# DOCUMENT READING SYSTEM FOR BLIND PEOPLE ("READING EYE")

19-20-J 17

Software Requirements Specification

IT16165762
P.S.N.Kularathne

Ms. Suranjini Silva

B.Sc. Special (Honors) Degree in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

September 2019

# DOCUMENT READING SYSTEM FOR BLIND PEOPLE

# ("READING EYE")

19-20-J 17

Software Requirements Specification

(SRS documentation submitted in partial fulfilment of the requirement for the Degree of Bachelor of Science Special (honors) In Information Technology)

IT16165762

P.S.N.Kularathne

Ms. Suranjini Silva

B.Sc. Special (Honors) Degree in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

September 2019

# DECLARATION

I declare that this is my own work and this system requirement specification does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any other university or institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgment is made in the text.

| Name | Student ID | Signature | Date |
|---|---|---|---|
| P.S.N.Kularathne | IT16165762 | | |

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

…….. …..…..…….……

Signature of the supervisor Date

…….. …..…..…….……

Signature of the co-supervisor Date

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# 1 Introduction

## 1.1 Purpose

The prime purpose of this software requirements specification (SRS) document is to provide a comprehensive description of the functional, non-functional requirements, dependencies and all the relevant specifics of the Analyzing and Detecting Mathematical Equations and Charts component of the proposed Document Reading System for Visually impaired People ("Reading-Eye").

The document will further depict the purpose, key features, product perspective, product functionalities flow, end-user characteristics, constraints under which the system should perform and various external interface requirements such as system interface, user interface, software interface, hardware interface and communication interface of the proposed component and also it will provide the details about the specific requirements and the supporting information of the relevant component.

The intended audience for this document is the supervisor, co-supervisor, the research team members, developers, testers and any other individuals with good computer literacy and all type of stakeholders interested in this research.

## 1.2 Scope

This document comprises the full depth description of the CEBCRD component which will guide the VI user to read and get the idea about the mathematical equations and the charts in the documents. Furthermore, the document illustrates all the tools and technologies, applied concepts for implementation, the flow of the system through use case diagrams, use case scenarios and other related Unified Modeling Language diagrams, libraries and external interfaces allied to the CEBCRD component.

The proposed component utilizes Machine learning(ML) techniques along with deep learning (DL) techniques, convolutional neural networks(CNN) techniques by training the pre-processed data sets by using relevant algorithms. Moreover, the Optical context of character recognition (OCR) is used with natural language processing (NLP) to identify and separate the charts based-contents and Equations based contents. Using a cross-platform mobile application outcome will present to the VI end-user through using azure text to speech API.

Therefore, this document will guide the proposed methodology to accomplish the target objectives of the proposed CEBCRD component.

## 1.3 Definitions, Acronyms, and Abbreviations

| Term | Definition |
|---|---|
| Reading-Eye | The name of the proposed system |
| Convolutional Neural Network | A specific type of artificial neural network that used to analyze data |
| TensorFlow | Open-source machine learning framework |
| Keras | High-level neural networks API which runs on top of TensorFlow |
| Tesseract OCR | Open source optical character recognition engine |
| Natural Language Processing | Artificial Intelligence-based component which makes computers to read the text, hear speech, interpret it, measure sentiment and determine which parts are important. |
| OpenCV | The library used for Image Processing |
| LetNet Architecture | A latest convolutional network designed for handwritten and machine-printed character recognition. |
| SoftMax Identification Function | A function that takes as input a vector of K real numbers, and normalizes it into a probability distribution consisting of K probabilities proportional to the exponentials of the input numbers. |

*Table 1- Definition for the terms used in this SRS*

| Acronym/Abbreviation | Definition |
|---|---|
| CEBCRD | Charts and Equations based contents Reading and Detection. |
| ML | Machine Learning |
| DL | Deep Learning |
| CNN | Convolutional Neural Networks |
| NLP | Natural Language Processing |
| OCR | Optical Context of character Recognition |
| VI | Visually Impaired |
| API | Application Programming Interface |
| R-CNN | Region of Convolutional Neural Networks |

*Table 2- Glossary of Acronyms*

## 1.4 Overview

### 1.4.1 Tasks of the system

The CEBCRD is a main component of the Reading-Eye System. This component will mainly analyze and detect the equations based contents and the charts based contents in the documents. In order to achieve this task, the implementation of the system is done using Deep Learning methodologies and Natural Language Processing which supported by mobile and cloud platforms. The existing data in the cloud storage is used by the component to get real-time updates.

