AUTOMATED DOCTOR APPOINTMENT MANAGEMENT SYSTEM FOR MEDICAL DOMAIN IN SINHALA USER(ADAMS).

2020-175

Project Proposal Report

U.L.N.P. Uswatte

B.Sc. (Hons) Degree in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology
Sri Lanka

March 2020

DIALOGUE MANAGEMENT & TEXT TO SPEECH PROCESS

2020-175

Project Proposal Report

B.Sc. (Hons) Degree in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology
Sri Lanka

March 2020

DECLARATION

We declare that this is our own work and this proposal does not incorporate without acknowledgement of 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	IT Number	Signature	Date
U.L.N.P. Uswatte	IT17029278		25/02/2020

Supervisor

-- ----------------------------

Prof. Koliya Pulasinghe Date

Co-Supervisor

-- ----------------------------

Ms. Vijani Piyawardana Date

ABSTRACT

This paper proposes an approach on implementation of a Text to Speech system for Sinhala language using trained voice modules by the arrangement with Sinhala texts on RASA framework and the dialogue management mechanism. This is implemented as a sub objective of **A**utomated **D**octor **A**ppointment **M**anagement **S**ystem for medical domain in Sinhala user (**ADAMS**). In this project, a set of rules for mapping text to sound were identified and proceeded with unit selection mechanism. The medical term datasets used for this study were gathered by interviewing patients and doctors in channeling centers.

A diphone database is created and natural language is implementation of natural language processing modules. They are described here. [1]. This approach is aimed at developing the synthesis Sinhala voice to convert text to speech using a unit selection mechanism in RASA framework [2]. In Sri Lanka, majority of people are confident in interacting in Sinhala language which is known as morphologically rich language. Therefore, implementing a TTS system for Sinhala is a challenging task [2].

This document will continue with the background and literature review, research gap with implemented researchers, evaluation of the methodology criteria, project requirements declaration according to the Software Development Life Cycle and budget justification.

Keywords -: RASA, Text to Speech, Unit Selection, Phonetics, Sinhala TTS

LIST OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

A Text to Speech (TTS) can be recognized as the computer-based system which is capable of converting text to its desired spoken form, considering a grapheme to phoneme [1] mapping. Initially the query, which was generated from the Natural Language Process is used to create the Sinhala text according to the user inputs. Important medical facts which are in the query will be used for this process. This query has to be converted into Sinhala language. Simultaneously generated Sinhala text will be sent to the RASA NLU. Through this process dialogue will be generated using that Sinhala text.

User's medical question and the generated response answer will be displayed as a conversation of mobile interface. As the result of the Text to Speech process Sinhala response text will be converted into the voice command. Response medical answer will be output as a Sinhala voice command.

1.1 BACKGROUND & LITERATURE SURVEY

There are few researches already done in this research area. "Festival-si" is the one of research that done already in this research area. The other research is done using "MaryTTS" framework in Sinhala language. Below explanations are about these researches.

The conversion of text to speech involves many important processes. These processes can be divided mainly in to three stages; text analysis, linguistic analysis and wave-form generation [1]. The text analysis stage is responsible for converting the non-textual content into text. This stage also involves tokenization and normalization of the text; identification of words or chunks of text.

Text normalization establishes the correct interpretation of the input text by expanding the abbreviations and acronyms. This is done by replacing the nonalphabetic characters, numbers, and punctuation with appropriate text-strings depending on the context. The linguistic analysis stage involves finding the correct pronunciation of words, and assigning prosodic features (e.g. phrasing, intonation, stress) to the phonemic string to be spoken.

The final stage of a TTS system is waveform generation which involves the production of an acoustic digital signal using a particular synthesis approach such as formant synthesis, articulatory synthesis or waveform concatenation [6]. The text analysis and linguistic analysis stages together are known as the Natural Language Processing (NLP) component, while the waveform generation stage is known as the Digital Signal Processing (DSP) component of a TTS System.

According to this implementation and evaluation of a Sinhala text-to-speech system based on the diphone concatenation approach. The Festival framework [1] was chosen for implementing the Sinhala TTS system.
TTS systems have been developed using the Festival framework for different languages including English, Japanese, Welsh, Turkish, and Hindi, Telugu among others. However, no serious Sinhala speech synthesizer has been developed this far [1].

Sinhala Phonemic Inventory and Writing System

1. The Sinhala phonemic inventory

   Sinhala is one of the official languages of Sri Lanka and the mother tongue of the majority - 74% of its population. Spoken Sinhala contains 40 segmental Sinhala.
   phonemes; 14 vowels and 26 consonants as classified below in Table 1 and Table 2 [1].

