A FRAMEWORK FOR INTELLIGENT VOICE-ENABLED E-EDUCATION SYSTEMS

BY

AZETA, Agbon Ambrose

(CUGP050134)

A THESIS

SUBMITTED TO THE DEPARTMENT OF COMPUTER AND INFORMATION SCIENCES,

SCHOOL OF NATURAL AND APPLIED SCIENCES, COLLEGE OF SCIENCE AND

TECHNOLOGY, COVENANT UNIVERSITY,

OTA, OGUN STATE NIGERIA,

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF

DOCTOR OF PHILOSOPHY

IN

COMPUTER SCIENCE

MARCH, 2012

CERTIFICATION

This is to certify that this thesis is an original research work undertaken by Ambrose Agbon Azeta with matriculation number CUGP050134 under our supervision and approved by:

Professor Charles Korede Ayo   ---------------------------

Supervisor   Signature and Date

Dr. Aderemi Aaron Anthony Atayero   ---------------------------

Co- Supervisor   Signature and Date

Professor Charles Korede Ayo   ---------------------------

Head of Department   Signature and Date

External Examiner   Signature and Date

DECLARATION

It is hereby declared that this research was undertaken by Ambrose Agbon Azeta. The thesis is based on his original study in the department of Computer and Information Sciences, College of Science and Technology, Covenant University, Ota, under the supervision of Prof. C. K. Ayo and Dr. A. A. Atayero. Ideas and views of this research work are products of the original research undertaken by Ambrose Agbon Azeta and the views of other researchers have been duly expressed and acknowledged.

Professor Charles Korede Ayo   ---------------------------
Supervisor   Signature and Date

Dr. Aderemi Aaron Anthony Atayero   ---------------------------
Co- Supervisor   Signature and Date

ACKNOWLEDGMENTS

TABLE OF CONTENTS   **PAGE**

CHAPTER ONE

CHAPTER TWO

CHAPTER THREE

3.0 THE INTELLIGENT VOICE-ENABLED E-EDUCATION FRAMEWORK 45

ABSTRACT

Voice-enabled e-Education system is the enabling technology that allows users to access Learning resources on a web using a telephone. Telephony applications in the area of education are developed mainly for the visually impaired to address the limitations of user interface of a mobile learner in accessing web learning resources. However, with the proliferation of learning resources on the web, finding suitable content (using telephone) has become a rigorous task for voice-based online learners to achieve better performance. The problem of Finding Content Suitability (FCS) with voice e-Learning applications is more complex when the sight impaired learner is involved. Existing voice-enabled applications in the domain of e-Education lack the attributes of adaptive and reusable learning objects to be able to address the FCS problem. Nevertheless, the adoption of artificial intelligence methodologies involving case-based reasoning (CBR) and domain ontology is required to provide recommender and adapted systems in voice-enabled learning towards solving the FCS problem. This study provides a framework for intelligent Voice-based e-Education Systems (*iVeES*) and an implementation of a system based on the framework with dual user interfaces – voice and web.

The framework was designed using schematic diagram, mathematical models and ontology; and implemented using PHP, VoiceXML and MySQL. A usability study was carried out in a visually impaired and a non visually impaired school using the International Standard Organization 9241-11 (1998) and Nielsen (2003) specification to determine the level of effectiveness, efficiency, user satisfaction, learnability and memorability. The result of the usability evaluation reveals that the prototype application developed for the school has "Good Usability" rating of 4.12 out of 5 scale. This shows that the application will not only complement existing mobile and web-based learning systems but will be of immense benefit to users, based on the system's capacity for taking autonomous decisions that are capable of adapting to the needs of both visually impaired and non-visually impaired learners.

CHAPTER ONE

INTRODUCTION

1.1   BACKGROUND INFORMATION

Mobile phone interface is limited by small screen size and the requirement of a light pen to enter data. Notwithstanding, there is a rapid movement from static Personal Computer (PC) based applications towards the use of mobile device as a medium of accessing web applications. Software applications developed using voice processing technologies allow "hands free" information input and retrieval and thus can help to overcome the limitations of the user interface of a mobile learner in accessing e-Education application. E-Education entails the use of modern information and communication technology (ICT) in teaching and learning to enhance traditional classroom education and achieve better results in distance education, mobile learning and web-based learning (Kastelic and Loncaric, 2007).

Over the last decade, there has been a change of focus from web-based learning systems that merely turn pages of content to systems which present learning materials in such a way as to satisfy the needs of learners. This change of focus is especially important in modern learning methods, which places strong emphasis on learners' previous knowledge. The uniqueness of a learner is met by making the web-based learning content adaptive (Oboko et al., 2008). The increasing number of Learning Management Systems (LMS) for online teaching, quiz, assignment delivery, discussion forum, email, chat, et cetera, means that dynamic educational online services will be needed for efficient management of all educational resources on the web. Selecting and organizing learning resources based on learner's interest is cumbersome (Gil and García-Penalvo, 2008). The process of selection may be easier with the normal users, but for certain category of learners with visual impairment, navigating a Voice User Interface (VUI) for the desired learning content is a strenuous task. Web-deployed VUI applications for educational purposes provide user accessibility to content via telephone. One of the tools used for developing VUI applications is Voice eXtensible Mark-up Language (VoiceXML or VXML).

VUI applications are primarily developed to cater for the visually impaired (Raghuraman, 2004), to address the problem of web accessibility associated with the use of m-Learning. The major problem of web access is that the services of reusable learning objects currently available to Internet-connected users are not available for the visually impaired (Holzinger et al., 2006), for reasons of constraint in their navigation abilities. Thus, the voice-based web applications mainly designed for the visually impaired lack adequate quality of using recommender and adaptable system for learning, which is a major requirement for this category of users as a result of their visual impairment. Recommender systems are software agents that recommend options for users. These agents can be very useful in an e-Learning environment to recommend actions, resources or links (Zaiane, 2005). The services provided by recommender systems are generally referred to as recommendation services. Adaptive web system monitors particular user's behavior and characteristics. Based on them, the system compiles document from larger universal source document and then produce adapted document to users (Bures and Jelinek, 2004).

The goal of adaptive voice-based web learning is to adjust the content of the learning objects to suit user's knowledge level, whereas recommendation services provide the most appropriate learning objects to users through voice. The inability of existing e-Education voice applications to meet these requirements has some far reaching implications such as limited accessibility for certain users especially the visually impaired, and usability issues as a result of lack of features for reasoning, adaptation and recommendation. A significant contribution would be to introduce the concept of reusing learner's previous experiences which is often neglected in existing voice-based e-Education systems, to make them more adaptable and provide recommendation services to users' needs. Thus, we have to further take into account that the reference guidelines for developing existing voice-based e-Education applications, which although important, lack intelligent component services to approach the problem. One of the ways of enhancing existing voice-based e-Education applications is to adapt their content to the needs of each student.

The needs of users differ when accessing learning content, such needs may require the ability of the system to reason, make decisions, be flexible and adapt to divers requests during interaction. These needs have placed new requirements in voice application development such as the use of advanced models, techniques and methodologies which

take into account the needs of different users and environments. The ability of a system to behave close to human reasoning for adaptability support is often mentioned as one of the major requirements for the development of intelligent voice applications. Intelligent systems belong to the domain of Artificial Intelligence (AI). AI is a scientific and engineering discipline devoted to understanding principles that make intelligent behavior possible in natural or artificial systems (Russell and Norvig, 2003). A system can also be said to be intelligent if it is able to improve its performance or maintain an acceptable level of performance in the presence of uncertainty. The main attributes of intelligence are learning, adaptation, fault tolerance and self-organization (Jain and Martin, 1998). If adaptability services must exist as a feature of voice-based e-Education system, then an intelligent agent will be required to power the application.

From the foregoing, it becomes obvious that there exists a need to provide a framework for improving the level of intelligence of voice-based e-Education applications. This study aims at addressing the needs as identified, and employing the framework so obtained for the development of a prototype Intelligent Voice-based E-Education System *(iVeES)*. The prototype application developed was tested in a school for the blind, and the result of evaluation reported. The framework suffices as a reference model for implementing intelligent voice applications in the domain of e-Education.

The type of e-Education application that will be provided through this research would include that of teaching, learning and multiple choice examination. Applications developed based on this framework would exhibit the necessary attributes of Case Based Reasoning (CBR), including adaptation and recommendation during interaction resulting in intelligence such as the ability of the system to take autonomous decision that will adapt to learners' requests based on requirements. The application will therefore be helpful for people with physical access difficulties engendered by their visual impairment.

1.2   MOTIVATION FOR THE STUDY

The equipment used as assistive technology for learning in school for the blind such as Slate and Stylus, Mathematics board and Figures, Braille, Typewriters, Abacus, etc are expensive to procure and maintain (Azeta et al., 2009b). As a result of the pace of technological

revolution, these equipment are gradually being replaced by the numerous applications developed to run on the Internet.

The Internet, certainly has received significant attention in recent years, but verbal communication is still the most natural form of interaction amongst human. Therefore the human speech is a very good medium of communication using telephone access. Given the naturalness and expressive power of speech, speech input and output have the potential for becoming key modalities in future interactive systems (Niels, 2000). There is an effort to use this kind of communication also in human – computer interaction and for data access on the web (Ondas, 2006). Mobile phone is one of the media for accessing web content. The limitations of mobile phones such as small screen size, small keypad, and limited battery power to effectively hold and display information does not pose any constraint to telephone calls and conversation, compared to when using it to access the web through Wireless Application Protocol (WAP). Data entry on the tiny keyboards of WAP-enabled phones or pointing with stylus is very uncomfortable and prone to errors. Another issue is that mobile devices are often used when a person is really "on the move". Operating in such conditions is either impeded or even prohibited, e.g. in the case of car driving (Zaykovskiy, 2006).

The natural way to overcome this problem is the use of voice technology. The possibility of using spoken language to interact with computers has generated wide interest in spoken language technology. Many universities and research laboratories have put their efforts to development of the spoken language technology (Lerlerdthaiyanupap, 2008). However, voice technology is limited to only one form of input and output – human voice (Kondratova, 2009).

The high popularity and wide spread use of telephone globally, with an estimate of about four billion active telephone lines (fixed and mobile phone) prompted the World Wide Web Consortium (W3C) speech interface framework (Larson, 2000) to create an avenue for users to interact with web-based services via key pads, spoken commands, prerecorded speeches, synthetic speeches and music. The interface framework led to the massive convergence of web and voice technologies to provide solutions in the areas of e-Health, e-Stock, e-Democracy, e-Learning and a lot more.

Majority of the present day e-Learning applications have support for the web user interface (WUI) through the use of PC, while others are based on mobile devices such as Wireless Application Protocol (WAP) phones and Personal Digital Assistants (PDAs). Given the limited input/output capabilities of mobile devices, speech presents an excellent way to enter and retrieve information either alone or in combination with other access modes. Furthermore, people with disabilities should be provided with a wide range of alternative interaction modalities other than the traditional screen-mouse based desktop computing devices. People who suffer disability whether temporary or permanent, either in reading difficulty, visual impairment, or any difficulty using a keyboard, or mouse can rely on speech as an alternate approach for information access (Farinazzo et al., 2010).

The need to reduce access barriers for the visually impaired has necessitated this research in the education domain with mobile and land phones as access point. The problems with the various conventional learning systems such as face to face, telephone, electronic mail, chat room, instant messaging, etc, are numerous, since using them becomes a more difficult task for those with disabilities.

Lack of accessibility in the design of e-Learning courses continues to hinder students with vision impairment. Despite significant advances in assistive technologies, blind and visually impaired Internet users continue to encounter barriers when accessing web content (Gerber and Kirchner, 2001; Williamson et al., 2001). E-Learning materials are predominantly vision-centric, incorporating images, animation, and interactive media, and as a result students with low or acute vision impairment do not have equal opportunity to gain tertiary qualifications or skills (Armstrong, 2009). A blind person cannot see or communicate through mail or electronic means that require the ability to see the screen. Lack of provision for voice in the conventional learning methods has restricted support for people with limited capabilities such as the visually and mobility impaired that affect either data entry, or ability to read and check what they have entered, since these applications are visual in nature and require the ability sight to see the blackboard or computer screen and manipulate the computer keyboard. More so, it excludes people who have medical history of Repetitive Strain Injury (RSI) as a result of their sedentary lifestyle using a peripheral device such as a keyboard. The web-based and mobile learning system does not have support for people with textual interpretive problems

e.g. dyslexia, to enter text verbally, and others who have problems of identification of characters and words.

An obvious limitation in providing Information Technology access to learners with low vision is the difficulty that these learners have in reading text. This difficulty is compounded when a visually impaired user is faced with computer technologies and concepts, as a first time user. Voice technology is a possible solution to this problem as it can enable students to learn literacy skills using a more natural human-computer interface (Walsh and Meade, 2003). However, existing voice-based learning application have not been able to adequately proffer solution to the problem of providing suitable learning content in the midst of numerous learning resources available on the web for both normal and visually impaired learners.

1.3   STATEMENT OF THE PROBLEM

In some Educational Institutions, particularly public schools in the developing nations, there is a growing shortage of academic staff and the demand for teachers has continued to increase. Demand for access to education remains much higher than the system's physical capacity to accommodate students in public schools (NUC, 2004). As a result, the ratio of teachers to students is constantly unbalanced with students having an increasing number. The process involved in conducting course registration, lecture delivery, examination and result is tedious, since many students are attended to by few teachers. More so, no adequate attention is given to students by their teachers in the areas of counseling, teaching, invigilation, etc. Currently, several Universities and Colleges are overcrowded with inadequate facilities to match the upsurge, thus resulting in deviation from the approved carrying capacity accomplished with poor academic performance (Bobbie et al., 2008).

While trying to address the aforementioned problems, many studies have shown the value of existing platforms such as web and m-Learning, but lack access to such applications for the large number of students through virtual learning. The number of e-Learning web portal has continued to rise as a result of the advancement in web technologies. Some of the popular web portal available for learning includes: Moodle, WebCT, Blackboard, Sakai and Claroline (Ermalai et al., 2009). Voice-based learning (Motiwalla and Qin, 2007) evolved to

address the problems of accessibility associated with m-Learning, arising from the limitation in using mobile phone to access learning content on the web.

With the proliferation of web learning resources, finding suitable content (using telephone) has become a rigorous task for voice-based online learners to achieve better performance. The biggest disadvantage of voice-based technologies is the rigid structure that they impose on the end user. While it is convenient to use mobile telephony application, it can be extremely slow when the user is forced to drill through several layers of options before finding exactly what he/she wants (Raghuraman, 2004). The Finding Content Suitability (FCS) problem with voice e-Learning applications is more complex when the visually impaired learner is involved. Thus, existing voice-enabled e-Education applications lack the necessary attributes of reasoning and adaptability during interaction to be able to address the FCS problem.

To solve the FCS problem and enhance voice-based learning, recommending and adapting the contents to the needs of each student is required. It is a necessity that spoken dialogue system performance be enhanced through adaptation (Pakucs, 2003). More so, the adoption of artificial intelligence methodologies using CBR and domain ontology to provide recommendation services to the different needs of students is essential (Abdel-Badeeh and Rania, 2005). Generally, there is dearth of voice application in the domain of e-Education and the few existing voice-based e-Education application are not intelligent to provide recommendation services beyond student's queries and request.

Thus, we take cognizance of the fact that there is no reference model containing intelligence upon which the existing voice-based e-Education applications were developed. Although the approach employed for developing some of them are the general software development life-cycle model, they do not constitute a framework to provide intelligence for voice e-Education applications.

These considerations will lead us to the following research questions:

(i) How can a framework be developed to improve the level of intelligence in voice-enabled e-Education applications towards providing learning content suitability for users?

(ii) How can the framework be adapted as a test bed for providing **i**ntelligent **V**oice-enabled **e**-**E**ducation **S**ystems *(iVeES)*?

We believe that answering these questions will solve the FCS problem and improve the level of flexibility and intelligence of voice-enabled e-Education applications.

1.4   AIM AND OBJECTIVES OF THE STUDY

The aim of this study is to provide a framework for improving the level of system intelligence for voice-enabled e-Education applications and adapt the framework for the development of *iVeES* that has the capability to improve learning processes using telephone and web-based technologies.

The objectives of this study are as follows:

i) To conduct a requirement elicitation and provide a framework for intelligent Voice-based e-Education System (*iVeES)* that will serve as a generic guideline for developing intelligent voice-based applications in the domain of e-Education.

ii) To provide a design and prototype implementation of intelligent voice-based application in e-Education with dual user interface (Voice and Web) as proof of concept.

iii) To evaluate the implemented *iVeES* application.

1.5   RESEARCH METHODOLOGY

The research methods adopted in this study include:

i)   Conducting extensive preliminary studies and developing pilot schemes in e-Education for experimentation purposes. The findings at this stage were presented in seminars and conferences where useful contributions were made.

ii)   The use of appropriate research instruments which include questionnaire and interview were developed for requirement elicitation, and the information acquired was used to develop a framework for *iVeES*. Protégé 3.4.1 and Web Ontology Language (OWL) Editor were used to construct domain ontology during the system design. Protégé is a free, open-source platform that provides a growing user community with a suite of tools to construct domain models and knowledge-based applications (Protégé, 2010). OWL is a software tool used for building ontologies (Moise and Netedu, 2009).

iii)   Implementation of *iVeES* using artificial intelligence methodologies involving CBR (Spasic et al., 2005) and Modified Porter Stemming (MPS) algorithm to realize the intelligent search agent. The original Stemming algorithm was developed by Martin Porter for reducing English words to their word stems. For example, the word "connections" would be reduced to its stem form "connect" (PorterStemming, 2007). The core of *iVeES* methodology is depicted using the schematic diagram in Figure 1.1. In the diagram, VUI at level one represent the client interface for connecting to the system. The various techniques and tools shown in the legend of Figure 1.1 were engaged in actualizing the objective of the system. The main contribution of the study – Adaptation and Recommendation are at level two and three respectively from the top of Figure 1.1. The box in each diagram represents a rectangular box with three compartments labelled Artefact, technique and tools. Technique and tools are the methods and development tools used to analysis, design and implement the system.

Tools	Platform
VoiceXML	Voxeo Speech Server
PHP, MySQL	Apache
OWL	Protege

Artifact
Technique
Tools
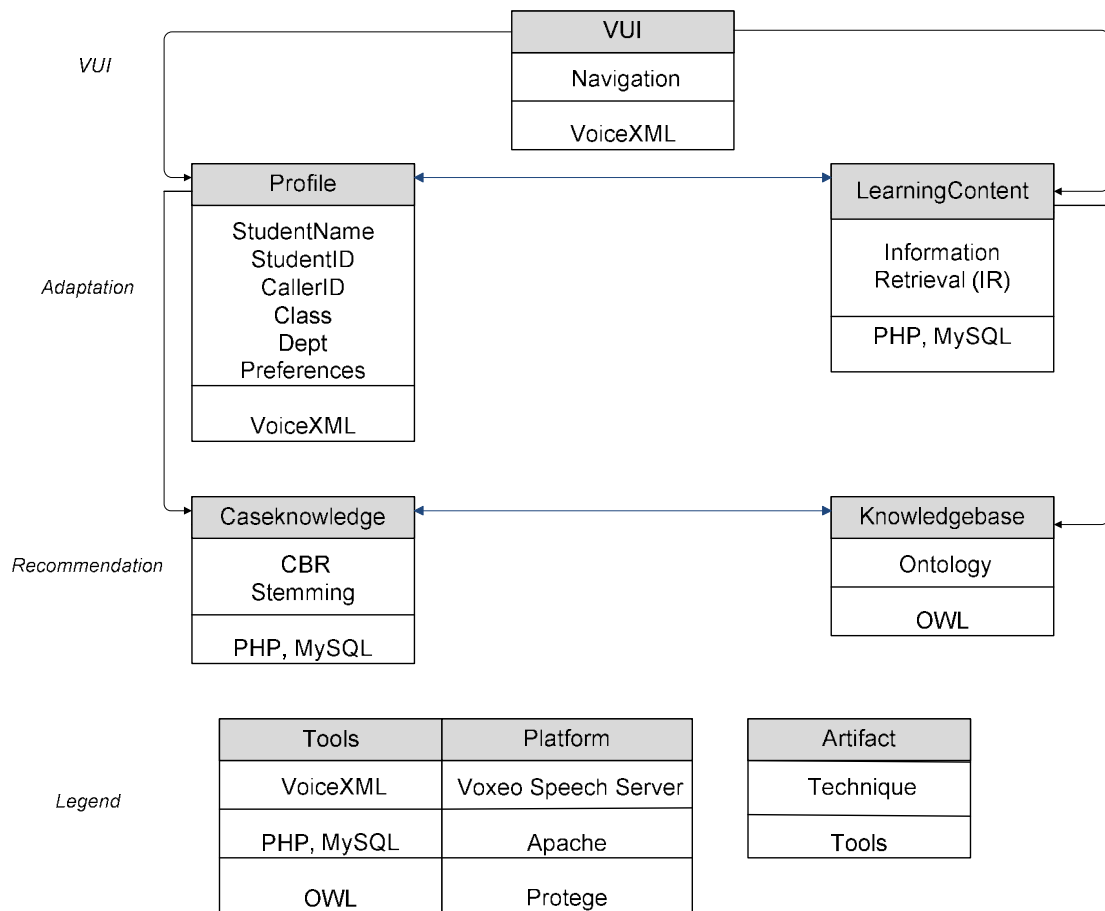
Figure 1.1: Core of *iVeES* methodology

CBR has been used in this study because we desire a system that can reason from experience to save the time of learners attending to large number of students. The ratio of students to teachers in most institutions is appalling especially in developing countries such as Nigeria. A system that will not only provide solution but recommend a viable solution is most appropriate to realize the aim of this research. The reason for using Porter Stemming algorithm is that the suffix stripping process will reduce the total number of terms in the Intelligent Reasoning (IR) system, and thus reduce the size and complexity of the data in the system, which is always advantageous (Singh, 2003). A smaller dictionary size of words results in a saving of storage space and processing time.

The tools used for the implementation of the prototype e-Education application are: VoiceXML for the Voice User Interface (VUI), PHP for Web User Interface (WUI), Apache as middle-ware, and MySQL database as back-end. The speech engine used was provided by voxeoprophecy (2003) for application development and testing. The choice of these tools is

due to their advantage as Free and Open Source Software (FOSS) (Siemens, 2003). VoiceXML's wide acceptance as a standard, huge industry uptake, suitability for multimodal interaction, and increased developer productivity clearly demonstrate its superiority over other tools for developing voice applications (Farrans, 2004).

The VoiceXML application was deployed on a Voxeo voice server (Datamonitor, 2006) on the web and accessed from a mobile phone, while a prototype part of the web application was deployed on a web server on the Internet (Divakaran, 2011) and was accessed using an Internet ready computer.

iv)   Evaluation of intelligent Voice-based e-Education System using ISO 9241-11(1998) and Nielsen (2003) usability criteria. Statistical Package for Social Sciences (SPSS) was used to generate the usability analysis and reports.

1.6   SIGNIFICANCE OF THE STUDY

This research work is significant in a number of ways:

i)   The framework will be useful for researchers in intelligent voice-enabled e-Education domain.

ii)   The e-Education application will be useful especially for the physically challenged students such as the visually impaired. Also, the mobility impaired who are immobile as a result of some forms of disability; and people who have medical history of reacting to repetitive strain injury (RSI) as a result of their sedentary lifestyle using the keyboard.

iii)   The e-Education system offers an alternative platform of learning without any physical disability. It complements the existing mobile and web-based learning system.

iv)   The e-Education system offers the possibility to learn anytime and anywhere there is telephone access.

v)   The application can reduce the expenses associated with the traditional face-to-face learning method which is common in most developing countries such as lecture halls and other learning delivery facilities associated with physical presence between the students and teachers.

vi) The application can provide economies of scale at tertiary institutions, as cost of each additional learner is negligible once the lecture materials have been developed and hosted in a central server.

vii) In solving the digital illiteracy problems common to developing countries, the voice-based e-Education application can improve accessibility to education, including distance learning for learners who are visually impaired in the school for the blind. By so doing, the target group will not be completely neglected in the scheme of promoting ICT in education and learning. Loss of sight is one of the most difficult disabilities to come to terms with. The assistive technology provided in this study has the ability to fundamentally change the way teaching and training is delivered to the students of the school for the blind.

1.7 CONTRIBUTIONS TO KNOWLEDGE

This study offers the following contributions to knowledge:

i) Provision of a reference model (framework) for intelligent voice-based e-Education applications which can be adapted to e-Education related project.

ii) Provision of an intelligent voice-enabled e-Education system (*iVeES*) for ubiquitous learning that is particularly suitable for the visually impaired and other forms of disabilities such as repetitive strain injuries (RSI), high spinal injury, etc, that make reading and writing difficult.

1.8 LIMITATIONS OF THE STUDY

The physically challenged learners includes learners that are visually impaired, partially sighted, mobility impaired and deaf. The main focus of this research is to provide a framework and voice-enabled assistive learning technology for the first two – the visually impaired and partially sighted learners using Nigerian English accent as a case study. To the normal learners (non-visually challenged), the study also serves as a complement to the existing learning methods such as PC, web and mobile learning.

The primary domain focus of this research is in the area of Education. An educational institution (a school for the blind) in Nigeria was considered as case study. The

components of the developed application that was used to test the *iVeES* framework includes: 1) Registration, for course registration by students, 2) Voice lecture, where audio streams of lecture notes were received by students through voice, 3) Tutorial, for receiving tutorial questions and answers, 4) Examination, for conducting examination questions and quiz, and 5) Result, for checking of examination results. However, intelligence was only applied to items 2 Voice lecture and 3 Tutorial. All the modules were accessed through voice interaction with the system.

The *iVeES* provided as a test bed in this research does not apply to classroom environment where the teacher would have to be physically present in a class with the students simultaneously receiving lectures through interactive voice response. The utilization of speech interaction by a mobile learner is provided and could significantly improve experience of his/her "learning on the go". Most frequently, these types of application utilize a constrained vocabulary of user commands and a constrained grammar of possible user entries (Kondratova, 2009). In *iVeES* just as with other voice applications, restricted grammar and vocabulary were engaged.

This research did not consider the development of speech engine comprising text-to-speech (TTS) or speech-to-text (speech synthesis). An existing free Voxeo speech server (Datamonitor, 2006) was used as our speech engine, although the limitations of using Voxeo were taken into consideration. Studies involving design and implementation of TTS or Automated Speech Recognition (ASR) can be found in Odejobi and Ogwu (2005) and some other researches.

1.9   GLOSSARY OF TERMS AND DEFINITIONS

The following are the basic terms used in e-Education and Speed technology (Cook, 2003):

m_Learning: m_learning is also known as mobile learning. It is a type of learning that enables the learner to move from one place to another and learn using a mobile device.

Case-Based Reasoning: CBR is a technique for solving new problems based on experiences of previous solutions.

e-Learning: The use of ICT to support learning processes.

H*idden Markov Model (HMM):* HMM are used for making probabilistic models of sequence analysis.

Speech: A speech is a sound signal used for language communication. Superficially, the speech signal is similar to a sound produced by a musical instrument, although it is more flexible and varied. When we speak, we push air from our lungs through the vocal chords, sometimes tightening the chords to make them vibrate as the air passes over them.

Utterance: An utterance is the vocalization (speaking) of a word or words that represent a single meaning to the computer. Utterances can be a single word, a few words, a sentence, or even multiple sentences.

Vocabularies: Vocabularies (or dictionaries) are lists of words or utterances that can be recognized by the speech recognition system. Generally, smaller vocabularies are easier for a computer to recognize, while larger vocabularies are more difficult. Unlike normal dictionaries, each entry does not have to be a single word. They can be as long as a sentence or two.

Accuracy: The ability of a recognizer can be examined by measuring its accuracy − or how well it recognizes utterances. This includes not only correctly identifying an utterance but also identifying if the spoken utterance is not in its vocabulary. Good ASR systems have an accuracy of 98% or more! The acceptable accuracy of a system really depends on the application.

Training: Some speech recognizers have the ability to adapt to a speaker. When the system has this ability, it may allow training to take place. An ASR system is trained by having the speaker repeat standard or common phrases and adjusting its comparison algorithms to match that particular speaker. Training a recognizer usually improves its accuracy. Training can also be used by speakers that have difficulty speaking, or pronouncing certain words. As long as the speaker can consistently repeat an utterance, ASR systems with training should be able to adapt.

Corpora: A speech corpus (or spoken corpus) is a database of speech audio files and text transcriptions in a format that can be used to create acoustic models (which can then be used with a speech recognition engine). A corpus is one of such a database.

1.10   STRUCTURE OF THE THESIS

This study explores a novel framework for intelligent voice-based e-Education system (*iVeES*) and it provides a prototype application. In chapter two of the report, the researcher provides Artificial Intelligence (AI) background information and review of relevant literature. The chapter also addresses the concept of voice-based systems in e-Education and usability measurement criteria.

The follow-up section – chapter 3 discusses the proposed framework and theoretical modeling of the components that make up the framework as well as architectural design of the system. Chapter 4 contains the systems design and implementation based on the proposed framework using Education as a test bed. User interfaces, testing and system deployment were also described. The system evaluation report is presented in chapter 5. The attributes used for the system evaluation was enumerated. Chapter 6 is the summary and conclusion of our findings. This section also includes recommendations on the need for government and other stakeholders to provide an enabling environment and support for voice-based e-Learning for both normal and visually impaired. Areas for future research are also highlighted in this chapter.

1.11   SUMMARY

In this chapter, the background and motivation of the study are described. This is followed by a discussion of the statement of the problem. Thereafter, the aim and objectives, methodology, significance, contribution and limitations of the study are presented. Finally, the glossary of terms and structure of the thesis are presented.

CHAPTER TWO

REVIEW OF RELEVANT LITERATURE

2.1   INTRODUCTION

In this chapter, the concepts of Artificial Intelligence (AI) and e-Learning content management approaches that are relevant to the work are discussed. A theoretical framework and enabling technologies of e-Education and voice technology including artificial intelligence, information retriever, e-Education, voice-based and usability are also presented.

To bring to bear the several methodologies and tools that aided in the conduct of this research and content development of this thesis, it is necessary to describe the techniques that were engaged in the study. This is necessary for a proper understanding of the review of previous but related works done in the area of e-Education. AI concepts are among the concepts that require a description and discussion.

2.2 AI CONCEPTS AND TERMINOLOGIES

AI is a system that mimics the behavior of humans. Its ultimate goal is a theory of intelligence that accounts for the behavior of naturally occurring intelligent entities, and this guides the creation of artificial entities capable of intelligent behavior (Abraham, 2005). AI is a term that in its broadest sense would indicate the ability of a machine or artifact to perform the same kind of functions that characterize human thought.

The term AI has also been applied to computer systems and program capable of performing tasks more complex than straightforward programming, although still far from the realm of actual thought. Several intelligent computing technologies are becoming useful as alternate approaches to conventional techniques or as components of integrated systems (Kalogirou et al., 2009).

2.2.1 Case-Based Reasoning

Case-Based Reasoning (CBR) is a subfield of artificial intelligence (Leake, 2000). CBR is a general AI paradigm for reasoning from Experience (Abdel-Badeeh et al., 2005). CBR is one of the techniques used for enabling recommendation services in AI projects. It is the process of solving new problems based on the solutions of similar past problems. The CBR paradigm has a framework which includes: case acquisition, case representation, case Indexing, case search, case retrieval and case reasoning.

There have been several research contributions in the application of CBR in e-Education. For instance, with CBR, students always tend to learn better than their peers that use the conventional methods of e-Learning (ie those without intelligent component services such as conventional web-based learning, m-Learning, etc). Moreover, teachers feel they are able to reach more of their students with CBR methodology (kolodner et al., 2005). Hisayoshi et al., (1999) developed a prototype system based on CBR for educational information database which supports teachers' activities of learning and teaching, especially concerning the domain of information technology education. Tsinakoset el al. (2003) presented a web-based automated process of replying to students' questions using CBR. Other available AI techniques includes (Abraham, 2005; Kalogirou, et al., 2009): Expert Systems (ES), Artificial Neural Networks (ANNs), Genetic Algorithms (GA), Fuzzy Logic (FL), Logic Programming (LP), Natural Language Processing (NLP), Computer Vision (CV), Robotics, Machine Learning (ML), and Hybrid Intelligent Systems (HIS). Hybrid systems are combinations of two or more of the branches mentioned previously.

2.2.2 Information Retrieval

The field of information retrieval (IR) is concerned with organizing and retrieving information from text stored in Computer (Chowdhary, 2005). IR is the science of searching for documents or information within documents. Intelligent IR has been variously defined by different people, but a consistent theme has been that of the machine (or program) doing something for the user, or the machine (or program) taking over some functions that previously had to be performed by humans. That is, a program which takes a query as input, and returns documents as output (Belkin, 2005).