The system enables visually impaired users to read and identify the type, size, color of the charts with the percentage of the colored area and also the operators, variables, and numbers in the complex equations accordingly. The output will be presented as a JSON file format. However, the VI end user will get a detailed description of the contents as an audio listing file through the support of Azure text to speech API.

### 1.4.2 Main goal of the system

The main goal of the proposed system is to facilitate visually impaired people to read printed documents that are not written using the braille system and help them to improve their reading capability as normally sighted people.

### 1.4.3 Specific goals

The specific goals for the CEBCRD component are as follows.

- Identify and detect the chart based contents separately by chart type, size, color and the resolution of the chart and make a detailed description according to the identified content.
- Identify and detect the mathematical equations based contents separately by the numbers, variables, and operators contained in the equations and make a detailed description according to the identified content

### 1.4.4 Users

The user segmentation for the proposed mobile application would be mainly based on three categories as psychographic, geographic and demographic. When considering psychographic segmentation it is mainly considered about the visually impaired people who have partial or complete blindness. The initial target market based on geographic segmentation would be Sri Lanka. The financial statuses of the users are considered under demographic user segmentation. This product mainly focuses on providing a cross-platform mobile application that would help the user to improve their reading ability in a friendly manner.

# 2 Overall Descriptions

Reading is not a difficult task for a person who has the ability to see. However, it is a comparatively much difficult task for a visually impaired person. Especially the documents with huge contents are not accessible in the braille system. Moreover, there is a limitation of identification and to get proper knowledge of complex mathematical equations and chart based contents for them because most documents are not written in the braille system the complex contents. Therefore in such a situation, anyone will prefer to have a guide with a proper document reading system that acts promptly.

**Problem**

Due to the major development of modern technology, there are many existing document reading apps are available for visually impaired people. However, the majority of software applications developed to detect and analyze text contents but most of them cannot detect chart-based contents and equations-based contents [1],[2],[3],[4].

Moreover, most of the equation detecting researches had found impediments as follows in document reading systems[5].

- The difficulty of recognizing the structure of the mathematical expression.
- The difficulty or an error on the recognition of the numbers, variables, and the operators (symbols) in the equation.
- The difficulty of identifying the region of the equation in the complicated layout.

When considering the chart based content detection, there were the least amount of results and unable to find comparatively numerous researches since this feature compromises new methodologies and techniques but still it not resolved with a proper solution.

The major challenges in chart image classification as found as follows in research called Convolutional Neural Network Based Chart Image Class [6].

- The difficulty to detect the variability in the structure of the chart.
- The difficulty of identifying the size, region, content and the color of the chart.
- The difficulty of identifying the visual appearance of each chart type.

**Solution**

The proposed solution for the identified problems would be to guide visually impaired users to get a detailed description of the mathematical equations and the chart based contents when reading those contents in the document through listening to audio.

Mainly identified chart based and equation-based contents divide into the sub-modules accordingly. After that take the cropped image of the relevant chart and equation-based contents by using the OpenCV library. The region identification of the above sub contents will be done using the Mask R-CNN. Moreover using CNN identify and categorize the variety of chart type, size, color, resolution, and the content.

The analyzation of the chart based content is done using a trained data set of neural networks inspired by the LetNet architecture. At the end of the model, the fully connected LetNet

architecture layers with the SoftMax activation function make use of learned high-level features to classify the input images into predefined chart types. After detecting the chart type using OCR technique model to provide chart content output as the title of the chart, percentages of the contained detailed of the chart with the defined colors in the chart. After the OCR the read data will take input to generate the text. Moreover, NLP will be used for generating cohesive sentences in a way the user can comprehend.

Same as before OCR will be used to recognize characters, the numbers, variables, and the operators (symbols) of the equation in the image. NLP is used to understand the meaning of these symbols and after all, generate meaningful sentences. At the beginning system will implement to read simple equations. It will be further developed to read quite complex equations later. When implementing the above tasks Keras and python library will be used as usual.

## 2.1 Product Perspective

Reading-Eye is a comprehensive type of a document reading application which targeted to resolve the impediments of the visually impaired people. Most of the available existing applications provide minimal solutions for the targeted CEBCRD component in the proposed system.

The proposed component will fulfill the major gap in existing solutions in the market. Amazon Kindle, BARD Mobile, KNFB Reader [1],[2],[3], Capti Voice [4], Schmoozer [7] are the some of the existing applications in the market which related to the proposed system.