	Front		Central		Back	
	Short	Long	Short	Long	Short	Long
High	i	i:			u	u:
Mid	e	e:	ə	ə:	o	o:
Low	æ	æ:			a	a:

Table 1. Spoken Sinhala vowed classification [1]

		Lab.	Den.	Alv.	Ret.	Pal.	Vel.	Glo.
Stops	Unvoiced	p	t̪		ʈ		k	
	Voiced	b	d̪		ɖ		g	
Affricates	Unvoiced					ʧ		
	Voiced					ʤ		
Pre-nasalized voiced stops		ᵐb	ⁿd̪		ⁿɖ		ᵑg	
Nasals		m		n		ɲ	ŋ	
Trill				r				
Lateral				l				
Fricatives	Unvoiced	f	s			ʃ		h
	Voiced	v						
Approximants						j		

Table 2. Spoken Sinhala constant classification [1]

There are four nasalized vowels occurring in two or three words in Sinhala. They are /a/, /a:/, /æ̃/ and /æ̃:/ [4]. Spoken Sinhala also has following Diphthongs; /ai/, /iu/, /eu/, /au/, /ou/, /au/, /ui/, /ei/, /oi/ and /ai/.

2. The Sihala writing System

The Sinhala character set has 18 vowels, and 42 consonants as shown in Table 3.

Vowels and corresponding vowel modifiers (within brackets): අ ආ(ා) ඇ(ැ) ඈ(ෑ) ඉ(ි) ඊ(ී) උ(ු) ඌ(ූ) ඍ(ෘ) ඎ(ෲ) ඏ(ෟ) ඐ(ෳ) එ(ෙ) ඒ(ේ) ඓ(ෛ) ඔ(ො) ඕ(ෝ) ඖ(ෞ)
Consonants: ක ඛ ග ඝ ඞ ඟ ච ඡ ජ ඣ ඤ ඦ ට ඨ ඩ ඪ ණ ඬ ත ථ ද ධ න ඳ ප ඵ බ භ ම ඹ ය ර ල ව ශ ෂ ස හ ළ ෆ ං ඃ
Special symbols: ු ෘ ් ඥ Inherent vowel remover (Hal marker): ්

Table 3. Sinhala Character set [1]

Sinhala characters are written left to right in horizontal lines. Words are delimited by a space in general. Vowels have corresponding full-character forms when they appear in an absolute initial position of a word. In other positions, they appear as ‘strokes’ and, are used with consonants to denote vowel modifiers.

3. Diphone Database Construction

Designing, recording, and labeling a complete diphone database is a laborious and a time-consuming task. The overall quality of the synthesized speech is entirely dependent on the quality of the diphone database. This section describes the methodology adopted in the construction of Sinhala diphone database [1].

4. Natural Language Processing Modules

When building a new voice using Festvox [1], templates of the natural language processing modules required by Festival are automatically generated as Scheme files. The NLP modules should be customized according to the language requirements [1].

The NLP modules should be customized according to the language requirements. Hence, the language specific scripts (phone, lexicon, tokenization) and speaker specific scripts (duration and intonation) can be externally configured and implemented without recompiling the system. The NLP related tasks involved when building a new voice are [1]:

- Defining the phone-set of the language
- Tokenization and text normalization
- Incorporation of letter-to-sound rules
- Incorporation of syllabification rules Assignment of stress patterns to the syllables in the word
- Phrase breaking
- Assignment of duration to phones
- Generation of f0 contour

A blockchain is a decentralized, distributed, immutable, shared & tamperproof data structure which is used to store a continuously growing list of the transaction. When consider Blockchain as a register containing transaction records into timestamp blocks. Each block has its own identity called cryptographic hash [1].

Each block is provided with the hash value of the block that came before it. Because of which a link is established between the blocks, thus creating a chain of blocks. We get a clear picture of how a Blockchain works only when we explore how a Blockchain network runs. It is a peer-peer network where each node holds the record of each transaction that's been carried out in the network.

To carry out transactions each node has its own wallet. The interaction between the user and the network is via a pair of private & public keys (Cryptographic keys). A private key is used to sign their own transaction whereas the public key is visible to all the nodes in the network [1].

Someone who wants to carry out a transaction should send a message by signing the transaction with their private key, when this is combined with the public key it forms a digital signature [1]. This transaction is broadcasted onto the Blockchain network where it is verified by the miners.

Miners are the nodes in the Blockchain with high processing power. Miners make the transaction unaltered & irreversible using a consensus algorithm called Proof of work. There is a competition among miners to generate a valid block and the one who generates a valid block is rewarded [1].

