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

19-20-J 17

Software Requirements Specification

E. M. D. D. Ekanayake

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)

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 acknowledgement is made in the text.

Name	Student ID	Signature	Date
E.M.D.D. Ekanayake.	IT16102156		

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 CONTENT

LIST OF TABLES

LIST OF FIGURES

1 Introduction

1.1 Purpose

The prime purpose of this document is to detail the functional, non-functional requirements, dependencies and Region Identification and separate to digital images. Also, analysis of the digital image and identify the text-based contents and explore a detailed JSON file. Those components of the proposed “Reading-Eye” Application. This document will further depict the purpose, key features, product perspective, functionality 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.

This document for the supervisor Mrs. Suranjini Silva, core-supervisor Dr. Anuradha Jayakody, the research team members, CDAP team and all types of stakeholders interested in Assistive technology research area.

1.2 Scope

This document covers the requirements for the Region Identification and separates to digital images and reading whole document of the “Reading-Eye” Application which will analysis and read the document in detail as human bean. Moreover, the document explains all the tools and technologies to be used, concepts and techniques for implementation and the flow of the system through use case diagrams and use case scenarios.

The component is expected to achieve the said goals by utilizing cloud computing technologies along with deep learning to intelligently region identify, reading text-based content and the mobile platforms (iOS and Android) to present the information to the end-user efficiently. This document will discuss the features and technologies of each application which will act as the reference to the developers and stakeholders in selecting the best design.

Here it is applied to limited types of book structures. But this can be further modified and improved to fulfill any type of documents reading purpose specially for vision infrared peoples.

1.3 Definitions, Acronyms, and Abbreviations

Table 1: Definition for the terms used in this SRS

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 by google
Keras	High-level neural networks API which runs on top of TensorFlow
Azure	Microsoft cloud computing services

Table 2: Glossary of Acronyms

Acronym/Abbreviation	**Definition**
SRS	Software Requirement Specification
IP	Image Processing
ML	Machine Learning
DL	Deep Learning
CDAP	Comprehensive Design & Analysis Project
RI	Region Identification
RTBC	Reading of Text-Based Content
CFAO	Creating Final Audio Output
UI	User Interface
UCS	Use Case Scenario
EUI	External User Interface
MTBF	Mean Time Between Failures

1.4 Overview

RI is an advanced algorithm based on computer vision which uniquely identify the type of the captured image document and divided the separate folders automatically. Semantic segmentation method to identify the area of text, image, chart, equation and table. Many convolutional neural network models have been introduced for semantic segmentation of an image like Fast-RCNNs, Faster-RCNNs.

The second module is RTBC. The module will be implemented to read text in a document image. In the previous module, we extract the text content using a Mask R-CNN. Now we have to read word by word of that extracted chunk. As this task, is to use Optical Character Recognition techniques.

Another module is CFAO. Train out the meaningful descriptive finalized and concatenated JSON files and sent those to the smartphone as well -structured detailed wise. Visually impaired person will be getting the access to listen detailed well-described single audio of the document generates.

1.4.1 Main Goals

- Reading and detecting text-based contents and Region Identification and generate audio output.

1.4.2 Specific Goals

- Develop a user-friendly cross-platform mobile application for vision-impaired people.
    - Analyze captured photos and auto-rotate the image in the actual direction where the captured image should be in the actual position.
    - Make a single command to start analyzing.
    - Make the option after analyzing the document
    - Autoplay starts and the user able to listen to the document.
- Identify text, images, charts, tables, and the equations uniquely and create another digital image for different identical matches.
- In text - Identify language patterns and create a detailed description.

- Generate a JSON file of that detailed description for the user to listening purpose.
- Concatenate all JSON files and generates a single audio clip using open source API service for the user to listen.

1.4.3 Users

- The main users of this system/application will be the vision-impaired people.
- Other users of this system/application will be the very busy smartest peoples to read the document.

1.4.4 Organization of SRS

This SRS document is focused on the research component modules are RI, RTBC and CFAO.

The document consists of four main chapters. The first chapter describes about purpose, scope and structure of the document.

The second chapter explains the product perspective with interfaces, constraints and limitations. It describes the product functions and informal requirements of the system and focuses on the assumptions and dependencies.

The third chapter discusses about the specific requirements of the system. It describes the external interfaces that communicate with this system and explains on the non-functional requirements.

