Voi-Web: A Mobile Tool Support For Browsing The World Wide Web By Voice And Text Messages

by

BABIRYE DIANA
Reg. No:2009/HD18/16929U
BIT (Mak)
Department of Networks
College of Computing and Information Systems, Makerere Univeristy
Email: dianababirye@yahoo.com,Tel: +256782953767

A Project Report Submitted to the School of Graduate Studies in Partial Fulfillment for the Award of Master of Science in Data Communications and Software Engineering Degree of Makerere University
(OPTION: Software Engineering)

April 1, 2013

Declaration

I Babirye Diana, declare that this report is my original work and has never been published or submitted for any other award to any university/institution of higher learning.

Signed:..... Date:.....

Babirye Diana

Approval

This project report has been submitted with my approval as supervisor

Signed: .....Date:.....

Dr. Benjamin Kanagwa, PhD

Department of Networks,

School of Computing and Information Technology,

Makerere University.

Dedication

This work is dedicated to the Lord Jesus Christ the author and perfector of our faith.

Acknowledgements

Special thanks goes to my supervisor Dr. Benjamin Kanagwa for every support accorded to me in completing this project.

I also thank in a special way the following angels in my life:
Dr.Macrine Nyisomeh, Pastor. Michael Nyisomeh, Mrs. Kasule Edith,Brother Joseph Besong, Dr. Elyanu Peter,
SMS Media Uganda, Mr. Atukunda Martin, Mr.Otandeka Simon Peter, Jjaja Eva Manjeri Nadiope, Ms. Margaret Ntege and Ms. Nankya May

May the Lord bless you all abundantly.

Contents

Abstract

The World Wide Web (WWW) is well established as one of the authoritative sources of information on many aspects. However, a large fraction of people based in developing countries like Uganda are not able to access this internet service due to economic and infrastructural constraints.

In this report we present an alternative tool called Voi-Web that avails the World Wide Web information as Voice through the Mobile Phone. The initial results enable a user to search the World Wide Web using voice or Short Message Service(SMS) through the mobile phone .

The user is required to speak his or her search query to the tool through a mobile phone connected to a telecomunicaion network. The speech is transcribed into text and the text is used to search the web through a Google API. The text results of the search query are then speach synthesised into an audio file that is relayed back to the mobile phone user through a mobile telecommunication network.

The component tools that have been used to realize this functionality were carefully selected to be free and open source and they include; Asterisk a VOIP server, Kannel an SMS Gateway Software, Espeak a text to speech tool, Google Speech recognition API and Google Search API Interface for the World Wide Web.

The tool was tested with 30 ugandans with no access to internet services and an accuracy of 58% was achieved by the tool from the spoken queries.

This tool is presenting a unique alternative of presenting WWW information to Uganda's mobile phone users.

Chapter 1

Introduction

1.1 Background

Internet is an essential human need as convened by the United Nations[27]. The World Wide Web(WWW) that runs on the internet is one of the most important tools for information [9]. Web information is important mainly because it provides information on many subjects that essential to the welfare of a ugandan. These subjects include; agriculture, education entertainment etc.

Unfortunately, nearly 30 million Ugandans do not have access to Web Information through the internet due to economic and infrastructural constraints[**?**].

Basing on the fact that the market penetration for voice in Uganda is at 33.5%[6] and total number of mobile phone subscribers was at 14 million in 2011[**?**], we propose the use of voice to extend this Web Information Service to Ugandans whose access to these services has been limited as stated above.

We introduce Voi-Web; A Mobile Tool Support that provides for access to Web Information through the Voice and SMS functionality of the Mobile Telecomunications Infrastructure.

The advantages of the Voi-Web mobile tool include ; it is easy to use because it reflects basic telephone usage, it is as accessible as the mobile telecommunication network, it can be utilized by the semi illiterate, those who prefer to consume voice information instead of text and the blind.

1.2 Definition of Terms

Below is brief description of the techinical terminologies that we have used in the description of this tool.

1.2.1 Web Information

This is information that is presented through the World Wide Web(WWW).

1.2.2 Voice Search

Voice Search allows the user to use voice to search the WWW [23].

The Voi-Web Mobile Tool provides a search functionality of information through voice as an option by capturing one's Voice Query(VQ) through an Interactive Voice Response(IVR) System.

1.2.3 SMS Search

The SMS Search functionality allows the user to use SMS to search the Internet. SMS communication is the exchange of written text messages between fixed-line phone or mobile phone devices over a network [26]. The search of information through SMS is another option provided for by the Voi-Web Mobile Tool.

1.2.4 Speech Synthesis

Speech synthesis is the artificial production of human speech[25]. The Information obtained is converted into speech through the Speech Synthesis process and later relayed to the mobile user through the IVR System.

1.2.5 IVR System

An IVR system is a server application for making, receiving and performing custom processing of phone calls.

1.3 Statement of Problem

Web Information is an essential need for the progress of every Ugandan[27]. However, more than one half Ugandan's do not have access to this rich information due to high costs and poor infrastructure[**?**] and yet over 14 million ugandans are subscribed to Mobile networks.

1.4 Objectives

1.4.1 General Objective

To extend Web Information Services to Ugandans by providing them with a tool that provides access to Web Information through the Voice and SMS functionality of the Mobile Telecomunications Infrastructure.

subsectionSpecific Objectives

(1) To write a Software Requirements Specification for the VMT.

(2) To write Software Design Descriptions for the VMT

(3) To develop and implement a functional VMT system.

(4) To test and analyze the results from the VMT system.

1.5 Significance and Justification

This tool is important,as it provides access to Web Information services to Uganda's Mobile Phone users by utilizing the Voice and SMS functionality of the Mobile Telecomunications Infrastructure.

This tool is significant mainly because it is ;

- It is accessible since its availability is dependant on the proven reliable mobile telecommunication network that has spread in the whole nation. - It is usable since it reflects basic telephone usage behaviour - It is also convenient for the mobile nature of mobile phone users since it presents the information as voice.

1.6 Scope

The Voi-Web Mobile tool has focused on providing accessibility of Web Information Uganda's constrained Mobile Phone Users.

We are considering the english language voice queries , a target audience of; low income earners with no access to internet, ugandans from areas with no internet infrastructure , those who prefer to consume voice information and the visualy impaired.

1.7 Key Contributions

This Voi-Web Mobile Tool introduces a unique approach to provision of access to Web Information as voice through the Mobile Phone.

The Voi-Web Mobile Tool also caters for the mobile nature of mobile users by providing the option to search by voice and text through the Mobile Phone.

The key deliverables include;

Requirements specifications for the VMT system.

System Design Specifications for the VMT system

A checklist of the selected tools for developing the VMT system

A functional VMT system.

Chapter 2

Literature Review

2.1 Mobile Applications Presenting Web Information

Web Information is accessed through various ways using different gadgets. This project aims at presenting Web Information as Voice through an IVR System based on the SMS or Voice queries received from Mobile Phone users.Various applications with similar objectives and technology setup have been explored and below are the different arguments with each application.

2.1.1 Voice Enabled Web Application

This system was developed by William M. Wolfe et al to generate and execute Voice enabled XML Web applications [10]. This system retrieves text based forms and acts on them via telephone. However it neither receives Voice queries nor provides for a search functionality of information on the WWW. We argue that provision of Web Information as voice and availing the WWW search functionality through the Mobile Phone are vital tools in the promotion of accessibility of Web Information to Uganda's Mobile Phone users.

Accessing of Web Information also has some challenges that are yet to be addressed. Some of these challenges are that much of the Web content created in the developing world doesnt follow W3C's Web Content Accessibility Guidelines. They also have a greater propensity to use images and animation which adds complexity to accessing online material as Voice Data [19].

2.1.2 Apples Voice Over Tool

VoiceOver is a built-in screen reader that describes aloud what appears on your computer screen: it speaks the text that is in documents and windows[21]. Apples VoiceOver Tool also facilitates browsing of Web Pages where one can access the scope and structure of the Web page and also navigate the Web Page Based on the summary of Web Statistics presented through Voice[21]. However, the VoiceOver tool cannot be utilized without an apple brand device, and it also requires that one is connected to the internet to browse for information . Both of these; the apple brand device and internet are not affordable and accessible to the economicaly constrained ugandans that donot have access to the internet.