IR systems engage Vector-Space Model (VSM) using Term Frequency-Index Document Frequency (*TF-IDF*) which is a matching technique for textual information that provides a measure of semantic equivalence between documents. The method consists of determining a numerical representation of the text, which results from a calculation of keywords and their frequencies in the text. Several other Information Retrieval (IR) techniques exist as follows (Abdel-Badeeh et al., 2005): Induction technique, retrieval using nearest-neighbor technique and stemming algorithm.

2.2.3   Stemming Algorithm

For grammatical reasons, documents use different forms of word, such as *organize*, *organizes*, and *organizing*. Additionally, there are families of derivationally related words with similar meanings, such as *democracy*, *democratic*, and *democratization*. In many situations, it seems useful for a search of one of these words from the received documents to contain another word in the set. The goal of stemming is to reduce inflectional forms and sometimes derivationally related forms of a word to a common base form (Term, 2008).

In the searching and retrieval of information, terms in the domain knowledge are usually identified by their word stems. Stemming reduces the storage requirements by decreasing the number of words maintained. Porter's stemming algorithm is probably the most widely used stemmer in IR research. The Porter stemming algorithm removes the commoner and morphological endings from English words. It does not usually matter whether the stems generated are genuine words or not – thus, "computation" might be stemmed to "comput" (Selouani et al., 2006). Examples of products using stemming algorithms are search engines such as Lycos and Google, and also thesauruses and other products using Natural Language Processing (NLP) for the purpose of IR (Singh, 2003). Some of the most popular stemming algorithms includes: Paice/Husk Stemming Algorithm, Porter Stemming Algorithm, Loving Stemming Algorithm, Dawson Stemming Algorithm and Krovetz Stemming Algorithm (WhatisStemming, n.d.)

2.2.4   Vocabularies and Ontologies

Vocabulary is the collection of words which the person knows and is personal to that person and he/she makes use of them every day. Vocabulary can always be enhanced. A vocabulary is analogous to a dictionary. It identifies all the possible attributes in a schema, which is similar to a database schema. The ontologies list the words and the relationships between them that allow the communication with other entities.

The term ontology has four compartments (Rajpathak et al., 2001). They are: the top-level upper-level) ontology, the task-related ontology, the application-related ontology and the domain-related ontology, organized in a hierarchy of the ontologies. Some examples of these types of the ontologies are: Cyc ontology, WordNet ontology and EuroWordNet ontology (these are lexical ontologies) and Sowa's ontology. A second category of the ontologies is the task-related ontology that consists of an ontology dedicated to some specific tasks. The application-related ontology is the third category and is used for specific application. Lastly, the ontologies dedicated to an area are called domain-related ontologies or simply domain ontologies and they are specific to a field. An example of this type is the ontology dedicated to the fields of education. Ontologies can be used to mediate the search of the learning materials on the Internet (Moise and Netedu, 2009). One that is of significant interest is the project Ontologies For Education (O4E) (Dicheva et al., 2005). They present the expectations of the usage of ontologies in education. Ontologies have been widely applied by researchers in the education domain. Dicheva et al., (2005) attempt to provide a general snapshot of the state-of-the-art research in the field of applying and developing ontologies for education.

There are standard languages and a variety of commercial and open source tools for creating and working with ontologies (Liu, 2008). One of such popular tools is protégé 3.4.1 which has been widely applied to e-Education.

2.3   THE GENERATIONS OF WEB TECHNOLOGY

The World Wide Web Consortium (W3C) have created some web technologies discussed as follows (Naik, 2008; Web2, 2010):

The Web – Web 1.0

Web 1.0 is popularly referred to as "The Web" and was developed in 1996 by W3C.
In web 1.0, a small number of web developers created web pages for several users. This made it possible for people to get information by going directly to the source. The WWW or Web 1.0 is a system of interlinked, hypertext documents accessed through the Internet (Naik, 2008).

The Social Web – Web 2.0

The Web 2.0 is the social Web and it was developed in 2006. The term "Web 2.0" is commonly associated with web applications that facilitate interactive information sharing, interoperability, user-centered design, and collaboration on the World Wide Web (Web2, 2010). With the conventional Web 1.0, users are limited to the passive viewing of information that is provided to them, as against the Web 2.0 that allows interaction, edit and update of the content of the website. Examples of Web 2.0 include web-based communities, hosted services, web applications, social networking sites, video-sharing sites, wikis and blogs (Web2, 2010).

The Semantic Web – Web 3.0

Web 2.0 is about social networking and mass collaboration with the blurring of lines between content creator and user whereas Web 3.0 is based on "intelligent" web applications using (Web2, 2010):

- Natural language processing
- Machine-based learning and reasoning
- Intelligent applications

The goal is to tailor online searching and requests specifically to users' preferences and needs. Although the intelligent web sounds similar to artificial intelligence, it's not quite the same. New technologies will make online search more intelligent and may even lead to a web 4.0. The Semantic Web and Intelligent Web were proposed for 2016 (Naik, 2008).

2.4   E-EDUCATION TECHNOLOGIES

E-education involves e-Learning and online education. It is the system of learning where computers are used at each stage of learning process including: enrollment, instruction design, content delivery, evaluation, assessment and support. This means that students and teachers do not have to sit in the classroom but instead they learn from anywhere in the world and any time. The different forms of learning environment can be categorized into five levels. They include:

Face to Face: The traditional form of education is the first level, here; teachers and students have to be physically present in a classroom setting. Face to face interaction is the most prevalent means of communication between teachers and students.

Distance Leaning: The second level is by distance education. With the development of communication and transportation technologies in the industrial age, teachers and students are separated by space and time. Communication between teachers and students is achieved by using traditional mail, phone, radio and television.

E-Learning: Electronic revolution and the invention of Internet led to e-Learning, which is the third level. Different types of media like text, audio, video, hypertext, simulations and two-way communication have been used commonly in education. web-based learning is one of the most prevalent types of learning as a result of the widespread use of Internet globally.

M-Learning: The mobile learning (m-Learning) constitutes the fourth level and it involves the use of mobile communication tools like mobile phone, pocket PC, etc. It is a learning paradigm that takes place anytime, anywhere with the help of a mobile device. The advantages and disadvantages of m-Learning and other learning systems are discussed by Devinder and Zaitun (2006).

Voice Learning (v-Learning): V-Learning is the fifth level and it entails the use of mobile phones or landline phone to access learning contents in the Internet or Intranet anytime and anywhere by dialing a telephone number. A typical telephone web e-Education application provides learning materials that can be accessed via the web and telephone. Some students

have used speech recognition systems successfully for their studies and for exams, and the use of this technology has helped them to overcome their difficulties and proceed to higher education (Paul, 2003).

Voice-based learning is a type of "assistive technology" used by the physically challenged. The World Wide Web Consortium (W3C) defines assistive technology as software or hardware that has been specifically designed to assist people with disabilities in carrying out their daily activities (Adaptive, 2005). These technologies aid the learning process for learners with disabilities. People who are partially sighted have difficulty accessing e-Learning systems due to small print or the inability to see the position of text blocks clearly on the screen. For the visually impaired, e-Learning systems are often inaccessible due to the nature of the process requiring sighted information.

2.5   ASSISTIVE TECHNOLOGY

People who are blind or have visual impairment often face information accessibility barriers. However, advances in computer technology are increasingly lowering these barriers (Modukuri and Morris, 2004). Assistive technology is a platform used by individuals with disabilities in order to perform functions that might otherwise be difficult or impossible (AccessIT, 2006). Assistive technology can include mobility devices such as walkers and wheelchairs, as well as hardware, software, and peripherals that assist people with disabilities in accessing computers or other information technologies. For example, people with limited hand function may use a keyboard with large keys or a special mouse to operate a computer, the blind may use software that reads text on the screen in a computer-generated voice, people with blurred vision may use software that enlarges screen content, the deaf may use a TTY (text telephone), or people with speech impairments may use a device that speaks louder as they enter text via a keyboard. Several types of assistive technology exist as follows (Raghuraman, 2004; Modukuri and Morris, 2004):

- **Alternative keyboards or switches:** Alternate keyboards or switches are hardware or software devices used by people with physical disabilities that provide an alternate way of creating keystrokes that appear to come from the standard keyboard. Examples

include keyboard with extra-small or extra-large key spacing, key guards that only allow pressing one key at a time, on-screen keyboards and eye gaze keyboards.

- **Braille and Refreshable Braille**: Braille is a system that uses six to eight raised dots in various patterns to represent letters and numbers that can be read using the fingertips

- **Screen magnifiers:** A screen magnifier is software used primarily by individuals with poor vision that magnifies a portion of the screen for easier viewing. At the same time screen magnifiers make presentation larger; they also reduce the area of the document that may be viewed.

- **Sound notification:** Sound notification is an alternative feature of some web-based applications that allow hearing impaired users to receive visual notification that a warning or error sound has been emitted by the computer.

- **Screen readers**: This refers to software used by individuals who are blind or who have Dyslexia, which interprets what is displayed on a screen and directs it either to speech synthesis for audio output, or to refreshable Braille for output.

- **Speech recognition**: Speech (or voice) recognition is used by people with some physical disabilities or temporary injuries to hands and forearms as an input method in some voice browsers.

- **Scanning software:** Scanning software is an adaptive software used by individuals with some physical or cognitive disabilities that highlights or announces selection choices (e.g., menu items, links, phrases) one at a time.

However, the aforementioned computer-based assistive technologies have the following problems (Modukuri and Morris, 2004):

- lack of availability and cost of updating older computers, software and hardware equipment;

- lack of Internet access or limited access to the Internet;
- lack of training in the use of computers and/or lack of confidence in the use of such equipment;
- problems in communicating with computer instructors and/or technical advisors regarding functioning of the computer.

As a way of advancing technological support from the current assistive technology to utilize the numerous technologies available on the Internet, several tools and standard are provided over time by different organizations to facilitate vocal interfaces and telephone access to the web.

2.6   VOCAL INTERFACES

The explosive growth of the Internet and wide spread use of mobile devices has forced providers of voice-based applications to offer their services through the web. Domains such as e-health, e-Learning, e-trade, m-trade or e-administration are some of the sectors for which simple and efficient vocal interactions are used (Rouillard and Truillet, 2005). Several standards for vocal interfaces have been growing in popularity including: VoiceXML, SALT, and eXtensible HyperText Markup Language (XHTML)+Voice (X+V). All of them aim at supporting multimodal applications (Jankowska, 2007).

2.6.1   Voice eXtensible Markup Language (VoiceXML)

Voice eXtensible Markup Language (VoiceXML) is based on the Extended Markup Language (XML) standard. It was developed by the VoiceXML Forum (a group founded by AT&T, IBM, Lucent Technologies, and Motorola) in 1999 (Abbott, 2002; W3C, 2005b; ZDNet, 2010). VoiceXML is also known as VXML. It is one of the tools for developing voice-enabled e-Education applications. It is a web-based markup language for representing human-computer dialogs, just like the HyperText Markup Language (HTML). But while HTML assumes a graphical web browser, with display, keyboard and mouse, VoiceXML assumes a voice browser with audio output (computer-synthesized and/or recorded), and audio input (voice and/or keypad tones) (Gallivan et al., 2002). VoiceXML technology allows a user to interact with the Internet through voice-recognition technology by using a

voice browser and/or the telephone. The major goal of VoiceXML is to bring the advantage of web-based development and content delivery to Interactive Voice Response (IVR) system (voiceportalwhitepaper, 2001).

2.6.2. Speech Application Language Tags (SALT)

SALT standard, created by the SALT Forum23, was designed to embed voice into existing web formats (e.g. WML, XML, HTML or XHTML and enables multimodal applications (SALTForum, 2005). Users interact with applications by providing information using speech, keyboard, keypad, mouse, or stylus. The output is generated as synthesized speech, audio, plain text, motion video, or graphics. Each of these modes can be applied independently or concurrently.

SALT includes a set of tags with associated attributes and Document Object Model (DOM) properties, events, and methods. It provides Dual-Tone Multi-Frequency (DTMF) and call control capabilities for browsers running voice-only applications. A typical SALT architecture consists of a web server, a telephony server, a speech server, and a client device. The web server generates pages containing markup language elements (e.g. from HTML or WML), SALT, and embedded scripts. These scripts control the dialog flow for voice interaction. The telephony server is responsible for connections to the telephone network and incorporates a voice browser interpreting the documents. The speech server is able to recognize speech, plays audio prompts, and sends responses to the user.

2.6.3. XHTML+Voice (X+V)

XHTML+Voice (X+V) (W3C, 2005a) adds voice interaction to traditional visual web interfaces. XHTML and XML events technologies are integrated with XML vocabularies developed as a part of the W3C Speech Interface Framework. X+V comprises voice modules that offer support for speech synthesis, speech dialogs, command control, and speech grammars. Event handlers that implement VoiceXML actions can be attached to XHTML elements.

Some of the popular companies and tools engaged in developing voice applications include: Speech Application Programming Interface (SAPI) of Microsoft, Java Speech markup Language (JSML) by Sun Microsystems, Voice Markup Language (VoxML) by Motorola and Microsoft Windows Vista (Kouroupetroglou and Mitsopoulos, 2000; Kurschl et al., 2007).

2.7 VOICE-ENABLED WEB PARADIGM

The advent of fast and accurate voice recognition in the late 1990s and the desire to provide self service solutions accelerated the move to self-service systems (Markowitz, 2005). The conventional practice is web content with links for navigating from one page to another, using Internet browser. Voice recognition technology has been able to integrate IVR, VoiceXML and the web together with voice portals and servers.

2.7.1 Voice Portals

A portal is a web system with functionalities to authenticate and identify users as well as provide them with an easy, intuitive, personalized and self-customizable web interface for facilitating access to information and services that are of primary relevance and interest to the users (HKU, n.d.). A voice portal is a web site that provides voice-enabled functionalities. A portal that is not voice-enabled has limited access to the Internet resources (Azeta et al., 2008a).

2.7.2 Voice Servers

Voice servers, also known as speech servers or speech engines provide middleware functionality between the users and application/database content. In the voice business value chain, some vendors that provide voice or speech server services include: Voxeo, BeVocal, VoiceGenie, Audium, Clarity, Entervoice, HeyAnita, Microsoft, Plum Voice, Tellme, VoiceObjects, Voxpilot, VoiceShot, Avaya, Cisco Systems, Envox Worldwide, Convergys, Dimension Data, Fluency Voice Technology, Genesys, HP, ICT Group, Intervoice, Nortel, Nuance, RightNow, Tuvox and West Corporation (Datamonitor, 2006; Palowireless, 2009)

2.7.2.1   Discrete and Continuous Speech Recognition Systems

Until the late 1990's, commercial speech recognition systems were "discrete" in nature (Kirriemuir, 2003). This required the user to speak in an unnatural, command-like manner, with a gap between each word (to allow time for the computer to process the input and to make it clear to the computer when individual words ended). Not surprisingly, these speech recognition systems were more oriented towards command-based applications. Entering and editing a large document, such as a book chapter, was very time consuming. The rapid increase in desktop computer processing power and better algorithms led to the development of "continuous" systems. These allowed the user to speak at near-normal speed, and still (after sufficient training) obtain a high accuracy rate. Contemporary continuous systems can give an accuracy of 90%-95%, depending on training (though this still means that one in ten or twenty words could be incorrect).

At the heart of a speech recognition software is the translation part. Most speech recognition software breaks down the spoken words into phonemes, the basic sounds from which syllables and words are built up. These are analyzed to see which string of the units best "fits" an acceptable phoneme string or structure which the software can derive from its dictionary. Since the early 1970s, a number of different approaches to Automatic Speech Recognition (ASR) have been proposed and implemented, including Dynamic Time Warping (DTW), template matching, knowledge-based expert systems, neural nets, and Hidden Markov Model (HMM) (Ehsani and knodt, 1998).

2.7.2.2   Hidden Markov Model (HMM)

The use of the Hidden Markov Model (HMM) for speech recognition was first proposed by Baker (Abu-Zaina et al., 1996). It has since become the predominant approach to speech recognition, superseding the Dynamic Time Warping (DTW) technique. It has been found that the overall recognition accuracy and performance of the HMM is comparable to that obtained using a DTW algorithm with much less computation and storage requirements. The most common and successful speech recognition methods are based on statistical modeling, and especially on HMM (Aarnio, 1999). In Aarnio's work, the theory of HMMs and the related algorithms were reviewed, and then applied to speech

recognition. Roux et al., (2010), describes the implementation of an HMM-based text-to-speech synthesis system (HTS) in a module devoted to the development of literacy and reading skills. It describes an ongoing international consortium project entitled Mobile E-Learning for Africa (MELFA), which integrates a text-to-speech (TTS) engine as part of the e-Learning platform. The MELFA project focuses on communicative aspects related to South African English as well as to an indigenous African language, isiXhosa, which is one of the eleven official languages of South Africa.

2.8   ARCHITECTURE FOR VOICEXML-BASED SYSTEMS

A VoiceXML platform is the foundation for developing and operating voice automation application (Rouillard, 2007). During the human-computer interaction, it executes the commands and logic specified by applications written in VoiceXML. VoiceXML standard is a scripting language based on Extended Mark-up Language (XML) and is used to design voice interactive information access systems supported by natural language and text-to-speech processing tools for interpretation and transcribing information. The voice-based services are designed and hosted in a web server using server side technologies that generate VoiceXML scripts for the information to be delivered, while the VoiceXML interpreter platform renders this information at a certain cost from anywhere by dialing a toll free telephone number.
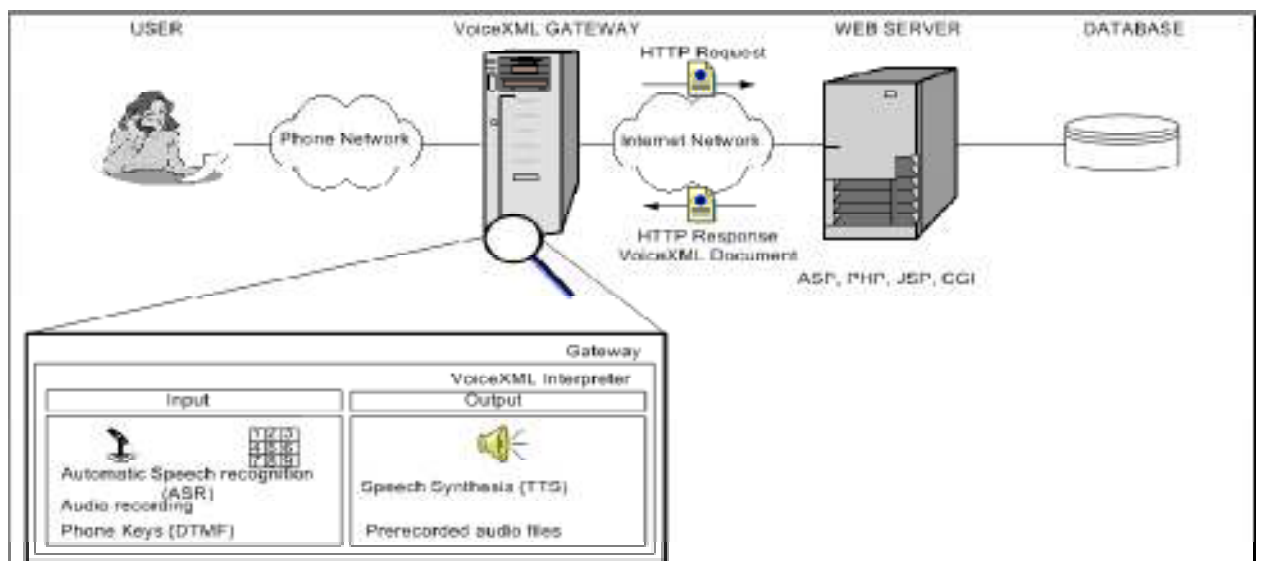
Figure 2.1: VoiceXML platform architecture   Source: (Rouillard, 2007)

Voice applications can now be developed based on VoiceXML platform technologies (VXML, 2004). The VoiceXML platform also provides the speech processing capabilities (speech recognition, speech synthesis, voice authentication, etc). As shown in Figure 2.1, VoiceXML platform architecture is based on hypertext transmission protocol (HTTP), and uses both phone and Internet. The web server is often connected to a database which the user can query, retrieve information from and update its content. VoiceXML is an extension of XML that defines voice segments and enables access to the Internet via telephones and other voice-activated devices. Voice technology, and, especially VoiceXML, is the enabling technology that allows people to interact with computers over the telephone using natural speech dialogues. VoiceXML technology allows people to access the web from anywhere they can make a call, thus providing increased access to goods and services (Kondratova, 2004). This family of markup languages takes also advantage of the power of the web to provide web access to people not being able to read and write or those that are visually or mobility impaired.

VoiceXML Gateway: The VoiceXML gateway is responsible for communicating with the user over the telephone and performs telephony tasks. The VoiceXML gateway pulls VoiceXML, grammar and "wav" files from the web server over the Internet or a dedicated link based upon application requirements. The voice browser in the VoiceXML gateway collect speech input from the user over the telephone and presents to the inner components of the VoiceXML gateway for processing. The voice browser also presents the output to the user in the form of spoken words over the telephone. Note that the VoiceXML code itself is neither generated nor stored by the VoiceXML gateway. The VoiceXML gateway fetches the VoiceXML code from the web server over the Internet. In some cases, voice server and application servers can co-exist in one server in order to reduce dependency on the Internet link between the two servers (Qureshi, 2007).

Voice Browser: A voice browser is the human-machine interfacing component of any speech driven voice application. A voice browser provides an interface between the caller and the different components of the voice server. The voice browser is just like the HTML browser which interprets the application server code and presents to the user and accepts inputs (HTML forms) from the user for processing by the server.

From architectural point of view, the main difference between a web and a telephone application is that in a web application, the browser is provided by the client machine. As a result of that, application designer can rely on client resources in using technologies like Javascript and flash to present and accept the data. The voice browser, on the other hand, is part of the VoiceXML gateway and does not reside on the client, which is the calling device (telephone) for voice applications. Also, efforts should be made to keep the voice session as short as possible for a given task.

Automatic Speech Recognition (ASR): ASR is called speech recognition engine, its main purpose is to convert human speech into the text form that the application can understand. It interprets callers spoken words and makes decisions based upon the spoken language and application rules defined in the grammar files.

Text-to-Speech (TTS): Also called TTS, a Text-to-Speech system provides exactly the reverse of the speech recognition process: that is converting text to speech. A TTS system announces text provided by the application to the caller in the form of spoken words.

Web Server: The web server houses the database and the necessary codes that are used to communicate with the user of the application and perform tasks related to business logic. The application can either reside in the web server or hosted in a separate server.

VoiceXML Grammar: A VoiceXML grammar identifies different words or phrases that a user might say, the recognized vocabulary and syntax are decided by the programmer. A grammar is the core of a VoiceXML application, since it determines what the user is able to say, or not. Speech grammar provides a template for valid user input. The speech recognition grammar format (SRGF) is embedded within VoiceXML for specifying speech grammar. Grammars can range from a single word or a list of possible words to form a sentence. A vocal grammar could be given in the body of the VoiceXML script (inline grammar), or in separate file (external grammar). Here, the important thing to understand is that user can only pronounce a sentence predefined in the VoiceXML application's grammar. For example, it means that it will be impossible for the computer to understand the word "attendance" or "matric", for instance, because those utterances are not available in the grammar. We have two types of grammar, the static grammar, which has been described

above using inline or external grammar. Dynamic grammar involves producing the content of the grammar file dynamically using the database. It means one can add or edit the content of the grammar files during program execution. VoiceXML requires specific skills in writing grammars and complex interactions for good usability to avoid bad user experiences with the understanding that users are not "visualizing" the content. An example of a VoiceXML document is contained in Figure 2.2.

```
<?xml version="1.0"?>
<vxml version="2.0" xmlns="http://www.w3.org/2001/vxml">
  <form>
    <block>
      <prompt>
        Hello students!
      </prompt>
    </block>
  </form>
</vxml>
```
Figure 2.2: Example of VoiceXML document

When interpreted by a VoiceXML interpreter this will output "Hello students".

Phone Network: One of the advantages of voice technology is that it uses already built infrastructure, including telephone lines and equipment. The telephone lines can be standard analog, digital Telephone 1 (T1), as well as digital ISDN lines. These lines are connected on one hand with the VoiceXML platform and on the other hand with the standard equipment – switches of the telephone company and VoIP gateways (Krastev et al., 2006).

2.9   REVIEW OF VOICE-ENABLED SYSTEMS

Voice-enabled systems allow users to access information on the Internet or Intranet through a telephone interface. It uses technologies such as speech recognition and text to speech (TTS) conversion to create a user interface that enables users to navigate through a dialogue system using telephone and voice commands (Gallivan et al., 2002). Voice technology has been widely applied in developing applications across different domains including: information providers ie stock, flight (Mult, 2004); financial institutions (Azeta et al., 2008b); e-health

and e-education. Voice is also applicable in the following areas: Customer service such as Cisco voice portal (Cisco, 2010), Envox voice portal (Datamonitor, 2006)), Customer Relationship Management (Atayero et al., 2009a), Job Recruitment (Azeta et al., 2008a), and a lot more.

Modukuri and Morris (2004) reviewed a voice interactive system design utilizing VoiceXML technology. Voice-based web services can be used as an alternative to other existing forms of assistive technology for use by visually impaired, as well as other forms of disability. For example, the visually impaired access news from a variety of local, regional or national newspapers and the only equipment required is a touch-tone telephone with no requisite knowledge of personal computers or specialized software.

A framework was proposed by Zambrano (2004) for the development of web applications that can be accessed from different types of devices, such as PCs, PDAs, mobile phones based on diverse technologies (like WAP and I-Mode) and conventional telephones that access the web through voice gateways and voice portals. The proposed framework serves as a guide for the development of this type of applications and it can be deployed to different server configurations and software development technologies. The framework was implemented using transportation sector as a test-bed.

The objective of the research by Liao and Li (2007) was to provide a VoiceXML-based solution for integrated dialogue management and realizing such a mechanism in a collaborative Multi-User Virtual Environment (MUVE). A dialogue scripting language called eXtensible Animation Markup Language – Voice Extension (XAML-V), based on the VoiceXML standard was designed, to address the issues that may affect applications such as dialogue management, etc, for multi-user interaction. The language was used to realize MUVE and evaluate the effectiveness of the system design. The research by Matzon (2005) described the implementation of a speech-driven receptionist. The receptionist was designed to be used by smaller Swedish companies. It answers calls coming into the company and direct the calls to an employee based on speech input from the user. It was programmed in VoiceXML and ColdFusion.

The study by Kenny et al., (2007) outlined the development of a voice-enabled application, which implemented a Locus of Control (LOC) scale specific to the domain of adolescent career development and choice. A 20-item scale representing expectations related to "Internality, Luck, Helplessness, Powerful and Others" were proposed as beneficial in the career guidance process as a means of indicating the extent to which pupils would engage and participate in carrier development process and assume responsibility for their own future. The newly developed Career LOC voice application allows greater accessibility and faster results than the traditional paper or web-based methods. According to the preliminary research conducted, it was believed that VoiceXML can indeed lead to improved system interface interaction, enabling people on the move to easily navigate computerized systems. Jumar et al., (2006) described the development of the first Slovak spoken language dialogue system. The dialogue system consists of the Galaxy hub and telephone, automatic speech recognition, text-to-speech, backend and VoiceXML dialogue management modules. It enables multi-user interaction in Slovak language. Its functionality was demonstrated and tested via two pilot applications: weather forecast for Slovakia and Timetable of Slovak Railways. Slovak is the language spoken in Slovakia, a country in Eastern Europe.

In co-operation with the city administration of Cottbus parliament, Fiedler et al., (2004) developed an e-Democracy portal SeSAM system which enables politicians as well as interested citizens to always be up-to-date in current parliamentary discussions. Depending on user's rights, the user has a role based access to the data of the members of parliament, the committees or parties. A user can read bills and motions that are currently under discussion and he/she can retrieve agenda of certain meetings. In addition to the web Interface for SeSAM, the authors developed a prototype for a voice-based access to the SeSAM functionality. The system interface has support for communication and coordination among the traveling parliamentarians. By making a simple phone call, the parliamentarians can track changes of appointments or agenda, and they can receive or send personal messages.

An alternative to traditional methods of interacting with a desktop computer using a speech recognition system was presented by Kirriemuir (2003). It uses voice input through a microphone, such that it can reduce the reliability on standard keyboard and mouse input. A speech recognition system based on desktop computers consists of the following: (i) a microphone, for the person to speak into, (ii) a speech recognition software, (iii) a computer

33

to take and interpret the speech, and (iv) a good quality soundcard for input and/or output. Using this method, a person can "speak", via a microphone to a computer. The computer translates the spoken words into either text, usually for automatic insertion into items such as word processing documents, or commands to execute functions in the computer e.g. print a document, open a file, etc.

To address the problem of large deployment of mobile applications, Fabbrizio et al., (2009) introduced the AT&T speech markup framework, an approach to speech services that leverages web services and cloud computing to make it easier to combine web content and speech processing. It was shown by Fabbrizio et al., (2009) that this new compositional method is suitable for integrating automatic speech recognition and text-to-speech synthesis resources into real multimodal mobile services.

E-education has received a growing popularity. Speech-enabled devices including Talking Scientific Calculators, Talking Dictionaries, Talking Braille Note-Takers, and talking (word) Translators have been successfully applied to reading/writing in e-Education (Freitas and Kouroupetroglou, 2004).

2.10   REVIEW OF EXISTING E-EDUCATION FRAMEWORKS AND SYSTEMS

This section discusses existing frameworks and systems in the area of e-Education. A framework provides reference guideline for developing applications. Existing framework for developing voice-based e-Education system such as VoiceXML application development life-cycle (VoiceXML, 2007) does not have an intelligent component services and it was modeled after the conventional software development life-circle model. However, VoiceXML applications developed using such framework serve as an alternative platform for non-visually impaired users.

In addition to the provision of alternative platform for normal users, voice-enabled e-Education systems can be helpful for people with physical access difficulties (e.g. repetitive strain injury, arthritis, high spinal injury) that make writing difficult (Donegan, 2000). It can also be effective for students with reading, writing or spelling difficulties, such as dyslexia (Husni and Jamaludin, 2009) and for those with visual impairment (Nisbet and Wilson, 2002).

Paul (2003) described some important factors to be considered when introducing and using speech recognition for students with disabilities. It was reported by Paul (2003) that speech recognition programs have been used successfully by students with a range of disabilities. The programs have been shown to provide an effective means of writing and recording work, and in some cases, they produced significant improvements in basic reading, spelling and writing skills. Chin et al., (2006) recommended that one can actually make use of VoiceXML technology to build speech applications that can serve educational purposes or in other words, to build an online learning system that provide better accessibility to users. One of the e-Education applications that can be provided using speech technology are those that deliver basic teaching by simply listening. For example, students can check their scores or other information by simply calling a particular number and get the information they want.

Several schemes for providing e-Education system have been proposed in literature. Chin et al., (2006) went further to develop an e-Learning prototype based on VoiceXML concepts. However, the prototype was not deployed on the real telephone environment. Instead, a microphone and a speaker were used to simulate the telephone environment.

Development of voice applications using VoiceXML for higher institutions of learning have remain an open area of research all over the world. As part of a larger project, a couple of tools for computer assisted language learning environment were developed (Kirschning, 2001). These tools are a first prototype for pronunciation verification and a bilingual dictionary, both for Spanish language students, whose native tongue is American English. Another very interesting aspect of the project is the use of speech technology-based systems to support language acquisition for deaf children.

Similarly, Selouani et al., (2006) developed a web-based learning tool called LINC (Learn IN Context). The learning material is delivered in the form of a collaborative and voice-enabled presentation. This mode combines face to face and distance approaches to education where an instructor meets with students in the classroom, and a resource-base of content material is made available to students through the web. LINC+, the voice enhanced version of LINC were described. The authors reported an evaluation experiments of LINC/LINC+ in a computer programming course taught in University de Moncton, Campus of shippagan, Canada. The findings showed that when the learning material is delivered in the form of a

collaborative and voice-enabled presentation, majority of learners seems to be satisfied with the new concept. Furthermore, Schindler (2005) developed a chat application for communicating between the deaf and blind. The goal of the project was to incorporate current speech recognition (speech-to-text) and speech synthesis (text-to-speech) technology into a chat room that is both free and does not require any additional equipment besides a desktop computer. The system was developed using C++ and run on client/server technology with a graphical user interface (GUI) for the client. The system could also be used in educational settings, regardless of students' or teachers' disabilities, as a teaching aid.