The final chapter includes all the supporting information such as references and appendices in this document.

2 Overall Descriptions

The first task is region identification. When the camera takes a photo, we cannot guarantee the picture will only contain one thing, as it will only contain text. It can contain images or charts as well. Maybe the picture will contain a combination of text, images and charts. First of all, before start analyzing we need to identify the contents in the captured photo. Identifying text areas because majority of the document will anyway be text. After that other contents should be identified and separated. We intend to do this separation by cropping the photo. For example, if a photo contains two paragraphs of text and one chart, the chart area will be cropped and saved and the same will happen to text as well. This is what will happen in the region identification.

The second task is reading text-based content. The module will be implemented to read text in a document image. Text content should extract using the Mask R-CNN. After all, read word by word of that extracted chunk using Optical Character Recognition techniques.

The final task is the final audio output generation. Train out the meaningful descriptive finalized and concatenated JSON files and sent those to the smartphone as well -structured detailed wise, meaningfully described JSON files. Using the headset visually impaired person will be getting the access to listen detailed well-described single audio of the document.

2.1 Product Perspective

There are other similar products available in the market. “Schmoozer” is one of them

Table 3: Existing products comparison

Features	Schmoozer	Reading-Eye
Smooth Real-Time Analysis	No	Yes
Using CNN for the detect the regions	No	Yes
Voice commands	Yes	Yes
Voice output speed can change	No	Yes
Voice output profile	No	Yes
Recent document saving	No	Yes
Using cross-platform mobile application	No	Yes

2.1.1 System Interfaces

The mobile application will once communicate with the cloud server whenever the driver enters the destination and preferred route, which will retrieve the appropriate dataset to the mobile device. The data gathered from the real-time detection module will also be stored in the cloud for further processing.

- Azure VMs

  Runs the server which centralizes all communications.

- Azure Storage

  Stores the collected dataset from the users and also for any dataset used for the preparation at the inception of the component.

- Azure Text to Speech

  Convert text to audio in near real time, play it back.

2.1.2 User Interfaces

Figure 1: User Interfaces 1

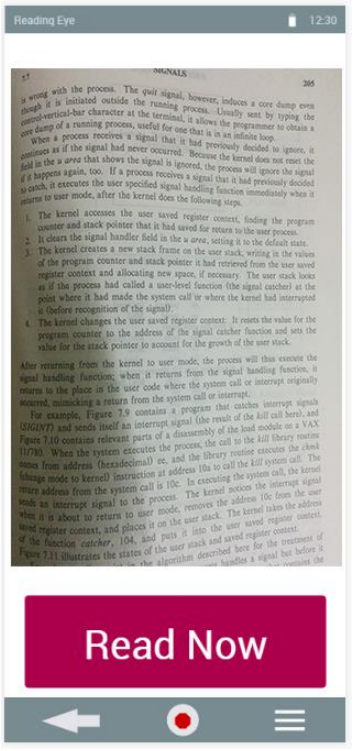

Figure 2: User Interfaces 2

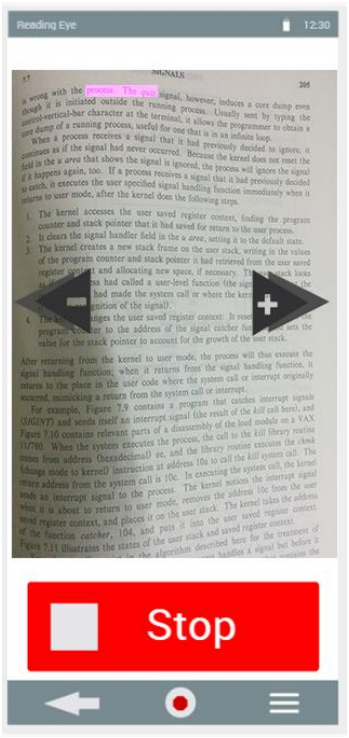

Figure 3: User Interfaces 3

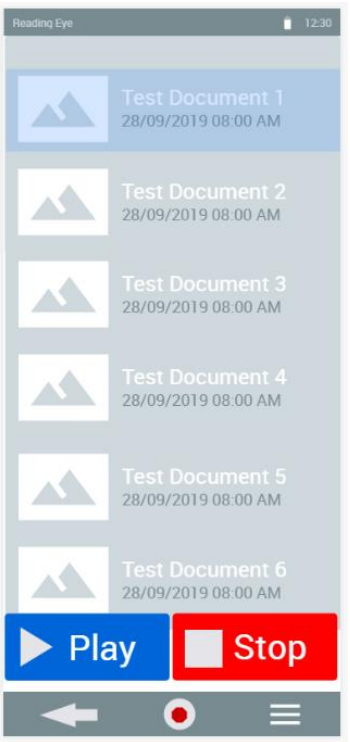

Figure 4: User Interfaces 4

2.1.3 Hardware Interfaces

No special hardware is required for the end-user. A smartphone with a decent 8MP rear-camera is sufficient to get the required information.