Below is the identified comparison of the proposed system features and the previous systems features.

| **Existing Product**<br>**Features** | Amazon Kindle | BARD Mobile | Capti Voice | KNFB Reader | Schmoozer | **Reading-Eye (Proposed System)** |
|---|---|---|---|---|---|---|
| Equations based contents Identification and Reading | ✗ | ✗ | ✗ | ✗ | ✓ | ✓ |
| Charts based contents Identification and Reading | ✗ | ✗ | ✗ | ✗ | ✗ | ✓ |
| Images' size reduction for efficient Communication | ✗ | ✗ | ✗ | ✗ | ✗ | ✓ |
| Voice output | ✓ | ✗ | ✓ | ✓ | ✓ | ✓ |

*Table 3 - Existing products comparison*

### 2.1.1 System Interfaces

The user can point the camera of his device to the printed document which he wants to read. Then the camera will capture the photo and the app will set it to correct orientation. The mobile application will provide the encryption and compress the image using the Microsoft Azure cloud computing platform. These details will be stored on the mobile phone. The system interfaces involved in this process are as follows,

- Azure VM – Centralize all the communication and run the server.
- Azure Storage – A cloud storage solution for modern data storage scenarios.
- Azure Text to Speech API - Converting text to speech with user experience and accessibility.

### 2.1.2 User Interfaces

The proposed system CEBCRD component basically deals with the Azure cloud storage by providing the final output as the JSON files. Therefore no need to use a user interface for the CEBCRD component process.

### 2.1.3 Hardware Interfaces

The Reading-Eye system presents the final output to the user through the smart mobile phone. Therefore the end-users do not need to have any special dedicated devices to run this system component.

### 2.1.4 Software Interfaces

The mobile app requires Android 4.1(API 16) or higher version and IOS 9.0 or higher version for the deployment.

### 2.1.5 Communication Interfaces

The Internet facility is needed to get connected to cloud services. Therefore for that purpose mobile data or Wi-Fi should be necessary.

### 2.1.6 Memory Constraints

Since the CEBCRD component process the information in the cloud beforehand, there is no need for the mobile phone to allocate much of memory for processing or storing. At least

200MB space is required to store the location detailed file in the mobile phone and approximately 2GB memory is required to run the app with the rest of the components.

### 2.1.7 Operations

The proposed system will be delivered to its end-users as a cross-platform mobile application. Therefore, the user should purchase the application from the play store and install it on an Android mobile device otherwise end-user can download it from the app store and install it in an IOS mobile device.

The user should always enable Wi-Fi or mobile data to make the internet connection continuously.

The mobile phone should place clearly to capture the photo from the camera to get a clear photograph. However, the captured photo is not clearly identified it will notify to end-user through the voice because all the alerts and notifications will be notified through voice using azure text to speech API.

### 2.1.8 Site Adaptation Requirements

The CEBCRD component will incorporate a training model to detect and read the contents of charts and equations.

Therefore, from the user's side, there won't be any special sessions to train the app to perform better. For the very first time, the user uses the Reading-Eye app to capture the photograph of the document that needs to read own.

The selection of the other components will operate the table, image, text category analyzations. Data will be fully encrypted for provide accurate results for the user. Therefore the Reading-Eye app will produce the results only for those components and if needed the user can change the settings later.

## 2.2 Product Functions

The CEBCRD is a component of the proposed Reading-Eye application. By analyzing the captured photograph of the document by the user system will encrypt the data and store it in the cloud storage. The captured image will separately divide into the relevant sub-modules as equations based contents and chart based contents etc. Moreover after that stage graphical image of the equations based contents and chart based contents identified separately.

The identification of the chart based contents is done by separating the chart size, type, color, resolution, etc. Moreover, the equations also separate from the variables, operators, numbers, etc.

After that separated identification proposed system will generate detailed descriptive text content which contained the contents of the charts and equations. As the final output of the component, a JSON file is generated for the user to listen.