Walsh and Meade (2003) presented the results of their experience by developing an experimental prototype for speech based literacy e-Learning and application architecture. In many developing countries, adults are functionally illiterate, that is, many people have difficulty in reading and writing everyday documents. The authors address this problem by allowing users to interact with speech enabled e-Learning literacy content using multimodal interfaces. In the same vein, Shin and Hong (2006) implemented a novel architecture of e-Learning system using VXML linked with PSTN (Public Switched Telephony Network). The newly proposed system supports telephone-based learning as well as computer-based learning. Learners can choose their learning tools from the existing state of things at each time. Learners can choose computers as a learning tool when they are in the house, on the other hand, learners can choose their telephones as a learning tool on the street

Gallivan et al., (2002) developed a voice-enabled web-based absentee system on TellMe voice portal. It was tested by the class of software engineering students; students who intended to miss a class called the VoiceXML telephone number and were led through an automated dialog to record their names, the date and time of the call, the courseID, and the date they would miss class in a database. The system provides the instructor and other administrators with a permanent record of absentees that can be accessed and displayed in various forms through a web interface. The absentee system application was developed basically for Pace University students to report class absences. The VoiceXML absentee system has been designed to include record keeping of absentee calls from students, faculty and university staff. This system made use of the general VoiceXML platform architecture in Figure 2.3.
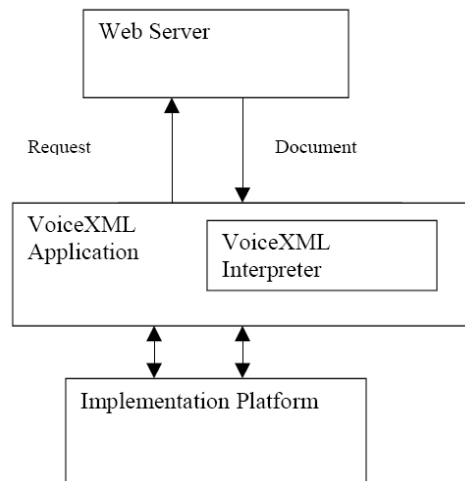
Figure 2.3: Architecture for voice-enabled web-based absentee system
Source: (Gallivan et al., 2002)

Likewise, Raghuraman (2004) designed, implemented and deployed a voice-enabled application called V-HELP system. In the V-HELP system, a portion of the Computer Science and Engineering (CSE) department website was voice enabled using VoiceXML to enable the visually impaired student population to access the information on it. The website also provides easy access to the information through voice, while the user is on the go, and if the user does not have a handheld with sufficient screen size or is constrained by the bandwidth to browse through the visual content. It can be used by visually impaired student population of the school to access information about the Computer Science and Engineering department. Though the primary purpose of this project is to utilize information technology in the area of assistive technology, it can also serve students, who do not have Internet connection to browse the text based website or who do not have access to a computer. Since this web portal can be accessed through a phone call, students will also be able to access CSE department information while they are on the go. Similarly, Kolias et al., (2004) described the design and implementation of an audio wiki application accessible via the Public Switched Telephone Network (PSTN) and the Internet for educational purposes. The application exploits World Wide Web Consortium standards such as VoiceXML, Speech Synthesis Markup Language (SSML) and Speech Recognition Grammar Specification (SRGS). The purpose of the application was to assist visually impaired, technologically uneducated, and underprivileged people in accessing information originally intended to be accessed visually via a Personal Computer. Users access wiki content via wired or mobile phones, or via a Personal Computer

37

using a Web Browser or a Voice over IP service. In its current form, the application supports pronunciation of the content of existing articles via synthetic speech. It does not support editing or creation of new articles due to the incapability of existing recognition engines to provide full speech recognition instead of recognition of words from a predefined set.

Motiwalla and Qin (2007) explored the integration of speech or voice recognition technologies into m-Learning applications to reduce access barriers. An educational online forum accessible through mobile devices was built, based on an m-Learning framework proposed in Motiwalla (2007). A customized interactive voice response (IVR) and text-to-speech (TTS) technologies to allow users to interact with the forum through voice commands was also developed. This voice-enabled discussion forum application does not only help normal users to avoid the cumbersome task of typing using small keypads, but also enables people with visual and mobility disabilities to engage in online education. The prototype forum was tested in two different blind institutions in Massachusetts with 10 users. The results from this study provide insights into how to improve accessibility of m-Learning. The research utilizes the framework in Figure 2.4 to develop a set of mobile phone accessible applications that can be used for the m-Learning environment.

	Personalized Content	Collaborative Content	
PUSH Mechanism	*Pedagogical Agents & Mentors*	*Communication Aids*	*SMS, IM, Alerts, Scheduling Calendars*
PULL Mechanism	*System Tools & Resources*	*Simulated Classrooms*	*WML websites, Discussion Boards & Chat Forums*
	Alerts, Scheduling Calendars, WML websites	*SMS, IM, Discussion Boards & Chat Forums*	**M-learning Applications**

Figure 2.4: An m-Learning Framework. Source: (Motiwalla, 2007)

The architecture in Figure 2.5 utilizes the m-Learning framework in Figure 2.4. The architecture contains 3 layers: back-end layer, middle-ware layer, and user interface layer. The back-end layer consists of core web applications such as forums, Web blogs, Wikis, as well as e-Learning applications such as WebCT, Blackboard, etc. The back-end also consists of the web/application server that supports those applications. The back-end provides source materials and basic services of the whole m-Learning system. The middle-ware layer consists of components that enable the users to access back-end e-Learning applications

through various means. For example, the speech engine may contain customized IVR and TTS technologies or commercial speech recognition software that allow users to interact with the e-Learning applications through voice commands. The mobile application server component enables users to interact with the e-Learning applications through WAP or SMS on their mobile devices. Finally, the user interface layer resides on user devices. Users can access the e-Learning applications on their PCs, telephones, or mobile devices. They can do so by typing messages in text browsers or issuing voice commands.
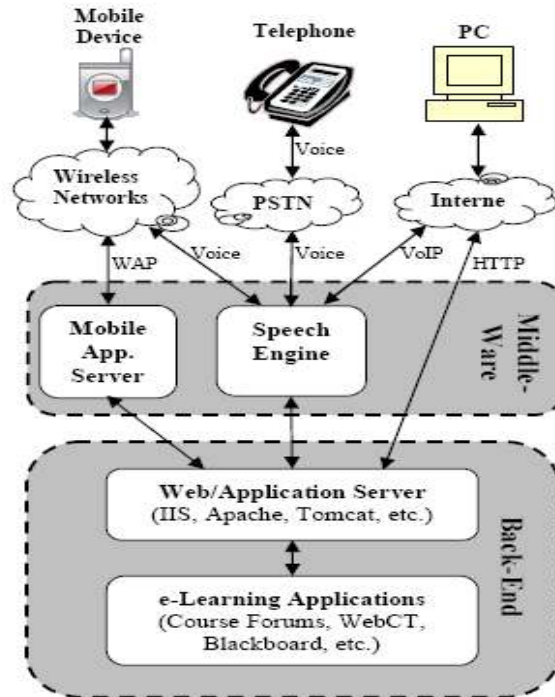
Figure 2.5: M-Learning Architecture with speech recognition enhancement.
Source: (Motiwalla and Qin, 2007)

Also, Azeta et al., (2008c) carried out a preliminary study of the *iVeES* project on the design and development of a prototype telephone-based e-Learning portal for course registration and examination by the students. Jianhua et al., (2009) employed the technology of VoiceXML to construct an open VoiceXML-based Mobile Learning System (VMLS). The hardware and software system structure design was achieved. The research results could not only be used in VMLS, but could also be used in other voice-based application system. Peter et al., (2008) presented an architecture to enable multi-channel interaction (Web, SMS/MMS and voice) with services that support collaboration on a blog server. The use of blogs to support learning can be found in some works by Lin et al., (2006) and Divitini et al., (2005).

The motivation for this work was to enable learners to access collaborative learning services regardless of their situation and location. Kondratova (2009) discusses issues associated with improving usability of user interactions with mobile devices in mobile learning applications. The focus is on using speech recognition and multimodal interaction in order to improve usability of data entry and information management for mobile learners.

A Voice-enabled Interactive Service (VoIS) was proposed by Motiwalla (2009) for e-Learning. The goal of VoIS platform for an e-Learning system is to better meet the needs of students and mobile learning communities via speech recognition technology. The pilot study explored this goal by demonstrating how an interaction tools like discussion forums can be voice-enabled using VoiceXML. The author designed, developed and evaluated a prototype application with blind and visually impaired learners from different organizations.

The study by Garcia et al., (2010) introduced ''Voice Interactive Classroom'', a software solution that proposes a middleware approach to provide visual and auditory access to web-based learning. The deployment architecture of the Voice Interactive Classroom is depicted in Figure 2.6. The approach has successfully enabled the integration of visual interfaces to voice dialogues and to provide auditory access to functionalities already present in various learning management systems such as Segui, Moodle and Sakai.
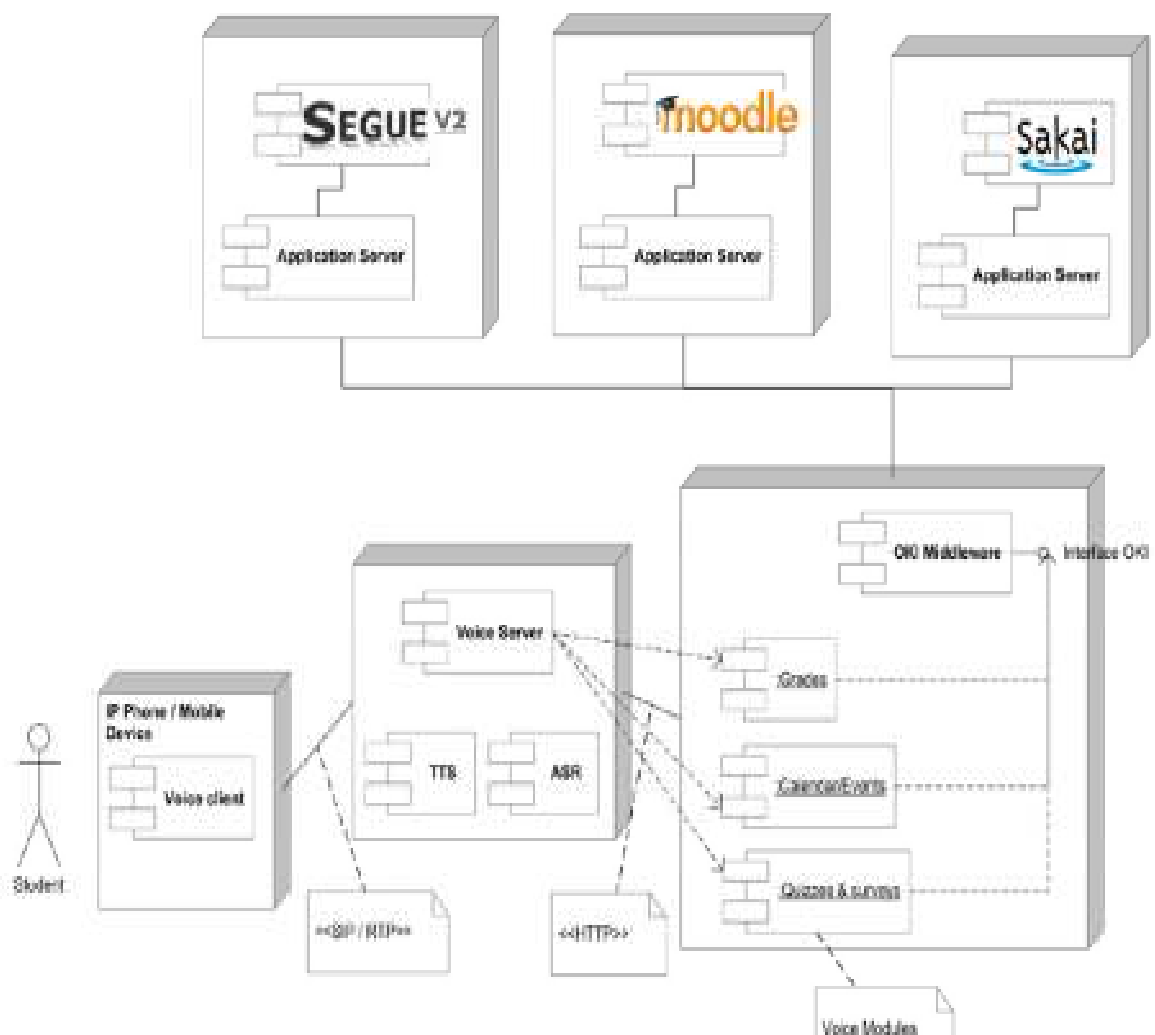
Figure 2.6: Deployment Architecture of the Voice Interactive Classroom
Source: (Garcia et al., 2010)

The educational contents and the framework in these systems lack support for intelligence such as the ability to take decisions that will adapt to users' request. Thus, we believe that if finding suitable content of learning resources can be introduced into a voice-based e-Learning system, including CBR and ontology-based properties, then the voice-based e-Learning system will have the capabilities of providing adaptable and intelligent learning to users.

This study also addresses some of the concerns expressed by Devedzic (2004), where it was noted that there are several challenges in improving web-based education, such as providing for more intelligence, and that a key enabler of this improvement will be the provision of ontological support. Other studies such as Holohan et al., (2005); Aroyo and Dicheva,

(2004); Aroyo et al., (2003) have used ontology to provide e-Learning Content Generation based on Semantic Web Technology. Meanwhile these approaches were not based on voice architecture.

2.11   CLIENT DEVICES FOR VOICE-BASED SYSTEMS

The client devices for the telephone-based application comprises of land phone, mobile phone and smart phone.

Land Phones: A landline or fixed-line is a telephone line which travels through a solid medium, either through metal wire or optical fibre. This is distinguished from a mobile cellular line, where the medium used is wireless and airwaves. Landlines usually cost less than cellular lines and are used where mobility is unnecessary or where the service of a cellular phone is not available. Even as mobile devices are shrinking in size, however, their capability and content are changing into more complex and diverse functionalities to meet user requests (Kwang and Roger, 2006).

Mobile Phones: The features common to modern mobile phones are basic contact, calendar and scheduling features (Howard, 2006). Several mobile phones provide additional features like a simple web browser that uses WAP technology (Kim et al., 2006), tools like a calculator and the ability to send SMS messages. They have small screens, typically monochrome or able to display 4096 colors at 96x68 pixel resolution (Nokia, 2009).

Smartphone: A Smartphone combines the functions of a cellular phone and a handheld computer in a single device. It differs from a normal phone in that it has an operating system and local storage, so users can add and store information, send and receive email, and install programs to the phone as they could with a PDA (Howard, 2006). A Smartphone gives users the best of hybrid features. It has much the same capabilities as a handheld computer and the communications ability of a mobile phone.

2.12   MEASURING THE USABILITY OF SYSTEMS

The evaluation of a product is a fundamental requirement in determining the practical usability of a product (Ikhu-Omoregbe, 2010). The usability of a system is measured to specify the features and attributes required to make the product usable using ISO's standard of usability (ISO 9241-11, 1998) as consisting of three distinct aspects:

Effectiveness is the accuracy and completeness with which users achieve certain goals. Indicators of effectiveness include quality of solution and error rates. *Efficiency* is the relation between 1) the accuracy and completeness with which users achieve certain goals; and 2) the resources expended in achieving them. *Satisfaction* is the users' comfort and positive attitudes towards the use of the system.

According to Nielsen (2003), usability can also be defined by the following five attributes: learnability, efficiency, memorability, errors and satisfaction:
- Learnability: how easy users accomplish their basic tasks the first time they encounter the design.
- Efficiency: how quickly can users perform a task using the interface of the system?
- Memorability: how easy is to memorize how to use the interface of the system and how easily users can reuse the system after a break?
- Errors: How many errors do users make using the interface of the system and how serious are these errors?
- Satisfaction: How do users like using the system's interface?

Methods of usability testing vary widely depending primarily on the amount of time and money available to the developer. Although there is a wide range of usability testing methods, the range of variables measured in such testing is more constrained. In a typical usability test, the tester wants to know how long it takes users to complete tasks, how many errors users make, how much help they need, users' subjective ratings of the system, and, if the application requires training, how much training was needed. To answer these questions, the tester conducts detailed observations of users completing tasks, recording all user interactions with the system (e.g., mouse clicks and keystrokes), when they occurred, what the user said during the test, request for help (e.g., references to a help manual, verbal

requests, etc.), facial expressions, and responses to post-surveys/interviews. These measures can be grouped into three categories: Errors, time, and subjective responses.

Errors are perhaps the simplest form of usability data to analyze because fewer errors almost certainly lead to improved usability. Time is often used to determine how long it takes users to complete tasks, find information, or answer questions related to the application under evaluation. The most critical (and the most frequently omitted) piece to assessing usability data is the context in which the data were collected and the type of users who participated in the usability tests. Finally, the usability of the learning environment is an important attribute because it influences the user's satisfaction, the ability to learn and to remember the content of the learning environment; in minimizing number of errors during interaction which leads to more effective and efficient learning (Ebibi et al., 2007).

2.13   SUMMARY

In this chapter, the components of artificial intelligence used in the research are discussed, as well as information retriever paradigm. Following this, the e-Education technologies involving traditional, distance learning, e-Learning, web and mobile voice learning were discussed. Furthermore, assistive technology, vocal interfaces, voice-enabled web paradigm, voice portals, voice servers and voice interfaces are presented. This is followed by the architecture for voice-based systems. Existing voice-based systems and voice-based frameworks in e-Education were reviewed. Finally, client devices for voice-based systems and the criteria for measuring the usability of systems were discussed.

CHAPTER THREE

THE INTELLIGENT VOICE-ENABLED E-EDUCATION FRAMEWORK

3.1   INTRODUCTION

This chapter presents requirement elicitation and *iVeES* framework. The components of *iVeES* framework including adaptation and recommendation services are discussed. An Education Ontology was designed using Protégé 3.4.1 and developed with Web Ontology Language (OWL). A VUI tree was designed for the interaction management component of the proposed framework. The chapter concluded with the Hidden Markov Model (HMM) framework.

3.2.   ELICITATION OF REQUIREMENTS

The development of a framework requires domain information from the potential users of the system. A school for visually impaired and non-visually impaired were visited to enable us acquire the required information. The investigation carried out in the visually impaired school shows that the conventional assistive equipment used are expensive to acquire and maintain. It is also difficult to organize the visually impaired learners into classes from their hostel without the assistance of attendants, and this makes the entire learning process to be cumbersome.

From investigation carried out in schools attended by the normal learners (i.e. students from non-visually impaired institutions), the existing learning methods in most public institutions in Nigeria are yet to sufficiently provide e-Learning access to students to enable them learn independently (i.e. on their own or on the move) irrespective of location. Students have to be physically present in the classroom to participate in class lectures. This does not favor most of the working class students who have to constantly leave the office for lectures, tutorial and examination. It is the opinion of students and management of one of the campuses/centers that a system that allows the students to learn on their own through distance learning would minimize the problem of over-enrollment of students and scarcity of teachers, among others in the centers.

The collated requirements and specification for the system include: registration of courses, audio streaming of lecture notes, tutorial questions and answers, course examination and checking of examination results. In order to activate the intelligent component services of the system, discussions were held with some students and teachers involved in the learning process to gather the required information that is relevant for CBR paradigm. The system requirement is presented in Figure 3.1, in which the outer circle is the actual needs of the learners. The inner circle comprise of Intelligent Voice-based e-Education System (*iVeES*), the platform upon which the framework was developed and deployed.
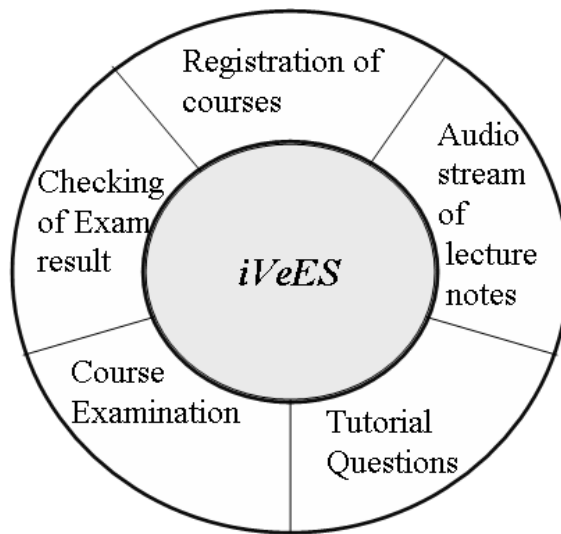
Figure 3.1: The system requirement for *iVeES* framework

The system will have two interfaces – the web and phone interface. The web interface will provide enrollment and access to the services provided by the application. The phone interface will allow users to call into the application to get service or record their request. The user must first enroll via the web prior to using the phone service. The user will enter some pertinent information and create a unique userId and password to subsequently enter the system. The information entered by the user during registration with the system include: studentName, studentID, callerID, class, department and preferences. The callerID is the telephone number where calls will be made from. These information should be entered through the web interface for the user to have access to the service components provided by the system. The system is to provide the following services:

- **Registration of courses:** Allow learners to registration for the courses he/she intend to offer.

- **Audio stream of lecture notes**: Allow learners to receive learning content through voice request and response, with support for content suitability.

- **Tutorial question:** Allow learners to ask tutorial questions and receive answers through voice, with support for suggestive functionalities.

- **Course examination:** Allow learners to participate in a multiple choice examination and quiz.

- **Checking of examination results:** Allows learners to dial and listen to the results of an examination that has been conducted.

3.3   THE FRAMEWORK

A user-centric design approach is adopted in order to develop a framework that is based on voice interaction. A survey was conducted to find out the requirements of an e-Education systems, and in particular that of the visually impaired. From the data collated from the survey questionnaire, significant differences were found between expert users, beginners and intermediate users. Expert users can do most of the typical tasks that users would normally do on the e-Learning application, participate in tutorials and check examination results. As was expected, beginners have little or no knowledge of some particular learning content.

Based on the information gathered, the following points were identified as major requirements for the framework and application: i) Ability to provide users with services involving course registration, voice lecture, tutorial, examination and result; ii) ability to accommodate different types of users based on their respective learning profile; iii) ability to differentiate the course contents into different levels such as expert, intermediate and beginners; and iv) selection of different course contents based on users' profile. The proposed framework for *iVeES* is shown in Figure 3.2.
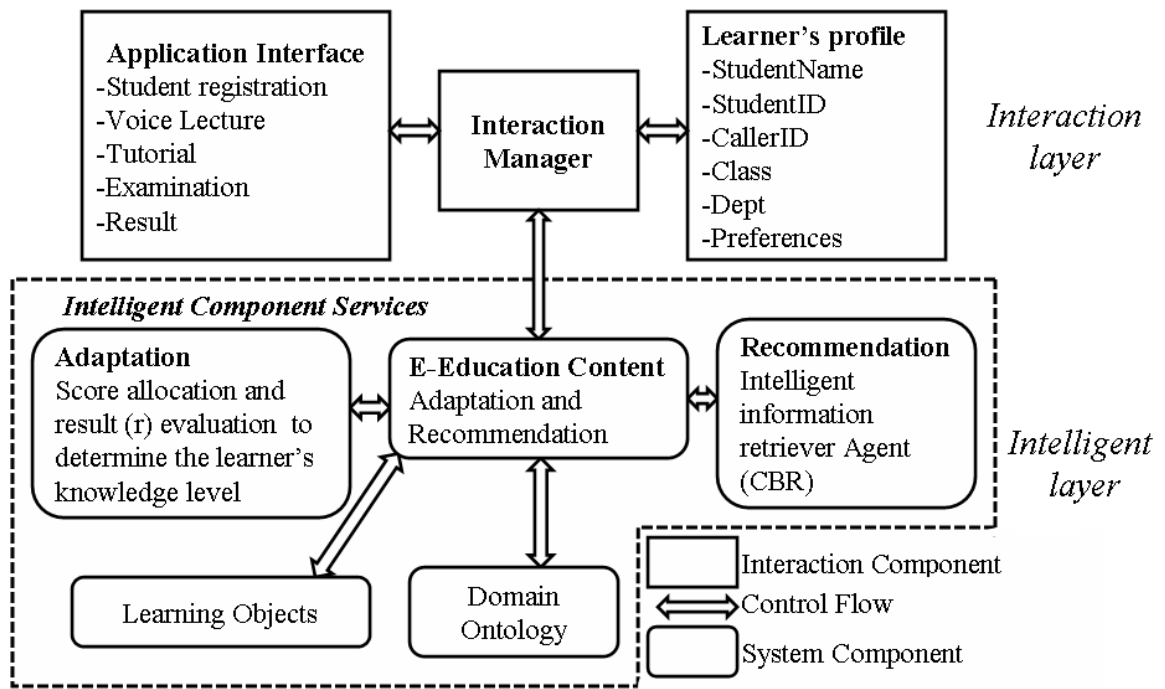
Figure 3.2: Intelligent voice-based e-Education system (*iVeES*) framework

The framework in Figure 3.2 comprises of interaction and intelligent layer. The intelligent layer provides services for the interaction layer. The system navigation process is contained in the interaction management. The classification as beginner, intermediate or expert level is done through content adaptation and recommendation using score allocation and result evaluation services. The intelligent information retrieval (IR) involving CBR and stemming is engaged to provide recommended answers to tutorial questions using previous experience of learners, and also to expand the search for answers using stemming algorithm.

Descriptions of the components that make up the *iVeES* framework in Figure 3.2 are discussed in the next sections. The components includes: Application interface, learning objects, learner's profile, Adaptation component service, Recommendation component service, Domain ontology and Interaction manager.

3.3.1 Application interface

The first point of contact when a user dials a telephone number to access the application is a welcome message. Thereafter, the system will navigate to the application interface which is the main menu comprising: student registration, voice lecture, tutorial, examination and result. The user is expected to say one of these options to get the desired service.

3.3.2   Learning objects

Learning objects consist of chunks of course materials in text format which allow information to be presented in several ways to users. Tutorial questions asked by previous learners and e-Education data are stored in case knowledge and domain ontology respectively.

3.3.3   Learner's profile

Learner's knowledge level is subdivided into three categories: beginner, intermediate and expert. Learner profile determines whether a learner should receive beginner, intermediate or expert content of lecture, tutorial, examination and result module. The learner's experiences and situation were captured in the learner's profile. By using this experience, the system is able to offer to the current learner the best suited learning content.

3.3.4   Adaptation component service

The adaptation component service uses the auto score allocation and result (r) evaluation model to determine the learner's knowledge level. The model is expected to create the most suitable course content for each learner and control the passage from one knowledge level to another. There is within a particular knowledge level, an activity (quiz) containing at least one question. Before moving from one knowledge level to another, the system must evaluate the learner's performance through set of evaluations. The evaluation criteria are represented in Figure 3.3.
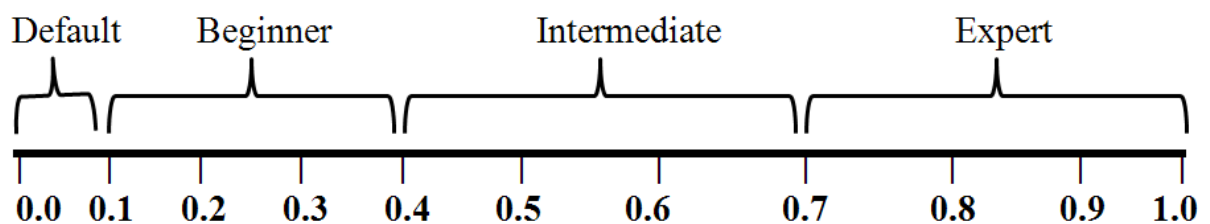
Figure 3.3: A model of score evaluation criterion

To be allowed to move from one knowledge level to another, the learner's result must satisfy the following transition criteria: $0.0 \leq R < 0.1$(Default), $0.1 < R \leq 0.4$ (Beginner level), $0.4 < R \leq 7$ (Intermediate level), and $0.7 < R \leq 1.0$ (Expert level). However, the

model operates under certain constraints as follows: If R < 0 then R = 0; If R > 1 then R = 1; If knowledge level = Beginner and R ≥ Intermediate upper limit, then R = Intermediate upper limit; and If knowledge level = Intermediate and R ≥ Expert level upper limit, then R = Expert level upper limit.

In the implementation of the model, collection of questions make up an activity and collection of activities make up knowledge level. Before a learner dial to access learning content, the system would set the default knowledge level to 0.0. Scores are assigned to each questions by dividing the maximum of score Max(r) obtained by the number of questions (Q) in each activity as shown in Equation 3.1. As the learner engage in the quiz/examination, the performance of the learner is determined by adding all the scores for the correct answers to the value of R and subtract the interval for knowledge level (I) from it to derive Y which is the evaluation result for each Activity(see Equation 3.2). The total value for a learner's knowledge level is computed by adding all the Activities result (Ys) together (see Equation 3.3).

Score allocated to each question:

$$Q = \sum_{k=1}^{q} Q_k \quad \text{and } S = \frac{Max\,(r)}{Q} \tag{3.1}$$

Evaluation result for each Activity:

$$Y = W - \left(\sum_{j=1}^{s} S_j \right) + R \tag{3.2}$$

Evaluation result for each knowledge level:

$$R = \sum_{m=1}^{y} Y_m \qquad \text{for } Y \geq 0 \tag{3.3}$$

The following definitions were made in formulating Equations 3.1, 3.2 and 3.3.

r = Expected result for score evaluation criteria (see Figure 3.3).

Max(r) = The largest value of r.

R = evaluation result for each knowledge level. Initial value = 0.

Q = total number of questions for each activity k where k = 1, 2, .., q.

S = score for each question j where j = 1, 2, .., s.

Y = evaluation result for each activity m where m = 1, 2, .., y.

W = Upper boundary of the next knowledge level.

Adaptation Model of *iVeES*

The general steps for adaptation are contained in the Unified Modeling Language (UML) sequence diagram of Figure 3.4. It was adapted from the study by Srimathi and Srivasta (2008). The learner's request for voice lecture based on learner's profile and the learner's characteristics are updated. The adaptation engine forms a rule and the rule is used to retrieve the adapted learning object. Thereafter, the suited content is provided to the learner through voice response.
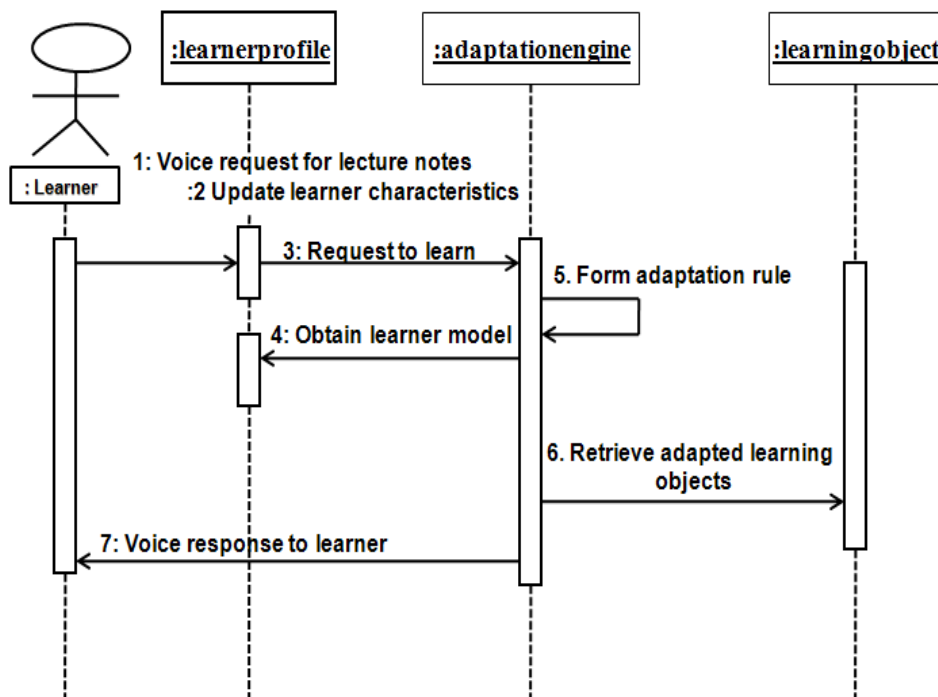
Figure 3.4: A UML sequence diagram for adaptation service

The flow of control for adapting learning content was modeled using Data Flow Diagram (DFD) as shown in Figure 3.5. From the initial state, a caller calls *iVeES* and the user's callerid is authenticated. The learner checks the learner's profile to determine the status of a

learner. This status is used to determine the user's learning content. The learning option is selected along with the course code. Thereafter, the learning content is retrieved. Each of the data flow names is represented by the following statement: Type of user=student, faculty and admin; Learning option = adapted learning, adapted quiz; Course code = Course to learn; and Knowledge level = Beginning, Intermediate and Expert.
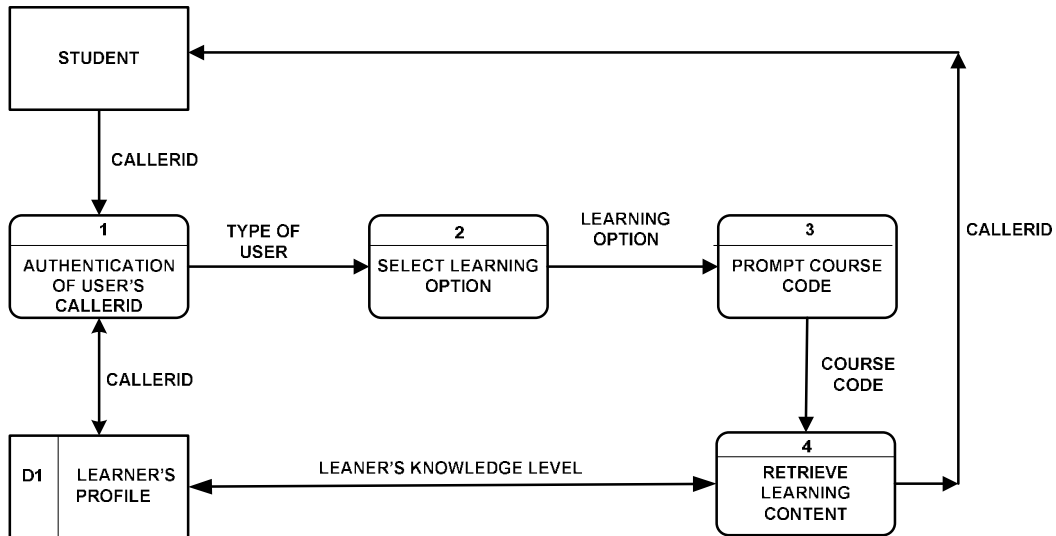
Figure 3.5: Data flow diagram for adapting learning content in *iVeES*

3.3.5 Recommendation component service

The recommendation component service is designed using CBR, stemming and domain ontology. One of the main challenges is to establish structures with which data can be represented as cases. This structure is presented in Figure 3.6. Alphanumeric data are stored as *text,* numeric data as *number* and date as *date* attribute. A typical data entry for the structure would be 011122, BIO200, "Introduction to Biology", "BiologyCase", "What is Biology", "Biology is the study of life and a branch of the natural sciences", "09-04-2010", "partial".