2.1.4 Software Interfaces

The mobile app mainly requires an Operating System. There are two options for end-user.

- Android 4.1(API 16) or above version
- IOS 9.0 or above version

Text to Speech SDK Release version 1.3.1 or above.

2.1.5 Communication Interfaces

The application requires an active internet connection some times. It can be used the Internet connected Wi-Fi connection or mobile data connection.

2.1.6 Memory Constraints

250MB of space is required to install the application with 1GB of memory to run it effectively to run the RI and RTBC modules in parallel and 150MB of cache memory for all required for CFAO module.

2.1.7 Operations

- The first step is installing the application to mobile. This product is cross-platform mobile application. Therefore, having two methods.
    - Android devices – “Reading-Eye” application from Play store
    - iOS devices – “Reading-Eye” application from App store
- Users must capture the photo from the mobile phone real-camera. If success the capture success massage delivers true the voice.
- The final step is the analyzing document. Users can “Analyze now” voice command to start analyze the captured document.

2.1.8 Site Adaptation Requirements

The user does not have to configure the RI and RTBC modules. After installing the mobile application properly, it will detect the mobile phone real-camera automatically and do all the configurations on its own.

2.2 Product Functions

2.2.1 High Level Diagram of the RI, RTBC and CFAO Modules

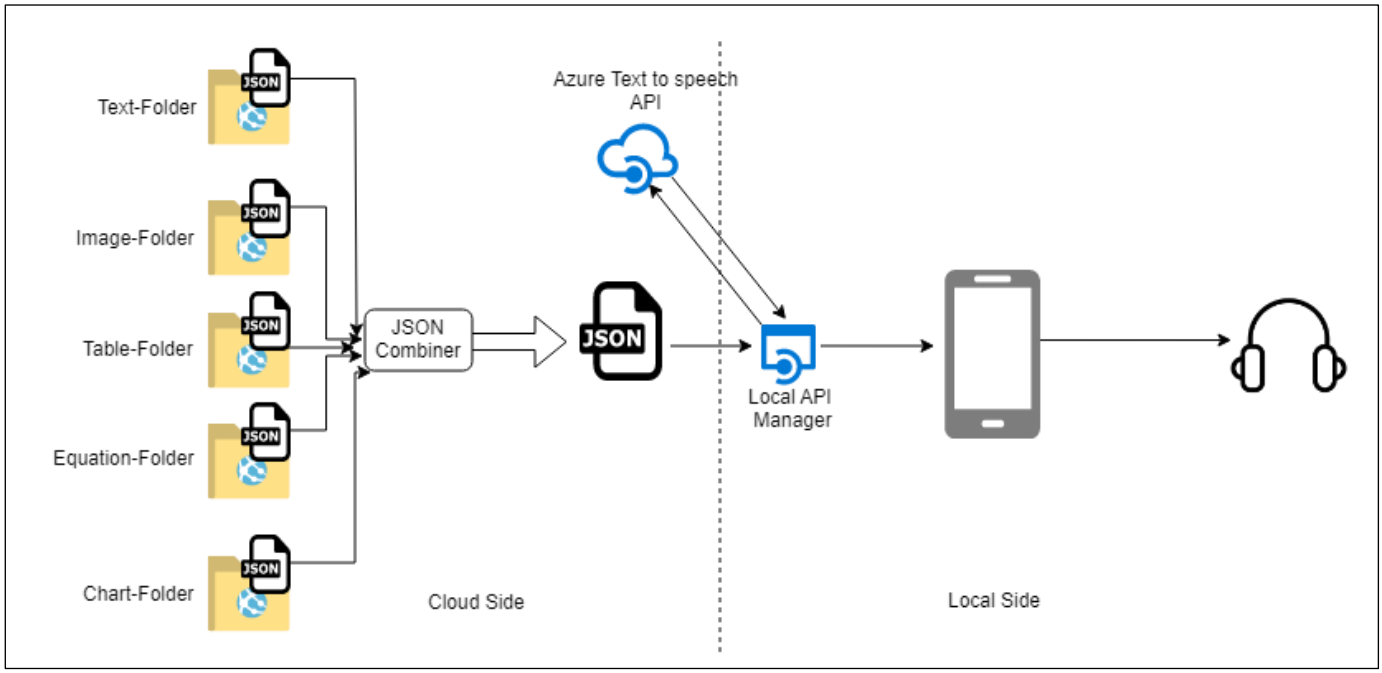