### 2.2.1 Diagram for the CEBCRD component

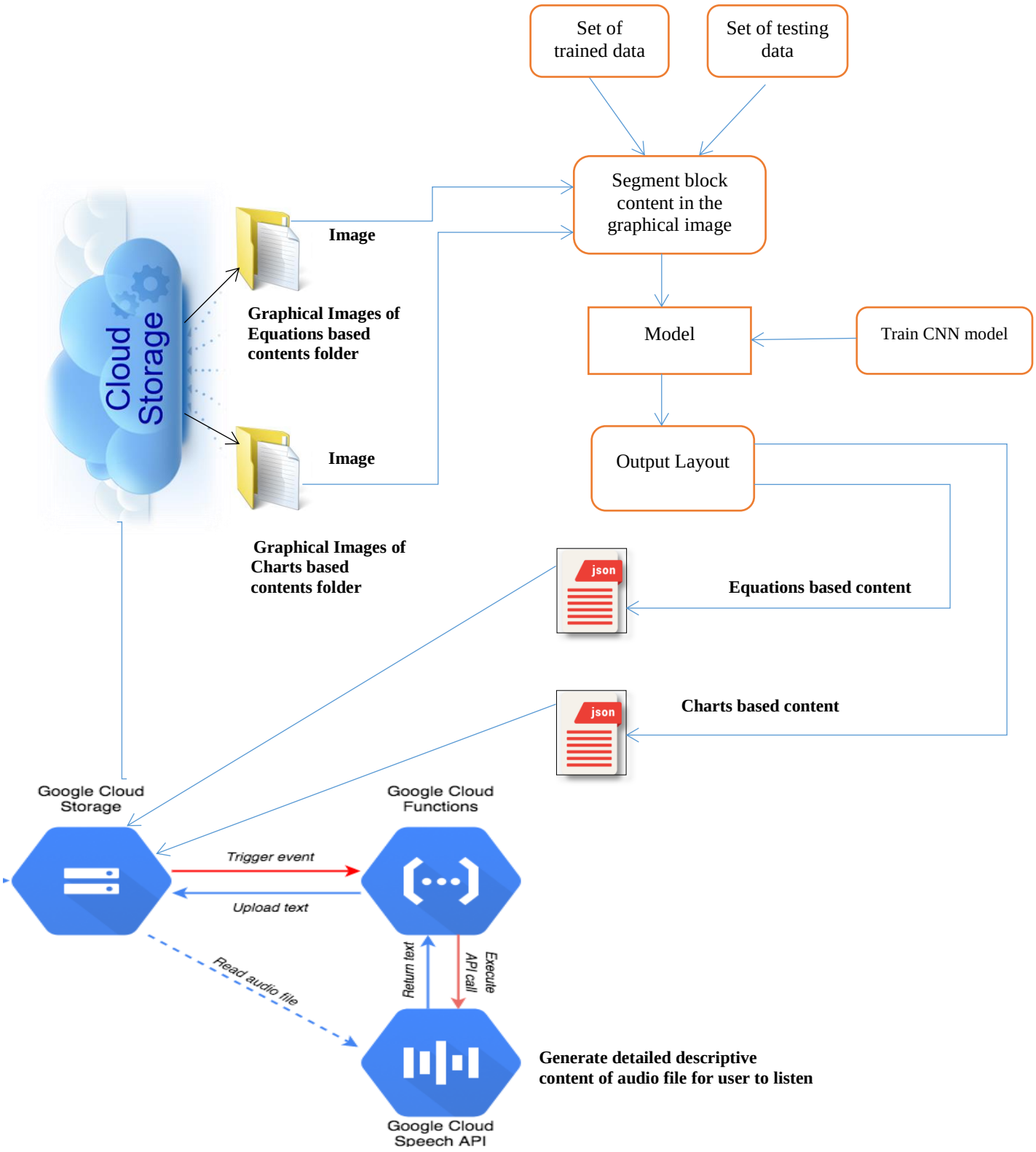


*Figure 1 - High-level architecture diagram for CEBCRD_component*

2.2.2 Use case diagram

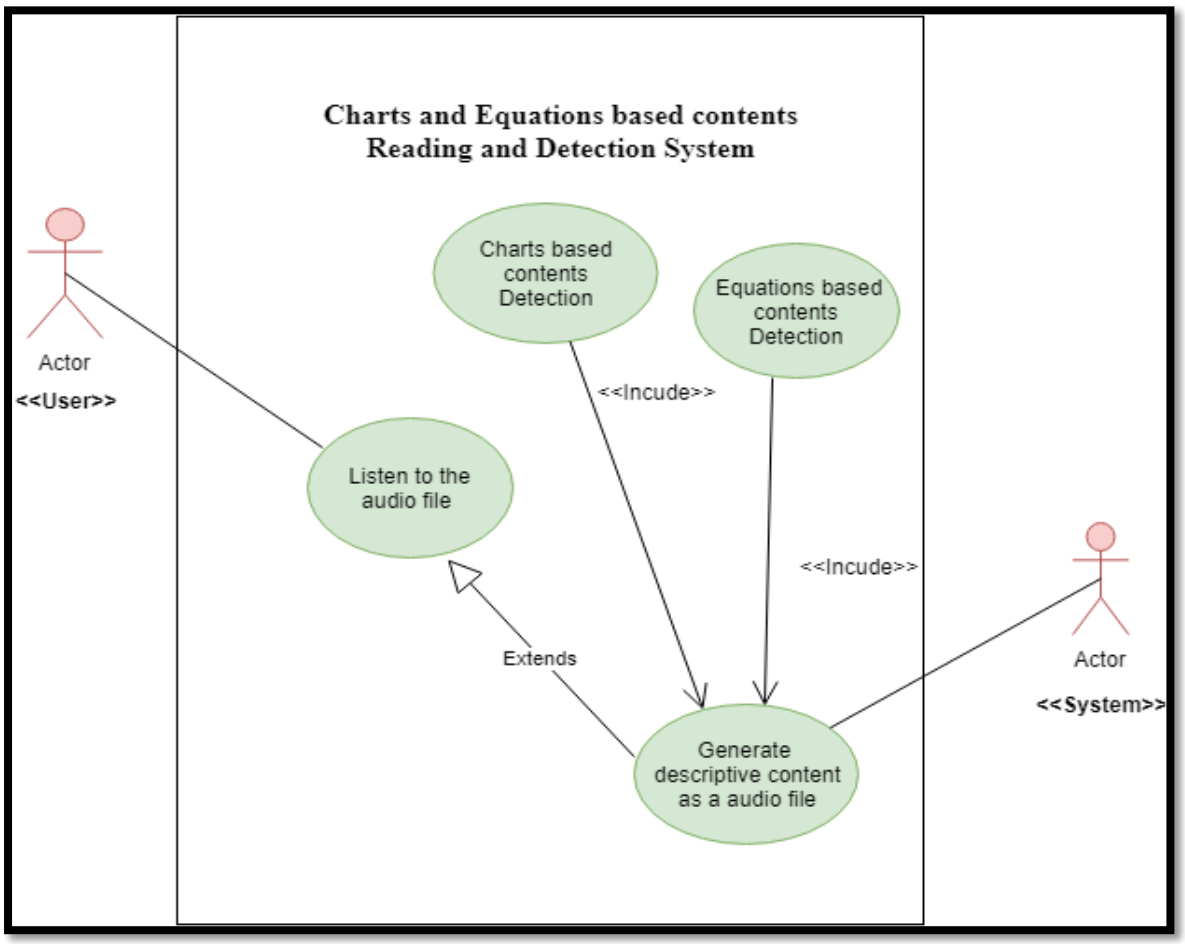


*Figure 2 - Use case diagram of CEBCRD_component*

2.2.3 Use case scenarios

| | |
|---|---|
| Use case ID | CE_UC1 |
| Use case Name | Retrieving files from the folders |
| Usecase Description | The system has to retrieve Charts and Equations from the respective folders. |
| Primary actors | System |
| Pre-conditions | 1. User should capture the photograph.<br>2. The system should be connected to cloud services. |
| Main flow | 1. The system should check the repository where the files are saved.<br>2.System checks Charts and Equations folders.<br>3. The system returns the relevant files for analyzing.<br>4.End of the Usecase. |

*Table 4 - Use case scenario 1*

| | |
|---|---|
| Use case ID | CE_UC2 |
| Use case Name | Charts based on contents detection. |
| Usecase Description | The system detects objects and areas in the chart based image content. |
| Primary actors | System |
| Pre-conditions | 1. The user should capture the photograph.<br>2. The system should be connected to cloud services.<br>3.The system should previously retrieve Charts from the respective folder. |
| Main flow | 1. The system analyzes the chart based image content.<br>2. The system compares the chart based image content with the data sets for object detection.<br>3. The system detects objects (size, type, etc) and areas separately.<br>4. The system saves the identified objects and areas temporary.<br>5.End of the Usecase. |
| Post-conditions | The chart based content output will be saved as a .txt file format for text generation. |

*Table 5 - Use case scenario 2*

| | |
|---|---|
| Use case ID | CE_UC3 |
| Use case Name | Equations based on contents detection. |
| Usecase Description | The system detects objects and areas in the equations based image content. |
| Primary actors | System |
| Pre-conditions | 1. The user should capture the photograph.<br>2. The system should be connected to cloud services.<br>3. The system should previously retrieve Equations from the respective folder. |
| Main flow | 1. The system analyzes the equations based image content.<br>2. The system compares the equations based image content with the data sets for object detection.<br>3. The system detects objects (numbers, operators, characters, etc) and areas separately.<br>4. The system saves the identified objects and areas temporary.<br>5.End of the Usecase. |
| Post-conditions | The equations based content output will be saved as a .txt file format for text generation. |