Case field	Attribute
StudentID:	number
CourseID:	text
CourseTitle :	text
CaseName:	text
Query:	text
Solution:	text
Date:	date
SearchMethod:	text

Figure 3.6: A Case structure for *iVeES*

CBR may be viewed as a cycle involving four steps (Spasic et al., 2005): *retrieve* the most similar case, *reuse* the case to solve the new problem, *revise* the suggested solution and *retain* the useful information obtained during problem solving, after the solution has been successfully adapted to the target problem. The retrieval stage is the first step. The procedure of case retrieval begins with identifying the most important features and using them in identifying cases to reuse. Additional domain information often improves results, i.e. a list of words and their synonyms or a dictionary provides comparable words. Our proposed system uses Ontology for Voice-based Education that represents specific knowledge, i.e., relationship between words used in Education (Biology) subject. The text tokenizer and Modified Porter Stemming (MPS) algorithm decomposes the whole textual information into sentences, and then into individual words with their stem for ease of retrieval of its synonyms. A Computation of the weight for every word and enhancement of term using ontology are required to improve retrieval effectiveness due to the huge amount of words.

Vector space models (VSM) is employed for performing query retrieval and assigning weights to the words in the query (Salton et al., 1975). Two main components of the weight are Term Frequency and Index Document Frequency (*TF/IDF*). *TF/IDF* estimates the importance of a term in a given document by multiplying the new term frequency (TF) of the term in a query by the term's inverse document frequency (IDF) weight. Weight of the ith term in j^{th} query is computed using the formula in Equation 3.4 (Kumar et al., 2006):

$$\mathrm{w}i,j = \mathrm{f}i,j * \log\left(\frac{n}{dfi}\right) \tag{3.4}$$

The weighting terms $W_{i,j}$ method calculates the weight of each term or word from the stored cases and the inputted user's query to perform further matching, where $f_{i,j}$ is the frequency of the ith term in j^{th} query; df_i is the number of queries that contain the word i and n is the total number of queries. To find the textual similarity between a stored case vector and a new case query vector, we apply cosine similarity function (Weber et al., 2005) for the textual information. This ratio is defined as the cosine of the angle between the vectors, with values between 0 and 1 and can be calculated using Equation 3.5.

$$\mathrm{Cos\theta}C_i = \mathrm{Sin}\,(Q, \mathrm{Ci}) = \frac{Q.Ci}{|Q||Ci|}$$

$$= \frac{\Sigma\, w_{q,j} w_{i,j}}{\sqrt{\Sigma_j w^2\, q,j}\, \sqrt{\Sigma_j w^2\, i,j}} \qquad\qquad (3.5)$$

where $Cos\theta\ Ci$ is the *cosine of the angle* between a stored case Q and a query case w. The dot product is calculated between the stored case and the query case by $Q.C_i$ where zero products are ignored; then vector lengths are calculated for a stored case and a query.

Recommendation Model of *iVeES*

Figure 3.7 shows the interaction between the various agents of the system using intelligent information retrieval agent (CBR and Stemming). The interaction follows the sequence of operation from 1 to 11 on the right hand side of the diagram.

1. Supply student query by voice
2. Match with case knowledge DB
3. Tokenized and stem query
4. Send reply to student by voice
5. Stemmed?
6. Check each token with term in domain knowledge DB for a match
7. Replace token with corresponding synonym
8. Add to unresolved special case DB
9. Faculty send reply to student by mail
10. Discard case
11. Enrich case knowledge DB
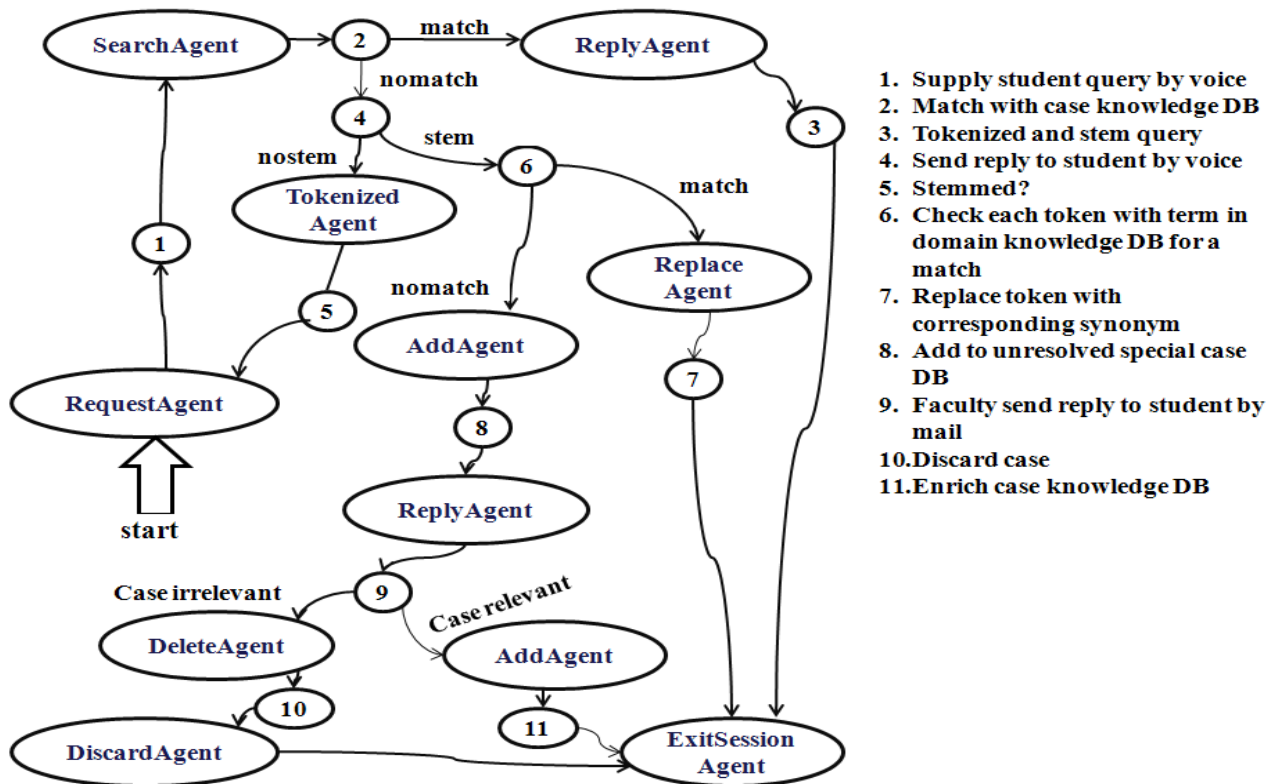
Figure 3.7: Recommendation Model of *iVeES* framework
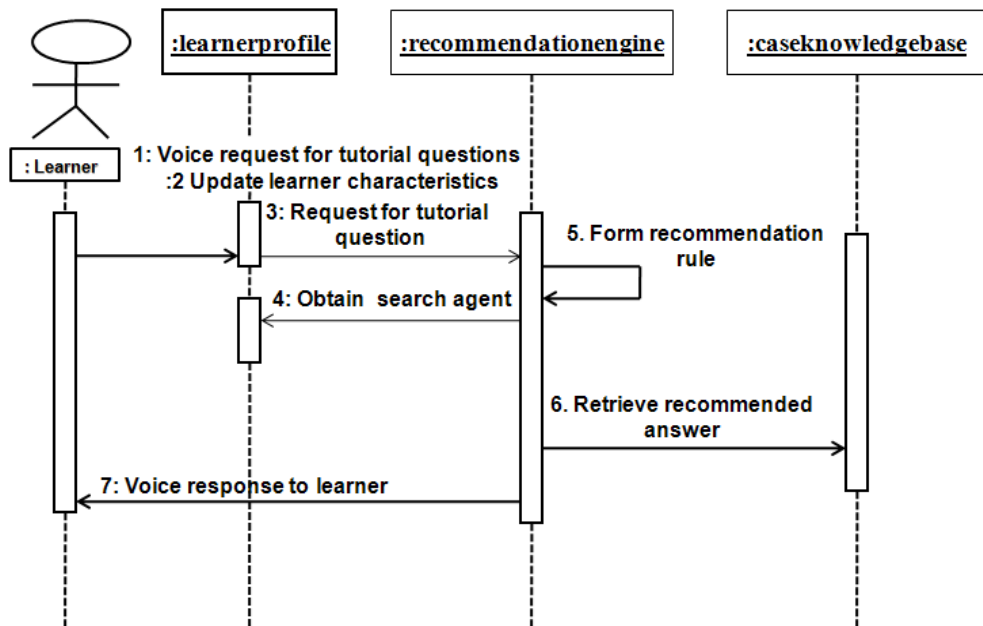
Figure 3.8: A UML sequence diagram for recommendation service

A UML sequence diagram for the recommendation service is shown in Figure 3.8. The learner initiates voice response for tutorial questions and the learners' characteristics are updated. The request for tutorial question is executed using search agent, which then provides the recommended answer using some rules. Thereafter, the recommended answer is passed to the leaner through voice response.

3.3.6 Ontology-based Model of *iVeES* Framework

Domain ontology is a detailed description about an application specific domain including concepts, entities, attributes and processes related to a given application domain. In this research, our domain is *Education Ontology*. The proposed ontology consists of e-Learning objects and their relationships. Therefore via these ontological relationships, the required learning resources can be located and their metadata can be retrieved. Then matching of user profiles and learning resource metadata can be done to ascertain the relevance of them. This ontology can be extended according to the changes to user profiles and learning content. When we try to satisfy learner requirements we need to match the user profiles with the learning objects in the system (Heiyanthuduwage et al., 2006). However, a direct matching between the learning objects and user profiles is not efficient enough and it will take lot of memory.

Enhancing the Term Vector using Ontology

Domain ontology is used to enhance the term vector. Figure 3.8 shows how the ontology helps to improve the weight vector. The dashes (----) in Figure 3.9 signifies continuation. Figure 3.9a contains a *Case query* extracted from *Education Ontology* comprising Biology lecture note, amongst others, and Figure 3.9b shows the *Domain Ontology*. The ontology is based on subject or course and it is named *Education Ontology*. The *Education Ontology* is represented in a tree form with several nodes (N). It is possible to treat each word of a case as a term which makes it easy to calculate the weight of each term for every case where terms of each case are satisfied with other case by exact matching, by synonym, antonym or hypernym. Meanwhile, some words or terms which have a complex relationship, for example, the term "nonexistence" and "organism" can be defined by ontology and the weight of those terms can be increased automatically for that case using the *Domain Ontology* created by the domain expert.

The weight of the vector terms can be enhanced for each case using the following conditions adapted from Begum et al., (2007):

Condition-1: If a term *TF* in the case is related to a term T_O in the ontology but the term T_O does not exist in the case, then the term T_O can be added as a *new* term with the same importance as the weight of the source term i.e. the score of *TF/IDF*.

Condition-2: If a term *TF* in the case is related to a term T_O in the ontology and also the term T_O exists in the case, then the strength of relationship between the term *TF* and T_O can be added to the original weight (i.e. score of *TF/IDF*) of those terms.

a.   Case query   b. Ontology
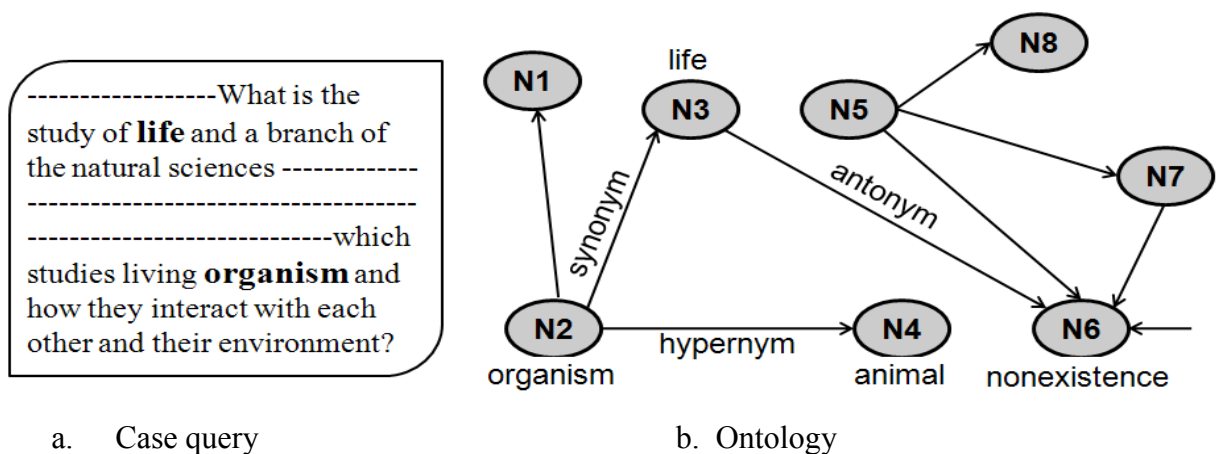
Figure 3.9: Enhancing the weight of a term using Ontology

The term "organism" appears both in the *Case query* and in *Ontology*, and has a relation with the term "animal" (N4). But the term "animal" does not exist in the *Case query*, so the term "animal" is important for this case and can be added according to condition 1. Again the terms "life" and "organism" both already exist in the *Case query* and have a relation in the *Ontology*, so the value of their strength of relationship for those two terms ("life" and "organism") will increase their importance (condition 2). Therefore, depending on the *Ontology*, the similarity measure of each term will be improved.

A number of tools have been used to design ontology including Unified Modeling Language (Cranefield et al., 2002). The e-Education repository ontology is shown in Figure 3.10 as a UML class diagram. Other ontology development tools include: TM4L-Topic Maps for e-Learning (Dicheva et al., 2005). In addition there is adequate ontology-processing tool support, e.g., Ontolingua (KSL, 2005), the Jena RDF Toolkit from Hewlett-Packard and the Protégé GUI ontology editor from Stanford University (Stanford, 2008; Holohan et al., 2005).
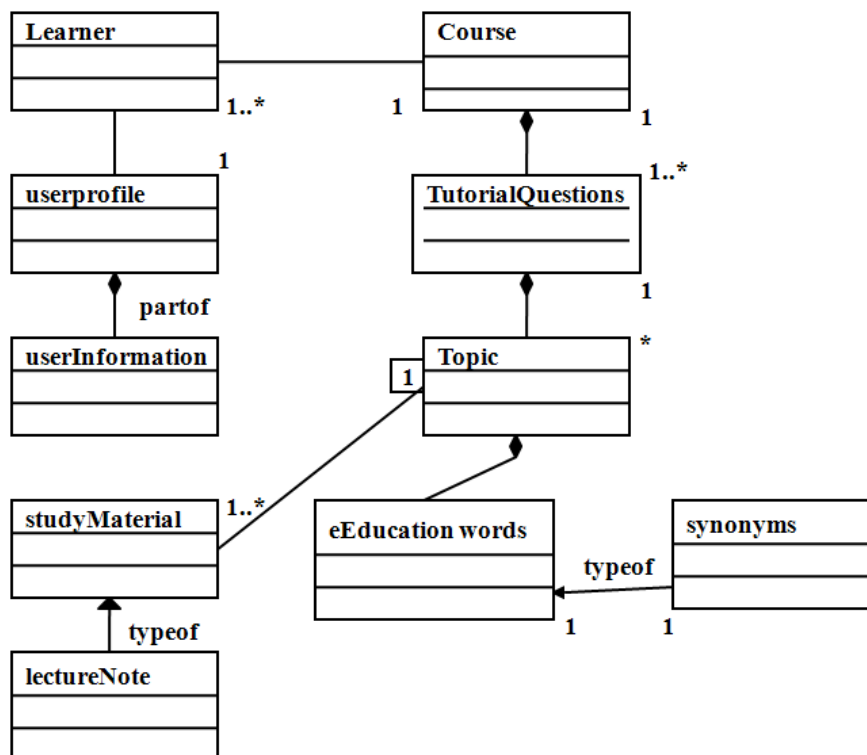
Figure 3.10: E-Education repository ontology as a class diagram

The Education Ontology development

Our first goal was to collect and structure the available information related to the use of ontologies in the field of education. In this study, we have to deal with subject domain ontology and not structure ontology. The *domain ontology* represents the basic concepts of the domain under consideration along with their interrelations and basic properties. The *structure ontology* defines the logical structure of the content. The Web Ontology Language (OWL) was used to design our ontology. We created an ontology named Education.owl and three classes that model separately the different components of the system. The ontology is encoded in OWL using Protégé 3.4.1. In the ontology there are classes, relationships and attributes related to the domain. New classes, relationships and attributes can be entered to the ontology e.g. by using Protégé 3.4.1. Each model can be extended to include as many relationships and classes as possible. The knowledge base resides in Apache HTTP Server. Figure 3.11 shows a snapshot of our Education ontology with the classes and properties. It has three classes; Learner, LearningContent and Assessment. A child node is an instance of a father node i.e. Course and Tutorial are instances of LearningContent. All the instances of Enumeration applies to all the classes. The ontology can be used for adaptive learning to retrieve the content of a course. The architecture of Education Ontology model is presented in Figure 3.12, with the three classes and the relationships between the classes.
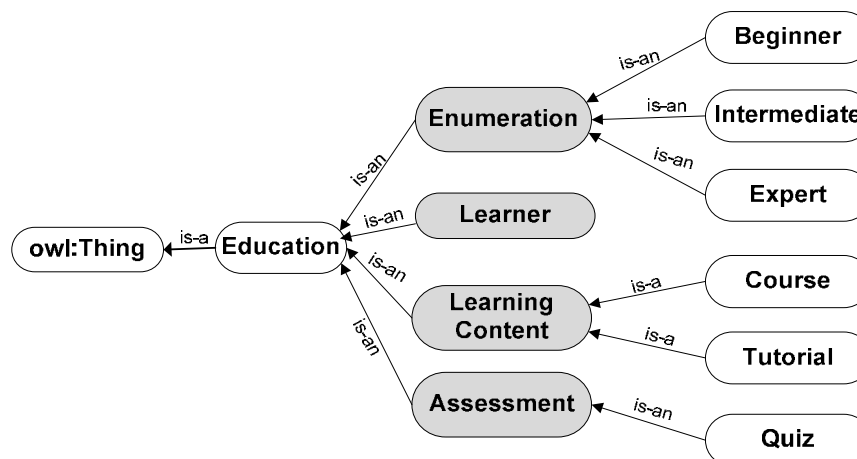
Figure 3.11: Main Classes of the Education.owl Ontology
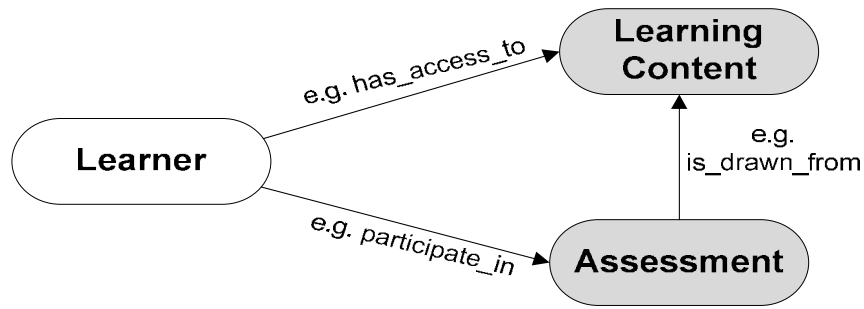
Figure 3.12: Architecture of Education Ontology Model

Once a class of the ontology is selected, the application delivers the instances related to that class. Figure 3.13 presents Classes in Education.owl Ontology and the properties/relations respectively using Protégé 3.4.1. The development of the Education ontology involves the creation of the classes, properties and relationship between deferent classes.

a.   b.
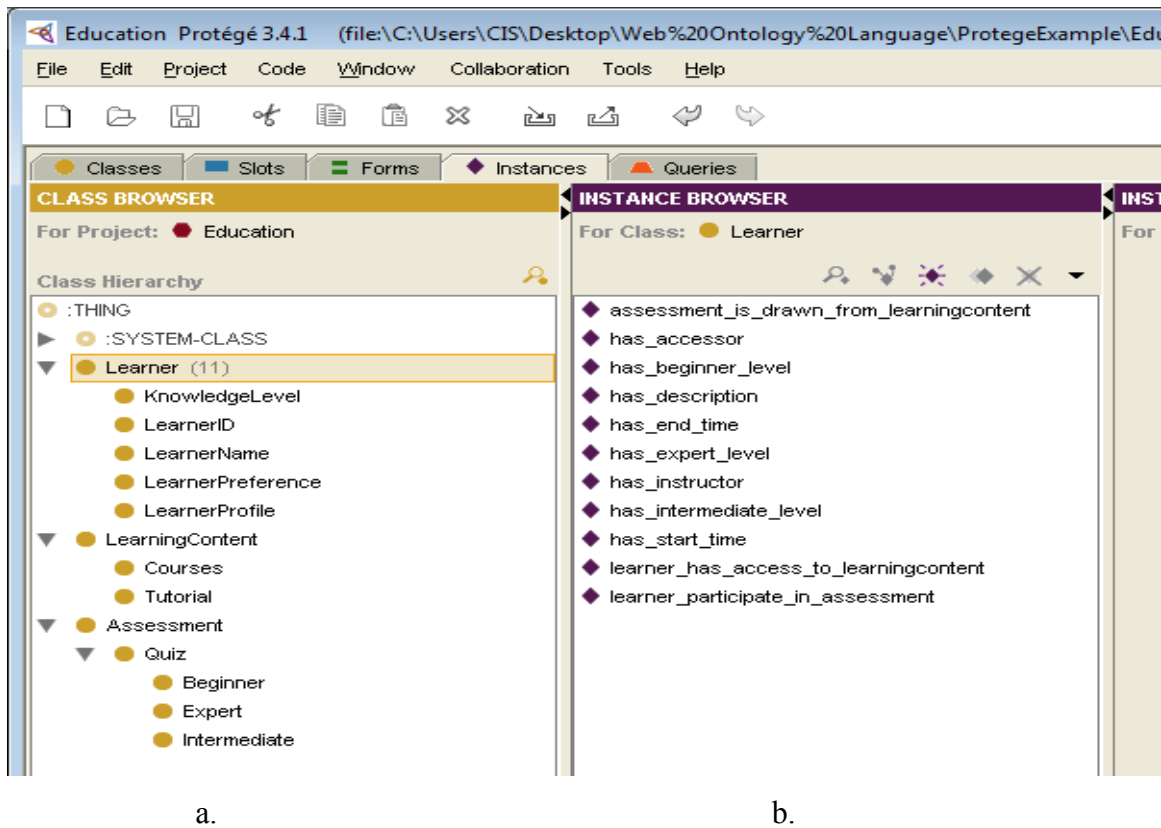
Figure 3.13: A Screenshot of Education Ontology design using Protégé 3.4.1

Figure 3.14 (a and b) shows an example of representing the same piece of information in RDF and in OWL. RDF is a W3C recommendation based on XML that offers support for semantic interpretation, to understand the document existing on web. RDF is a framework to represent data and a model for representing data about "things on the web" (resources). It

comprises a set of triples (O,A,V) that may be used to describe any possible relationship existing between the data – Object, Attribute and Value (Ghaleb et al., 2006). The RDF and OWL code shown in Figure 3.14 was automatically generated using the Protégé 3.2.1 in Figure 3.13.

```
<rdfs:Class rdf:about="&rdf_;LearnerID"
        rdfs:label="LearnerID">
        <rdfs:subClassOf rdf:resource="&rdf_;Learner"/>
</rdfs:Class>
<rdfs:Class rdf:about="&rdf_;LearnerName"
        rdfs:label="LearnerName">
        <rdfs:subClassOf rdf:resource="&rdf_;Learner"/>
</rdfs:Class>
<rdfs:Class rdf:about="&rdf; LearningContent"
        rdfs:label="LearningContent">
        <rdfs:subClassOf rdf:resource="&rdfs;Resource"/>
</rdfs:Class>
```
(a)

```
- <owl:Class rdf:ID="LearnerID">
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">LearnerID</rdfs:label>
    <rdfs:subClassOf rdf:resource="#Learner" />
  </owl:Class>
- <owl:Class rdf:ID="LearnerPreference">
    <rdfs:subClassOf rdf:resource="#Learner" />
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">LearnerPreference</rdf
s:label>
  </owl:Class>
-   <owl:Class rdf:about="#Assessment">
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Assessment</rdfs:label>
  </owl:Class>
- <owl:Class rdf:about="#LearningContent">
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">LearningContent</rdfs:l
abel>
  </owl:Class>
- <owl:Class rdf:ID="KnowledgeLevel">
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">KnowledgeLevel</rdfs:l
abel>
    <rdfs:subClassOf rdf:resource="#Learner" />
  </owl:Class>
```
(b)

Figure 3.14: (a) A piece of RDF code and (b) The same information in OWL code for

iVeES

3.4   INTERACTION MANAGEMENT

Advances in speech recognition and VUI systems have motivated the provision of a VUI design framework for interacting with vocal applications. For most developers, designing a VUI is a relatively new experience. Today, good VUI frameworks are few mainly because VUI design is novel and requires usability criteria for users' friendliness and ease of navigation during interaction. This section presents an enhanced VUI tree framework for intelligent Voice-enabled e-Education System (*iVeES*) using Multi-RootNode Tree to enhance speed of access and users' flexibility while interacting with the system. The proposed system will be of immense benefit to the normal users and the visually impaired learners since it will make navigation, access speed and interaction process with the system a lot easier using a mobile phone.

The proliferation of mobile devices in the last decade has made access to the telephone network even easier. As discussed by Tucker (2003), every successful voice-enabled application is built on three solid foundations: Firstly, universal user interfaces guidelines and usability principles. Secondly, the project guidelines for the application and system requirements. Thirdly, the interaction guidelines for the personality that will be portrayed by the Voice User Interface (VUI). This foundation requires that a standard VUI framework be followed as a guideline for building a voice interface application.

Providing a voice-based interface with the web interface is an advantage to the visually impaired who are unable to use a visual interface. It is also possible to use the application for accessing a web-based interface even while on the move through a mobile phone, which is much easier to carry around than a personal computer. Phone applications are more spontaneous. Most people these days always have their phone on their hips. In many cases, the phone transaction can be completed before the PC boots or one can log in. The number of telephone users is far greater than the number of people that use personal computers or the Internet.

In voice-enabled telephony systems, users interact with machines in a conversational dialog over a phone connection. The design of such dialogs affects the quality of service offered by the system. Hence, VUIs should be carefully designed with high quality interactive

management techniques to attain a positive usability attribute particularly in the area of user's satisfaction, ease of navigation and accessibility. One of the most significant elements of a good user interface is visibility of the system's status. In an ideal situation "What you see is what you get" (WYSIWYG), but with telephony system, "What you hear is what you get" (WYHIWYG). That is, if software VUI interface hears the word *exam* and *exam* is understood, then *exam* happens. Users must also notice immediately what's going on behind the scenes and whether their actions have actually led to the expected results (Smashing magazine, 2009).

Voice-enabled applications are currently being applied in e-Education. Thus, by using VoiceXML applications, more students can be reached than is possible with the Internet. This is so because the educational environment is one of the most popular sectors where majority of people carry mobile phones. One of the goals of an e-Education system includes providing quality learning objects to students at a considerable low cost through a voice-enabled learning platform. This goal can easily be achieved with a voice-enabled application. One way of building quality into technology-supported-education is by providing a successful VUI that is user-friendly and easily navigated to encourage learners to continue using the e-Education platform. Designing a VUI is very different from designing other user interfaces. Most user interfaces usually offer some visual information for users to interact with. With VUI, users have to listen, memorize, and speak to interact with the system. The VUI cannot provide rich content information because users are not able to memorize pages of verbal information. The design has to be simple, with short and clear dialogs so that users can navigate through the service without any problem.

There are several voice-enabled frameworks in the areas of e-Education. However, there is dearth of a user friendly VUI framework for interaction management and the speed of accessing voice applications is considerably low. The depth of menu-driven navigation for VUI application has been a problem for certain category of users. However, designing a VUI is a relatively new experience for most developers. The depth of menu-driven navigation for *iVeES* is depicted using process flow diagram in Figure 3.15. It shows a model of interaction among the various VoiceXML documents that make up *iVeES*.
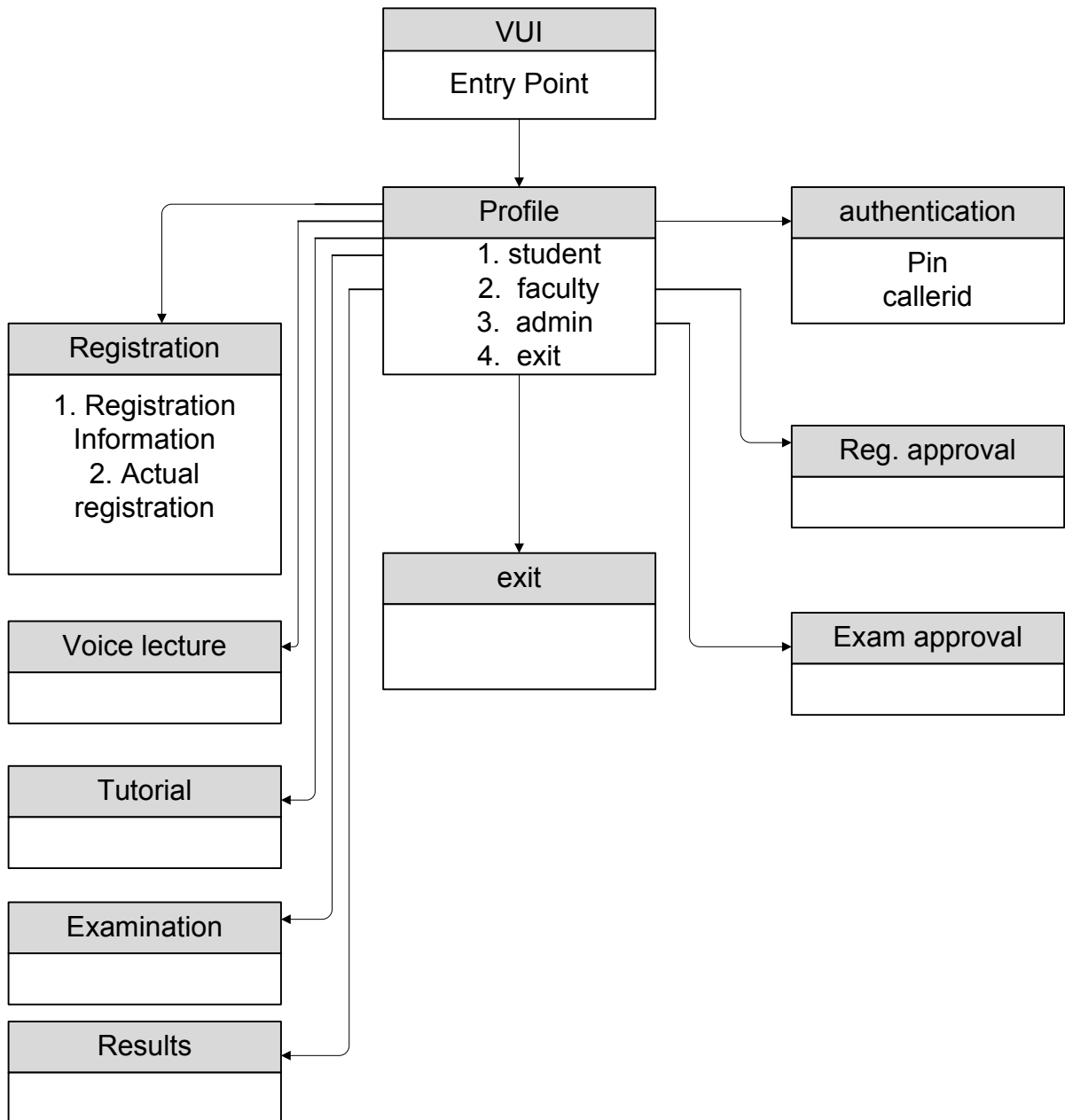
Figure 3.15: Process flow diagram of *iVeES*

There are very few good VUI guidelines available today mainly because VUI design is so new and requires sufficient usability testing (Yankelovich, 2000).