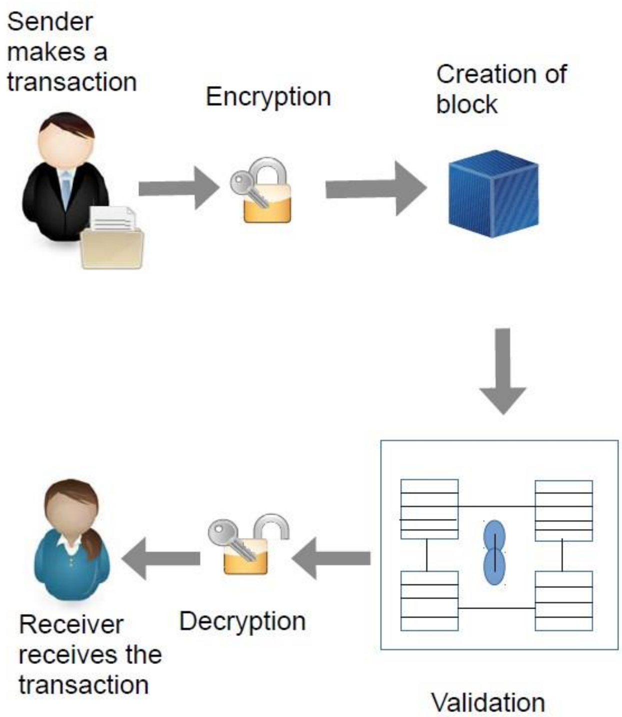

Figure 1. Working of Block-Chain Technology [3]

A block of the transaction is approved only when it is verified by all the miners in the network and if more than 50% of the miners validate the transaction then this block is considered as a valid block and is added to the longest Blockchain [3].

1.2 RESEARCH GAP

Using Festival framework [1] and MaryTTS framework [2] already done some of the features which are mentioned below. Those are features implemented in a desktop application in Sinhala language. Through ADAMS all the developments under mobile application are done using the RASA framework in medical domain.

None of these applications didn't able to develop the Sinhala terms dictionary to gather Sinhala medical terms knowledge. Also didn't able to display whole conversation on an interface.

Features	“Festival-si” [1] in Festival framework (desktop version)	MaryTTS [2] framework (desktop version)	RASA Framework on Flutter mobile application
Create Sinhala voice module	✔	✔	✔
Train Sinhala voice modules	✔	✔	✔
Sinhala medical facts dictionary	✘	✘	✔
Implement Sinhala text using medical terms knowledge	✘	✘	✔
Phoneme mapping	✔	✔	✔
Dialogue generation & display the conversation in mobile interface	✘	✘	✔
Sinhala Text to Speech conversion	✔	✔	✔
Sinhala voice output	✔	✔	✔
Used of Block-Chain mechanism to manage historical and updating data	✘	✘	✔

Table 4. Research gap between existing system and ADAMS

2. RESEARCH PROBLEM

Nowadays in Sri Lanka, all of the e-channelling systems are based on the English language. Mostly e-channelling systems are used by people in western province comparing to other provinces. The reason is the erudition that they have is very impecunious in the English language.

Sometimes, people having diseases such as Short-Term and Long-Term Incapacitation and people with dyslexia, are incapable of using an E-channelling system to communicate.

Most of the Sri Lankan e-channelling systems are web-predicated and utilizing the web-predicated system in mobile to minimize celerity, sometimes the browser support is very impuissant and not facile to utilize on a mobile. It is not user-friendly for a mobile.

Most of the patients don't know the specialization of the doctors, and who are the best doctors of their diseases. They only know the diseases that they have at that moment.

Some of the patients know doctor's name but they will not know the details about the doctor, such as which hospital the doctor is available, time schedules of the doctors.

Some patients know all the details in order to get an appointment with doctors, but they couldn't make it, because they will face interaction problems with a system, the system will perform in English then they can't understand it.

Patient have to pace on above difficulties in this kind of a challenging situation.

3. OBJECTIVE

3.1 MAIN OBJECTIVE

Dialogue management and the Text to Speech process. Have to implement the conversation between patient's medical question and the response medical answer.

3.2 SPECIFIC OBJECTIVES

Implementation of the Sinhala text

By gathering all the important medical facts of the user's input, query will be generated. Using that query it will have to implement the medical answer for the user's medical question. These important facts of the medical question are used to identify what is the problem that the user has. According to these medical facts, response medical answer will be generated through the system. At the end of this process system Sinhala text will be created.

Dialogue management

User's medical question and the response medical answer will be set as a conversation. Through this whole conversation between patient and the bot, it has to be managed as a dialogue.