2.1.3 Korea's Voiceyes PC mate

This is as usable as audio book reader and also a real time TTS reader and scanner. This device uses 2.D barcodes to allow a large amount of text information to be compressed into a single barcode, which can be read into a device and relayed as audio data[19]. Newspapers in Korea have already started providing barcodes for accessible information in these formats, and this is now a national standard for the visually-impaired. It can be noted, that this solution is practical as regards to transfering of Web Information as its compressed and the data can be sent through an SMS. Nevertheless one would require a smart Mobile Phone and some expertise to retrieve this information in audio format. Again this limits the economically constrained Ugandans.

2.1.4 The Mobile Augmented Messaging application

This application was developed to test the concept of sharing audio messages on a mobile platform [1]. The application allows non-real-time, asynchronous communication using audio messages, which are shared among members of groups. Audio messages are recorded using the microphones of the ARA headset and are then stored on a database server.

For situations where the recording of audio messages is not possible or desirable, messages can be input as text and converted to audio using text-to-speech synthesis [1]. This application used eSpeak speech synthesizer to implement speech synthesis. One of the reason it was selected is because it

offers the possibility to modify synthesis parameters, such as speed, volume, pitch and pitch range [1].

The Voi-Web Mobile Tool has also utilized eSpeak speech synthesizer to implement speech sysnthesis but is mainly focusing on presenting Web Information and not chat messages .

2.1.5 Google Mobile App

In November 2008 Google introduced Google Mobile App(GMA) for iPhone that included a search by voice feature . GMA search by voice extended the paradigm of multi-modal voice search from searching for business on maps to searching the entire world wide web. A multi-modal experience has some distinct advantages compared to the IVR (voice-only) system. First, the output modality can be visual rather than spoken, allowing much richer information flow [8].

The Voi-Web Mobile Tool will also search the entire WWW based on the voice or text queries received but will mainly utilize the traditional IVR systems. This is with reference to the fact that visual out-put is good but does not solve the problem of the non technical people who cannot operate their mobile phones or the visually impaired, therefore not being able to deal with the problem adequately.

An examination of the category distribution by input method of the query was made and the following salient differences were shown [8].

(1) Voice searches are more likely to be about an on-the-go topic: Mobile queries and voice searches in particular have a much greater emphasis on categories such as food, drink and local businesses.

(2) Voice searches are less likely to be about a potentially sen- sitive subject: Categories that consist of sensitive content (adult themes, social networking, and health) are avoided by voice search users, relatively speaking. This may be because they wish to preserve their privacy in a public setting.

(3) Voice searches are less likely to be for a website that requires significant interaction: Voice searches are rarely about the sort of topics that require significant interaction following the search, such as games and social networking.

Another observation made was that short queries, in particular one and two word queries, are relatively more frequent in voice searches than in typed searches [8]. The factors that influence whether a user continues to search by voice were measured and found that recognition accuracy is the most important factor [8].

Google's mobile voice features begin with Speak now so Users know that they should now start to speak [8].

Based on this knowledge the Voi-Web Mobile tool has also utilized the Speak now feature so that user can know when they can start to speak.

2.1.6 IVR Systems

The Voi-Web Mobile Tool will relay search results and information through an IVR system . We have therefore explored other IVR systems to review their adaption and implementation so as to learn from the previous work done .

Some of the systems we explored include ; IVR System for population based infectious disease reporting : The authors of this work aimed to evaluate the web and Interactive Voice Response phone service . Fourteen thousand subjects were randomly selected from the Swedish population to register all respiratory tract infections.

Participants were classified as belonging to the Web or IVR group based on their choice of technology for initial registration . In all 1,297 individuals registered via IVR while 2044 choose the Web [2].

Based on this adaption analysis we argue that there is a particular niche in society that can be satisfied by IVR systems considering that Sweden is more developed than Uganda but over one thousand two hundred individuals chose IVR.

2.1.7 Google SMS

In September 2009 Google introduced Google SMS for Mobile Phone Users on the MTN network in Uganda.Google SMS provides searching for Information on the WWW or other sources through SMS for topics such as health , sports,farming etc [12]. This solution is practical as it provides access to WWW information. It is however limited by the number of characters that

	Presents Web Information As Voice	Relayed Through An IVR System	Supports SMS Queries	Supports Voice Queries
Voice Enabled Web Application	Yes	Yes	No	No
Apples Voice Over Tool	Yes	No	No	No
Korea's Voiceyes PC mate	Yes	No	No	No
The Mobile Augmented Messaging Application	No	No	No	Yes
Google Mobile App	Yes	No	Yes	Yes
IVR Systems	No	Yes	No	No
Google SMS	No	No	Yes	No

can be carried in an SMS . The Voi-Web Mobile tool in this case is presenting a better solution as voice relay through the Mobile Phone is not limited by number of characters.

2.2 Automatic Speech Recognition

Automatic Speech Recognition (ASR) is a task of finding the most likely set of words for a given acoustic signal. There are three key models in ASR, namely, pronunciation models, acoustic models and language models. In the past three decades, there has been tremendous technology progresses towards higher ASR accuracy, speed, and robustness [20].

An acoustic model (AM) is created by taking audio recordings of speech and their transcription as input and compiling them into a statistical representations of the sounds that make up each word [20].

State-of-the-art ASR systems use Hidden Markov Models (HMMs) for the acoustic model. The challenge is the presence of ambient noise in mobile environment significantly impacts ASR accuracy [20].

Language models(LM) represent the probability of the sequence of words. LM in voice search often have very high perplexity and large vocabulary sizes [20]. Pronunciation models(PM) translate words into phonemes. Most large vocabulary speech recognition systems make use of a dictionary with multiple possible pronunciation variants per word, which are either provided by trained developers or generated from letter-to-phoneme (L2P) rules [20].

Since speech recognition is an error-prone process and Voice search is essentially an integration of automatic speech recognition (ASR) and text or database search. Researchers have proposed various approaches to tightly coupling the two components for better overall system performance [20].

The Voi-Web Mobile Tool has tightly coupled Voice and SMS for better overall system performance.

2.3 Web Search and Presentation of Results

2.3.1 Web Search Queries

There are four broad categories of most web search queries[28]. These categories include;

(1) Informational : These queries cover broad topics [28].

(2) Navigational: These queries seek a specific website or webpage[28].

(3) Transactional: These are related to purchase of a given service or product[28].

(4) Connectivity: These report on the connectivity of the indexed web-graphs [28].

The Voi-Web Mobile Tool concetrates Informational search queries to achieve the goal of provinding access to information to non privildeged Ugandans.

The findings in Information Processing and Management article showed that "more than 80 percent of Web queries are informational in nature, with about 10 percent being navigational and transactional" [3].

From the above findings, the Voi-web is able to facilitate 80 % of Web search needs.

Mobile Phones are able to support various queries which include [29];

(1) Textbased keyword queries: The GoogleSMS Mobile Application takes text messages as keyword queries to a local mobile operater shortcode and responds with details by short message [12].

(2) Image queries: Images can be searched through recognizing characters or codes that are present in the images and searching the images directly by their content known as content based image retrieval(CBIR) [29].

(3) Spoken queries: Some of the audio search products that aim to provide Web Information that is informational search and directory information[29] include PromptU [18] which allows users to find content with spoken content and Google's GOOG411[11] that receives a spoken query and transfers the call to the business free of charge or provide the details by short message[29].

The Voi-Web Mobile supports Textbased and Spoken queries.

2.3.2 Presentation of Search Results

Search results from Web searches of the different queries on the Mobile Phones have been presented invarious formats as discussed below;

Text Messages: The GoogleSMS Mobile Application presents search results as text messages through the Mobile Phone. [12].

Audio Format: The GOOG411 product also presents the results from the search as audio when it forwards the caller to a business directory [29].

Images Format: As image queries are searched from the Web some of the information returned are image [29]. Information in form of images can also be returned with text and audio queries.

Video Format:Information is also presented in video format. Video information can be returned with text queries as it is with Youtube [30].

2.4 Conclusion

As we have subsquently argued, the tools available are practical solutions for different markets but not necessarily for our constrained Ugandans hence the need for the Voi-Web Mobile Tool.

Chapter 3

Methodology

3.1 Introduction

As stated in the previous sections, the VMT solution has been implemented to improve accessibility of Web Information to the ugandans.