3.4.1 The VUI Tree Framework

This section explores the design of a VUI tree framework using an enhanced Multi-RootNode tree structure to improve the users' mode of interaction. The algorithm engaged

to improve and restructure the tree is known as node promotion/demotion algorithm (Chang, 2007).

The VUI Navigation was modeled using VUI tree structure as shown in Figure 3.16. When a user named catman uses a telephone and makes a phone call to *iVeES*, s/he reaches the root node of the VUI tree and hears a greeting message, which may look like "Welcome to intelligent voice enabled e-Education Application". If you would like to select student menu say "student", for faculty menu say "faculty" and admin menu say "admin". User catman may say faculty and traverse the VUI tree to N12. The interaction between user catman and the system will be confirmed when catman navigates to level two of the VUI tree. At this point he can listen to the registration.

From Figure 3.16, the root node represents the main menu of *iVeES*. The leaf node signifies the functional components that make up the application. The intermediate node represent the submenus that leads to the leaf node. The broken line (--------) indicate the path for the most frequently accessed leaf node. In our case study, lecture and tutorial module is the most frequently access by the learner.
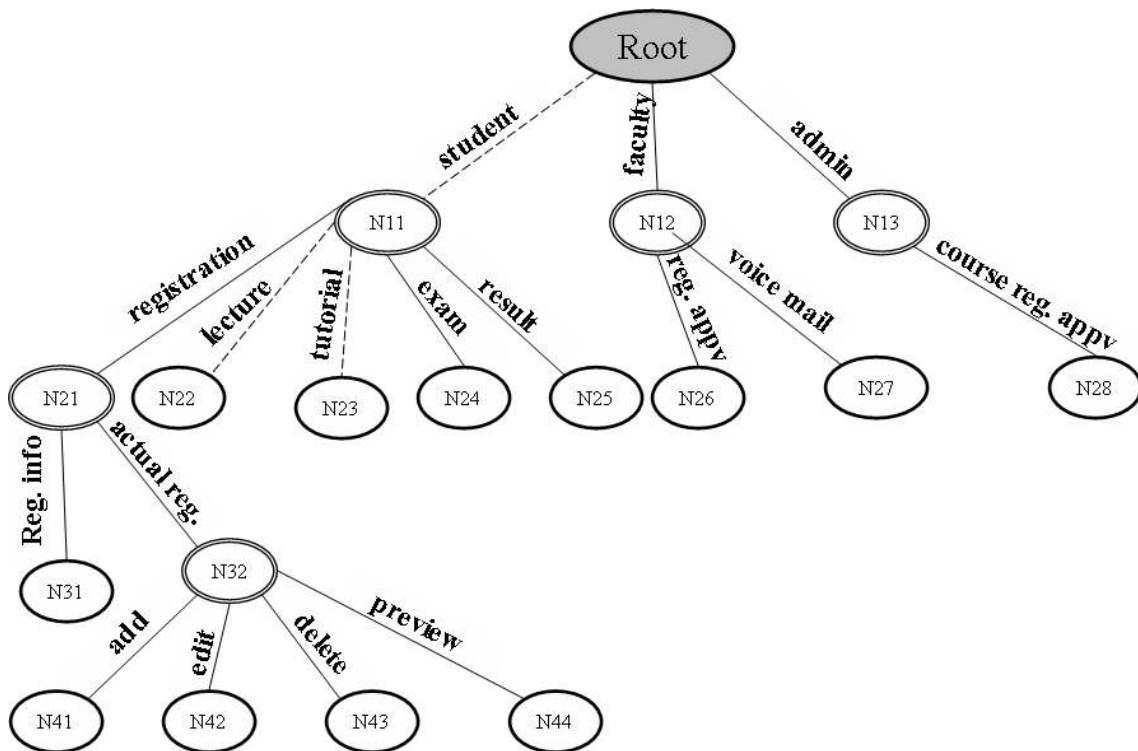
Figure 3.16: A VUI tree framework for the system

Table 3.1: Interpretation of Nodes in the VUI tree

SN	Node	Node shape	Node Type	Interpretation on *iVeES*
1	Root	Shaded oval	No Predecessor	Main menu
2.	Leaf	Single line oval	No Successor	Executable menu
3.	Interior	Double line oval	Non-Leaf	Sub menu

The nodes in Figure 3.16 are explained in Table 3.1. The root node is represented by shaded oval and signifies a main menu. Leaf node is represented by single line oval and has no successor. Leaf node implies an executed menu. Interior node signifies double line oval and it represents a non-leaf node or sub menu.

VUI Optimization and Enhancement

One approach for optimizing VUI in Chang (2007) is the operation of node promotion/demotion (by moving some nodes to higher/lower levels of the VUI tree). This approach uses a Genetic Algorithm based optimization technique. In this model, *root* node is at level 0, and N_{ij} is the *j*th node at level *i*. *Node access time* is the time length of audio heard by users during each visit to the node. *Node access count* is the frequency users navigate through and access to the node in a specific term. *Leaf node access time* is the total access time users navigate from *Root* to the leaf node. *Total time* is the summation of every *leaf node access time* which multiplies *leaf node access count*. In a voice site, users' destination must end at leaf nodes; therefore, we focus on the leaf nodes in the model. It means only leaf nodes have the property of "access counts."

The properties of the nodes used in this example are described in Table 3.1. In this case, the values of *node access time* are randomly created in the range between 3 and 8 seconds, the values of *node access count* are only applicable to leaf nodes and randomly generated between 40 and 100, and each value of the *leaf node access time* is derived from the values of *node access time* of all nodes on the path from the *root* navigating to that particular leaf node. For example, the *leaf node access time* of *N22* is the summation of *node access time* values of the *root*, N11, N21 and N22 (i.e., 5 + 6 + 5 + 3 = 19), and the *leaf node access time* of *N23* is the summation of *node access time* values of the *root*, N11, N21, N22, N23

and N23 (i.e., 5 + 6 + 5 + 3 + 6 = 25). Using the VUI optimization mechanism, we can restructure the VUI tree and calculate the new total time. The new tree structure is shown in Figure 3.17 (nodes having asterisk [*] are promoted), and the properties of the nodes of the restructured tree are described in Table 3.1. Note that while the values of *node access time* and *node access count* on all nodes are unchanged, the values of *leaf node access time* on some leaf nodes are changed because the VUI tree is restructured. In this sample case of optimization model in Chang (2007), the improvement is about 57% i.e. (3830-1660=2170, (2170/3830*100) = 57%. Using our enhanced model, the improvement is about 79% i.e. (3830-810=3020, (3020/3830*100) = 79%, even though this is just a simple illustration.

The calculation of total access time, optimized time and enhanced time is contained in Table 3.2. The leaf node access time *(t)* for *N22 has changed from 19 to 8 and for *N23 from 25 to 11 as a result of the optimization. Similarly, the leaf node access time *(t)* for **N22 has changed from 8 to 3 and for **N23 from 11 to 6 as a result of the enhancement. The values were derived using the formula in Equation (3.6), (3.7) and (3.8):

$Total\ access\ time\ = \sum_{i=1}^{L} Nij(t) * Nij(c),$ $j = 1..n$ (3.6)

$Optimized\ access\ time\ = \sum_{i=1}^{L} Ni(t) * Ni(ot),$ $j = 1..n$ (3.7)

$Enhanced\ access\ time\ = \sum_{=1}^{L} Ni(t) * Ni(et),$ $j = 1..n$ (3.8)

Therefore total time = 19 x 70 + 25 x 100 = 3830, Optimized total time = 8 x 70 + 11 x 100 = 1660, and Enhanced total time = 3 x 70 + 6 x 100 = 810. The broken lines (-----) in Figure 3.17 represent the path to the nodes that have been promoted through restructuring. Several other approaches for optimizing VUI tree include the creation of extra links (for providing short-cuts to VUI tree navigation). Changing the order of child nodes for every non-leaf node can create many different restructured VUI trees, and every VUI tree represents a possible solution. One demerit of node order change is that for a VUI tree with hundreds of nodes, the potential number of restructured VUI trees can be very high.
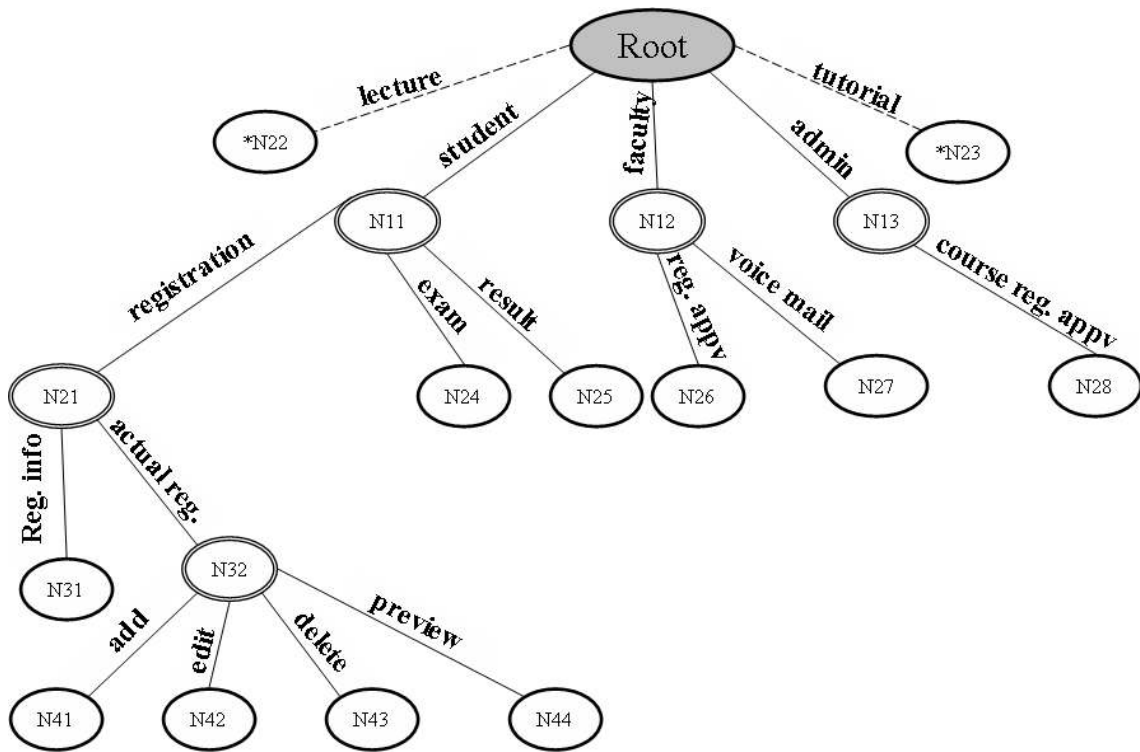
Figure 3.17: The restructured VUI tree (Node marked * were promoted)

Table 3.2: Properties of VUI tree nodes

Node	Root	N11	N12	N13	N21	N22	N23	N24	Total Time
Node access time (a)	5	6	5	4	5	3	6	-	-
Leaf node access time (t)	Na	na	na	na	na	19	25	-	-
Node access count (c)	Na	na	na	na	na	70	100	-	-
Total time	-	-	-	-	-	1330	2500	-	3830
Optimized Leaf node access time (ot)	Na	na	na	na	na	8	11	-	-
Optimized total	-	-	-	-	-	560	1100	-	1660
Enhanced Leaf node access time (et)	Na	na	na	na	na	3	6	-	-
Enhanced total	-	-	-	-	-	210	600	-	810

The Enhanced Multi-RootNode Tree Model

With the VUI enhanced Multi-RootNode approach, we restructured the VUI tree and calculate the new total time as shown in Table 3.2. The new tree structure is shown in Figure 3.18 (nodes having double star mark [**] are restructured). The node label lecture and tutorial have been moved to the same level with the root node.
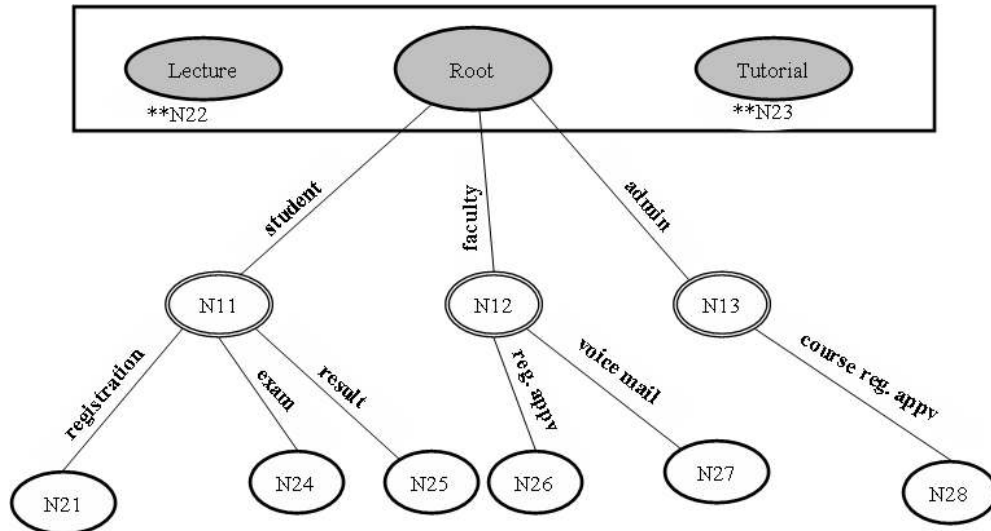
Figure 3.18: The VUI tree (Node marked ** were restructured)

Multi-Root Node VUI Call flow

Figure 3.19 shows a partial Call flow of the proposed multi-root node model where voice lecture and tutorial menu have been promoted up to be among the root node.

iVeES: Welcome to intelligent Voice enabled e-Education System
iVeES: Voice lecture
iVeES: Root
    iVeES: Student menu
    iVeES: Faculty menu
    iVeES: Admin menu
iVeES: Tutorial
--
iVeES: Welcome to Voice lecture menu
iVeES: Mention the subject code to study
iVeES: BIO200
--
iVeES: Welcome to student menu.
iVeES: registration
iVeES: examination
iVeES: result
iVeES: exit
--
Figure 3.19: Partial Call flow for the system

3.4.2 The Hidden Markov Model (HMM) Framework

Hidden Markov Model (HMM), sometimes referred to as Finite State Automata (FSA), is a mathematical apparatus for modeling sequence analysis, including speech recognition. In the study of speech technology, speech is viewed as a sequence of sound signals. HMM is an extension of Markov chain. In mathematics, a Markov chain, named after Andrey Markov, is a discrete-time stochastic process with the Markov property, that, given the present state, future states are independent of the past states. In other words, the description of the present state fully captures all the information that could influence the future development of the process (Radev, 2008). Future states will be reached through a probabilistic process instead of a deterministic one. That is, given the present, the future is conditionally independent of the past. Nothing that has happened in the past can influence or determine the outcome in the future, the future is all possibilities. At each time instant the system may change its state from the current state to another, or remain in the same state, according to a certain probability distribution. The changes of state are called transitions, and the probabilities associated with various state-changes are termed transition probabilities.

HMM could be used to predict the next VoiceXML document or dialogs from a particular state. A VoiceXML-based application forms a conversational finite state machine (FSM) (Raghuraman, 2004). FSM typically move through series of states and produce some kind of output either when the machine has reached a particular state or when it is moving from state to state. A HMM consists of a set of states connected by a number of possible transition paths (see Figure 3.20) for a three-state HMM. Each transition path has an associated transition probability labeled by a_{ij}. A symbol is produced on arrival (or indeed re-entry) into a state, assuming for now we are working with a discrete observation HMM. Each state in the HMM has associated with it a set of output symbols with its own probability distribution. The notation $b_i(k)$ indicates the observation probability that symbol k is produced given state i.
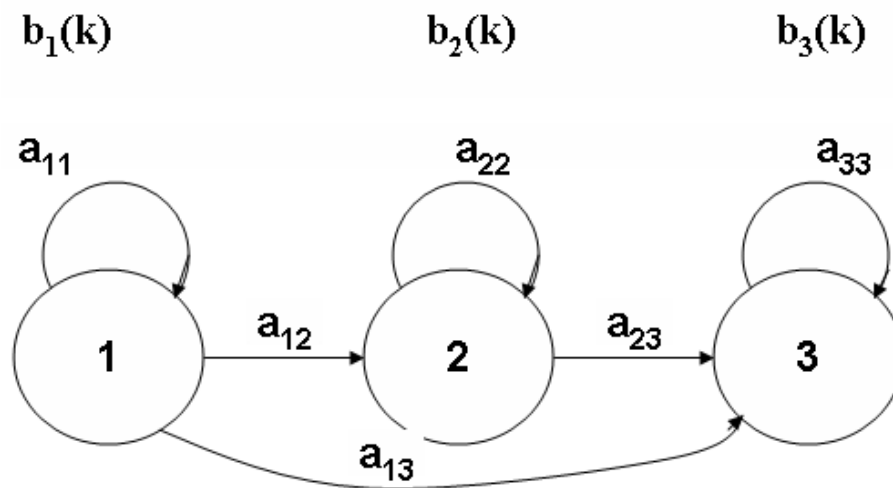
Figure 3.20.   **Example of a three-state HMM**   **(Source: Burke, 2007)**

The following elements are required to fully define HMM (Narada, 1996):

- The number of states of the model, N.

- The number of observation symbols in the alphabet, M.

- A set of state transition probabilities, satisfying the constraints,

$$\sum_{j=1}^{N} ai_j = 1, \quad 1 \le i \le N$$

      where $aij \ge 0, \quad 1 \le i, j \le N$

- A probability distribution in each of the states,

       $b_j(k)$ satisfying the constraints

$$\sum_{1}^{m} b_j(k) = 1, \quad 1 \le j \le N$$

where, $B_j(k) \ge 0$ $1 \le j \le N, \quad 1 \le k \le M$

- The initial state distribution.

Properties of Hidden Markov Models

In the theory of HMMs, the following properties are applicable:

The Markov assumption: The next state is dependent only upon the current state.

The stationarity assumption: Here, it is assumed that state transition probabilities are independent of the actual time at which the transitions take place.

The output independence assumption: This is the assumption that current output (observation) is statistically independent of the previous outputs (observations). The state transition diagram and probability of *iVeES* is depicted in Figure 3.21 and Table 3.3 respectively.
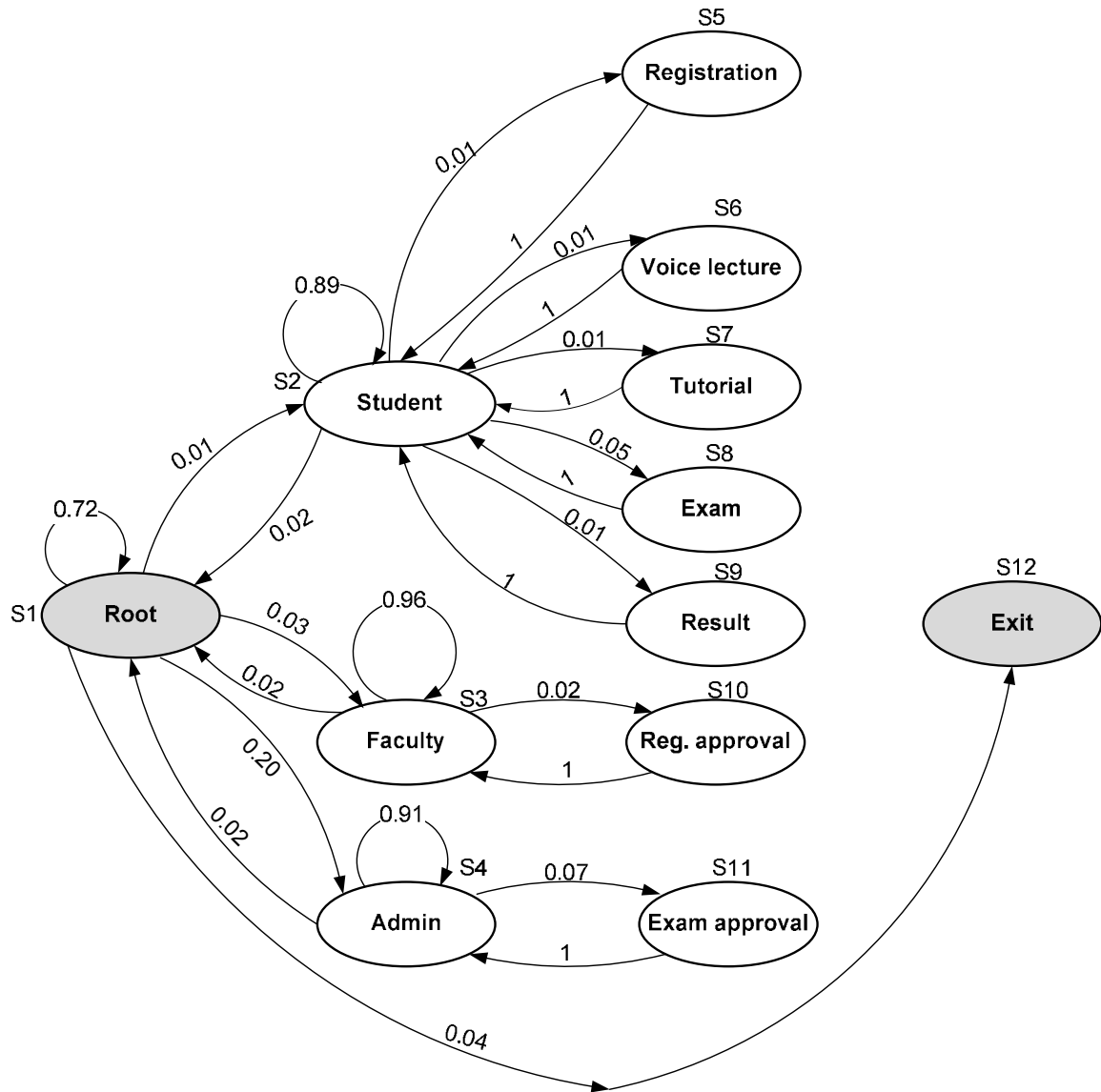
Figure 3.21: The state transition diagram of *iVeES*

The learner is at conversational state S1. Each dialogs determine the next dialogs to transit. Figure 3.22 contains the format of state transition probabilities. The probability of transiting to the next state is computed in Equation 3.9, 3.10, 3.11 and 3.12 with the values shown in Table 3.3. For example, the probability of moving from state S1 to S2, S3, S3 and S4 is 0.20. Transitions are specified using URL, which define the next document and dialogs to use. If

an URL does not refer to a document, the current document is assumed. The program will end when a dialogs does not specify a successor or when it encounters an exit command.

A11   A12   ..   A1N
A21   A22   ..   A2N
..   ..
AN1   AN2   ..   ANN

Figure 3.22: Format of state transition probabilities

<u>Definitions</u>

En_i = total number of edges leaving state i, $\qquad i = 1,…,n$

Ok = observation, $k=1…r$

CS = current state

W_j = weights assigned to alphabets Wj are [a,b,c,d,e,f..]=[1,2,3,4,5,6..], where j=1,2,..,26

Ps_i = Probability of state i, $\quad i = 1,…,n$

$P(En_i)$ = Probability of En_i

$$P(En_i) = \frac{1}{En_i} \qquad\qquad\qquad\qquad\qquad (3.9)$$

$$Ps_i(Ok) = \frac{P(En_i)}{W_j}, \ i = 1,..,n. \ \ j = 1,…,m. \qquad\qquad (3.10)$$

$$total_{Ps} = \sum_{i=1}^{n} Ps_i(Ok), \qquad k = 1,…,r \quad \text{and} \ i = 1,…,n \qquad\qquad (3.11)$$

$$Ps_i(CS) = 1 - totalps \qquad\qquad\qquad\qquad\qquad (3.12)$$

Table 3.3:   State transition probabilities

States	S1 Root	S2 student	S3 faculty	S4 admin	S5 register	S6 lecture	S7 tutorial	S8 exam	S9 result	S10 reg. approval	S11 exam approval	S12 exit
S1	0.72	0.01	0.03	0.20	0	0	0	0	0	0	0	0.04
S2	0.02	0.89	0	0	0.01	0.01	0.01	0.05	0.01	0	0	0
S3	0.02	0	0.96	0	0	0	0	0	0	0.02	0	0
S4	0.02	0	0	0.91	0	0	0	0	0	0	0.07	0
S5	0	1	0	0	0	0	0	0	0	0	0	0
S6	0	1	0	0	0	0	0	0	0	0	0	0
S7	0	1	0	0	0	0	0	0	0	0	0	0
S8	0	1	0	0	0	0	0	0	0	0	0	0
S9	0	1	0	0	0	0	0	0	0	0	0	0
S10	0	0	1	0	0	0	0	0	0	0	0	0
S11	0	0	0	1	0	0	0	0	0	0	0	0
S12	0	0	0	0	0	0	0	0	0	0	0	0

HMM can also be described in a mathematical style (without a diagram) generally known as formal technique. A formal technique has been used to model the transition from state to state. A state transition function pr gives the new state where pr(current state, observation) = new state e.g. Pr(S1,student) = S2, Pr(S3, faculty)=S10.

The state transition diagram of Figure 3.21 was used to provide answers to some vital e-Education questions. Sample of such questions are provided as follows:

Question 1. What is the probability that a student dial a telephone number to use *iVeES* and carry out course registration before attending lectures?

Answer: Here, we have Ok={S1, S2, S5, S2, S6}

$$P[Ok|model] = P[S1,S2,S5,S2,S6|model]$$

$$= P[S1].P[S2|S1].P[S5|S2].P[S2|S5].P[S6|S2]$$

$$= (1).(0.02).(1).(0.01).(1)$$

$$= 0.0002$$

$$= 2.0 \times 10^{-4}$$

Explanation: In a normal school setting, students should register for courses before attending classes for the courses registered. The probability of 0.0002 is very low in this

circumstance, meaning that using our model assumptions, only few cases may be recorded of students carrying out course registration before attending classes for lectures. This is not a normal situation and students could be made to explain why he/she attended lectures without course registration.

Question 2. What is the probability that a student did examination without course registration?

Answer: Here, we have Ok={S1, S2, S8}

$$P[Ok|model] = P[S1,S2,S8|model]$$

$$= P[S1].P[S2|S1].P[S8|S2]$$

$$= (1).(0.02).(1)$$

$$= 0.02$$

$$= 2.0 \times 10^{-2}$$

Explanation: In a normal situation, students should register for courses, attend classes for lectures and participate in examination. The probability of 0.02 is very low in this circumstance, meaning that using our model assumptions, only few students did examination without registering for the course. This is close to a normal situation. It would have been absurd if many students are writing examination without first registering for the course.

3.5   SUMMARY

In this chapter, requirement elicitation and the proposed framework are presented. The intelligent component services of the framework were modeled using adaptation, recommendation and ontology. Protégé and OWL were used to model and design the Education ontology. Finally, *iVeES* interaction management was described using VUI tree and Hidden Markov Model (HMM) framework.

CHAPTER FOUR

iVeES DESIGN AND IMPLEMENTATION

4.1   INTRODUCTION

This section presents the design and implementation of the system based on the proposed framework.

4.2   ARCHITECTURAL DESIGN

4.2.1   The Physical Architecture

The architectural design consists of a layered and deployment architecture. The layered architecture of the system is presented in Figure 4.1. It consists of the client devices, middleware and data. The database is separated from the client through the middleware. The middleware is also referred to as the business logic tier.

Client devices

In the client device layer, students are able to connect to the e-Education application. The components of the clients' interface are i) Voice, and ii) Web. These units do not store or process any form of data. They only provide an interface for the middle ware and the data layer. Data or files or voice browsers are not stored on the mobile phones due to resource constraints associated with hand-held devices. The voice browser runs on the speech gateway and receives any call into the application and submits them to speech server for further processing.

Middle ware

The speech server communicates with the speech server component of the middle ware through the voice browser. The middle ware comprises of speech server, web server and application server. Users access the application from various mobile telephone devices and land line telephone, anywhere, anytime. Once a user has been authenticated, the user's query is translated by the automated speech recognition (ASR) to text and it is passed to the middle ware for execution. The text-to-speech (TTS) does the reverse of translating text to speech. A

user can only access the services for which he or she is authorized. The client application interfaces with the middle ware using the ASR and TTS.

Data

This layer provides data services and data base management system function. The database is a key component of *iVeES* architecture. It holds information regarding the learning content (registered users, tutorial, examination and result information). The data layer is responsible for changing, adding, or deleting information in the database within the system. MySQL was used as database for the implementation of the data layer.

Client Devices   **Middle Ware**   **Data**
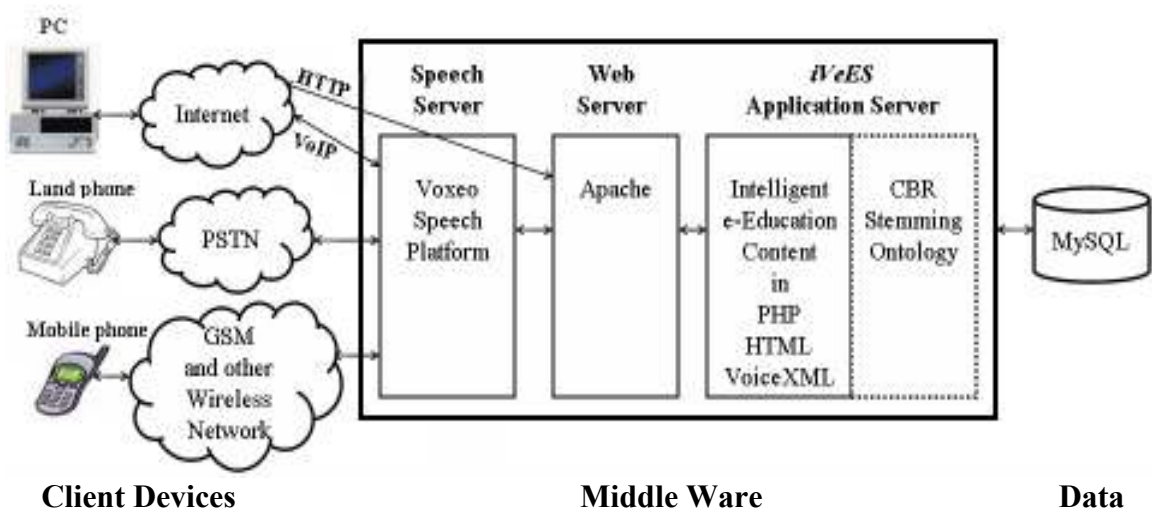
Figure 4.1. Intelligent voice-based e-Education architecture

The user dials a telephone number and through PSTN connects to a speech server. Speech server accepts the calls and connects with a Hypertext Transfer Protocol (HTTP) protocol to *iVeES* application server. The *iVeES* application server which comprises of e-Education content and intelligent services makes query to databases to retrieve information.