Development of the Sinhala Voice Modules

Sinhala voice modules will be developed using Sinhala voice commands. Corresponding sentences were recorded by a professional speaker. By using those Sinhala medical terms Sinhala voice module will be developed.

Mapping between Sinhala Voice Modules and the Sinhala Phoneme [1] (Text to Speech conversion)

The conversion of text to speech involves many important processes. These processes can be divided mainly in to three stages: text analysis, linguistic analysis and wave-form generation [1].

The text analysis stage is responsible for converting the non-textual content into text. This stage also involves tokenization and normalization of the text, identification of words or chunks of text. Text normalization establishes the correct interpretation of the input text by expanding the abbreviations and acronyms. This is done by replacing the nonalphabetic characters, numbers, and punctuation with appropriate text-strings depending on the context [1].

The linguistic analysis stage involves finding the correct pronunciation of the words, and assigning prosodic features (e.g. phrasing, intonation, stress) to the phonemic string to be spoken [1].

The final stage of a TTS system is waveform generation which involves the production of an acoustic digital signal using a particular synthesis approach such as formant synthesis, articulatory synthesis or waveform concatenation [1].

Sinhala Voice output

At the end of the above three processes Sinhala voice output will be presented as a voice command. Sound waves will be here in smooth, quality and better speed that will be a healthy level to the humans.

Implementation of the Block-Chain related to the user registration functionality

The users who are involved with the system has to register to the system as a patient or as a medical agent. All the records related to the users are maintained through a block-chain mechanism. Through this mechanism user registration details and the historical data will be managed.

When consider Blockchain as a register containing transaction records into timestamp blocks. Each block has its own identity called cryptographic hash [3]. Each block is provided with the hash value of the block that comes before it. Because of which a link is established between the blocks, thus creating a chain of blocks [3].

4. METHODOLOGY

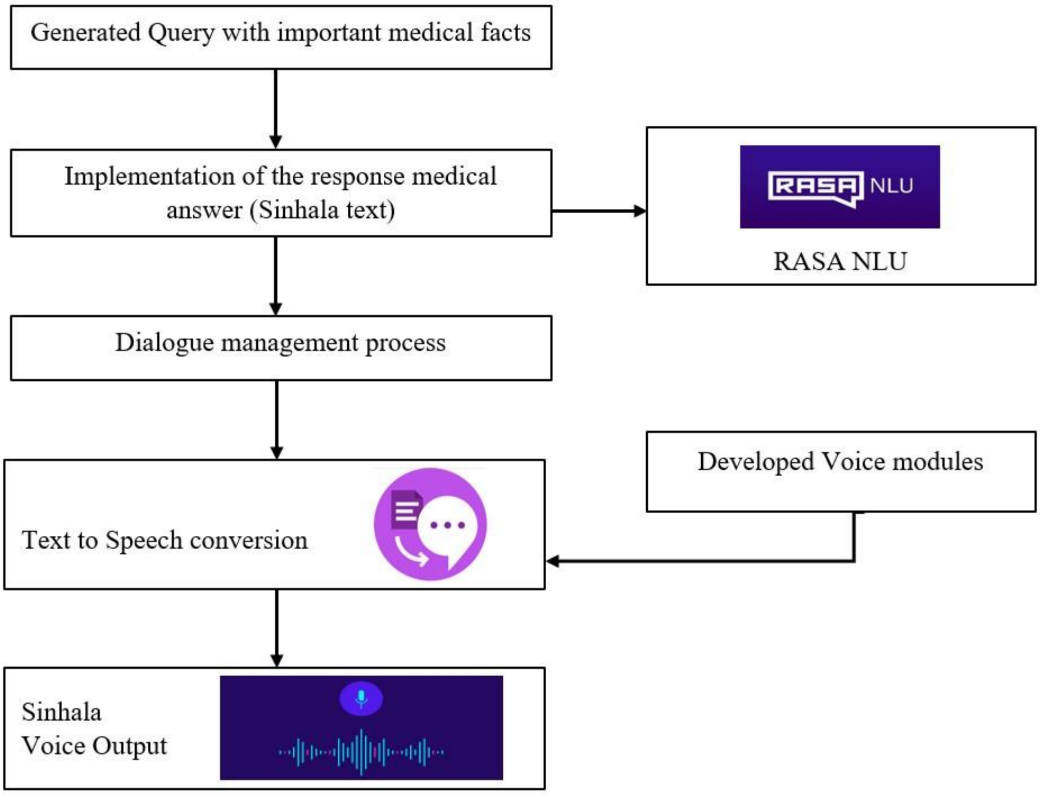

Figure 2. System Overview Diagram

The knowledge of the trained medical terms is used to generate the query. Using that query it has to implement the medical answer for the user's medical question. These important facts of the medical question use to identify the what is the problem that the user has. According to these medical facts, response medical answer will be generated through the system. At the end of this process system Sinhala text will be created.