Figure 5: High Level Diagram for RI and CFAO Modules

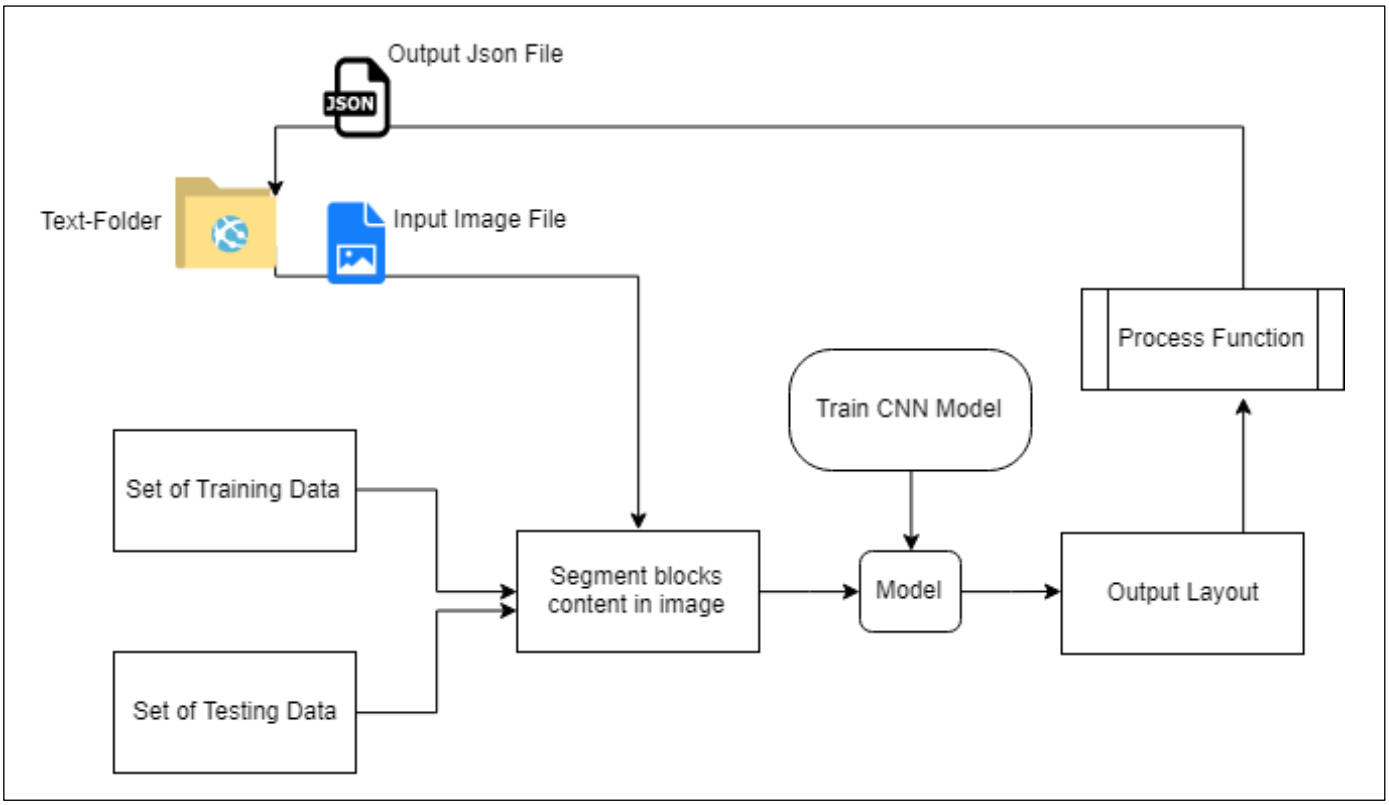

Figure 6: High Level Diagram for RTBC Module

2.2.2 Use case diagram for RI, RTBC and CFAO Modules

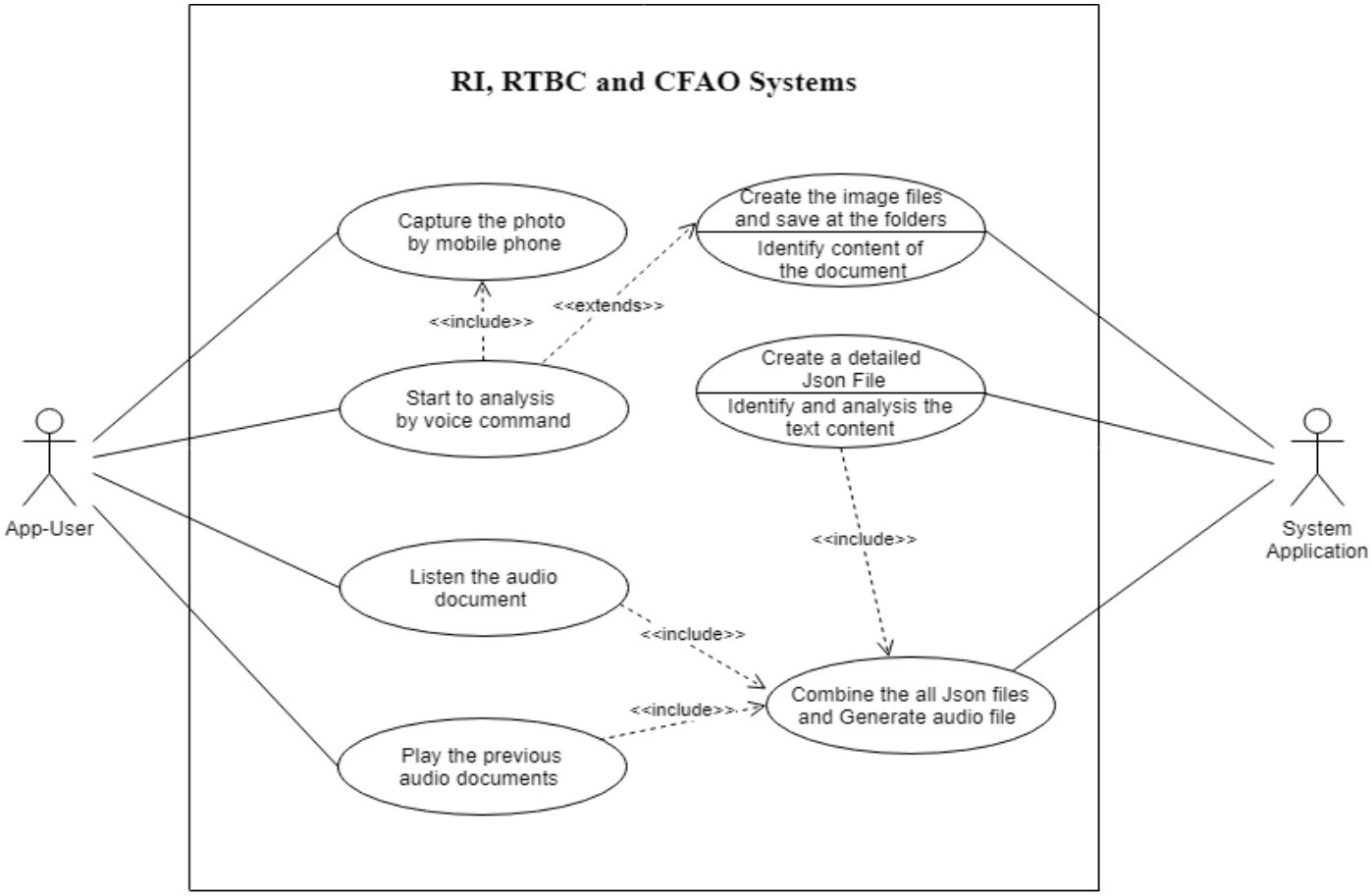

Figure 7: Use case diagram for RI, RTBC and CFAO Modules

2.2.3 Use case Scenarios for RI, RTBC and CFAO Modules

Table 4: Use case Scenario ID - RI_UCS_1

Use case ID	RI_UCS_1	
Use case name	Capture the photo	
Goal in context	Capture the photo on the smartphone by the application user	
Pre-condition	None	
Post-condition	Voice command by the application user	
Primary actor	Application user	
Secondary actor	None	
Main flow	**Step**	**Action**
	1	The use case begins when the Application user opens the application and automatically opened the camera.