The vertical view of the architecture of voice-based e-Education application is presented in Figure 4.2. It shows the location of each of the modules in the application and it consists of presentation tier, business logic tier and data tier. The business logic tier separates the presentation tier from the data tier. The layered architecture is the generic platform for the proposed framework. The database is separated from the client by the logic-tier. In the presentation tier, users are able to connect to the e-Education application. The components of

the clients' interface are i) Personal Computers (PC) for the WUI, and ii) Land phone and Mobile phone for the VUI.
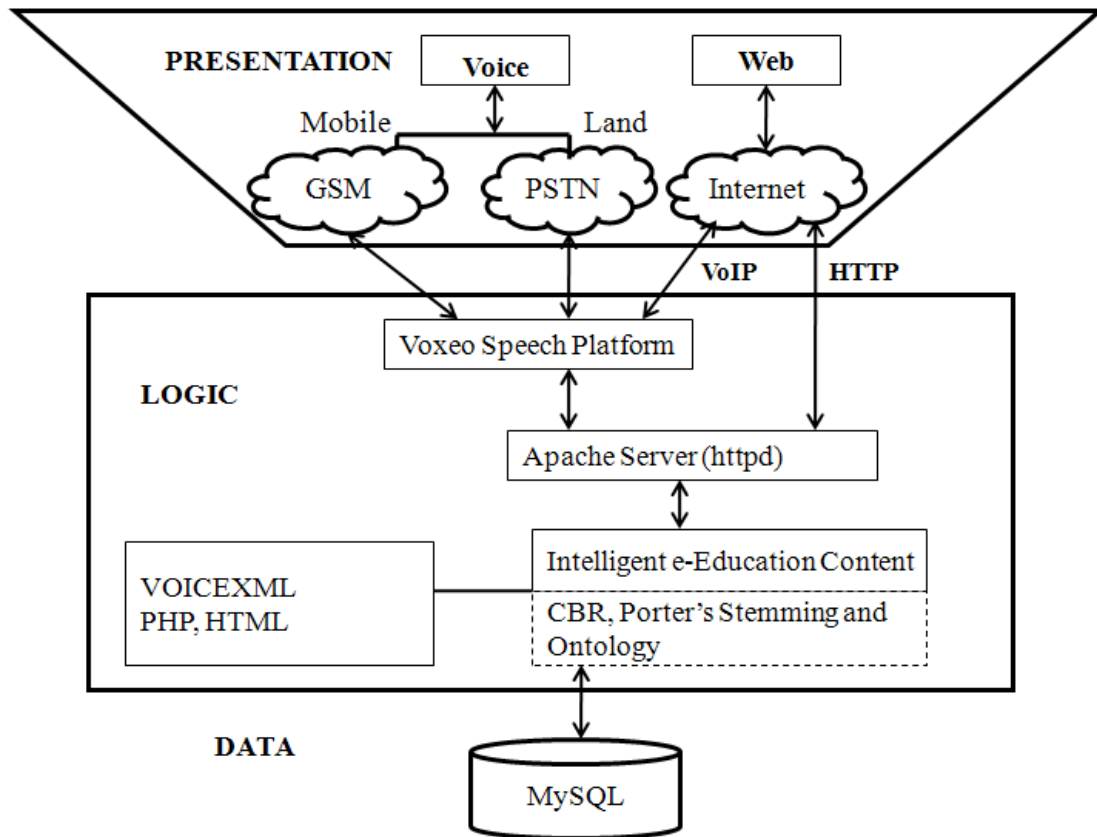
Figure 4.2: An instance of intelligent voice-enabled e-Education Architecture

4.2.2. *iVeES* Application Architecture

The VoiceXML *iVeES* architecture (see Figure 4.3) consists of some VoiceXML software suites including Registration, Voice lecture, Tutorial, Examination and Results. For this prototype implementation, we used free Voxeo hosting service to provide a quick way to get started in order to eliminate huge capital expenses in subscription.
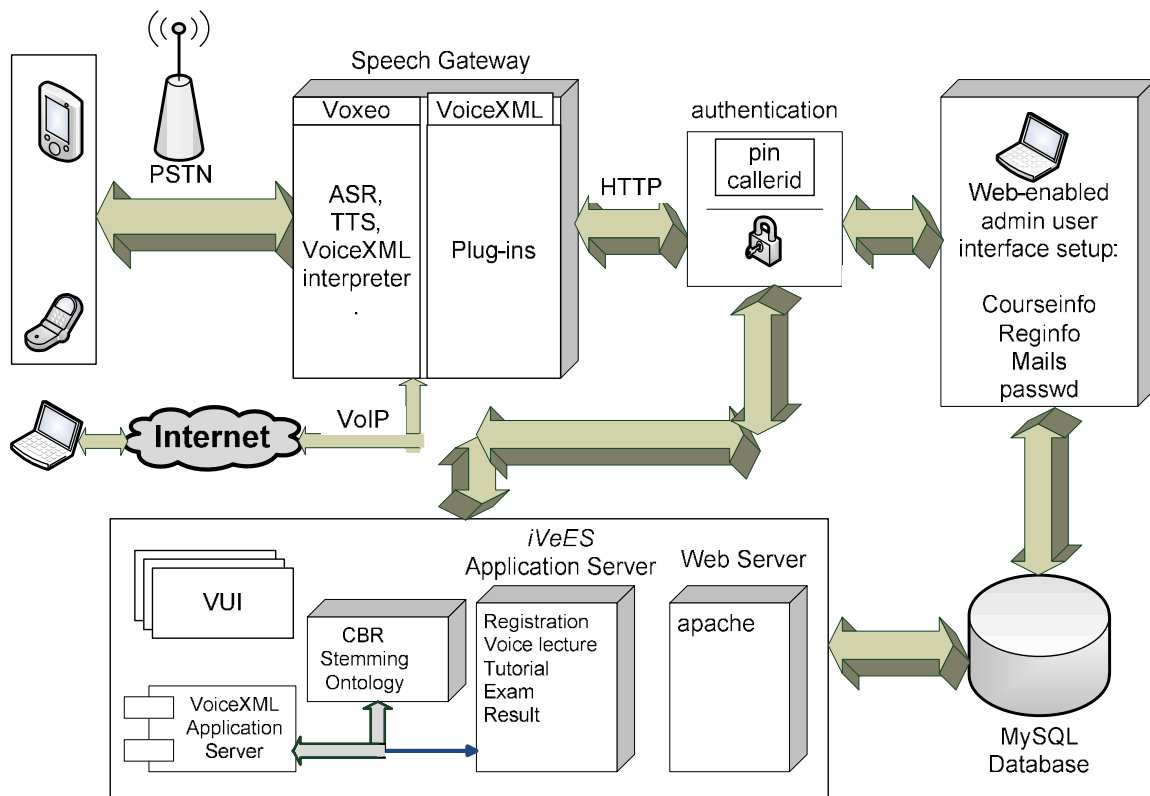
Figure 4.3: *iVeES* application architecture

The *iVeES* application architecture shown in Figure 4.3 was drawn using modular architectural design. The description of the architecture is presented as follows:

- The learner connects through a telephone and PSTN to the VoiceXML interpreter through Voxeo Speech gateway.

- The VoiceXML interpreter executes the call interaction with a caller using the instructions of a VoiceXML script supplied by the voice commands in application server.

- The interpreter calls TTS and ASR as plugins to complete its tasks.

- The VoiceXML interpreter communicates via web protocols (HTTP) to VoiceXML application server.

- The VoiceXML application server delivers the application including VoiceXML text pages.

- The web application server queries the database via apache to dynamically retrieve information and the VoiceXML interpreter TTS speaks with the caller.

4.2.3   The UML Deployment Architecture

The deployment architecture is represented using a UML deployment diagram and it includes clients, servers and database (see Figure 4.4).

Clients

The client interface comprises of mobile phone, land phone and PC phone. In institutions where students are not allowed to use mobile phones or where the students may not afford to own and maintain one, then the alternative device is the use of PC phone such as Skype through Voice over Internet protocol (VoIP).

Servers

The voice server, application server and web server make-up the servers. The e-Education application is stored in the servers and is used to maintain Internet connectivity with the voice gateway. The robust web server provides real-time access to the e-Education database.

Databases

The database stores the student profile and other information related to the learning content. MySQL is a server application (for database) that is able to carry out a great number of SQL commands.
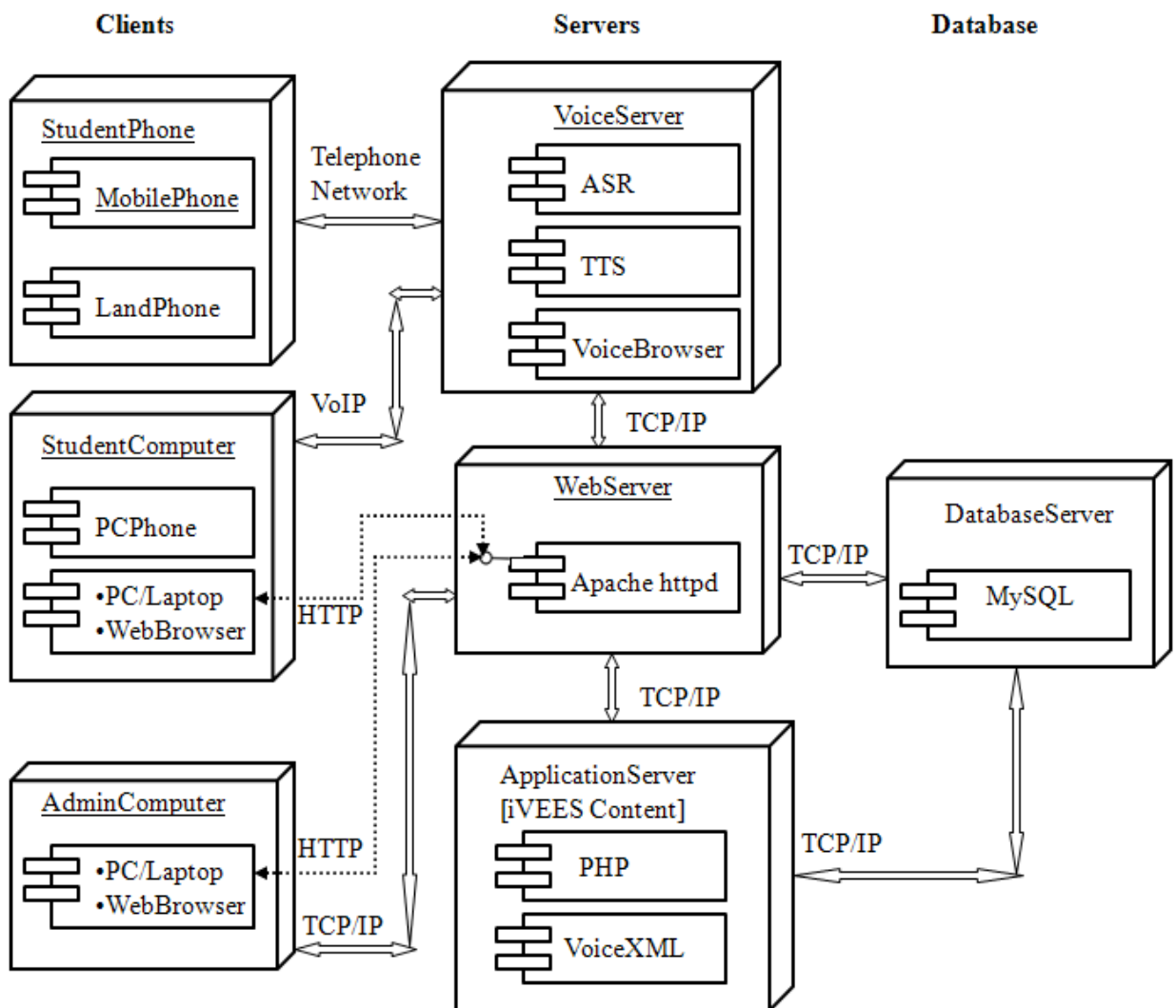
Figure 4.4: A UML Deployment scheme for *iVeES*

4.3   A DESCRIPTION OF CBR AND STEMMING APPROACH

The process model of tutorial questions shown in Figure 4.5 contains the intelligent IR system using CBR and stemming algorithm. The symbols used in the model are Business Process Modeling (BPM) notations which are also close to flowchart symbols. In Figure 4.5 a rectangle represented an activity, rhombus as decision, rectangle with round edges as action, circle as event and directed arrow as sequence of process flow. The user submits a query and the query is compared with the case-knowledgebase terms for a match. If the query is found, the corresponding solution is retrieved, and the system terminates, otherwise the query is tokenized and stemmed, and the process restarts. Generated tokens of the query are

80

compared with the generated index key keywords in the domain knowledge for a match. If token is found then token is replaced with index keywords, and the process restarts. If token is not found, then send token to stemming algorithm and the stemmed token is compared with index keywords and the process restarts.
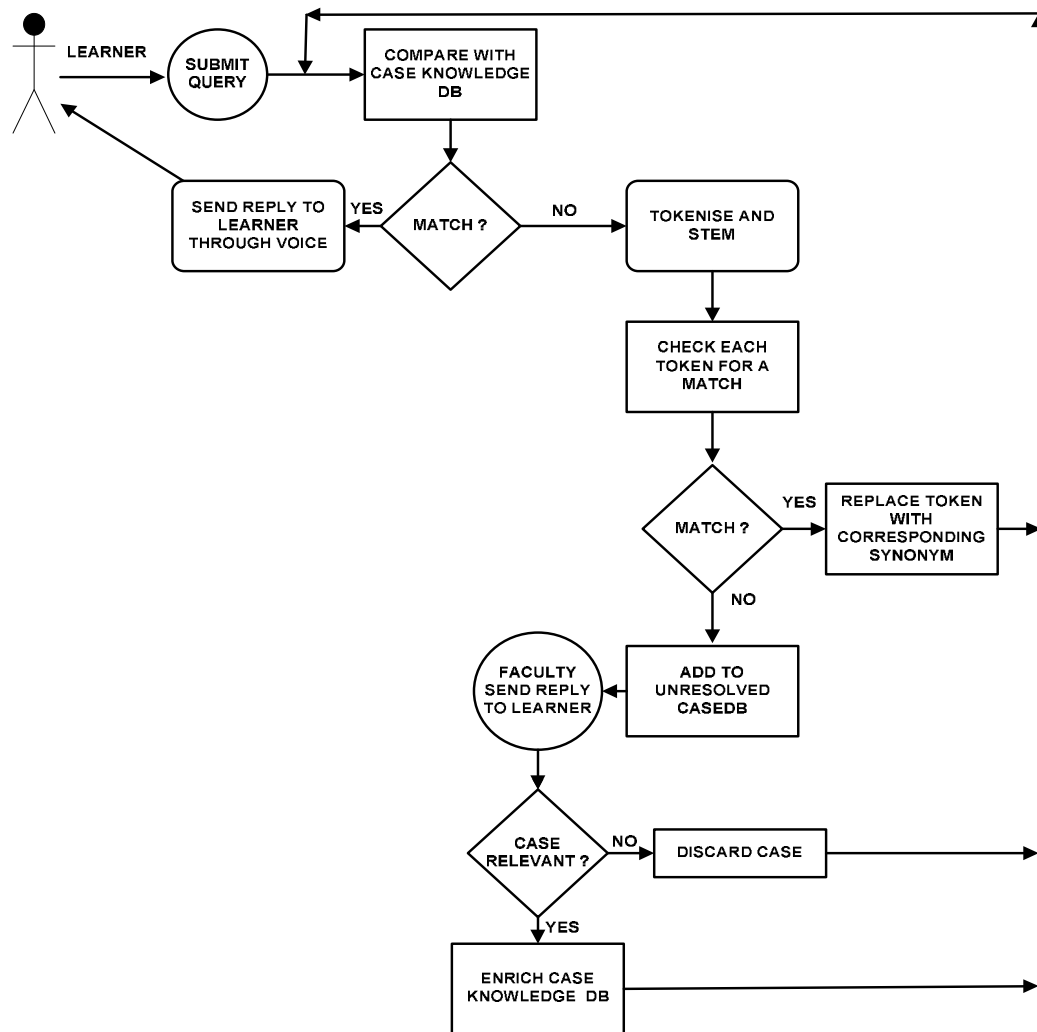
Figure 4.5: A process model of tutorial questions for *iVeES*

4.4   PSEUDOCODE FOR *iVeES*

4.4.1   Algorithm Development

Porter Stemming algorithm was modified and adapted to realize our information retriever. The Modified Porter's Stemming (MPS) algorithm for *iVeES* is shown in Figure 4.6. The

algorithm shows the sequence of activities used for the design of the system. The algorithm has a main program and a function segment called "getmatchingsolution".

    Begin
      Input demand query Qd, supply query Qs
         Input demand search method Md, supply search method Ms
         Found := getmatchingsolution(Qs, Ms)
         //Replace term in q with similar term in domain, where q ∈ Qs & m ∈ Ms
            If there exist m such that m := partial ***then***
               ti := adv-count-words (q) //count terms in query Qs
               Tokens := tokenizestring(Qs);
               //separate the string into tokens. Found f ∈ (0,1)
               //Search each term in r and replace with corresponding term in
                Domain knowledge.
            endif
            If there exists f such that f:=0 ***then***
               Print 'string is not found'
               Add q to specialcasedb;
            endif
            If there exists f such that f:=1 ***then***
               Print 'string is found' & ***return*** result;
            endif
          --
    End
    Function getmatchingsolution (Qs, Ms)
          //normal search
        Foreach q in Qs
         Do
            compare source query q with term t
        enddo
        If q := t ***then*** fetch the corresponding solution C
          //stem source string q and compare with destination term s.
          ti := adv-count-word (Qs) // count terms in query Qs
          Tokens = adv-count-words (Qs)
        endif
        Foreach i = 1 to ti step 1
         do
           Stemsource (i) = stemtokens(i)
           Compare each of stemsource(i) with destination string r
        enddo
        If there is a match ***then***
          report and fetch solution C
        endif
          //stem destination string before search
          //stem source and destination string before search
         --
    Return found f

Figure 4.6: Modified Porter Stemming (MPS) algorithm for *iVeES*

Variables used in MPS algorithm

Q_d: question demanded by *iVeEs*

Q_s: Question supplied by learner

M_s: Search method

q: a term in Qs

m: a search method in Ms

f: signifies found or not found

t: term stored as caseknowledgebase in the system

C: represent a solution that is found

i: subscripted variable for stem source string

r: destination string

4.4.2. Voice User Interface (VUI) Design

The design of VUI application requires not just programming skill but the ability to represent voice input data and grammar for flexibility and easy navigation. The design of the system is such that every authenticated user of the application will undergo a question and answer session, which will be matched against the database content and the result, received by the user through voice output.

In the course of designing *iVeES,* the following were considered for getting learning content through TTS: initial wordings, retry, time out and exit session of VoiceXML document (see Table 4.1 for details). In a situation were recognition of wordings is a problem, the system will exit with the message 'I had trouble hearing you. Good bye', after several attempts.

Table 4.1: VUI Design for getting learning content in *iVeES*

Get learning Content		
Prompts: Welcome to Intelligent Voice-enabled e-Education Systems		
Type	**Wordings**	**Source**
Initial	Please say the course code of the learning content you would like to hear	Get_Content
	on your phone	TTS
Initial	Please say the course code of the learning content you would like to hear	Get_Content
	on your phone	TTS

Retry 1	Please say the course code of the learning content you would like to hear	Get_Content
	on your phone	TTS
Retry 2	Please say the course code of the learning content you would like to hear	Get_Content
	on your phone	TTS
Timeout 1	No Input! I'm sorry, I didn't hear anything. Could you please try again? Please say the course code of the learning content you would like to hear	Get_Content
	on your phone	TTS
Timeout 2	No Match! I'm sorry, I didn't understand you. Could you please try again? Please say the course code of the learning content you would like to hear	Get_Content
	on your phone	TTS
Exit	I had trouble hearing you. Good bye.	Get_Content

The voice interaction for listening to learning content from source VoiceXML document is shown in Table 4.1. It involves initial voice prompt, several retry attempts, and timeout/exit session depending on the reliability of voice recognition and the noise level of the learning environment of the user. Voice interactions with *iVeES* will eventually result to a timeout and exit session if the recognition level of speech server is low or the noise level in the atmosphere is high.

4.4.3   Sample (call flow) of the system

Figure 4.7 contains a sample call flow for the VUI of *iVeES*. When a user connects to the system, after the welcome message, a voice response will say "To login into the system, you must enroll and register, to obtain user name and password at http://www.e-education.com. The user is expected to follow the instructions to achieve a learning goal.

iVeES:Welcome to The intelligent Voice-based e-Education System (*iVeES)* application.
To login into the system, you must enroll and register, to obtain user name and password at www dot e-education dot com.

iVeES: What is your user name.

Caller: catman

IF NOINPUT then *iVeES:*   I did not hear any user name. Please try again.

IF NOMATCH then

iVeES The user name you entered is not recognized.
Please try again.

IF NOMATCH then

IVR: The user name you entered is not recognized.
Please try again.

ENDIF

IF MATCH then

IVR: You have successfully logged in.

ENDIF

iVeES: Select any of the following options.

say Course registration

say Voice lecture.

--

say Examination.

Say Result.

IF one is selected then

IVR: for the following, listen and say:

Add for Adding of registered courses.

Edit for editing of already registered courses.

Delete for deleting of already registered courses.

Preview for previewing of already registered courses.

ENDIF

-

[3 selected]

iVeES :   Welcome to Tutorial Session.

iVeES: Select a course

iVeES: BIO200

iVeES: Ask your question.

iVeES: What is Biology

iVeES: The study of science.

--

[7 selected]

iVeES: Welcome to exit session. Good bye and have a nice day.

Figure 4.7: Sample call flow of *iVeES*

4.5   SYSTEMS IMPLEMENTATION

There are several platforms used to develop, deploy and test applications for free. Voxeo

(Datamonitor, 2006) is one of such platforms which we have used in hosting the prototype

iVeES.

Voxeo is used to build and deploy voice applications according to the architecture specification standardized by the W3C. VoiceXML development platforms are nothing but Integrated Development Environment (IDE) for developing VoiceXML applications.

4.5.1 Voxeo Platform

The Voxeo Prophecy IVR Platform is a pure software, standards-based foundation for any IVR, VoIP, or other telephony application. The flexible and reliable platform has a proven track record of successful, real-world deployments by Voxeo. A typical platform of Voxeo prophecy is shown in Figure 4.8. Prophecy IVR Server works with existing telephony and web-based infrastructure and supports traditional touch-tone IVR, speech recognition, text to speech (Datamonitor, 2006). Hosting with Voxeo platform is free for non commercial purposes.

Figure 4.8: Voxeo IVR hosting platform

Source: (Datamonitor, 2006).

4.5.2 Web User Interfaces of *iVeES*

A prototype application has been developed using the *iVeES* framework. The components of the application comprising design, implementation and development have been reported in Azeta et al., (2009b) and Azeta et al., (2009c). The WUI of the application presented in Figure 4.9 shows the screen where teachers or administrators upload lecture notes. Figure 4.10 contains screen shot of a sample tutorial question used to demonstrate the recommendation component service of the application. The faculty or course lecturer select course code, course title and type tutorial question for which answers are sought. The date is automatically selected by the system. It is the responsibility of the faculty to select the search method as either partial of perfect. Partial means displayed answers will include both the

exact match and matches that are similar, whereas perfect search showcases answers that are exact. Figure 4.11 shows a screen shot of unresolved tutorial questions. The WUI is developed for system administration purposes such as upload of tutorial questions and lecture notes. Voxeo speech platform (Datamonitor, 2006) was engaged as the speech server while CBR and Modified Porter Stemming (MPS) algorithm were used to provide intelligent services.
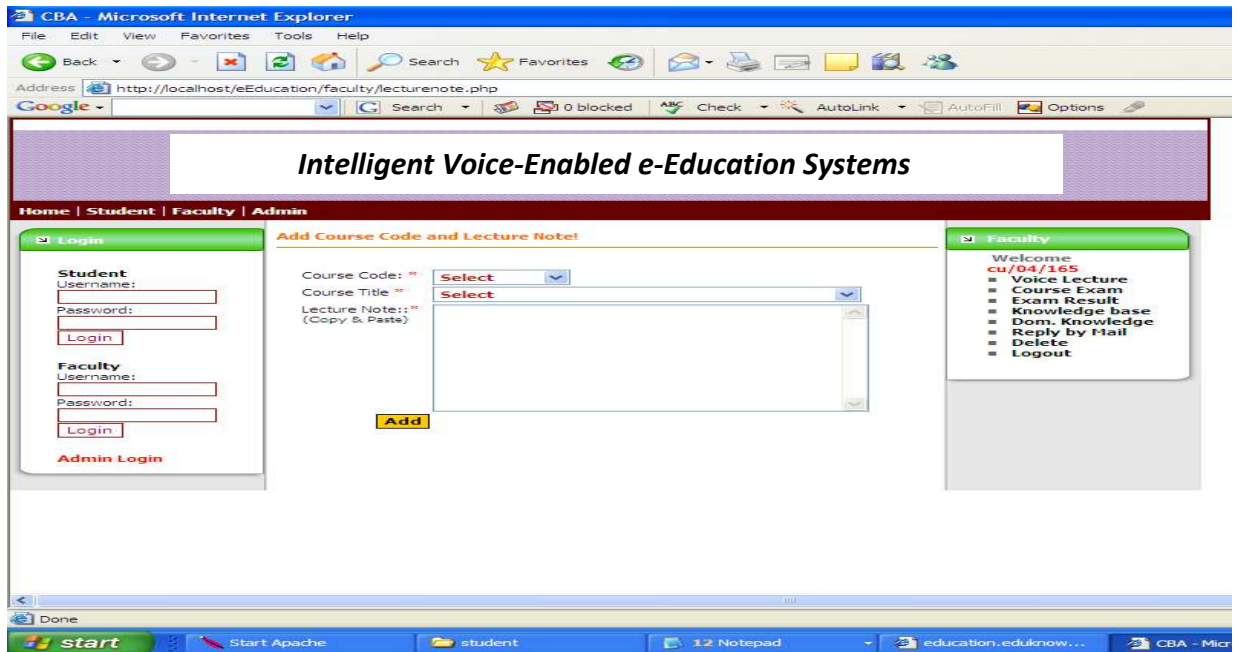
Figure 4.9: A screenshot of lecture note web interface
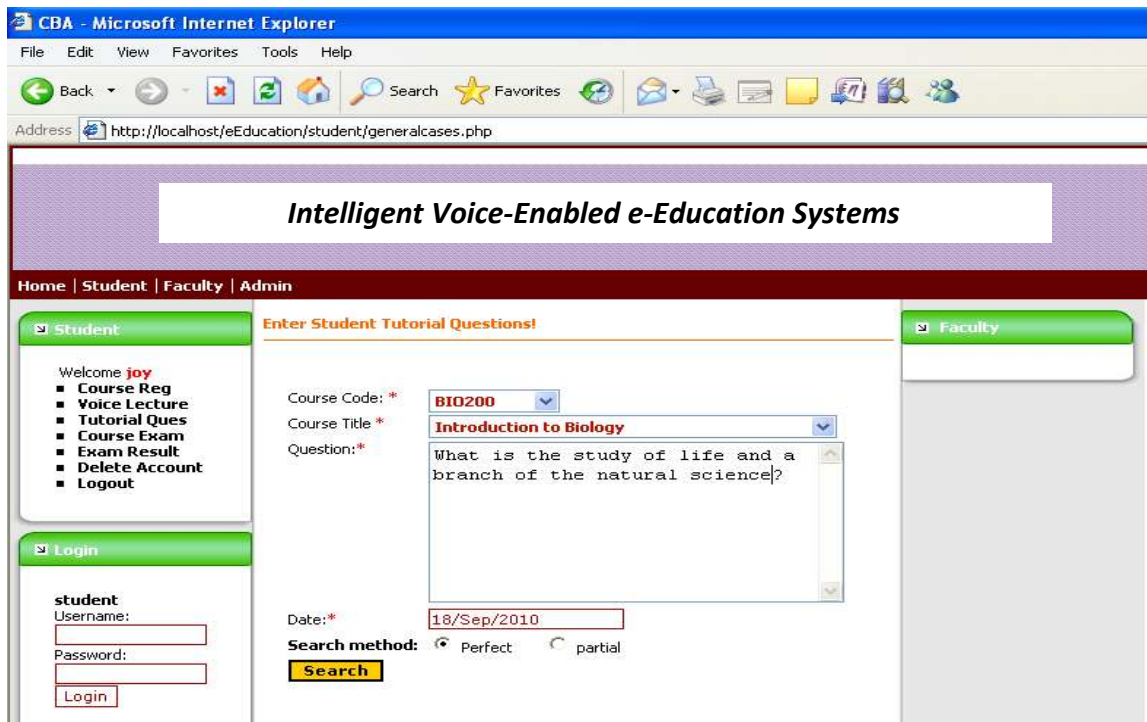
Figure 4.10: A screen shot of tutorial questions web interface
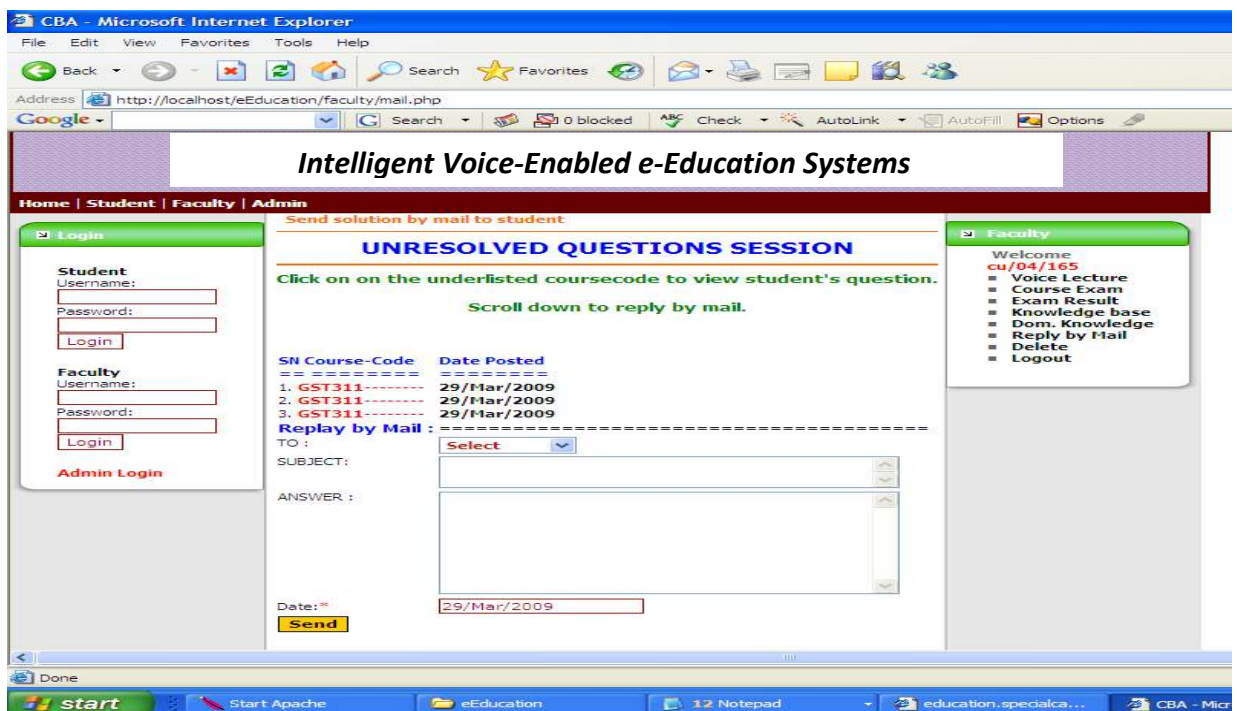
Figure 4.11: A screenshot of unresolved questions web interface

4.5.3 Testing and Deployment

The process of testing involves running and debugging to ensure that the initial requirements are well captured. The prototype VoiceXML-based application was deployed and tested using sample student data. The user logs onto the application using "pin" and "callerid" specific to their registration profile. The developed prototype application can be accessed through a local computer. It was also deployed for access using telephone/web.

4.5.4 Accessing the application

The VUI interface is accessible by dialing a land or mobile telephone number while the WUI is accessible using the Internet. When initiating the dialogue, the application goes to the profile of learners and authenticates the CallerID, which in this case is the learner telephone number. Thereafter, the system asks the learner the User-ID. Once the process of user identification has been completed, it will be presented with a welcome message and a list of learning options. The relevant information will be provided. The WUI is used by the system administrator for user management, account management and configuration management. All users (students) access the application through the VUI while the administrator accesses the application through the WUI.

Local computer

A Voxeo Prophecy phone emulator (VoxeoProphecy, 2003) was downloaded and installed on a local computer and used to develop the voice user interface (VUI) application. A headset was connected to the local computer for the caller to get voice response and also be able to supply voice input. Clicking the Dial button from the Voxeo Prophecy SIP Softphone keypad (see Figure 4.12) allowed connection to the application for access to the various services provided by the application. Once connected, the application prompts with a welcome message and go ahead to authenticate the user's name and password before any transaction can take place. The application will ask for the services demanded by a student and go ahead to process the request.