Below is a set of principles,guidelines and procedures that have been applied in the implementation and analysis of the Voi-Web product deliverables i.e the Software requirements specification, Software design specification , VMT functional application and the Analysis of the test results.

3.2 VMT Software Requirements Specification

The VMT requirements have been modeled using an object-oriented approach and the layout of the SRS has been structured to conform to the 830-1998 - IEEE Recommended Practice for Software Requirements Specifications.

The IEEE practise has been chosen because of it describes the content and qualities of a good software requirements spe cification.

Some of the characteristics of a good SRS as recommened by the IEEE practise include;

[label=]Correct Unambiguous Complete Consistent Verifiable Modifiable Traceable

3.3 VMT Software Design Descriptions

The layout of the VMT SDD has been structured to conform to the 1016-1998 - IEEE Recommended Practice for Software Design Descriptions .

The IEEE practise has been chosen because it specifies the necessary information content, and recommended content for a Software Design Description(SDD).

Some the SDD IEEE practise recommendations include considering the following for producing a SDD;

3.3.1 Software Lifecyle

The lifecycle approach is effective and provides a model for a context within which to discuss the use of the SDD.

3.3.2 SDD Within The Lifecycle

For all software systems , it is important to ensure that the design and implementation satisfy the requirements of that system.

3.3.3 Purpose of An SDD

The SDD is the translation of requirements into a description of the software structure, software components, interfaces and data necessary for implementation.

3.4 Voi-Web Mobile Tool Software

The Voi-Web Mobile Tool(VMT) development has been broken down into the following interfaces described below;

3.4.1 VMT Speech Recognition Interface

The Speech Recognition Interface receives Voice Information from the IVR Interface, converts this information into text and then submits the text to the Search API .

3.4.2 VMT SMS Receiver Interface

The SMS Receiver Interface receives SMS queries through the telecommunication SMS gateway and submits the SMS to the search API.

3.4.3 VMT Reporting Web Interface

The Voi-Web Reporting Web Interface keeps track of the inforamtion flow from one module to another and reports the VMT usage statitics.

3.4.4 VMT Search API

The Search API takes text queries from both SMS and transcribed Voice queries and returns the relevant url with related information. The information from this url is retrieved and cleaned and converted to a file which is delivered to the speak Speech Synthesis module.

3.4.5 VMT Speech Sysnthesis Interface

The Speech Synthesis Interface receives a file with the Web Information searched and produces a gsm file. The gsm file is relayed through the IVR Interface to the Mobile Phone user.

3.4.6 VMT IVR Interface

The IVR Interface receives Voice queries from Mobile Phone Users through the Telecom networks infrastructure and submits them to the Speech Recognition Interface.

The IVR Interface also receives Audio gsm files from the Speach Synthesis Interface with search results and relays this information back to the Mobile Phone user.

3.4.7 VMT Implementation Environment

The VMT implementation environment consits of the following;

(1) Operating System: Linux Open Suse 12.1

(2) Web Server: Apache2 http/http 2.1

(3) Telephone Card

(4) Telecom Network SMS Gateway connection

(5) Telecome Network Voice Gateway connection

(6) Internet Connectivity

3.5 The VMT Results Analysis

Thirty low income Mobile Phone users from different regions in the country were indentified in the Slums of Mengo-Kisenyi a Sububurn area in Kampala City.

Each of these mobile phone users was recorded saying one of the most searched information queries in Uganda according to google's report for 2012. [**?**]. These queries include;

(1) What is love

(2) What is technology.

(3) What is planning?

(4) What is ethics?

(5) What is language?

These recordings have been taken through the Voi-Web tool and the results have been discussed in the implementation section.

The resultant data has been analysed with the confusion matrix tool[**?**].

The confusion matrix tool was choosen because it provides for examination of performance classifiers.

3.6 Conclusion

The usage of each of the techiniques discussed above has been explained further in the implementation section.

Chapter 4

Implemenation

4.1 Introduction

This chapter describes the implementation process of the Voi-Web Mobile Tool(VMT) and the analysis of results from the tests carried out with the tool.

4.2 VMT Requirements and Design

4.2.1 Purpose

The purpose of this section is to specify the VMT software requirement specifications,it outlines a brief description about what functions that are to be performed; what data to is to be captured; and what outputs/results the system will produce.

4.2.2 Product Functions

PF-1: The VMT supports reception of SMS queries from Mobile Phone users.

PF-2: The VMT supports reception of Voice queries from Mobile Phone users.

PF-3: The VMT supports searching of Web Information on the WWW.

PF-4: The VMT supports converting of the Web Information to audio files.

PF-5: The VMT module relays the audio information of the Web Information back to the Mobile Phone user.

PF-6: The VMT supports an Interface to manage User profile information adding, deleting, editing and classification.

PF-7: The VMT supports reporting of the VMT Mobile Phone users' usage statistics.

4.2.3 User Classes and Characteristics

UC-1: Administrators

They are responsible for adding, editing and classifying the users of the system.

They are also responsible for view the Mobile Phone users' usage statistics of the VMT.

Users under this category must have sufficient knowledge about computers and basic system management.

UC-2: End Users (low privileged users)

These are the end users that utilize the VMT to search for Web Information.

This category of users must attain a basic level of Mobile Phone usage i.e making a call and sending an SMS.

4.2.4 Operating Environment

The system will operate in a Client - Server model;

OE-1: The VMT runs on Linux Operating systems with their supported hardware platforms.

OE-2: The Server hosting the VMT is publicly accessible over the Internet.

OE-3: The VMT operates on a back-end relational database management system.

OE-4: The VMT operates with connectivity to the Mobile networks.