	2	Auto detect the document and given commands to the user by the application. (Ex. Move the right, Move the left, Etc.)
	3	When the user touched the mobile phone screen, automatically capture the document image.
	4	If an image captures successfully, notify the user and showing analysis button.

Table 5: Use case Scenario ID - RI_UCS_2

Use case ID	RI_UCS_2	
Use case name	Voice input command	
Goal in context	Voice command by the application user to start analyze	
Pre-condition	"RI_UCS_1" was successes	
Post-condition	Upload the image to cloud	
Primary actor	Application user	
Secondary actor	System application	
Main flow	**Step**	**Action**
	1	Application waiting the user voice input.
	2	When user tell "Analysis Now" command application starts to light weight and upload to cloud.

Table 6: Use case Scenario ID - RI_UCS_3

Use case ID	RI_UCS_3
Use case name	Identify content of the document
Goal in context	Create the image files and save at the cloud folders for the individual analyzing purpose.
Pre-condition	"RI_UCS_2" was successes
Post-condition	Individual analyzing Text, Equations, Images, Tables, Charts concurrently.
Primary actor	System application

Secondary actor	None	
Main flow	**Step**	**Action**
	1	Identify content of the cloud uploaded digital image document.
	2	Create the new image files and save in the relevant cloud folders.
	3	Maintained the original document structure in array structure.

Table 7: Use case Scenario ID - RTBC_UCS_1

Use case ID	RTBC_UCS_1	
Use case name	Identify text contents of the document	
Goal in context	Analyze the texts contents and create a detailed a Json File	
Pre-condition	"RI_UCS_3" was successes	
Post-condition	Combine all Json files	
Primary actor	System application	
Secondary actor	None	
Main flow	**Step**	**Action**
	1	Identify text contents of the cloud digital image document.
	2	Analysis the identified texts.
	3	Create the detailed Json file and save in the "Text-folder".

Table 8: Use case Scenario ID - CFAO_UCS_1

Use case ID	CFAO_UCS_1
Use case name	Combine all Json files
Goal in context	Combine the all Json files and Generate audio file to be listing purpose
Pre-condition	Analyze the individual contents and create detailed Json Files
Post-condition	Listening audio document

Primary actor	System application	
Secondary actor	None	
Main flow	**Step**	**Action**
	1	Get Json files from the relevant cloud folders.
	2	Compare with document structure and arrange the Json file accordingly.

Table 9: Use case Scenario ID - CFAO_UCS_2

Use case ID	CFAO_UCS_2	
Use case name	Listening the document	
Goal in context	Listening the audio describe the document	
Pre-condition	"CFAO_UCS_1" complete successfully	
Post-condition	Save the audio file in the device storage	
Primary actor	Application user	
Secondary actor	System Application	
Main flow	**Step**	**Action**
	1	Get JSON file and pass to text to speech API.
	2	Play the audio file and save it in local device storage.

2.3 User Characteristics

Basically, the application user should have the ability to understand the English language and should need a proper listing and hearing ability without disabilities.

Users can be arrangeable of using a smartphone.

2.4 Constraints

- Proposed application modules will be provided in the English language.
- Since the application runs three real-time modules that use the camera, it requires more free memory to run effectively and might have slight latency depending on the device.
- The module requires a fairly adequate rear-camera.

- The CFAO module requires internet access once for every route entry to retrieve route information.

2.5 Assumptions and Dependencies

- The result of each component and any voice assisted applications on the device will not overlap.
- The back-end server will not crash or go down.
- The network connection is actively connected to internet.
- Mobile phone rear-camera is no any errors and photo capturing environment having good lightning conditions.

2.6 Apportioning of Requirements

The requirements mentioned in sections One and Two of this SRS documents are primary specifications. Requirements mentioned in section Three are referred to as requirements (or functional) specifications. The two levels of requirements are intended to be consistent.

The proposed product will be the outcome of a research project, therefore the procedure to attain the goals might differ. The major components and their outcomes mentioned here will not change. In spite of that, the technologies and the methodologies mentioned are likely to change in order to make the results more feasible and accurate.