The web user interface (WUI) can be accessed by typing the URL - http://localhost/eEducation on a local computer running Apache as middleware.
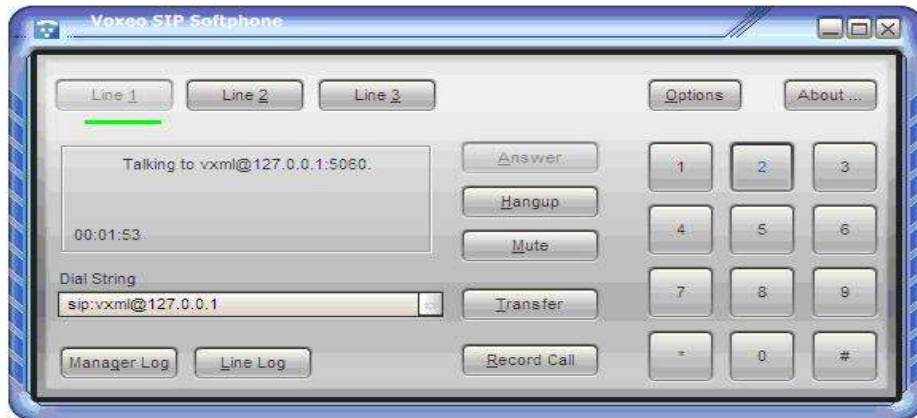
Figure 4.12: Voxeo Prophecy SIP Softphone keypad

Source:(VoxeoProphecy, 2003)

Telephone/Web

A prototype part of the VoiceXML application (voice user interface) was deployed on a Voxeo voice server on the web and accessed from a mobile phone and land phone using the format:<source country int. dial out #> <destination country code><destination area code><generated voice network 7 digit #>. Dial: 009-1-312-3805870 or 009-1-412-5284985 from any mobile or land phone from Nigeria (009) will connect to, and execute the application. To access the application from another country, replace 009 with source country international dial out #. The default username for student login is cartman. Figure 4.13 shows a model of mobile phone interaction with the system. A user dial a telephone number in Figure 4.13a. Once a connection is established (Figure 4.13b), the user can then commence conversation (Figure 4.13c) with *iVeES*.
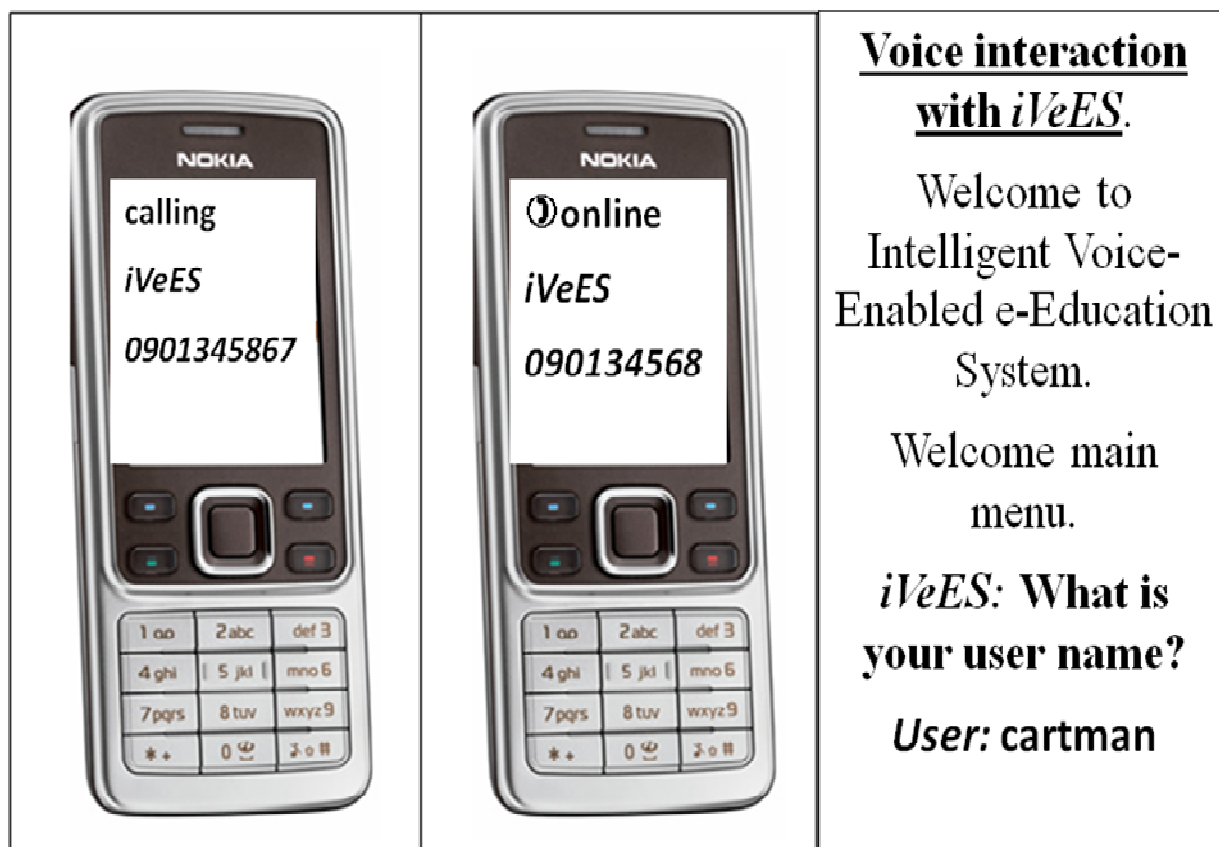
| a. | b. | c. |

Figure 4.13: A model of mobile phone interaction with *iVeES*

Similarly, a prototype part of the web application (web user interface) was deployed on a web server on the Internet and can be accessed using an Internet ready desktop PC by visiting the URL - http://www.emma.byethost24.com/eEducation. After deploying the application, the teachers and students went through some training session to guide them on how to use the prototype application. One of the instructions given to the students and teachers was that *iVeES* allows only multiple choice questions.

4.6   BENEFITS OF *iVeES*

The *iVeES* system offers numerous benefits:

- The *iVeES* application can reduce the related expenses for the traditional face-to-face learning method which is common in most developing countries such as lecture halls and other learning delivery facilities associated with physical presence between the students and teachers.

- The *iVeES* application can provide economies of scale at tertiary institutions, as cost of each additional learner is negligible once the lecture materials have been developed and hosted in a central server.

- In solving the digital illiteracy problems common to developing countries, the *iVeES* application can improve accessibility to education.

- The able (non-physically challenged) is offered an alternative learning platform; and there will be drastic reduction in the illiteracy level since mobile phone which is widely used device all over the world can be used for learning (m-Learning).

4.7   SUMMARY

In this chapter, an architectural design of the system is presented. Thereafter, a description of CBR, stemming, pseudocode and algorithm design and sample call flow of the application were described. Finally, in implementation, which is the follow-up section, some web user interface of *iVeES* as well as testing and deployment techniques were discussed.

CHAPTER FIVE

iVeES EVALUATION

5.1   INTRODUCTION

The *iVeES* application was evaluated for usability to determine the level of effectiveness, efficiency, users' satisfaction, learnability and memorability. The evaluation of a product is a fundamental requirement in determining the practical usability of a product (Ikhu-Omoregbe, 2010). The usability of the e-Learning application was measured to specify the features and attributes required to make the product usable using ISO 9241-11 (1998) and Nielsen (2003) standard of usability.

The telephony application was tested using user survey to evaluate the usability attributes. The method proposed in Walker et al., (1997), called PARAdigm for DIalogue System Evaluation (PARADISE) could also be used to evaluate the system. The choice of ISO and Nielsen standard of usability criteria is due to our prior experience of respondents in mobile phone usage, to compare the findings of this study with that obtained from the pilot implementation of *iVeES* project (Azeta et al., 2009b), and to determine the level of proficiency and memorability of the users. The users who evaluated the application are not first time users; they garnered some experience of VUI application during the pilot implementation of the project.

5.2   RESEARCH DESIGN

In conducting the usability evaluation, questionnaire were designed and administered. The teachers and students from a school for the blind and partially sighted in Lagos, Nigeria were some of the respondents. The survey questionnaire was divided into five sections namely: background information, user experience with mobile phone and the system, effectiveness of the system, efficiency of the system and user satisfaction with the system, learnability of the system and memorability of the system.

Each section of effectiveness, efficiency, user satisfaction, learnability and memorability contains five Questions represented by Q1, Q2, Q3, Q4 and Q5. The questionnaire aims at eliciting information from the school in order to measure the usability of the voice-based e-Learning application provided. A total of 70 questionnaires were administered on teachers and students but only 63 responses were received, analyzed and reported (see Table 5.1). The questions were designed using five-point likert-scale where 1= strongly disagree, 2 = disagree, 3= undecided, 4 = agree and 5= strongly agree.

Table 5.1: Descriptive Statistical Analysis of Questionnaire Data

Summary of study variable						
Usability Measures	*Responses from teachers*		*Responses from Students*		*Total # of respondents*	*Total Mean Rating*
	Visually impaired teachers	*Normal teachers*	*Visually impaired students*	*Normal students*		
Effectiveness	5	5	47	6	63	3.98
Efficiency	6	6	47	6	63	4.29
Satisfaction	6	6	47	6	63	4.13
Learnability	5	5	46	7	63	4.07
Memorability	5	6	45	7	63	4.11

5.3   LESSONS LEARNT

There were issues during the deployment of the application in the school as a result of certain problems they encountered such as poor Internet service, cost of calling the application from a mobile phone and speech recognition delay (Azeta, et al., 2009b). In attempting to resolve some of these issues, the following were done. In a situation where students could not afford a mobile phone or where cost of calling the speech platform was not affordable, a PC phone (i.e. Skype) through Voice over Internet Protocol (VoIP) was chosen as an alternative.

To improve the accuracy of speech recognition, the application was modified to accept numeric characters as against alphanumeric characters as voice input for some of the menu items. For instance, there were some initial delay for the speech engine to recognize course codes such as 'say BIO200' or 'say ENG201' unless the caller repeatedly say the course code many times before the application could recognize and process the request. One of the reasons observed was the combination of alphabets and numeric characters as a voice input

94

in VoiceXML code. It was easier for the speech server to recognize terms whose characters are either alphabets or numbers, and not combination of alphabets and numbers as the case of BIO200. The course codes were placed in a menu form and requested the caller to say '1' for BIO200, say '2' for ENG201, etc. This resulted in the "Good Usability" of 4.12 scale out of 5 scale as against the 3.35 recorded in the pilot evaluation during the first trial of the application. The resistance from the school experienced during the pilot implementation was not noticed in the second phase of the project that resulted in 4.12 scale. A sample of the questions from each section in the questionnaire is presented in Appendix III.

5.4   EVALUATION RESULTS

An overall score of all the learners was computed for each of the usability dimensions by averaging all the ratings on the questionnaire that was used. With the assistance of some of the non-visually impaired teachers, the respondents were taken through a short training on how to use a mobile phone to dial a number that will connect the learners to the application and show how to navigate within the application. The average (AVG) ratings, standard deviation (SD) and variance (VAR) to determine the mean and dispersion of data collected for the usability attributes are presented in Table 5.2. In terms of ownership/usage of mobile phones, users in our sample own/use mobile phones for almost 6 years which resulted to 35 as those having and 28 as those not having experience. In the area of acquiring ICT training to be able to use the application, 51 out of 63 said they would need training. The average effectiveness is 3.98, average efficiency is 4.29 and average users' satisfaction is 4.13, average learnability is 4.07 and memorability is 4.11.

Some learners were concerned with the time it took for *iVeES* to recognize a human voice as a result of noise distortion, and that accounted for the low score of 3.98 out of 5 in the question that say *"The voice request & response I got from the system was clear"*. The recognition rate of Voxeo ASR engine was ineffective some of the times arising from noise interference. The VoiceXML interpreter for TTS had problems compiling strings stored in the database with single and double quotes ('and"). The quotes were edited out of the text to record success in TTS translation. Notwithstanding, the user survey results were encouraging, particularly, the Question two (Q2) under efficiency in Table 5.2 that asked '*I was able to complete my task on time'* resulting to 4.21.

Survey Questions	AVG	SD	VAR
Background Information			
Do you own/use a mobile phone? [yes] [no]	35/28		
Would you be able to afford a mobile phone to call the e-Learning application? [yes] [no]	39/24		
Would you support the use of mobile phone for e-Learning in your school? [yes] [no]	42/21		
Do you need more computing skills/training/time to be able to use the system? [yes] [no]	51/12		
Effectiveness			
Q1	I was able to complete my task successfully and correctly using the application	4.00	0.92	0.84
Q2	The system did not show error message(s) while using it.	4.02	0.85	0.73
Q3	I was able to recover from my mistakes easily.	4.16	0.83	0.68
Q4	I feel comfortable using the application.	4.21	0.83	0.68
Q5	The voice request & response I got from the system was clear	3.49	0.56	0.32
Average Effectiveness	**3.98**	**0.80**	**0.65**
Efficiency			
Q1	Using the system saves me time of learning	4.08	0.81	0.65
Q2	I was able to complete my task on time	4.21	0.72	0.52
Q3	I was well able to navigate the voice user interface on time when using the system	4.41	0.74	0.55
Q4	I didn't have to carry out too many/difficult steps before completing my Task	4.35	0.74	0.55
Q5	The learning content I requested for suited my needs	4.38	0.77	0.59
Average Efficiency	**4.29**	**0.75**	**0.57**
User Satisfaction			
Q1	The system was easy to learn	4.05	0.83	0.69
Q2	The system was easy to use and user friendly	4.13	0.75	0.56
Q3	I am satisfied using the system	4.11	0.84	0.71
Q4	I feel the system met my need	4.21	0.79	0.62
Q5	I am satisfied with the performance of the system in accomplishing my tasks	4.13	0.79	0.62
Average User Satisfaction	**4.13**	**0.80**	**0.64**
Learnability			
Q1	The system provides clarity of wordings.	4.10	0.82	0.67
Q2	The grouping and ordering of menu options is logical for easy learning.	4.06	0.76	0.58
Q3	The command names are meaningful.	4.05	0.83	0.69
Q4	I could perform tasks on a proficient level as a first time user	4.13	0.81	0.66
Q5	As new user, I was able to orient myself with the system	4.02	0.79	0.63
Average Learnability	**4.07**	**0.80**	**0.65**
Memorability			
Q1	I gain the same level of skill as the last time I used the system	4.02	0.81	0.66
Q2	I was able to remember the commands I performed last time	4.13	0.73	0.53
Q3	I was able to reestablish my skills after a long time of using the system	4.11	0.76	0.58
Q4	I could recall every commands I performed moments earlier	4.14	0.76	0.58
Q5	I could remember the ordering of menu structure after a long time of using the system	4.13	0.73	0.53
Average Memorability	**4.11**	**0.76**	**0.58**
Resultant Average Rating	*4.12*	*0.78*	*0.62*

Table 5.2: User survey results N=63

Most of the users were visually impaired. The duration of their disability was an average of 8 years. The users preferred use of *iVeES* to other assistive technology such as Braile, Stylus, etc, which are very expensive to purchase and maintain. The ASR was sensitive to environmental speech interference during the recognition process.

Frequency Distribution of *iVeES* Usability Evaluation

In the five questions that make up the efficiency questionnaire category, question five (Q5) contains the question "*The learning content I requested for suited my needs*". This section presents the statistical analysis of Q5 as shown in Figure 5.1, 5.2 and 5.3.

N	Valid	63
	Missing	0
Mean		4.3810
Std. Deviation		.77102
Variance		.594

Figure 5.1: Mean rating for Q5 (Efficiency)

Q5_Efficiency

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2.00	1	1.6	1.6	1.6
	3.00	8	12.7	12.7	14.3
	4.00	20	31.7	31.7	46.0
	5.00	34	54.0	54.0	100.0
	Total	63	100.0	100.0	

Figure 5.2: Frequency distribution for Q5 (Efficiency)
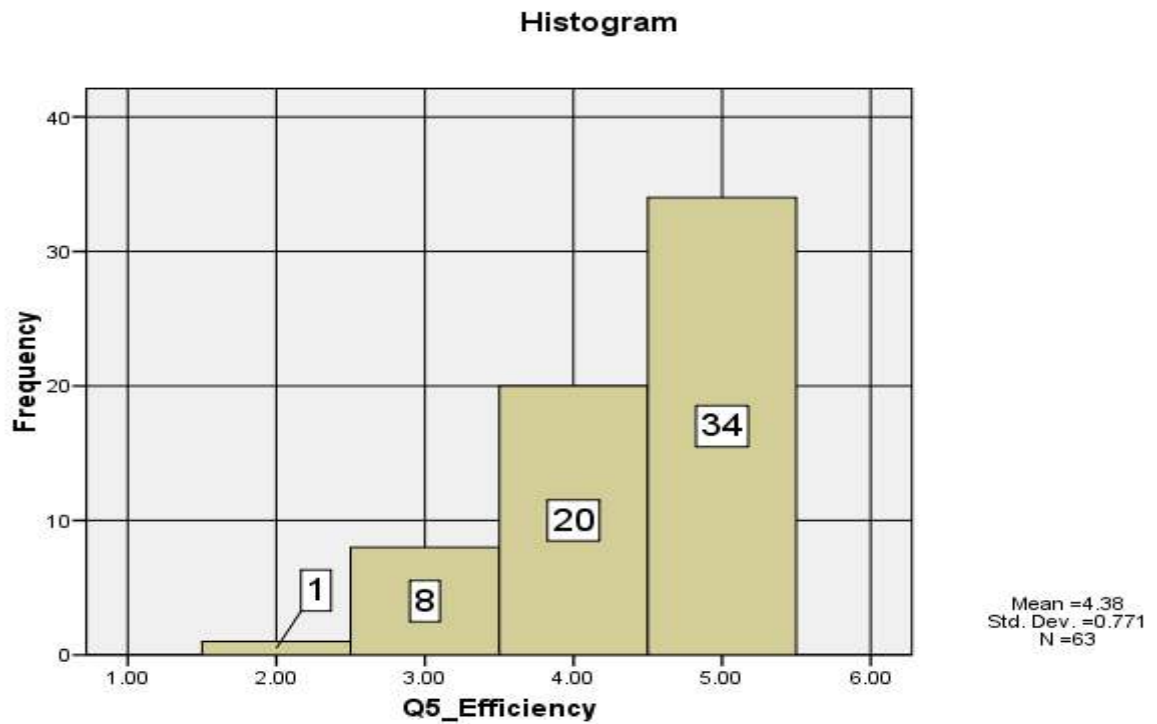
Figure 5.3: Frequency data for Q5 (Efficiency) Histogram

The usability attributes analysis of the evaluation for effectiveness, efficiency, user satisfaction, learnability and memorability are shown in Figure 5.4 with the scale and attributes for each of the usability criteria shown on Table 5.2.
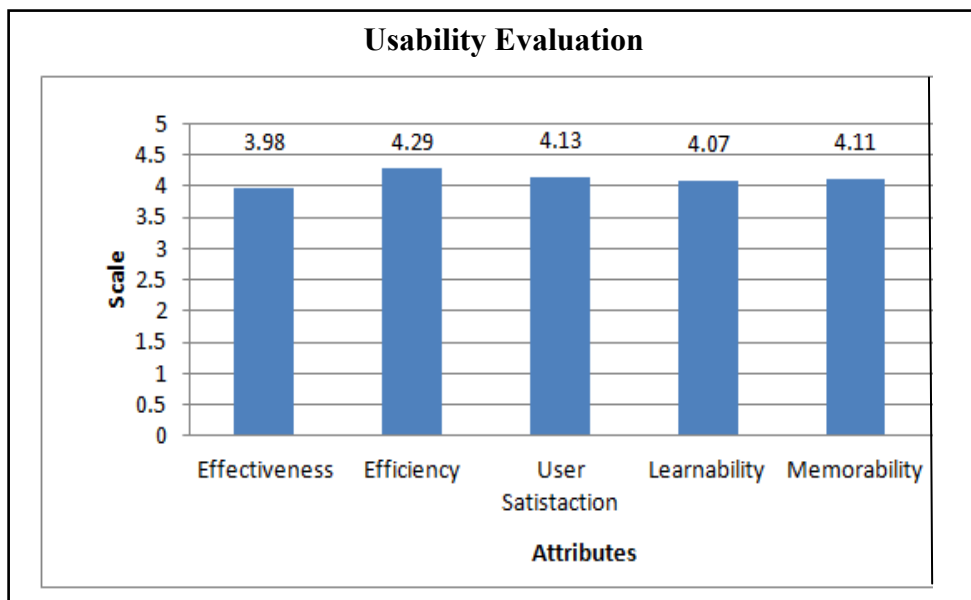
Figure 5.4: Usability attributes analysis of *iVeES*

Several usability studies suggest the system with "Very Bad Usability" should have 1 as mean rating, "2 as Bad Usability", 3 as Average Usability, "4 as Good Usability" and "5 as Excellent Usability". Sauro et al., (2005) proposed that "Good Usability" should have a mean rating of 4 on a 1-5 scale and 5.6 on a 1-7 scale. Therefore, it can be concluded that the prototype application developed for the school has "Good Usability" based on the average (AVG) total rating of 4.12.

5.5   DATA ANALYSIS

For all the learners, an overall score was computed for each of the usability dimension by averaging all the ratings on the questionnaire that was used. Statistical Package for Social Sciences (SPSS) was used to generate the frequency distribution and mean and all the relevant charts for the ratings.

Discussions

The ratings for the usability attributes for effectiveness,efficiency, user satisfaction, learnability and memorability are presented in Figure 5.5, Figure 5.6, Figure 5.7, Figure 5.8 and Figure 5.9.

Effectiveness: The "effectiveness" was evaluated for each of the tasks performed by each learner. The average mean rating (average effectiveness) for "Effectiveness" is 3.98 (See Figure 5.5).

Mean = 3.98, N = 63
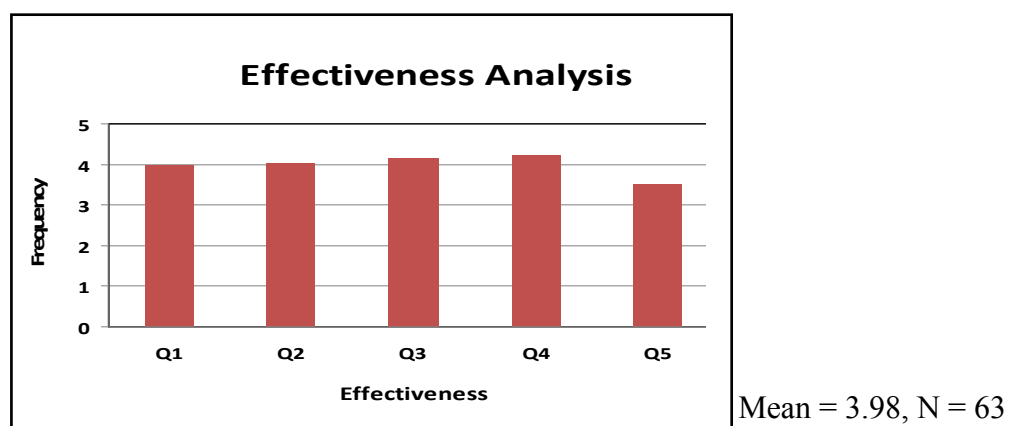
Figure 5.5: Effectiveness Analysis

Efficiency: The rating for "Efficiency" indicates the time it takes to achieve a task, most of the learners were able to realize their tasks on time as indicated in Figure 5.6: with a mean rating of 4.29.
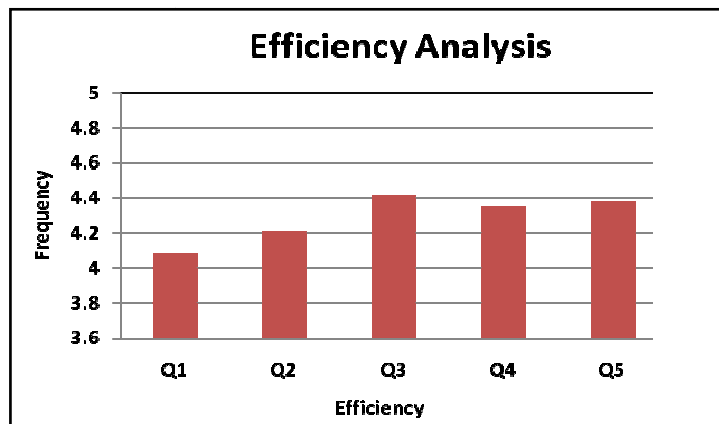
Mean = 4. 29, N = 63

Figure 5.6: Efficiency Analysis

Satisfaction: The system has a mean rating of 4.13 for "Satisfaction" as shown in Figure 5.7.
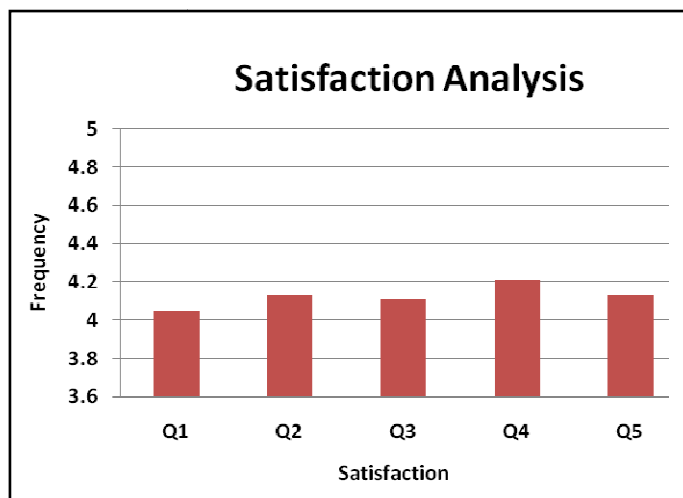
Mean = 4.07, N = 63

Figure 5.7: User Satisfaction Analysis

Learnability: The system has a mean rating of 4.07 for "Learnability" as shown in Figure 5.8.
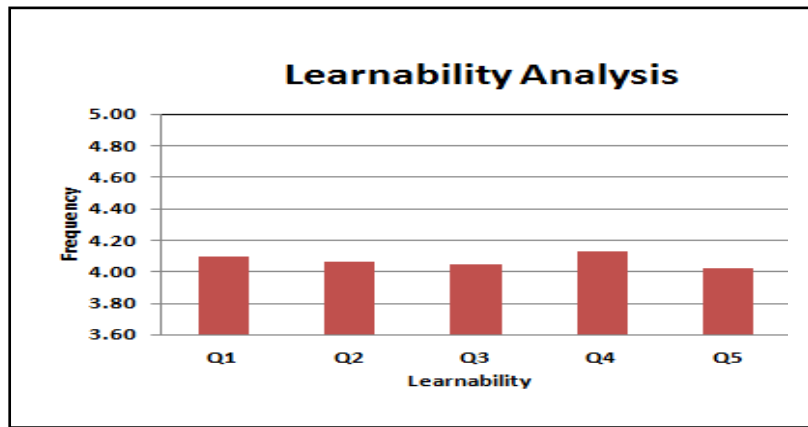
Mean = 4.07, N = 63

Figure 5.8: Learnability Analysis

Memorability: The system has a mean rating of 4.11 for "memorability" as shown in Figure 5.9.
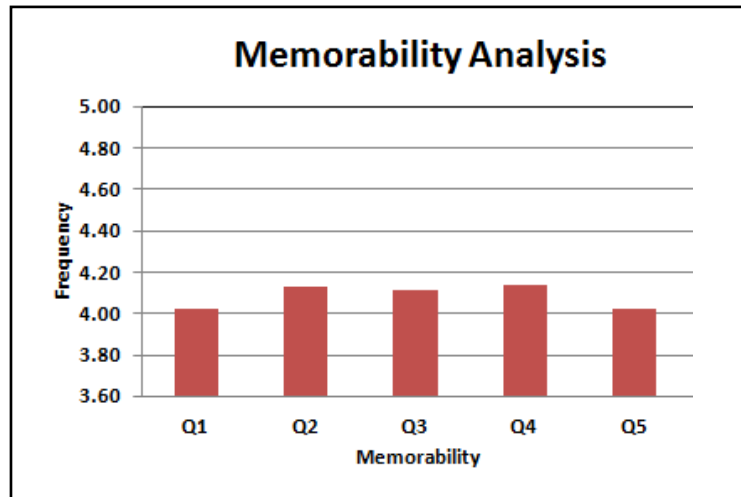
Mean = 4.11, N = 63

Figure 5.9: Memorability Analysis

The ratings for the five usability attributes (effectiveness, efficiency, user satisfaction, learnability and memorability) are depicted in Figure 5.10.
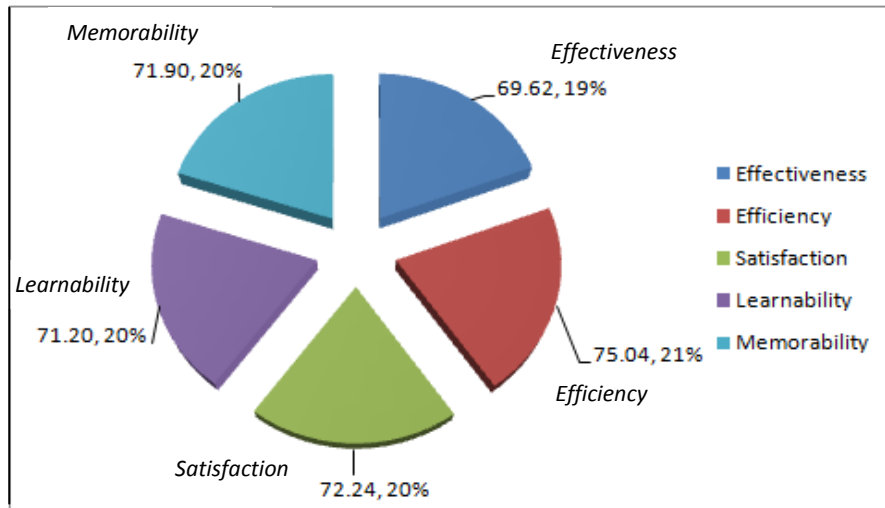
Figure 5.10: Usability Attributes Analysis

5.6 SUMMARY

This chapter presents the evaluation of *iVeES* using three usability criteria such as effectiveness, efficiency, user's satisfaction, learnability and memorability. The research design and evaluation results were discussed. Finally, the statistical analysis of data was presented.

CHAPTER SIX

SUMMARY, CONCLUSION AND RECOMMENDATIONS

6.1   SUMMARY

This study explored the learning experiences of students to bring to light the problems encountered by visually impaired learners for effective learning support. We have provided a framework and prototype voice-based e-Learning application using the VoiceXML to proffer a solution that will complement the existing learning methods in the school used as case study.

The voice-based e-Learning application provided for the blind has many implications. Learning can be realized faster and more efficiently than existing learning methods. The application can achieve just-in-time learning with greater reach irrespective of location (whether on the move, at home or work), speed of response and consistency of message. The response time of the use of intelligent Voice-based e-Education system is faster than a non-intelligent voice-based e-Education system. While the claim that voice learning is an expensive form of education can be misleading, voice learning can reduce the expenses incurred in the traditional face-to-face system such as lecture halls and other learning delivery facilities associated with physical presence between the students and teachers. It can also provide economies of scale at higher levels, as cost of each additional learner is negligible once the lecture materials have been developed and hosted in a central server.

The voice-based e-Learning technology will improve accessibility to education, including distance learning for learners who are visually impaired in the school for the blind. By so doing, the target group will be accommodated in promoting ICT in education. Loss of sight is one of the most difficult disabilities to come to terms with. The assistive technology reported in this thesis has the ability to fundamentally change the way teaching and training is delivered to the students of the school for the blind used in this case study and other similar schools.

ICT driven revolutionary change in the education sector has created ever-changing

knowledge and skills requirements, which the traditional approaches to learning are struggling to keep up. The face-to-face method of learning will provide the needed support for school teachers in teaching and learning. Meanwhile, there are growing staff shortage in the school while the demands for teachers continued to rise. Voice learning as a component of e-Learning has been deployed in the school for the blind to help enhance growth in the school with limited resources. The voice learning was able to provide the visually impaired learners with a more participatory educational experience.

6.2 CONCLUSION

In this thesis, a framework for intelligent voice-based e-Education system has been provided to offer solution to the problem encountered by learners who use telephone to access web learning content. This study attempts to solve the problem of using telephone access to find content suitability (FCS) among the numerous learning resources on the web. The *iVeES* framework was used as a generic work guideline to develop a prototype intelligent voice-based e-Education application for the *iVeES* project. The Intelligent component services of the application were realized using CBR, Modified Porter Stemming (MPS) algorithm and ontology. The prototype application was tested in a school for the blind, and the result of evaluation shows that the prototype application developed for the school has "Good Usability" rating of 4.12 out of 5 scale. The findings show that the users are enthusiastic about using a voice-based telephone learning as another form of assistive technology to compliment the conventional learning methods for the visually impaired.

The framework would serve as a reference model for implementing telephone-based e-Education applications for normal and visually impaired learners. The application will also assist people with physical access difficulties (e.g. repetitive strain injury, arthritis, high spinal injury) that make writing difficult. It could equally be useful to students with reading or spelling difficulties (e.g. dyslexia). It is also vital in the area of ubiquitous learning. The e-Learning application will be useful especially for students who are physically challenged such as the visually impaired. It will also be useful for people who have medical history of reacting to repetitive strain injury (RSI) as a result of their sedentary lifestyle using the keyboard. Therefore this system gives every learner equal right and access to quality education irrespective of his or her physical disability.