Hardware Interfaces

HI-1: Server has the follow specifications

320 GB Hard Disk Space

Intel icore5 Processor

6GB RAM

4.2.5 Use Case Activity Diagrams

Web Information search by SMS Activity Diagram

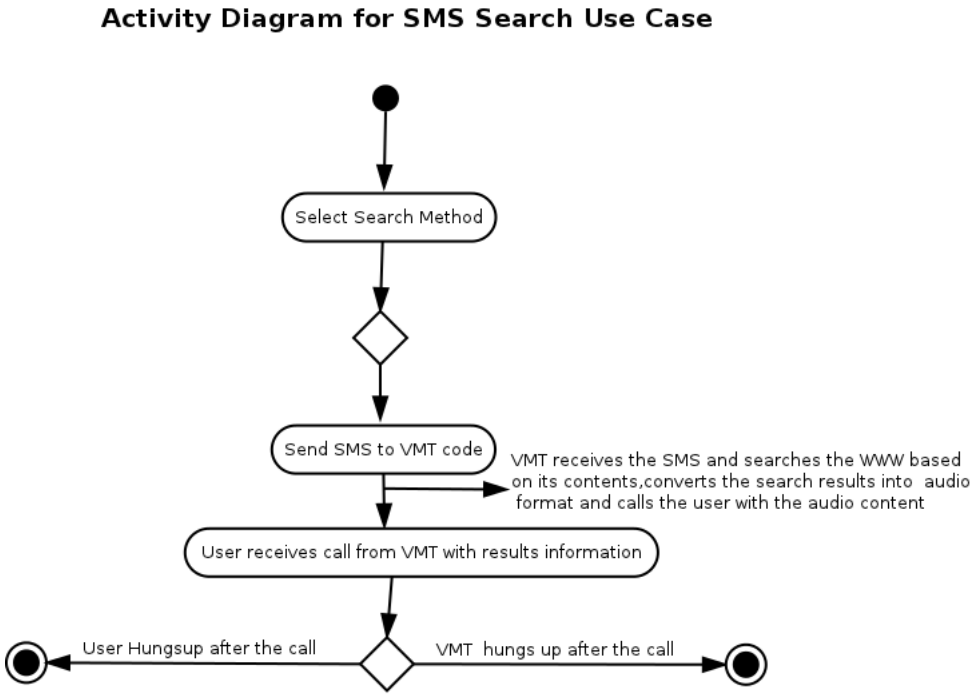


Figure 3: SMS Search Activity Diagram

Web Information search by Voice Call Activity Diagram

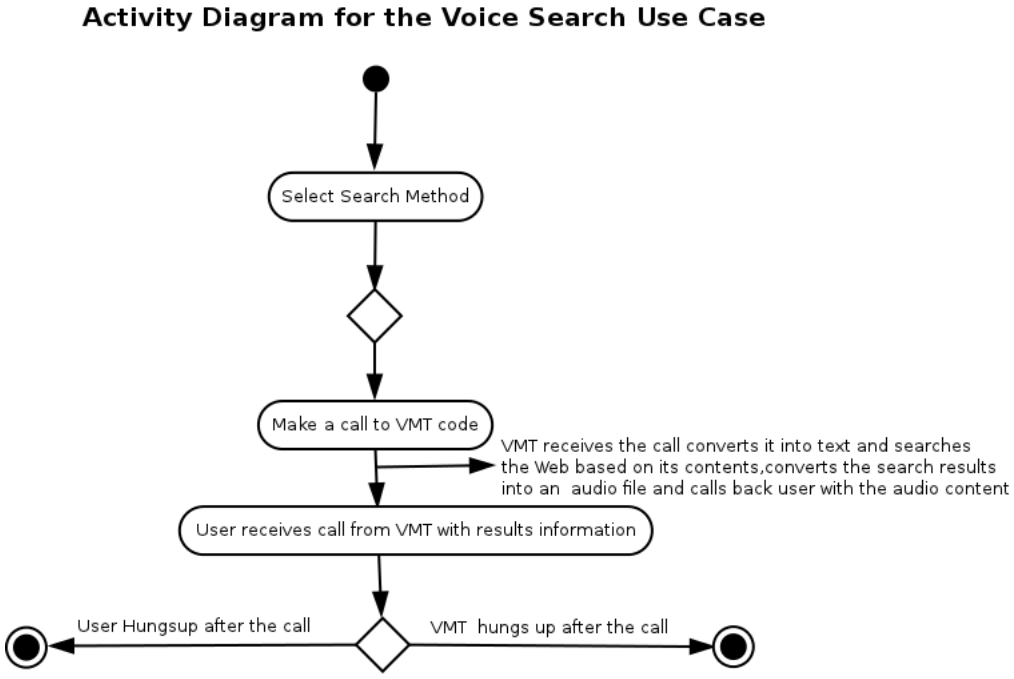


Figure 4: Voice Search Activity Diagram

Login Activity Diagram

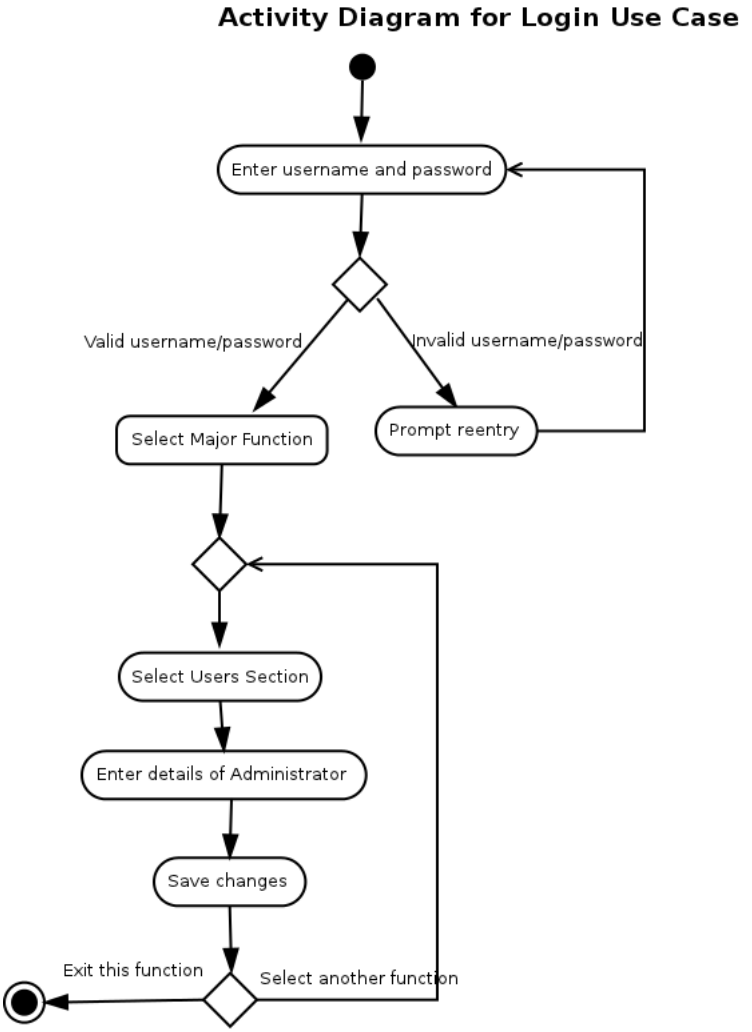


Figure 1: Login Use Case Activity Diagram

Usage Reports Activity Diagram

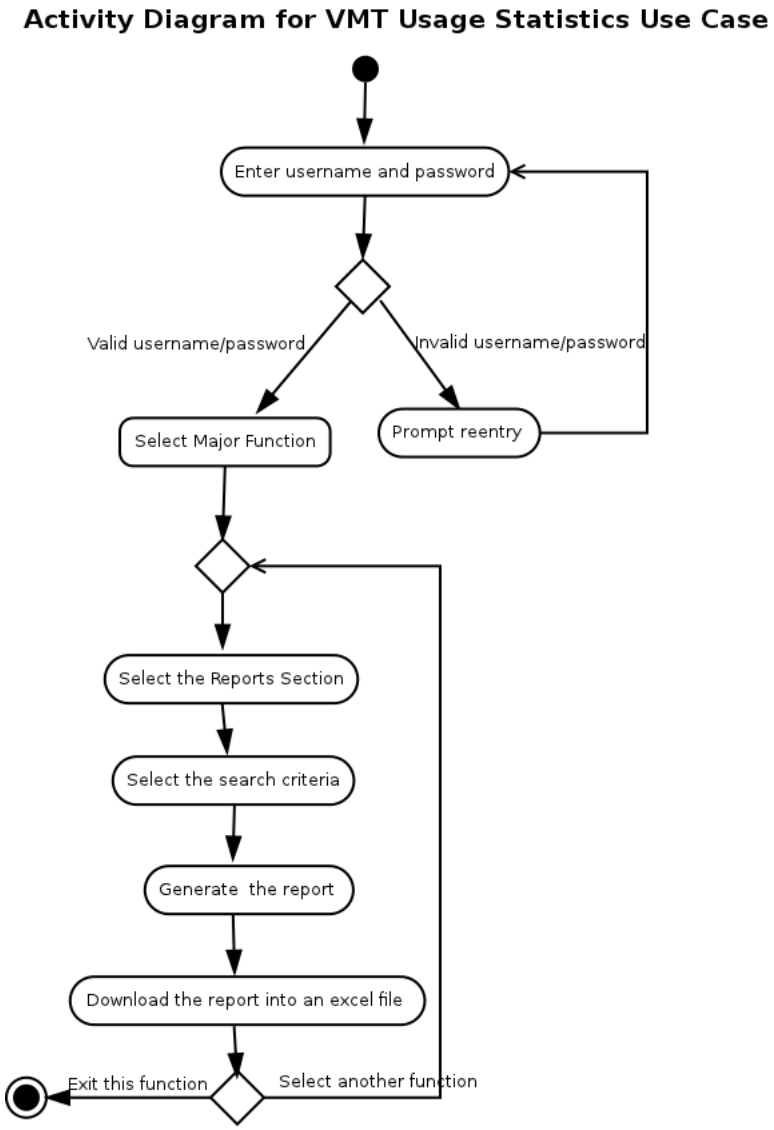


Figure 2: Usage Statistics Activity Diagrams

4.2.6 System Features

Authentication

Description and Priority This feature manages high privileged user access to the system, through password authentication. This is a high priority feature since access rights and privileges are determined at this level. **Stimulus/Response Sequences**

Stimulus: User provides username and password

Response: System validates username and password

If the username and password are found, the system shall then allow access to the user with the corresponding privileges Else a login failure message will be triggered off

Functional Requirement

REQ-1: System must allow access to only authorized users.

REQ-2: System must allow authentic users to change passwords based on a defined set of rules.

REQ-3: System must allow system administrator to change authentic user passwords or deactivate user accounts.

Use case: Login

Primary Actor: Admininstrators

Goal in context: To authenticate access to the system.