*Table 6 - Use case scenario 3*

| | |
|---|---|
| Use case ID | CE_UC4 |
| Use case Name | Generate Descriptions |
| Usecase Description | The system generates detailed descriptions for chart based and equations based contents. |
| Primary actors | System. |
| Pre-conditions | 1. The user should capture the photograph.<br>2. The system should be connected to cloud services.<br>3.The system should previously retrieve Charts and Equations from the respective folders.<br>4. The system should previously detect the objects of Charts and Equations. |
| Main flow | 1. The system retrieves temporary text files that were saved earlier when the object detection.<br>2.A detailed description will be generated by taking text files as the input data by using NLP techniques.<br>3.The generated detailed description of files will be saved in the JSON file format.<br>4.End of the Usecase. |

*Table 7 - Use case scenario 4*

| | |
|---|---|
| Use case ID | CE_UC5 |
| Use case Name | Generate Audio Files |
| Usecase Description | The system generates audio files by using the generated description of files. |
| Primary actors | System |
| Pre-conditions | 1. The user should capture the photograph.<br>2. The system should be connected to cloud services.<br>3. The system should previously retrieve Charts and Equations from the respective folders.<br>4.The system should previously detect the objects of Charts and Equations.<br>5.The system should previously generate a detailed description of the files. |
| Main flow | 1. The system concatenates all the JSON files that were generated earlier.<br>2. Using the Microsoft Azure text to speech API generates an audio file.<br>3.End of the Usecase. |

*Table 8 - Use case scenario 5*

| | |
|---|---|
| Use case ID | CE_UC6 |
| Use case Name | Listen to the audio file. |
| Usecase Description | The user will listen to the audio file narration. |
| Primary actors | User |
| Pre-conditions | 1. User should turn on the mobile data. |
| Main flow | 1. The Usecase starts when the user captures the photograph of the document.<br>2. After processing the system will generate an audio file.<br>3. The user will listen to the audio narration.<br>4. End of the Usecase. |
| Post-conditions | The app will be ready to play the audio. |

*Table 9 - Use case scenario 6*

## 2.3 User Characteristics

The segmentation of the users in the proposed application mainly based on geographic, psychographic and demographic criteria.

A basic understanding of using a smartphone is required as a user. The system is designed to give instructions in the English language but can be changed to beeps to fit the user's preference. Therefore the user should have the ability to understand the English language and should need a proper listing and hearing ability without disabilities.

## 2.4 Constraints

- The app will be developed as a cross-platform mobile application.
- Instructions and audio will be provided in the English language.
- A proper Internet connection is required to run the application.
- A major dataset needs to be trained in order to get the accurate output. Therefore, the required time for training will be very high.
- Since there are four major components in the app, achieving higher performance will be a constraint.
- Security will be highly enhanced by data encryption.

## 2.5 Assumptions and Dependencies

- The internet connectivity is always active and of good connection speed.
- The server will not crash or go down.
- The mobile battery doesn't drain quickly.
- The result of each component will not get overlapped when presented in an audio format.

## 2.6 Apportioning of Requirements

Since the proposed product is an outcome of a research project, the path to reach the objectives might differ. The major components and their outcomes mentioned in this document will not change in the future. Nevertheless, the methodologies and the technologies mentioned to achieve those outcomes might change in order to make the results more accurate, reliable and efficient.

The requirements mentioned in chapter 1 and 2 in this document are the primary requirements and is very unlikely to be changed in the future. The requirements mentioned in chapter 3 are desirable requirements which will be taken into consideration in the current version's latter stages of implementation, probably in upcoming versions of the proposed application.

# 3 Specific Requirements

## 3.1 External Interface Requirements

External Interface Requirements section will discuss the remotely hosted backend of the system. In order to make the data centralized, the mobile application needs to store and retrieve data and information. The web service or the backend would cater to this requirement through the server which communicates using the HTTP requests.

### 3.1.1 User Interfaces

The web service does not have any user interfaces, as it is a web service that caters HTTP requests using its protocol.

### 3.1.2 Hardware Interfaces

The mobile device's in-built camera will be taken use of for the real-time image capturing. The camera module will be used through the native API by the module which runs using the CNN model. A smartphone is a sufficient condition since this is a fairly demanding task. Therefore No, other advanced hardware interfaces are required.

### 3.1.3 Software Interfaces

System development of the application is done using the React Native. The mobile app requires Android 4.1(API 16) or higher version and IOS 9.0 or higher version for the deployment. In order to perform to the detection of the images, CNN will be utilized.