User's medical question and the response medical answer will be set as a conversation. Through this whole conversation between patient and the bot, it has to be managed as a dialogue. To output Sinhala voice command there should be a Sinhala voice module which is developed already. There is mapping mechanism between trained voice modules and the generated Sinhala response text.

Sinhala voice modules will be developed using Sinhala voice commands. Corresponding sentences were recorded by a professional speaker. Recorded Sinhala words according to the medical domain is used to mapping mechanism. By using those Sinhala medical terms Sinhala voice module will be developed. There is a voice module train mechanism used to convert this whole text to speech process. The conversion of text to speech involves many important processes. These processes can be divided mainly in to three stages: text analysis, linguistic analysis and wave-form generation [1].

The text analysis stage is responsible for converting the non-textual content into text. This stage also involves tokenization and normalization of the text, identification of words or chunks of text. Text normalization establishes the correct interpretation of the input text by expanding the abbreviations and acronyms. This is done by replacing the nonalphabetic characters, numbers, and punctuation with appropriate text-strings depending on the context.

The linguistic analysis stage involves finding the correct pronunciation of the words, and assigning prosodic features (e.g. phrasing, intonation, stress) to the phonemic string to be spoken. The final stage of a TTS system is waveform generation which involves the production of an acoustic digital signal using a particular synthesis approach such as formant synthesis, articulatory synthesis or waveform concatenation [1].

At the end of the above three processes Sinhala voice output will be presented as a voice command. Sound waves will be heard in smooth, quality and better speed which is a healthy level to the humans.

5. PROJECT REQUIREMENTS

Require engineering is a process of establishing

- The services that a customer requires from a system and
- The constraints under which it operates and is developed.

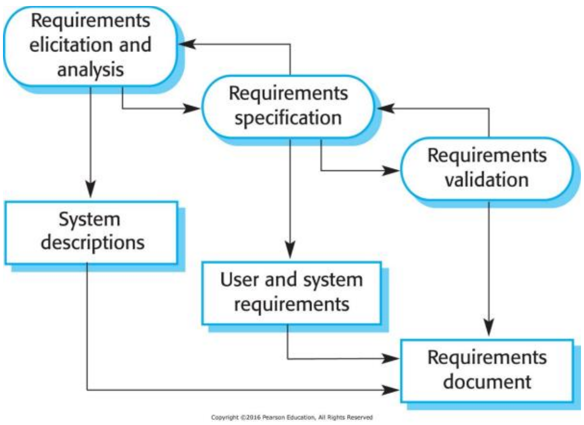

Figure 3. Requirements Engineering Process

Two types of Requirements Levels

1. User requirements
   - Platform to patient to e-channeling
   - Platform to dispensary/ hospital to register their system
   - Platform to doctor to register their system

2. System requirements

   A structured document setting out detailed descriptions of the system's functions, services and operational constraints

   System requirements are the configuration that a system must have in order for a hardware or software application to run smoothly and efficiently. Failure to meet these requirements can result in installation problems or performance problems.

   The former may prevent a device or application from getting installed, whereas the latter may cause a product to malfunction or perform below expectation or even to hang or crash.
   There are two types of System Requirements

   1. Functional requirements
   - ❖ Statements of services of the system should provide, how the system should react to inputs and how the system should behave in particular situations.
   - ❖ Describe functionality or system services
   - ❖ Depend on the type of software, expected users and the type of system where the software is used
   - ❖ Functional user requirements may be high-level statements of what the system should do
   - ❖ Functional system requirements should describe the system services in detail

     - Patient registration to the system
     - Patient assistant through the system
     - User interact with the system using Sinhala voice command
     - Doctor appointment for the patients
     - Patients can get the information about doctors, hospitals and channeling
     - Doctor channeling through the system
     - Medical service centers registration to their mobile application

   2. Non-functional requirements
   - ❖ Constraints on the services or functions offered by the system often apply to the system rather than individual features or services. These define system properties and constraints

❖ Non-functional requirements may seem to be more critical than functional requirements. If these are not met, the system may be useless.