Trigger: Need to add,delete,edit an administrator's details.

Scenario:

1. The Administrator log on to the Voi-Web Reporting Interface.

2. The Administrator enters his username and password.

3. The VRWeb Information selects the 'Users Section' on the interface.

4. The Administrator enters details of the administrator to be added,deleted or edited.

5. The Administrator saves the changes.

6. The Administrator can logout or move to the 'Reports Section' on the interface.

Priority : High priority

Usage Reports Interface

Description and Priority

This feature provides for generation of usage statistics of VMT.

Stimulus/Response Sequences Stimulus: Administrator enters search criteria. Response: System presents VMT according to the search criteria.

Functional Requirement

REQ-1: The system to provide an interface to generate the VMT usage statistics.

Use Case: Generation of VMT usage Statistics

Primary Actor: Administrator

Goal in context: To generate VMT usage statistics

Trigger: Need to generate VMT usage statistics

Scenario:

1. The Administrator log on to the Voi-Web Reporting Interface.

2. The Administrator enters his username and password.

3. The VRWeb Information selects the 'Reports Section' on the interface.

4. The Administrator enters search criteria of usage statistics.

5. The Administrator generates the reports .

6. The Administrator downloads the reports as excel files

6. The Administrator can logout or move to the 'Users Section' on the interface.

Prioty : High priority

Web Information search by SMS

Description and Priority

This feature provides for search of Web Information through SMS

Stimulus/Response Sequences Stimulus: Mobile Phone user sends an SMS to the VMT number containing the search query. Response: System responds with a call relaying information generated from the search.

Functional Requirement

REQ-1: The system to provide for searching the WWW through SMS .

Use case: Searching the WWW through SMS

Primary Actor: Mobile Phone User

Goal in context: To get Web Information

Trigger: Need for Web Information.

Scenario:

1. The User sends an SMS containing search query to the VMT code.

2. The System receives the SMS and searchs the Web based on its contents.

3. The System converts the search results into an audio file.

4. The System the calls back the Mobile Phone user through the IVR system and relays the audio data of the search results.

5. The Mobile Phone user receives the calls listens to the content and hungs up.

Prioty : High priority

Web Information search by Voice Call

Description and Priority

This feature provides for search of Web Information through Voice Calls

Stimulus/Response Sequences Stimulus: Mobile Phone user makes a call to the VMT number saying the search query. Response: System responds with a call relaying information generated from the search.

Functional Requirement

REQ-1: The system to provide for searching the WWW through Voice Calls .

Use case: Searching the WWW through Voice Calls

Primary Actor: Mobile Phone User

Goal in context: To get Web Information

Trigger: Need for Web Information.

Scenario: 1. The User makes a call relaying the search query to the VMT code.

2. The System receives the call , converts it into text and searches the Web based on the converted text.

3. The System converts the search results into an audio file.

4. The System the calls back the Mobile Phone user through the IVR system and relays the audio data of the search results.

5. The Mobile Phone user receives the calls listens to the content and hungs up.

Prioty : High priority

4.2.7 Other Non-Functional Requirements

OR-1: The system should be simple to use and understandable.

OR-2: The system should provide relevant information with reference to the search query.

OR-3: The system should be avaialble always.

OR-4: The system should always call back with feedback.

4.3 VMT System Implementation

The VMT system is composed of five modules i.e. Speech Recognition Module, SMS Receiver Module, Search API, Speech Sysnthesis Module, Voi-Web Reporting Web Interface(VRWI) and the IVR Module. The VMT infrastructural design and architecturial design indicated in the methodology were used define the external entities interacted with by the software interacts and the nature of interaction [17].

4.4 VMT Infrastructure

The VMT works with the Tele-communication Infrastructure as illustrated in figure 5 below.

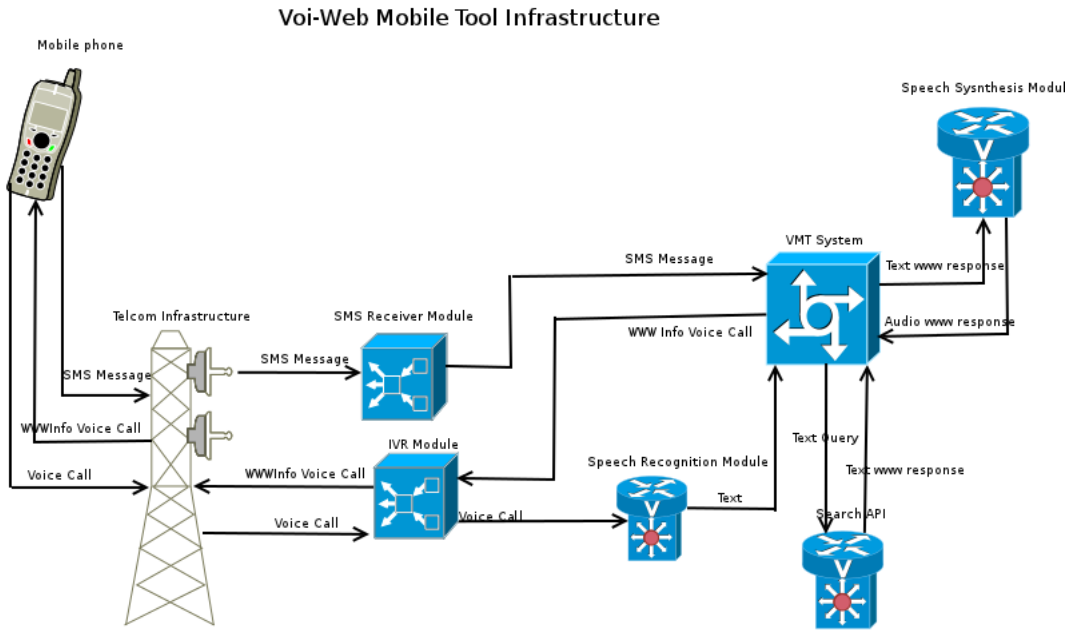


Figure 5: Voi-Web Mobile Tool Infrastructure

4.5 VMT Architecturial Design

The VMT architecturial design consists of the components specified in figure 6 below.

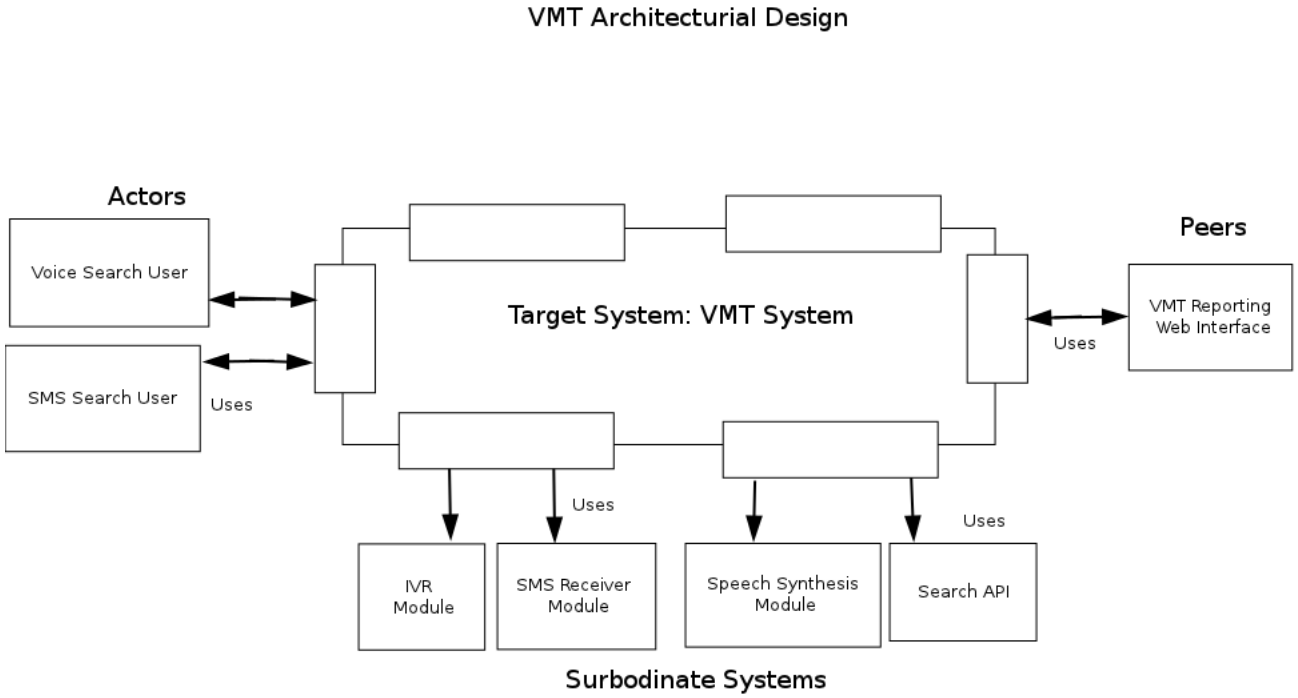


Figure 6: VMT Architecturial Design

Below is the description of the systems that constitute the VMT solution Architecture. Refering to Figure 6, systems that interoperate with the VMT system for which architecturial designed has been developed are represented as ;

4.5.1 Subordinate Systems

These are systems/modules have been used by the VMT system to provide data and processing to complete the VMT functionality[17]. These modules include; IVR Module, SMS Receiver Module, Searh API and the Speech

Synthesis Module.