The Azure cloud services will be leveraged for computation for the backend of the application, including training the model and handling requests from the mobile application. Training of the model will use Python 3.0 with Anaconda python distribution, Keras 2.1.6, TensorFlow 1.8 and Spider IDE.

### 3.1.4 Communication Interfaces

The Internet is required to get connected to the cloud services in order to retrieve the dataset containing the cloud storage and get the audio using the API. Therefore, mobile data or portable Wi-Fi should be used for this purpose.

## 3.2 Classes/Objects

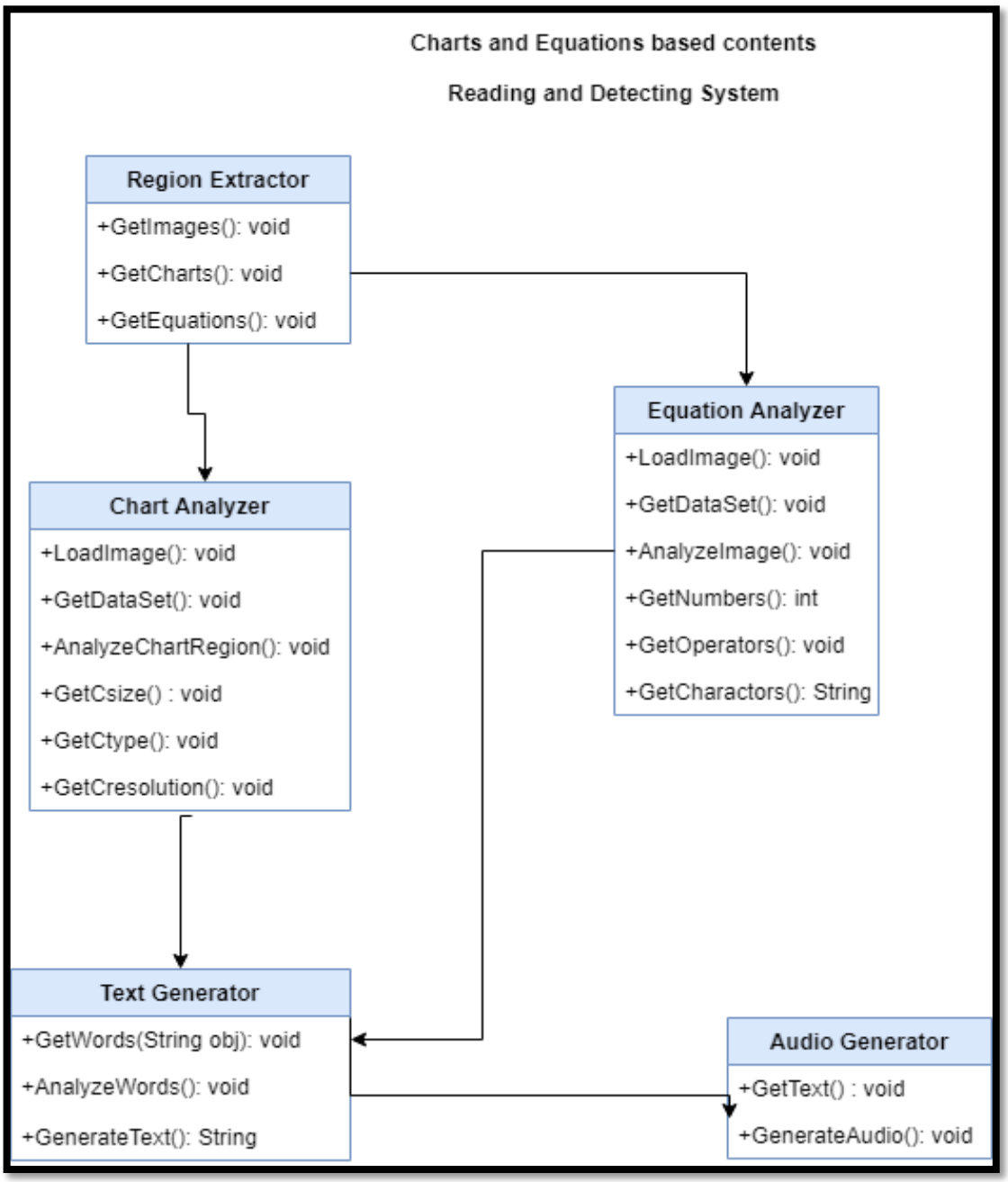


*Figure 3 - Class diagram of CEBCRD_component*

## 3.3 Performance Requirements

As this system is a collaboration of mobile device and cloud storage is the main rating factor, A good internet connection is a must in order for the system to operate at its optimum level to get real-time updates properly, Therefore data must be sent to the servers as fast as possible a crawling internet connection, Moreover, between the device to the server and the server to the device must have a good internet connection.