- Automated conversational mobile application
- The quality of the Sinhala Voice commands and the performance like System response time, throughput, utilization, static, volumetric
- Reliability of the system data is the most important fact for both Patients and the Medical Service Centres
- System and its whole functionality (doctor channelling, information accessibility) should be available anytime for the patients.
- User's convenience when interact with this system's functionalities
- Total cost for the system implementation, developments and the use of the services
- How the system localization for Sinhala users with interaction of Sinhala voice commands

Property	Measure
Speed	Processed transactions/second User/event response time Screen refresh time
Size	Mbytes Number of ROM chips
Ease of use	Training time Number of help frames
Reliability	Mean time to failure Probability of unavailability Rate of failure occurrence Availability
Robustness	Time to restart after failure Percentage of events causing failure Probability of data corruption on failure
Portability	Percentage of target dependent statements Number of target systems

Table 5. Metrics for specifying non-functional requirements

6. WORK BREAKDOWN STRUCTURE

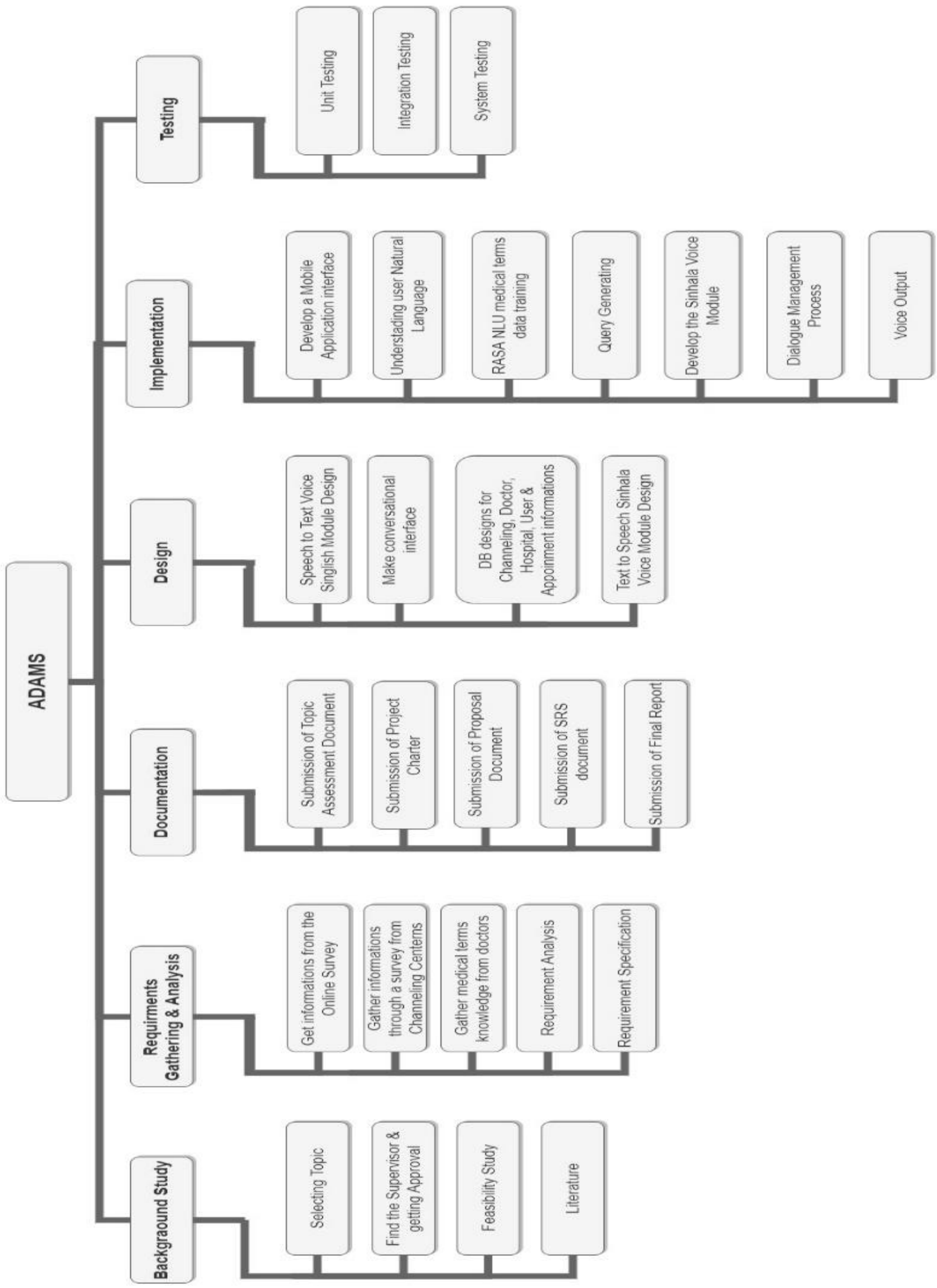

Figure 4. Work Breakdown Structure

7. DESCRIPTION OF PERSONAL AND FACILITIES

- Understanding user Natural Language
- Natural Language Process
- Dialogue Management
- Connecting an NLU into Rasa framework

This project would be done by a group of four members and the research and development workload of the project is being distributed among all the members. Detailed explanation of the assigned components has been discussed in the previous sections of the document.