4.5.2 Peer-level systems

The VMT Reporting Web Interface (VRWI) is also a sub system that interacts on a peer-to-peer basis with the VMT system. The VMT logs information with the VRWI.

4.5.3 Actors

These are entities that interact with the VMT system by producing or consuming information that is necessary for requisite processing[17].These entities include Voice search users and SMS search users.

4.5.4 Speech Recognition Module

The Speech Recognition Module has been developed with the help of a Google API for speech recognition[**?**].The Google API has been choosen because it is platform independent and it recogonizes the voice queries better compared to all the speech recogition toolkits that have been explored.

CMUSphinx[5] is available, free, open source and is able to interact with Asterisk the VOIP Server to be utilized. However it has lower recognizing ability in comparison to the Google API.

Julius: "Julius is a high-performance, two-pass large vocabulary continuous speech recognition (LVCSR) decoder software for speech-related researchers and developers" [22] However Julius is only distributed with Japanese models though English acoustic models are being created for it by VoxForge

Shout: "a large vocabulary continuous speech recognition toolkit" [16]. However there was on instance of Shout working with Asterisk in all its examples .

It should has been noted however, that all modern descriptions of speech are to some degree probabilistic. That means that there are no certain boundaries between units, or between words. Speech to text translation and other applications of speech are never 100

4.5.5 SMS Receiver Module

The SMS Receiver Module was development with Kannel an Open Source SMS Gateway [14]. The Kannel SMS Gateway was the only SMS gateway considered mainly because it is a free open source SMS Gateway that is easy to use.

4.5.6 Search API

The search API utilizes the Google custom search API [7]. The google search API was chosen because statistics according to Stat Owl[**?**]show the google search engine as the leader at an average percentage of 80.4 % of the market share, as of November,12.

4.5.7 Speech Synthesis Module

This Module has been developed with Espeak an open source tool kit for speech synthesis and sox an open source tool kit that converts wav audio files into gsm audio files . Another tool that were considered was Festival, however Festival lacked the functionality to adjust the speech speed which Espeak could, so Espeak was chosen over Festival.

4.5.8 IVR Module

The IVR has utilized Asterisk an open source telephony applications platform distributed under the GPLv2 [4]. Astersik was mainly considered because its open source, runs on linux platforms, free and easy to implement. Asterisk is a software implementation of a telephone private branch exchange (PBX); it allows attached telephones to make calls to one another, and to connect to other telephone services including the public switched telephone network (PSTN) and Voice over Internet Protocol (VoIP) services [24].

The Asterisk software includes many features available in proprietary PBX systems: voice mail, conference calling, interactive voice response (phone menus), and automatic call distribution [24]. By supporting a mix of traditional and VoIP telephony services, Asterisk allows deployers to build new telephone systems, or gradually migrate existing systems to new technologies. Some sites are using Asterisk servers to replace proprietary PBXes; others

to provide additional features (such as voice mail or voice response menus, or virtual call shops) or to reduce costs by carrying long-distance calls over the Internet [24].

Critical Asterisk Architectural Concepts

Channels : A channel in Asterisk represents a connection between the Asterisk system and some telephony endpoint.[4]

Channel Bridging : Channel bridging is the act of connecting channels together for the purpose of passing media between them. The media stream is most commonly an audio stream [4].

Asterisk Frames

Communication within the Asterisk code during a call is done by using frames, which are instances of the astframe data structure. Frames can either be media frames or signalling frames. During a basic phone call, a stream of media frames containing audio would be passing through the system [4].

By default, all modules found in a predefined Asterisk modules directory on the filesystem will be loaded when the main application is started. This approach was chosen for its simplicity.

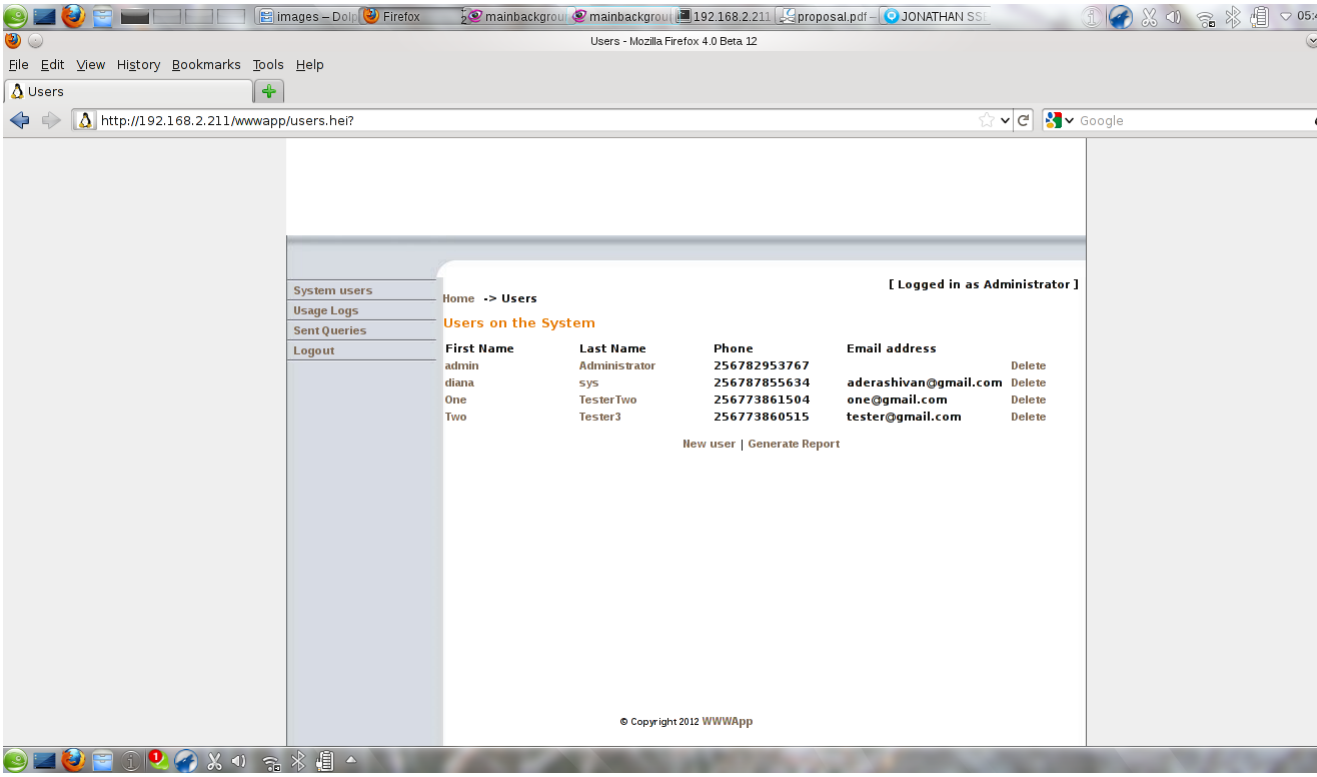
However, there is a configuration file that can be updated to specify exactly which modules to load and in what order to load them. This makes the configuration a bit more complex, but provides the ability to specify that modules that are not needed should not be loaded [4].