3 Specific Requirements

3.1 External Interface Requirements

3.1.1 User Interfaces

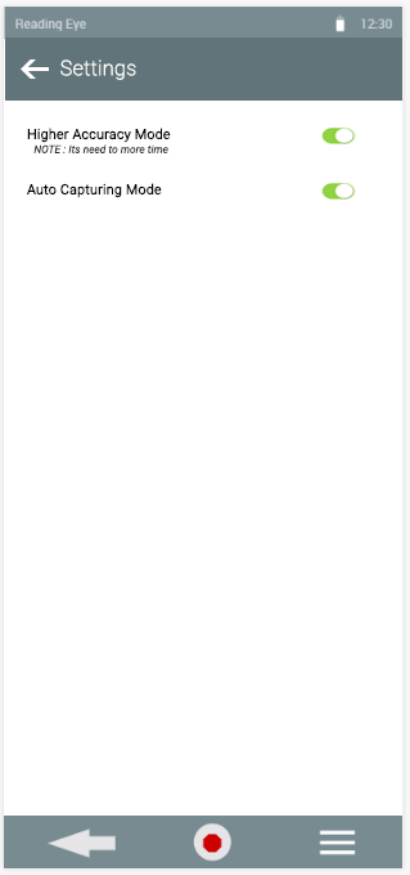

Figure 8: External User Interface 1

This application module having only one external user interface. Figure 8 mentions that the EUI.

If the “Auto Capturing Mode” enable means, not needed to capture photo to user. The document image will be captured by application automatically.

If the “Higher Accuracy Mode” enable means, the application always connects to the cloud server and always data transfer by the network. Therefore, application accuracy is higher. But the network delay should be adding to process time. Therefore, the document analyzing time is higher.

3.1.2 Hardware Interfaces

The mobile device's in-built real camera will be taken use of for the photo capturing in "Reading-Eye" application. Therefore, no any other advanced hardware interfaces are required.

3.1.3 Software Interfaces

The backend of the "Reading-Eye" application will be on in the Azure cloud space. The Training of the CNN model will use Python 3.6, Keras 2.0.8 and TensorFlow 2.0 which will run in Azure VMs.

3.1.4 Communication Interfaces

The software application requires an internet connection (mobile data/ Wi-Fi) to analysis the document image. Mainly used for Natural language processing.