Member	Component
D.D.S Rajapakshe	Understanding user Natural Language
Kudawithana K.N.B.	Natural Language Process
U.L.N.P. Uswatte	Dialogue Management
Nishshanka N.A.B. D	Connecting an NLU into Rasa framework

Table 6. Member details with main objective

8. BUDGET AND BUDGET JUSTIFICATION

This will be the budget that is going to be estimated in this project when implemented this functionality. In the Research Project is Firebase Realtime database is planned to be used. Because it’s a mobile compatible database. Firebase has two types of database planning. First one is Spark Plan and the second one is Blaze Plan. Spark Plan is a free plan, but it can’t customize. Then the best plan is Blaze Plan. Because it can customize as the developer wish.

Firebase Cloud Storage:
For the purpose of store Sinhala voice modules in cloud storage is the use of Firebase Cloud Storage. 5GB free data has already been provided for the cloud storage. Mainly in this research wants This cloud storage also provides a 5GB free data. This cloud storage also provides us a 5GB free data. Mainly store two voice modules then we want some high capacity storage. Mainly we want to store two voice modules then high capacity storage must be needed. Then 95 GB extra storage must be added. Then total storage capacity is 100 GB. Then $2.47must be paid additionally. 30 GB free storage for data transfer and 2,100,000 operations are provided additionally.

Budget		**Cost**
1. Firebase Cloud Storage		
1.1. Storage(100GB)	$2.47	$2.47
2. Stationaries (12 months)		
2.1. A4 sheets	$1.38	
2.2. Rough sheets	$0.55	
2.3. Binding	$1.38	
2.4. Pens & Pencils	$0.83	
2.5. Other Stuff	$1.10	$5.24
3. Communication (12 months)		
3.1. Internet service	$13.78	
3.2. Phone Cost	$16.54	$30.32
4. Printing Cost (12 months)		
4.1. Reports	$2.76	
Total Cost		**$38.03**