4.5.9 Voi-Web Reporting Web Interface

The Reporting Web Interface has been developed with Heitml Scripting Language [15] for the Application Interface and Postgresql [13] for the Database.

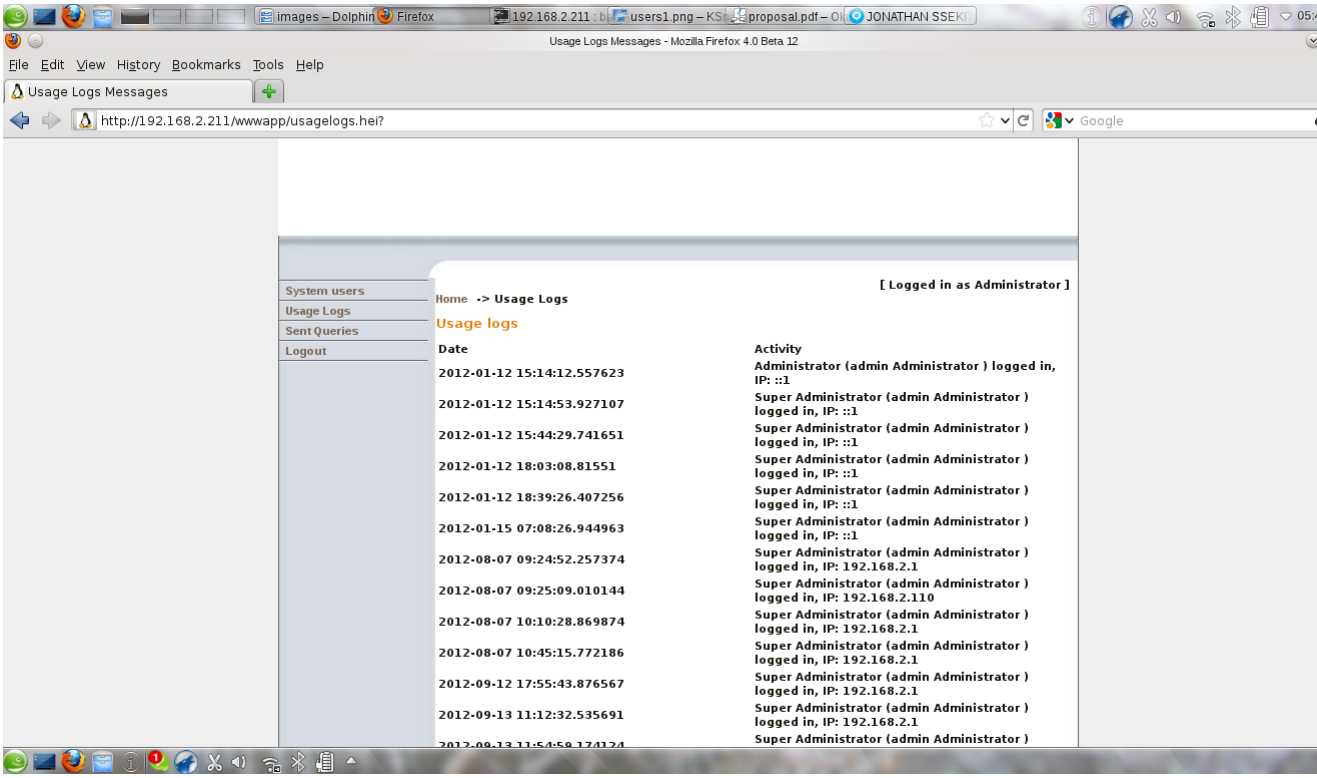
Administrators Management Interface

The Administrators Management Interface facilitates the authorized Administrator to add, modify and delete the Voi-Web Reporting Web Interface Administrators.


Users - Mozilla Firefox 4.0 Beta 12
File Edit View History Bookmarks Tools Help
Users
http://192.168.2.211/wwwapp/users.hei?
Google
System users
Usage Logs
Sent Queries
Logout
[Logged in as Administrator]
Home -> Users
Users on the System
First Name Last Name Phone Email address
admin Administrator 256782953767 Delete
diana sys 256787855634 aderashivan@gmail.com Delete
One TesterTwo 256773861504 one@gmail.com Delete
Two Tester3 256773860515 tester@gmail.com Delete
New user | Generate Report
© Copyright 2012 WWWApp

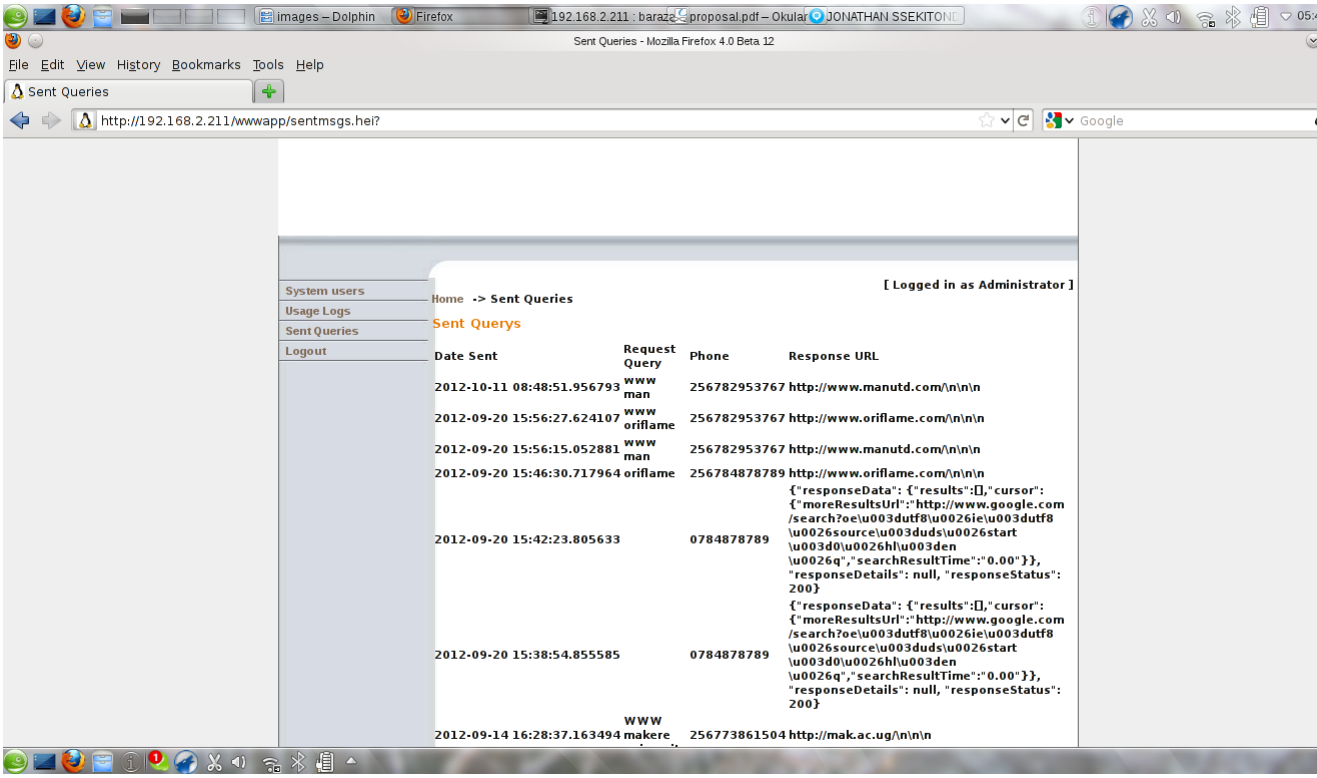
Usage Log Management Interface

The Usage Logs Management Interace facilitates monitoring the activity of each Administrator. A report of Usage logs can be generated into an excel file for further analysis with the Generate Report button.



Sent Queries Management

The Sent Queries Management interface facilitates viewing SMS Messages and Transcribed Voice Queries that are sent to the Voice-Web system from the Mobile phone user and URLs of the responses that were relayed back to the Mobile Phone User. A report of Usage logs can be generated into an excel file for further analysis with the Generate Report button.



4.6 The VMT Results Analysis

Here we discuss the Voi-Web results from the recordings we got from the sampled Ugandans. The recordings are information queries based on the most searched information queries in Uganda according to google.

