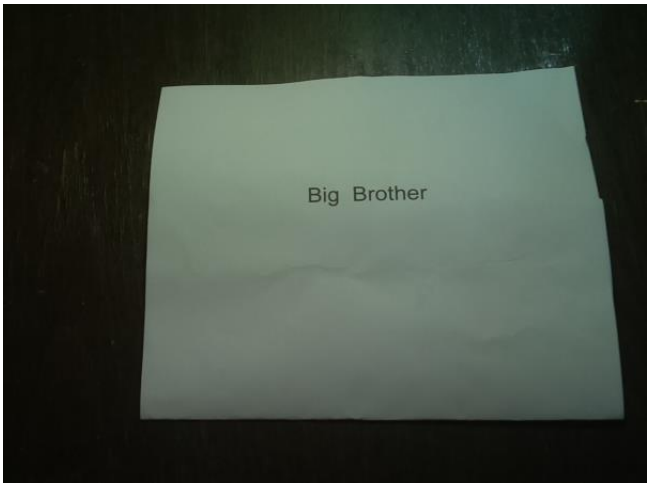

Figure 11 Input image (720x720)

OCR output: “Big Brother”.
Text Correction Output: “Big Brother”.
Page Area: 72974.5 pixel

Table 1 Time execution for each step

Step	Time (second)
Page Localization	1.0585
Server Side	8.0566
Total	9.7547

Example 2: table 2 shows the execution time for each step for the input image (720x720) shown in the figure 12.

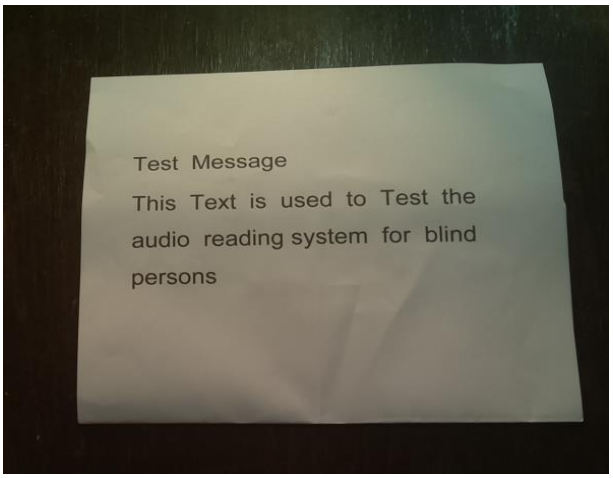

Figure 12 Input image (720x720)

OCR output: “Test Message This Text is used to Test the audio reading system for blind persons”.
Text Correction Output: “Test Message This Text is Used to Test the Audio Reading system for Blind Persons”.
Page Area: 87727.5 pixel

Table 2 time execution for each step

Step	Time (second)
Page Localization	0.8589
Server Side	10.2472
Total	11.9071

Example 3: the system asks the blind to take another image because page localization incorrect position. The input image shown in figure 13.

Figure 13 Input Image (720x720)

Conclusion

In this paper, proposed a real time audio reading system to read printed message for helping blind person. Due it is blind person, the successfully system should be having the following steps: Text detection step to speed up the execution time, page localization step to check the complete view page in image captured, and text correction step.

Finally, suggestion to apply the data mining algorithm and recommendation system in a data centralization which is collected in cloud.

References

Shirke, S.I., & Patil, S.V. (2018). Portable Camera Based Text Reading of Objects for Blind Persons. *International Journal of Applied Engineering Research, 13*(17), 12995-12999.

Goel, A., Sehrawat, A., Patil, A., Chougule, P., & Khatavkar, S. (2018). Raspberry Pi Based Reader for Blind People. *International Research Journal of Engineering and Technology*, *5*(6), 1639-1642.

Gurav, M.D., Salimath, S.S., Hatti, S.B., Byakod, V.I., & Kanade, S. (2017). B-LIGHT: A Reading aid for the Blind People using OCR and OpenCV. *International Journal of Scientific Research Engineering & Technology (IJSRET), 6*(5), 546-548.

Barathkumar, K., Balaji, S., Desikan, K.N., & Iyappan, S.P. (2019). Raspberry Pi based Smart Reader for Visually Impaired People. *International Journal of Research in Engineering, Science and Management, 2*(2), 204-207.

binti Harum, N., Zakaria, N.A., Eimran, N.A., Ayop, Z., & Anawar, S. Smart Book Reader for Visual Impairment Person using IoT Device. *International Journal of Advanced Computer Science and Applications, 10*(2), 251-255.

Sahu, N., Raut, A., Sonawane, S., & Shaikh, R. (2020). Prescription Reading System for Visually Impaired People using NLP. *International Journal of Engineering Applied Sciences and Technology, 4*(12), 265-268.

Ravi, A., Khasimbee, S., Asha, T., Joshna, T.N.S., & Jyothirmai, P.G. (2020). Smart Reader for Blind People. *Journal of Analysis and Computation (JAC), 8*(8), 221-230.

Bhagat-Thakre, M.P. (2020). *Smart Interaction System for Blind and Dumb.*

Zhou, X., Yao, C., Wen, H., Wang, Y., Zhou, S., He, W., & Liang, J. (2017). East: an efficient and accurate scene text detector. *In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 5551-5560.

Luscombe, A., Dick, K., Duncan, J., & Walby, K. (2020). *Access to Information and Optical Character Recognition (OCR): A Step-by-Step Guide to Tesseract.*

Tesseract OCR, https://github.com/tesseract-ocr.

Clausner, C., Antonacopoulos, A., & Pletschacher, S. (2020). Efficient and effective OCR engine training. *International Journal on Document Analysis and Recognition (IJDAR), 23*(1), 73-88.

Christanti, V.M., & Naga, D.S. (2018). Fast and Accurate Spelling Correction Using Trie and Damerau-levenshtein Distance Bigram. *TELKOMNIKA, 16*(2), 827-833.

Santosoa, P., Yuliawati, P., Shalahuddin, R., & Wibawab, A.P. (2019). Damerau levenshtein distance for Indonesian spelling correction. *Jurnal Informatika, 13*(2), 11-15.

Zhao, C., & Sahni, S. (2019). String correction using the Damerau-Levenshtein distance. *BMC bioinformatics, 20*(11).

Pooja, D., 2019. Cloud Computing-Overview and its Challenges.

Openstack, https://www.openstack.org/