3.2 Classes/Objects

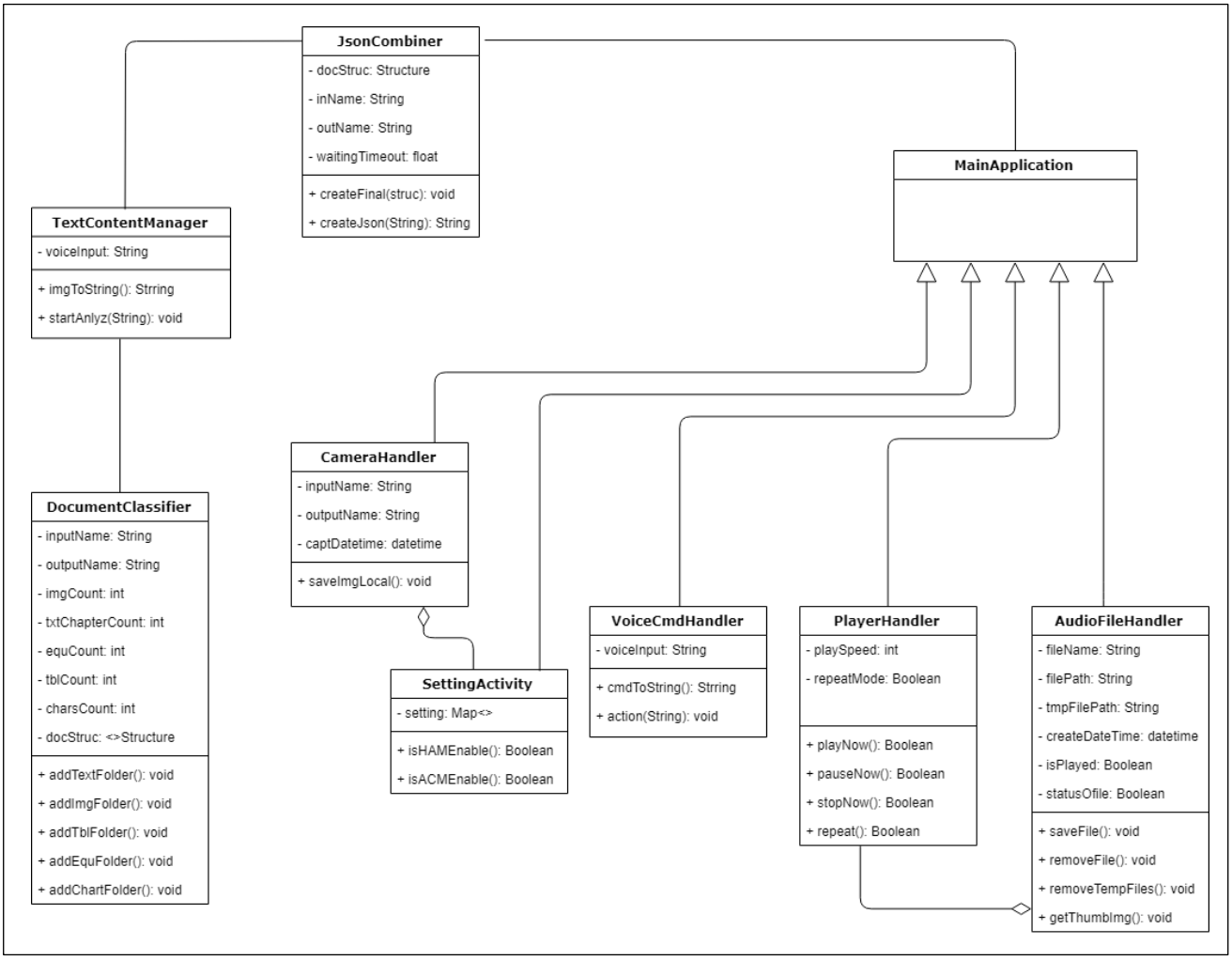

Figure 9: Class Diagram For modules

3.3 Performance Requirements

All three modules except for the “Reading-Eye” application, require higher computational power from the mobile device, but efficient ways of handling and reusing objects could minimize future bottlenecks in the Android and IOS operating systems at runtime.

Minimum requirements would include at least 1.5GB of free RAM, 1.4GHz CPU and 8MP primary camera for effective detection without any slack in performance. Higher internet connection required for analyzing progress.

3.4 Design Constraints

The “Reading-Eye” application mainly focus the vison impaired community. Therefore, UI is very simple and easy to use their community.

Moreover, the users should be given the ability to turn the module on or off. Except for these, there are no known limitations imposed by external factors on this module.

3.5 Software System Attributes

3.5.1 Reliability

Our system proposed to use the CNN model. The training model should be trained for large number of datasets and should process the result in a short period of time. When, providing result on appropriate time is equally important as providing an accurate result. No matter how accurate the result is if it is not delivered at the appropriate time. Through these strategies, the reliability of the system can be achieved.

3.5.2 Availability

The product should have a very nominal mean time between failures (MTBF), with 99.9999% uptime to serve the back-end services. The service back-end plays the vital role of providing with the detailed digital audio document to the android and IOS mobile devices, thus should have the very same MTBF.

3.5.3 Security

The proposed system having additional security mechanisms must be imposed on the dataset like encryption and data validation. Moreover, using cloud computing, having more security. Minimum server downtime and proposed system used server crash alert manger.

3.5.4 Maintainability

- Proposed system using microservices architecture. Therefore, a one of service has stopped but application works normally.
- The system may use logger services. It helps remotely errors and bug checkups.

3.6 Other Requirements

The requirements mentioned here will not be considered in the current release of the software product but rather will be implemented in later versions for better performance and feasibility. Therefore, no having other requirements.

4 Supporting Information

4.1 References

[1]"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: 08- Sep- 2019].

[2]"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: 08- Sep- 2019].

[3] "How to do Semantic Segmentation using Deep learning", AI & Machine Learning Blog,2019. [Online]. Available: https://nanonets.com/blog/how-to-dosemantic-segmentation-using-deep-learning/. [Accessed: 02- Sep - 2019].

[4]"Natural Language Processing is Fun!", Medium, 2018.[Online]. Available:https://medium.com/@ageitgey/natural-language-processing-is-fun-9a0bff37854e. [Accessed: 09- Sep - 2019].