Below are the queries that were recoreded;

(1) What is love

(2) What is technology.

(3) What is planning?

(4) What is ethics?

(5) What is language?

The Ugandans sampled were given the liberty to choose which query to say and most of them choose 'What is love'.

4.6.1 Confusion Matrix

Performance of classifer system is commonly evaluated using the data in the confusion matrix. The entries in the confusion matrix have the following meaning in the context of the study :

'a is the number of correct predictions that an instance is negative, b is the number of incorrect predictions that an instance is positive, c is the number of incorrect of predictions that an instance negative, and d is the number of correct predictions that an instance is positive'[**?**].

The accuracy (AC) is the proportion of the total number of predictions that were correct. It is determined using the equation:

A=a+d/a+b+c+d

The recall or true positive rate (TP) is the proportion of positive cases that were correctly identified, as calculated using the equation:

T=d/c+d

Finally, precision (P) is the proportion of the predicted positive cases that were correct, as calculated using the equation

$$P = d/b + d$$

The true positive values and true negative values have been calculated for the accuracy of the voice recognition based on this criteria.

T: Number of words in the initial query A: Number of word recognized from the recording of the initial query

True Positive = A/T

True Negative = T-A/T

Below are the conversions that were made with each recording.

Transcribed Query	Actual Recorded Query	T	A	True Positive	True Negative
what what is the Connecticut	what is technology	3	2	0.66666667	0.333333333
what is the mood	what is technology	3	2	0.66666667	0.333333333
what does the cannot adjust	what is technology	3	1	0.33333333	0.666666667
but the United	what is technology	3	0	0	1
whats love	what is love	3	3	1	0
whats it up	What is love	3	1	0.33333333	0.666666667
love	What is love	3	1	0.33333333	0.666666667
whats the	What is love	3	1	0.33333333	0.666666667
what is loud	What is love	3	2	0.66666667	0.333333333
love	What is love	3	1	0.33333333	0.666666667
okay whats planning	what is planning	3	3	1	0
what is wrong with you	what is language	3	2	0.66666667	0.333333333
Priceline way	what is technology	3	0	0	1
language	language	1	1	1	0
love	love	1	1	1	0
now	love	1	0	0	1
what is busy	What is ethics	3	2	0.66666667	0.333333333
what is up	What is love	3	2	0.66666667	0.333333333
love	love	1	1	1	0
what is up	what is love	3	2	0.66666667	0.333333333
what friends	what is planning	3	1	0.33333333	0.666666667
what is love	What is love	3	3	1	0
what is it off	What is love	3	2	0.66666667	0.333333333
what to pick Knology	what is technology	3	1	0.33333333	0.666666667
what is love	What is love	3	3	1	0
what do you say 6	What is ethics	3	1	0.33333333	0.666666667
by sticking already	what is technology	3	0	0	1
what do love you	what is planning	3	1	0.33333333	0.666666667
what is love	What is love	3	3	1	0
what is technology	what is technology	3	3	1	0
				17.3333333	**12.66666667**

4.6.2 Speech Synthsis Results Accuracy Calculations

Matrix Values

We have not been able to test the Voi-Web with negatives instances because search queries cannot be negative. a = 0; b = 0; c = 12.33; d = 17.33;

A=a+d/a+b+c+d

A=0+17.33/0+0+12.33+17.33

Accuracy Percentage = A * 100

Accuracy Percentage = 58 percent

Basing on the analysis from the confusion matrix the VMT speech synthesis test results achieved an accuracy of 58 percent.

4.7 Conclusion

The VMT Solution has been implemented to improve accessibility of Web Information to Uganda's mobile phone users that are constrained infrastructuraly and economically.

The VMT solution achieved an accuracy of 58

The welfare of Uganda's MP users will be improved with the VMT tool through the provision of Web Information as Voice, this will be so mainly because this tool is mobile and can used any where with the telephone network coverage ,queries are received through voice calls and relayed back through voice calls so even the illiterate can access Web Information, it is not limited by the small screen of Mobile Devices because information is relayed through voice, for areas that don't have internet provision facilities and some times electricity installed they too can still have access to Web Information with their Mobile Devices and the tool is easy to use because it reflects telephone usage behaviour.

4.7.1 Challenges

The Voi-Web Mobile Tool was developed with the infrastructure and tools specified in the sections above, however the speech recognition was not accurate mainly because of the difference in accents with the European accents that were developed with the Google API.

4.7.2 Recommendations

The following recommendations were made as tests were being carried out on the VMT Tool.

(1) To have the VMT Tool send text messages for the SMS Search as well as the Voice Information.

(2) To make the VMT Tool more Audible.

(3) To improve on the Speech recognition functionality.

(4) To have an Acoustic Model setup for African/Ugandan accents i.e have the information readout in a Ugandan accent that can understood by most Ugandans.

On the overall the Voi-Web system is a relevant tool that will be utilized even with the above short commings.

The ultimate aim of this project was to improve accessibility of Web Information to Uganda's Mobile Phone users by providing them with a tool that receives voice and SMS queries and presents speech synthesized Web Information to their Mobile phones as Voice calls. This goal has been achieved with admist various challenges as mentioned above,however on the overall the Voi-Web Mobile Tool is a relevant tool that will be used to access Web Information.

Bibliography

[1] R. Albrecht. Messaging in mobile augmented reality audio. 2011.

[2] C. H. S. O. S. Annika. Interactive voice response and web-based questionnaires for population-based infectious disease reporting. *Springer Science+Business Media B.V. 2010*, 2010.

[3] A. S. Bernard J. Jansen, Danielle L. Booth. Information processing and management, volume 44, issue 3, pages 1251-1266. May 2008.

[4] A. Brown and G. Wilson, editors. *The Architecture of Open Source Applications*. March 2011.

[5] cmusphinx. Basic concepts of speech, November 2011.

[6] U. C. Commission. 2009/10 post and telecommunications market review. 2011.

[7] G. Developers. Google custom search apis and tools. `https://developers.google.com/custom-search/`, 2012.

[8] J. doug Fracoise Bill Ciprian Mike Maryam Brian. Googe search by voice google inc.

[9] I. Engholm. Design history of the www.:website development from the perspective of genre and style theory. 2007.

[10] W. M. W. et al. Voice ip approval system using voice-enabled web based application server. (6507817), Jun 2003.

[11] Google. Google search app for your phone. `http://www.google.com/mobile/google-mobile-app/`, 2011.

[12] Google. Google sms applications. `http://www.google.co.ug/mobile/sms/`, 2011.

[13] P. G. D. Group. Postgresql. http://www.postgresql.org, 2012.

[14] T. K. Group. Kannel:open source wap and sms gateway. http://www.kannel.org, 2012.

[15] H.E.I. Heitml scripting language. http://www.radpage.com, 2012.

[16] M. Huijbregts. Shout). http://shout-toolkit.sourceforge.net, 2010.

[17] R. S. Pressman. *Software Engineering. A Practitioner's Approach*. 2010.

[18] PromptU. Pronto treno. http://www.promptu.com/prontotreno.php, 2012.

[19] J. M. M. Rakesh. Assistive technology for vision-impairments: An agenda for the ictd community. *WWW 2011 – Web for Emerging Regions*, pages 513–524, April 2011.

[20] J. M. Srinivas. Speech and multimodal interaction in mobile search. 2011.

[21] M. O. System. *VoiceOver*. Macintosh Operating System.

[22] J. D. Team. http://julius.sourceforge.jp.

[23] Wikipedia. Voice search. http://en.wikipedia.org/wiki/Voice_search, 2010.

[24] Wikipedia. Asterisk (pbx), June 2011.

[25] Wikipedia. Speech synthesis. http://en.wikipedia.org/wiki/Speech_synthesis, 2011.

[26] Wikipedia. Text messaging. http://en.wikipedia.org/wiki/Text_messaging, 2011.

[27] Wikipedia. Right to internet access. http://en.wikipedia.org/wiki/Right_to_Internet_access, 2012.

[28] Wikipedia. Web search query. http://en.wikipedia.org/wiki/Web_search_query, 2012.

[29] M. J. H. L. F. S. W.-y. Xing Xie, Lie Lu. Mobile search with multimodal queries, vol 96, no 4. April 2008.

[30] Youtube. You tube. http://www.youtube.com, 2012.