Table 7. Budget Plan

9. TIME PLAN

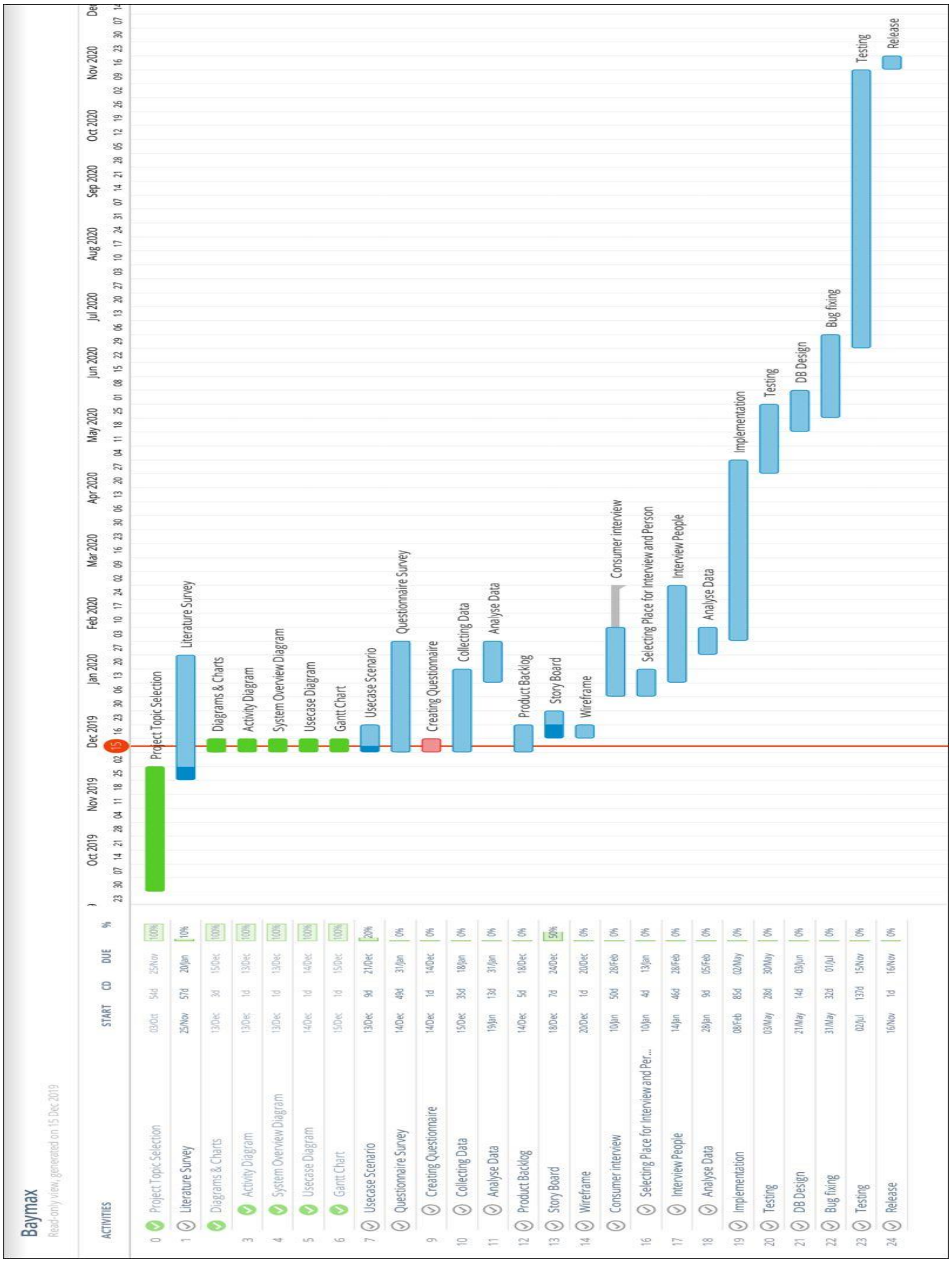

Figure 5 : Gantt Chart

(https://drive.google.com/open?id=1GXsOwinBr2liyusBCX8MbBiUnvq0dlom)

10. REFERENCES

[1] A. W. V. W. a. K. G. Ruvan Weerasinghe, "www.researchgate.net," University of Colombo School of Computing, 05 June 2014. [Online]. Available: https://www.researchgate.net/publication/221151931_Festival-si_A_Sinhala_Text-to-Speech_System. [Accessed December 2019].

[2] C. L. P. T. V. T. N. R. P. R. W. SinhalaLakshika Nanayakkara, “www.researchgate.net,” University of Colombo School of Computing, January 2018. [Online]. Available: https://www.researchgate.net/publication/328068273_A_Human_Quality_Text_to_Speech_System_for_Sinhala. [Accessed December 2019].

[3] D. S. A. S. C. B. S. R. Sudheera Palihakkara, “pdfs.semanticscholar.org,” University of Moratuwa, [Online]. Available: https://pdfs.semanticscholar.org/ec7d/90a15b08aee7d66d902e95c86824c2544a12.pdf?_ga=2.116270148.334219768.1580010100-1805892879.1580010. [Accessed January 2020].

[4] S. D. U. H. R. M. R. K. Harshini V M, “www.researchgate.net,” REVA University, Bangalore, India, April 2019. [Online]. Available: https://www.researchgate.net/publication/336439409_Health_Record_Management_through_Blockchain_Technology. [Accessed January 2020].

[5] S. T. T. S. S. E. L. Essay, “www.ukessays.com,” November 2018. [Online]. Available: https://www.ukessays.com/essays/english-language/sinhala-text-to-speech-stts-english-language-essay.php?vref=1. [Accessed January 2020].

[6] D. K. T. M. N. G. D. D. Wageesha Manamperi, “www.researchgate.net,” University of Moratuwa, May 2018. [Online]. Available: https://www.researchgate.net/publication/327517552_Sinhala_Speech_Recognition_for_Interactive_Voice_Response_Systems_Accessed_Through_Mobile_Phones. [Accessed January 2020].

[7] T. K. R. D. G. J. B. W. B. M. S. H. B. W T V L Gunarathne, “http://ir.kdu.ac.lk/,” General Sir John Kotelawala Defence University, 2015. [Online]. Available: http://ir.kdu.ac.lk/bitstream/handle/345/1667/001%281%29.pdf?sequence=1&isAllowed=y.

[8] Onilne, “www.academia.edu,” [Online]. Available: https://www.academia.edu/2818533/Sinhala_Text_to_Speech_System?auto=download. [Accessed January 2020].

[9] V. V. Sandra V A, “www.academia.edu,” Master of Computer Applications, College of Engineering, Trivandrum, Kerala, India, June 2019. [Online]. Available: https://www.academia.edu/40173703/IRJET-_Quinn_Medical_Assistant_for_Mental_Counseling_using_Rasa_Stack. [Accessed January 2020].