## 3.4 Design Constraints

The software product is designed to be deployed on Android 4.1 or higher version and IOS 9.0 or higher version. There could be a limitation as a mobile application that should adhere to provide a better experience to the end-user.

Therefore, in an exceedingly mobile, it's not solely the projected software package. However, there'll be varied alternative apps running.

Moreover, if all the four parts within the projected app are to be enforced as a period of time components then the performance of the mobile is curtailed and it would not accommodate to supply the output for sure. Therefore, the system is intended in such a way that, the outcomes of all four components and to be operated in an exceedingly mobile application to supply results for sure.

This android application would also be used by a visionless user, and from their perspective of use, the functionalities should be extremely concise, where their interactions with the device should be more audible than visual.

## 3.5 Software System Attributes

### 3.5.1 Reliability

The reliability of the proposed software product heavily depends on the accuracy level of the results at the appropriate time. This is applicable for CEBCRD component as well. In order to achieve maximum accuracy, the training model should be trained for enough number of datasets and should process the result in a short period of time. In that purpose, the accuracy of results is proportionate to the size of the data set.

Providing results on appropriate time is equally important as providing an accurate result. No matter how accurate the result is if it is not delivered at an appropriate time. Through these strategies, the reliability of the system can be achieved and the system must be reliable for the users to trust what they see or hear from the system.

### 3.5.2 Availability

The software product should have an uptime of 99.9999%. Availability also refers to the availability of information to the user through the system. Since the CEBCRD_component stores the charts based and equations based contents in the cloud storage, the information will be always available for the user to access. The only prerequisite for the user is to capture the photograph to get the final audio output which fetches the details from the cloud to the mobile device. Therefore, the system ensures that the mobile app along with sufficient details will be available in times of necessity.

### 3.5.3 Security

The system will be implemented with 0.001% of bugs. The system does not require any user authentications. The data security is highly enhanced through data encryption by using one main component of the system.

### 3.5.4 Maintainability

The Reading-Eye system is expected to be added with new features and enhancements for the prime four modules in the following releases. Therefore, modularity should be maintained in the different components for increasing the maintainability. Therefore providing real-time, reliable and accurate data to the end-user tends to improve over time by the use of cloud services with its services which will later be incorporated/ updated in the Android application. This requires good maintainability to achieve with minimal effort.

## 3.6 Other Requirements

There are no other requirements identified or provided as far as the phase is considered.

# 4 Supporting Information

## 4.1 References

[1] "Accessible Mobile Apps | American Foundation for the Blind", Afb.org, 2019. [Online]. Available: https://www.afb.org/blindness-and-low-vision/using-technology/assistive-technology-products/mobile-apps. [Accessed: 20- Aug- 2019].

[2] "An In-depth Evaluation of the BARD Mobile App from the National Library Service for the Blind and Physically Handicapped | AccessWorld | American Foundation for the Blind", Afb.org. [Online]. Available: https://www.afb.org/aw/15/2/15607. [Accessed: 20- Aug- 2019].

[3]Play.google.com. [Online]. Available: https://play.google.com/store. [Accessed: 21- Aug- 2019].

[4]"Read text aloud. PDF reading books speak text to voice, read it out loud", Captivoice.com.15[Online].Available:https://www.captivoice.com/captisite/public/entry/explore. [Accessed: 22- Aug- 2019].

[5]"A Simple Equation Region Detector for Printed Document Images in Tesseract - IEEE ConferencePublication", Ieeexplore.ieee.org,2019.[Online].Available:https://ieeexplore.ieee.org/iel7/6627713/6628563/06628621.pdf. [Accessed: 02- Sep- 2019].

[6] J. Amara, P. Kaur, M. Owonibi and B. Bouaziz, "Convolutional Neural Network Based Chart ImageClassification." Pdfs.semanticscholar.org,2019.[Online].Available:https://pdfs.semanticscholar.org/5fe7/c360e851f648095bb630eec742c7b809cf2b.pdf. [Accessed: 23- Aug- 2019].

[7] N. Gamage, K. Jayadewa, S. Senanayake, K. Udeshitha and J. Jayakody, "Audio Assistance for Vision Impaired Individual To Recognize Graphical Content on Print Disable Documents", Annual conference 2017 - IET - Sri Lanka, 2017.[Online].Available: http://theiet.lk/wp-content/uploads/2017/10/24-p9.pdf. [Accessed: 23- Aug- 2019].

## 4.2 Appendices