The e-Learning system offers an alternative platform of learning for learners without disability. It complements the existing e-Learning systems such as web-based learning, m-Learning, etc. The e-Learning system offers the possibility to learn anytime and anywhere there is telephone access, regardless of the availability of Internet services.

6.3   RECOMMENDATIONS

iVeES has several unique features that can appeal to the non-visually impaired and visually impaired users alike. The success of *iVeES* has the potentials to attract a wide variety of applications that can be voice-enabled in future. The benefits of using *iVeES* include: (1) No formal training. The only training required is how to use a phone, (2) CallerID login. Registered users can login with telephoneID and password, (3) Customizable VUI. The welcome message, (4) VOIP and PSTN phone access, and (5) Available anytime anywhere there is telephone access.

Other e-Learning application developers who try to develop a similar voice application for the blind will need to consider several factors. To deliver and access e-enabled voice-based learning materials will require a new level of competence and awareness of ICT on the part of the users. There are complexities with some users having to develop the basic ICT skills before using voice learning. There should be blended model with the face-to-face sessions at the initial deployment stage to prepare users for voice learning, a practice often referred to as parallel changeover from the existing learning methods to the new voice-based learning. This will minimize resistance from users. Resistance from students and teachers arise because they are used to attending classroom sessions physically. Such cultural expectations will change over time.

Another level of resistance may come from the perception of restrictions of mobile devices in terms of output and input capabilities. Mobile devices often have limited screen display sizes and limited capacity to support audio and video data. The remedy to this issue is the fact that voice application only requires dialing a telephone number to connect to an application, which does not require any additional resource overhead on a mobile phone. However, some visually impaired learners may still prefer to hear the teacher in a class room setting.

Generally, the problem of resistance may be more severe in developing countries where human capacity development may be less robust and the economy less stable. One further step to addressing the resistance issue is to understand why some learners and teachers resist. There may be a variety of reasons which include 1) the teacher's fear that the technology will make them obsolete and may lose their jobs, and 2) lack of familiarity with technology and the fear that they will look stupid before others if they do not use it correctly.

While voice learning can appear to be the best option for specific learning requirements such as the visually impaired, cost of development can be prohibitive if made available through a service provider. The cost of developing the smallest voice-based e-Learning application is enormous, that most schools for the blind may not be able to afford. A cost effective alternative would be to develop materials and content in-house and employ the services of an ICT administrator to manage the learning content that will be accessed by calling a telephone number. This will serve as a remedy to the issues of cost and skill since most schools for the blind in developing countries such as Nigeria do not have the required funds, skills and knowledge to make it work effectively.

The biggest challenge with speech user interfaces is that the interactive objects are invisible. Therefore, users must know what objects and operations are available in each system, and what they are called (Hakulinen, 2006). Learners, particularly the visually impaired users should be taught how to use any voice application that is provided for them. There is need to increase training and development capacity for students and teachers, particularly in the school for the blind, for national development. It is unlikely that classroom-based delivery alone can provide sufficient capacity to bridge the gap between the developed and developing countries in the area of technology supported learning for the blind. The digital divide challenges facing developing countries in terms of ICT deployment, especially the neglected populace such as the school for the blind are numerous. These include digital illiteracy, lack of adequate infrastructure such as electricity, etc, and lack of suitable ICT legal framework to support the visually impaired learners. In the developed and developing countries, several methodologies exist for implementing e-Learning applications. Meanwhile, not many of them consider the plight of the visually impaired during the analysis and design stages of these applications. There is need to address these issues so that the visually impaired can be part of the global ICT revolutionary change through e-Education.

6.4. FUTURE WORK

The future research direction, for this study is two-fold. First, evaluation of the system using PARAdigm for DIalogue System Evaluation (PARADISE), and second, voice biometric techniques based on speech data will be added as additional means of security mechanism to enhance the authentication of candidates for examination. One key identification technique is the use of voiceprint of the caller. An improved system should therefore be provided as further research to combine the conventional pin/telephone number authentication and biometric to increase security of the system.

With these additional features, the voice learning application will be able to provide a more participatory and secured voice-based educational experience for the students.

APPENDIX I: LIST OF PUBLICATIONS FROM THE THESIS

[1]   **Azeta A. A**., Ayo C. K., Atayero A. A. and Ikhu-Omoregbe N. A. (2009), A Case-Based Reasoning Approach for Speech-Enabled e-Learning System", *Proceedings of 2nd IEEE International **Conference** on Adaptive Science & Technology (ICAST). December 14 – 16, 2009, Accra Ghana. pp. 211-217. ISBN: 978-1-4244-3523-4 ISSN: 0855-8906*. Available online at IEEE Xplore database tab 4: http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=5375737

[2]   **Azeta A. A**., Ayo C. K., Atayero A. A. and Ikhu-Omoregbe N. A. (2009) "Application of VoiceXML in e-Learning Systems", *Cases on Successful E-Learning Practices in the Developed and Developing World: Methods for the Global Information Economy*. Chapter 7, a **Book** Published in the United States of America (USA) by Information Science Reference (an imprint of IGI Global). Pp. 92-108. Available online at: http://www.igi-global.com/reference/details.asp?ID=35211&v=tableOfContents

[3]   **Azeta A. A**., Ayo C. K., Atayero A. A. and Ikhu-Omoregbe N. A. (2010), "Intelligent Voice- Based E-Education System: A Framework and Evaluation", *International **Journal** of Computing, 2010, Vol. 9, Issue 4.Ukraine. ISSN: 1727-6209. pp. 327 – 324.* Available online at http://www.computingonline.net

[4]   **Azeta A. A**., Ayo C. K., Atayero A. A. and Ikhu-Omoregbe N. A. (2008), "Development of A Telephone-based e-Learning Portal", *1st International **Conference** on Mobile Computing, Wireless Communication, e-Health, M-Health and TeleMedicine* (FICMWiComTelHealth '08), 18th – 20th Nov. 2008 at Ladoke Akintola University of Technology(LAUTECH), Ogbomosho, Oyo State, Nigeria (pp. 141-149)"

APPENDIX II: LIST OF ABREVIATIONS AND SYMBOLS

AI	Artificial Intelligence
ANN	Artificial Neural Networks
ASR	Automatic Speech Recognition
BPM	Business Process Modeling
CBR	Case-Based Reasoning
CV	Computer Vision
CSE	Computer Science and Engineering
DTMF	Dual-Tone Multi-Frequency
DOM	Document Object Model
DTW	Dynamic Time Warping
ES	Expert Systems
FCS	Finding Content Suitability
FL	Fuzzy Logic
FOSS	Free and Open Source Software
FSA	Finite State Automata
GA	Genetic Algorithms
GUI	Graphical User Interface
HIS	Hybrid Intelligent Systems
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
HMM	Hidden Markov Model
ICT	Information and Communication Technology
IDE	Integrated Development Environment
iVeES	intelligent Voice-enabled e-Education Systems
IR	Information Retrieval
IVR	Interactive Voice Response
JSML	Java Speech Markup Language
LOC	Locus of Control
LMS	Learning Management Systems
LP	Logic Programming
MELFA	Mobile E-Learning for Africa

ML	Machine Learning
MPS	Modified Porter Stemming
MUVE	Multi-User Virtual Environment
NLP	Natural Language Processing
OWL	Web Ontology Language
PARADISE	PARAdigm for DIalogue System Evaluation
PC	Personal Computer
PDA	Personal Digital Assistant
PSTN	Public Switched Telephony Network
RSI	Repetitive Strain Injury
SAPI	Speech Application Programming Interface
SD	Standard Deviation
SPSS	Statistical Package for Social Sciences
SALT	Speech Application Language Tags
SSML	Speech Synthesis Markup Language
SRGS	Speech Recognition Grammar Specification
TF/IDF	Term Frequency/Index Document Frequency
TM4L	Topic Maps for e-Learning
TTS	Text-to-Speech
VAR	Variance
VoiceXML	Voice eXtensible Mark-up Language
VoIP	Voice over Internet Protocal
VoIS	Voice-enabled Interactive Service
VoxML	Voice Markup Language
VUI	Voice User Interface
VSM	Vector-Space Model
WAP	Wireless Application Protocol
WUI	Web User interface
WYSIWYG	What you see is what you get
W3C	World Wide Web Consortium
XHTML	Extensible Hypertext Markup Language
XML	Extensible Markup Language
<	Less than

≤	Less than or equal to
=	Equal to
≥	Greater than or equal to
-	Subtraction
+	Addition
*	Multiplication
∑	Summation
√	Square root

APPENDIX III: SAMPLE QUESTIONNAIRE

System Evaluation Questionnaire for Voice-Based e-Learning

INTRODUCTION

Good day Sir/Madam,

This questionnaire aims at eliciting information from your institution in order to measure the usability of the voice-based e-Learning application developed for your school as a case study for providing learning support services. Please answer the question honestly by ticking (√) or writing the answer that best expresses your view. We would like to assure you of the confidentiality of the information you provide. Thank You.

SECTION A: Background Information

1. Name of respondent: _____

2. Name of institution: _____

3. TEACHER: Position of respondent: _____

    No of years as a teacher [0-3 Years][4-6 Years] [7-10 Years] [More than 10 Years]

    STUDENT: Level/Class of respondent _____

. 4. Gender: Male [ ]   Female [ ]

SECTION B: User Experience with Mobile phone and the system

5.   Do you own/use a mobile phone? [Yes] [No]

6. If Yes to question 5, for how long? [Past 6 Month] [Past 2 Year] [Past 3 Years] [Past 4 Years] [4 Years and above]

7.   If No to question 5, then state reason(s) _____

8.   Would you be able to afford a mobile phone to call the e-Learning application? [Yes] [No]

9.   How would you rate your experience/skill in the use of mobile and land phones?

    [Novice]   [Average]   [Good]   [Expert]

10. Would you support the use of mobile phone for e-Learning in your school? [Yes] [No]

11. If your answer to question 10 is No, please underline **ALL** your reasons

    [A] Reliability   [B] Security   [C] Cost   [D] Usability

    [E] Inadequate Training   Others (Please Specify)_____

12. Are you excited at using the e-Learning application [Yes] [No]

13. If No to question 13, then state reason(s) _____

14. State the challenge(s) you have using the e-Learning application (if any)

    _____

15. Do you need more computing skills/training/time to be able use the system? [Yes] [No]

16. Which task/module did you use (tick one or both)? [Registration] [Examination]

17. Which device did you use to connect to the system?

    [Mobile phone]   [Land phone]   [PC phone: Skype, others_____

Please answer the following questions after using the application. The information below provides the code and meaning for each of the option to be ticked.

1- Strongly Disagree   2- Disagree   3- Undecided   4- Agree   5- Strongly Agree

SECTION C: Effectiveness of the System

18. I was able to complete my task successfully and correctly using the application

                1   2   3   4   5

    Strongly Disagree   ☐ ☐ ☐ ☐ ☐   Strongly Agree

19. The system did not show error message(s) while using it.

                1   2   3   4   5

    Strongly Disagree   ☐ ☐ ☐ ☐ ☐   Strongly Agree

20 I was able to recover from my mistakes easily.

                1   2   3   4   5

    Strongly Disagree   ☐ ☐ ☐ ☐ ☐   Strongly Agree

21. I feel comfortable using the application.

	1	2	3	4	5	
Strongly Disagree	☐	☐	☐	☐	☐	Strongly Agree

22. The voice input and response I got from the system was clear

	1	2	3	4	5	
Strongly Disagree	☐	☐	☐	☐	☐	Strongly Agree

SECTION D: Efficiency of the System

23. Using the system saves me time of learning

	1	2	3	4	5	
Strongly Disagree	☐	☐	☐	☐	☐	Strongly Agree

24. I was able to complete my task on time

	1	2	3	4	5	
Strongly Disagree	☐	☐	☐	☐	☐	Strongly Agree

25. I was well able to navigate the voice user interface on time when using the system

	1	2	3	4	5	
Strongly Disagree	☐	☐	☐	☐	☐	Strongly Agree

26. I didn't have to carry out too many/difficult steps before completing my task

	1	2	3	4	5	
Strongly Disagree	☐	☐	☐	☐	☐	Strongly Agree

27. The learning content I requested for suited my needs

	1	2	3	4	5	
Strongly Disagree	☐	☐	☐	☐	☐	Strongly Agree

SECTION E: User Satisfaction with the System

28.   The system was easy to learn

              1   2   3   4   5
Strongly Disagree ☐   ☐   ☐   ☐   ☐ Strongly Agree

29.   The system was easy to use and user friendly

              1   2   3   4   5
Strongly Disagree ☐   ☐   ☐   ☐   ☐ Strongly Agree

30.   I am satisfied using the system

              1   2   3   4   5
Strongly Disagree ☐   ☐   ☐   ☐   ☐ Strongly Agree

31.   I feel the system met my need

              1   2   3   4   5
Strongly Disagree ☐   ☐   ☐   ☐   ☐ Strongly Agree

32.   I am satisfied with the performance of the system in accomplishing my tasks

              1   2   3   4   5
Strongly Disagree ☐   ☐   ☐   ☐   ☐ Strongly Agree

SECTION F: Learnability of the System

33.   The system provides clarity of wordings

              1   2   3   4   5
Strongly Disagree ☐   ☐   ☐   ☐   ☐ Strongly Agree

34.   The grouping and ordering of menu options is logical for easy learning.
              1   2   3   4   5
Strongly Disagree ☐   ☐   ☐   ☐   ☐ Strongly Agree

35.   The command names are meaningful

1   2   3   4   5

Strongly Disagree   ☐ ☐ ☐ ☐ ☐   Strongly Agree

36.   I could perform tasks on a proficient level as a first time user

1   2   3   4   5

Strongly Disagree   ☐ ☐ ☐ ☐ ☐   Strongly Agree

37.   As new user, I was able to orient myself with the system

1   2   3   4   5

Strongly Disagree   ☐ ☐ ☐ ☐ ☐   Strongly Agree

SECTION G: Memorable of the System

38.   I gain the same level of skill as the last time I used the system

1   2   3   4   5

Strongly Disagree   ☐ ☐ ☐ ☐ ☐   Strongly Agree

39.   I was able to remember the commands I performed last time

1   2   3   4   5

Strongly Disagree   ☐ ☐ ☐ ☐ ☐   Strongly Agree

40.   I was able to reestablish my skills after a long time of using the system

1   2   3   4   5

Strongly Disagree   ☐ ☐ ☐ ☐ ☐   Strongly Agree

41.   I could recall every commands I performed moments earlier

1   2   3   4   5

Strongly Disagree   ☐ ☐ ☐ ☐ ☐   Strongly Agree

42.   I could remember the ordering of menu structure after a long time of using the system

1   2   3   4   5

Strongly Disagree   ☐ ☐ ☐ ☐ ☐   Strongly Agree

APPENDIX IV: SAMPLE VOICEXML SOURCE CODE

```
<?xml version="1.0" encoding="UTF-8"?>
<vxml version = "2.1">
 <nomatch>
      No Match! I'm sorry, I didn't understand you.  Could you please try again?
      <reprompt />
   </nomatch>
   <noinput>
      No Input! I'm sorry, I didn't hear anything.  Could you please try again?
      <reprompt />
   </noinput>
<var name="callerid" expr="session.telephone.ani"/>
<form id="main2">
  <block>
 <prompt bargein="false">
Welcome to intelligent Voice enabled eEducation system.
</prompt>
<goto next="#mainmenu"/>
    </block>
</form>
<menu id="mainmenu">
<prompt bargein="false">
Your telephone number stroke caller id is.
<value expr="callerid"/>
Welcome to main menu.
   </prompt>
 <prompt>
   Say one of: <enumerate/>
 </prompt>
 <choice next="#student">
   Student menu
 </choice>
```

```
  <choice next="#faculty">
    Faculty menu
  </choice>
  <choice next="#administrator">
    Admin menu
  </choice>
<choice next="#exit1">
    exit
  </choice>
  <noinput>Please say one of <enumerate/></noinput>
</menu>
<form id="student">
<property name="inputmodes" value="voice"/>
    <field name="username">
      <prompt>
Welcome to student menu. What is your user name.
      </prompt>
      <grammar mode="voice" root="text">
        <rule id="text" scope="public">
         <one-of>
 <item>cartman</item>
            <item>eji</item>
 <item>joy</item>
 <item>mat</item>
 <item>key</item>
 -
</one-of>
</rule>
</grammar>
</field>
<filled>
<submit     next="http://localhost/eEducationVui/studentlogin.php"     method="POST"
namelist="callerid username"/>
```

```
 <goto next="#menu"/>
    </filled>
 </form>
<form id="voicelearning">
    <property name="inputmodes" value="voice"/>
<field name="answer1">
      <prompt>
  Welcome to Voice learning session.
The system will Mention the course code you want to study. say yes to accept or no to reject.
BIO200, say yes or no.
      </prompt>
     <grammar mode="voice" root="text">
       <rule id="text" scope="public">
        <one-of>
          <item>yes</item>
      <item>no</item>
          </one-of>
        </rule>
      </grammar>
</field>
<field name="answer2">
      <prompt>
        BIO201, say yes or no.

      </prompt>
      <grammar mode="voice" root="text">
        <rule id="text" scope="public">
         <one-of>
           <item>yes</item>
       <item>no</item>
          </one-of>
        </rule>
```

```
        </grammar>
    </field>
      <filled>
<submit      next="http://localhost/eEducationVui/voicelearning.php"      method="POST"
namelist="answer1 answer2 answer3"/>
 </filled>
 </form>
</form>
</vxml>
```

REFERENCES

[1]   Aarnio T. (1999), "Speech Recognition with Hidden Markov Models in Visual Communication", Department of Mathematical Sciences, University of Turku, 1999. Available online at: http://research.nokia.com/files/Speech%20Recognition%20with%20Hidden%20Markov%20Models%20in%20Visual%20Communication.pdf

[2]   Abbott, K(2002).: Voice Enabling Web Applications: VoiceXML and Beyond, A Press, Berkeley, 2002. ISBN: 1-893115-73-9.

[3]   Abdel-Badeeh M. S. and Rania A. (2005), "A Case Based Expert System for Supporting Diagnosis of Heart Diseases", *AIML Journal*, Volume (5), Issue (1), March'05

[4]   Abraham A. (2005), *"Oklahoma State University, Stillwater, OK, USA Handbook of Measuring System Design*, edited by Peter H. Sydenham and Richard Thorn. 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8. Pp 893-900

[5]   Abu-Zaina H., Kirk D. and Elmer A. (1996), "A Speech Recognition System For Operation Of A Visual Inspection Workstation", RESNA '96 Proceedings, pg 143. Hoyer Institute for Rehabilitation Research and Service Wichita State University Wichita, KS U.S.A., available online at: http://www.dinf.ne.jp/doc/english/Us_Eu/conf/resna96/page143.htm

[6]   AccessIT(2006),"The National Center on Accessible Information Technology in Education, University of Washington. Funded by the National Institute on Disability and Rehabilitation Research of the U.S. Department of Education (grant #H133D010306)

[7]   Adaptive (2005). Adaptive and assistive technologies in e-learning. Retrieved February 20, 2009, from

http://www.adcet.edu.au/StoredFile.aspx?id=1353&fn=adptve_assstv_techn_elearning.Pdf

[8]   Armstrong H. (2009), "Advanced IT Education for the Vision Impaired via e-Learning", Journal of Information Technology Education, Volume 8, 2009. PP 244 – 256. Retrieved 14th February 2010 from http://jite.org/documents/Vol8/JITEv8p243-256Armstrong685.pdf

[9]   Aroyo, L. and Dicheva D. (2004). The New Challenges for E-learning: The Educational Semantic Web. In *Educational Technology & Society, 7(4), pp. 59-69, 2004.*

[10]   Aroyo, L., Mizoguchi, R., and Tzolov, C. (2003). OntoAIMS: Ontological Approach to Courseware Authoring. *International Conference on Computers in Education (ICCE'03).*

[11]   Atayero A. A, Ayo C. K. Ikhu-Omoregbe N. A, Azeta, A. A. (2009a), "Implementation of 'ASR4CRM': An Automated Speech-Enabled Customer Care Service System", *IEEE Xplore. 2009. Eurocon'09 IEEE International Conference*, Saint-Petersburg, Russia. May 18 – 23 2009. pp. 1712-1715. Program available online at:
www.ewh.ieee.org/r8/russia/nw/cem/eurocon2009/files/Schedule/May_20_Schedule.doc

[12]   Azeta A. A, Ikhu-Omoregbe N. A, Ayo C. K and Atayero A. A(2008a), "A Software Engineered Voice-Enabled Job Recruitment Portal System", *European Journal of Scientific Research (EJSR)*, ISSN 1450-216X Vol.21 No.1 July 2008, pp. 95-103. Available online at: http://www.eurojournals.com/ejsr%2021%201.pdf

[13]   Azeta A. A., Ikhu-Omoregbe N. A., Ayo C. K., and Atayero A.A. (2008b), "Development and Deployment of VoiceXML-Based Banking Applications", *Journal of Nigeria Computer Society*, Vol. 15 No.1 June '08 Edition, pg 59-72.

[14]   Azeta A. A., Ayo C. K., Atayero A. A. and Ikhu-Omoregbe N. A. (2008c), "Development of A Telephone-based e-Learning Portal", *1ˢᵗ International Conference on Mobile Computing, Wireless Communication, e-Health, M-Health and TeleMedicine* (FICMWiComTelHealth '08), 18ᵗʰ – 20ᵗʰ Nov. 2008 at Ladoke Akintola University of Technology(LAUTECH), Ogbomosho, Oyo State, Nigeria (pp. 141-149)"

[15]   Azeta A. A.,Ayo C. K., Atayero A. A. and Ikhu-Omoregbe N.A. (2009b), "Application of VoiceXML in e-Learning Systems", Cases on Successful E-Learning Practices in the Developed and Developing World: Methods for the Global Information Economy. Published by IGI Global. Available online at: http://www.igi-global.com/reference/details.asp?ID=35211&v=tableOfContents

[16]   Azeta A. A., Ayo C. K., Atayero A. A. and Ikhu-Omoregbe N.A. (2009c), "A Case-Based Reasoning Approach for Speech-Enabled e-Learning System", *2nd IEEE International Conference on Adaptive Science & Technology (ICAST)*. December 14 – 16, 2009, Accra Ghana. Available online at IEEE Xplore database tab 4: http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=5375737

[17]   Belkin N. J. (2005), "Intelligent Information Retrieval: Whose Intelligence", Understanding and Supporting Multiple Information Seeking Strategies, a TIPSTER Phase III Research Project, School of Communication, Information and Library Studies, Rutgers University

[18]   Begum S., Ahmed M. U., Funk P., Xiong N., Schéele B. B (2007), "Similarity of Medical Cases in Health Care Using Cosine Similarity and Ontology", in the proceedings of the 5ᵗʰ Workshop on CBR in the Health Sciences, p 263-272, Springer LNCS, Bel-fast, Northern Ireland, August, 2007 Retrieved 26 March 2010 from *www.cbr-biomed.org/workshops/ICCBR07/iccbr07wk-Begum.pdf*

[19]   Bobbie P. O., Fordjour I.. Aboagye D. O.,Dzidonu C., Darkwa O. and Nyantakyi K. (2008) "Technology Enablers for collaborative Virtual Education", 3rd International

Conference on ICT for Development, Education and Training. AICC, Accra, Ghana. May 28 – 30, 2008. pp. 191-193.

[20]   Bures I. M. and Jelinek I. (2004), "DESCRIPTION OF THE ADAPTIVE WEB SYSTEM FOR ELEARNING", IADIS International Conference e-Society 2004. Volume: 2 . Retrieved 9th Nov. 2010 from http://www.iadis.net/dl/final_uploads/200402C042.pdf

[21]   Burke D. (2007), a textbook titled: "Speech processing for IP Networks, media resource control protocol (MRCP)". Publisher: Chichester, West Sussex : John Wiley & Sons, 2007.

[22]   Divakaran D. (2011), "Create and Setup a Website in 10 mins". Retrieved 7th Oct. 2011 from http://www.deepeshmd.com/wp-content/uploads/Create-and-Setup-a-Website-in-10-mins.pdf

[23]   Chang S. E. (2007), "An Adaptive and Voice-enabled Pervasive web System", -International Journal of Enterprise Information Systems ,Volume 3, Issue 4. pp. 69-83 IGI publishing

[24]   Chowdhary K. R.(2005) "Current Trends in Information Retrieval", available online at: http://ncsi-net.ncsi.iisc.ernet.in/gsdl/collect/icco/index/assoc/HASH014f.dir/doc.doc, accessed 5th July 2008.

[25]   Chin C. C., Hock G. T. and Veerappan C. M. (2006), "VoiceXML as Solution for Improving Web accessibility and Manipulation for e-Education", *School of Computing and IT, INTI College Malasia*. Available online at: http://intisj.edu.my/INTISJ/InfoFor/StaffResearch/10.pdf accessed July 2008.

[26]   Cisco (2010), Cisco Unified Customer Voice Portal 8.0, Printed in USA. Page 1-5. Retrieved online 6th September from 2011http://www.promero.com/pdf/CiscoUnifiedCustomerVoicePortal8.0.pdf

[27]   Cook S. (2002), Speech Recognition HOWTO, available at: http://tldp.org/HOWTO/Speech-Recognition-HOWTO/software.html accessed July 2008.

[28]   Cranefield S., Hart L., Dutra M., Baclawski K. and Kokar M., Smith J. (2002), "UML for ontology development", The Knowledge Engineering Review (2002), 17: 61-64. Published online by Cambridge University Press 21 Aug 2002

[29]   Datamonitor (2006), "An Introductory Guide To Speech Recognition Solutions", *Understanding the technology, the vendors and the market Industry white paper* published Datamonitor. (pp 1- 20)

[30]   Devedzic, V. (2004). Education and the Semantic Web. In *International Journal of Artificial Intelligencein Education 14 (2004) 39-65.* IOS Press.

[31]   Devinder S. and Zaitun A.B (2006)," Mobile Learning in Wireless Classrooms", Malaysian Online Journal of Instructional Technology (MOJIT) Vol. 3, No.2, pp 26- 42, August 2006, ISSN 1823:1144

[32]   Dicheva D., Sosnovsky S., Gavrilova T., Brusilovsky P. (2005) "Ontological Web Portal for Educational Ontologies", International Workshop on Applications of *S*emantic *W*eb technologies for *E-L*earning (SW-EL). July 18-22 2005. Amsterdam. The Netherlands. Retrieved 18th November 2010 from http://www.win.tue.nl/SW-EL/2005/swel05-aied05/proceedings/4-Dicheva-final-full.pdf

[33]   Divitini, M., Haugalokken, O. and Morken, E.M.(2005), "Blog To Support Learning In The Field: Lessons Learned From A Fiasco. *In Proceedings of the Fifth IEEE International Conference on Advanced Learning Technologies ICALT '05.* pg 219-221

[34]   Donegan, M. (2000) *BECTA Voice Recognition Project Report. BECTA,*
         Available online at :
         http://www.becta.org.uk/teachers/teachers.cfm?section=2&id=2142
         accessed July 2008.

[35]   Ebibi M., Fetaj M., Loskoska S. and Fetaj B. (2007), "Usability of Virtual
         Learning Environment for Learning JAVA"., BCI 2007, Sofia, Bulgaria P523-534.
         Retrieved 27 February 2012 from http://halshs.archives-
         ouvertes.fr/docs/00/19/00/66/PDF/523-534.pdf

[36]   Ehsani F. and knodt E. (1998), Speech Technology in Computer-aided   Language
         Learning: Strengths and Limitations of a new Call Paradigm, *Language Learning
         &Technology, J*uly 1998, volume 2, number 1, Pp. 54-73. Retrieved online 14th
         February 2010 from http://llt.msu.edu/vol2num1/article3/

[37]   Ermalai I., Andone D., Vasiu R. (2009), "Study Cases on e-Learning Technologies
         Used by Universities in Romania and Worldwide", WSEAS TRANSACTIONS on
         COMMUNICATIONS, Issue 8, Volume 8, August 2009. pp 785 – 794. Retrieved 8th
         November 2010 from
         http://www.wseas.us/e-library/transactions/communications/2009/29-640.pdf.

[38]   Fabbrizio G. D.,Okken T. and Wilpon J. G.(2009), "A Speech Mashup Framework
         for Multimodal Mobile Services", ICMI-MLMI'09, November 2–4, 2009,
         Cambridge,   MA,   USA..   Retrieved   18   Mar   2010   from
         http://www.difabbrizio.com/papers/icmi174-difabbrizio.pdf

[39]   Farrans J. (2004), "COMMUNICATIONS OF THE ACM July 2004/Vol. 47, No. 7",
         Piscataway, NJ

[40]   Farinazzo V., Salvador1 M. Luiz A. S. Kawamoto and João Soares de Oliveira Neto
         (2010), "An Empirical Approach for the Evaluation of Voice User Interfaces" User
         Interfaces, Book edited by: Rita Mátrai, ISBN 978-953-307-084-1, pp. 270, May

2010, INTECH, Croatia, downloaded from SCIYO.COM and *www.intechopen.com/download/pdf/pdfs_id/10804*

[41] Fiedler G. and Schmidt P. (2004), "Developing Interactive Voice Response Interfaces for Large   Information Systems". Pp 182-184. Available online at: http://www.mgovernment.org/resurces/euromgov2005/PDF/19_S042FG-S04.pd accessed July 2008.

[42] Freitas,D. and Kouroupetroglou G. (2004), "Speech technologies for blind and low vision Persons", Technology and Disability 20 (2008) 135–156. IOS Press

[43] Gallivan P., Hong Q., Jordan L., Li E., Mathew G., Mulyani Y., Visokey P. and Tappert C., (2002), "VoiceXML Absentee System, *Proceedings of MASPLAS'02.* The Mid-Atlantic Student Workshop on Programming Languages and Systems Pace University. Retrieved online 10th January 2010 from http://csis.pace.edu/csis/masplas/p10.pdf

[44] Garcıa V. M. A., Ruiz M. P. P., and Perez J. R. P(2010), "Voice interactive classroom, a service-oriented software architecture for speech-enabled learning", Journal of Network and Computer Applications, 33 (2010) 603–610, Elsevier. pp. 603-610.

[45] Gerber, E. and Kirchner, C.(2001): Who's surfing? Internet access and computer use by visually impaired youths and adults. J. Vis. Impair. Blind. 95, 176–181 (2001)

[46] Ghaleb F. F. M, Daoud S. S., Hasna A. M., Jaam J. M. and El-Sofany H. F. (2006), "A Web-Based E-Learning System Using Semantic Web Framework", Journal of Computer Science 2 (8): 619-626, 2006. 2006 Science Publications. pp. 619-626.

[47] Gil A. and García-Penalvo F. J. (2008), "Learner Course Recommendation in e-Learning Based on Swarm Intelligence", Journal of Universal Computer Science, vol. 14, no. 16 (2008), pp. 2737-2755.

[48]   Hakulinen J. (2006), "Software Tutoring in Speech User Interfaces. Dissertations in Interactive Technology, Number 6, Tampere 2006.

[49]   Heiyanthuduwage S. R. and Karunaratne D. D. (2006) , "A Learner Oriented Ontology of Metadata to Improve Effectiveness of Learning Management Systems", Proceedings of the Third International Conference on eLearning for Knowledge-Based Society, August 3-4, 2006, Bangkok, Thailand. Pp 42.1 – 42.6

[50]   Hisayoshi I. and Toshio O. (1999), "Case Based Reasoning System for Information Technology Education", IEIC Technical Report (Institute of Electronics, Information and Communication Engineers) 1999.

[51]   HKU (n.d.), "Development of the HKU Campus Portal", University of Hong Kong. Available online at: http://www.hku.hk/cc_news/ccnews100/portal.htm, accessed, November, 2010.

[52]   Holohan E., Melia M., McMullen D., Pahl C.(2005), "Adaptive E-Learning Content Generation based on Semantic Web Technology", International Workshop on Applications of *S*emantic *W*eb technologies for *E-L*earning (SW-EL). July 18-22 2005. Amsterdam. The Netherlands. Retrieved 18th November 2010 from http://www.win.tue.nl/SW-EL/2005/swel05-aied05/proceedings/5-Holohan-final-full.pdf

[53]   Holzinger A., Smolle J. and Reibnegger G. (2006),"Learning Objects (LO): an Object-Oriented Approach to Manage e-Learning Content", In: Lazakidou, A. (Ed) Encyclopedia of Informations in Healthcare & Biomedicine. Hershey (PA), Idea Group Reference, 89-98IGI Book Chapter.

[54]   Howard S. (2006), "Interactive Training on Mobile Devices Training Using Text-to-speech and Speech Recognition, on Pocket PC PDA", MSc Multimedia Computing.

[55] Husni H., Jamaludin Z. (2009),'ASR technology for children with dyslexia: Enabling immediate intervention to support reading in Bahasa Melayu", US-China Education Review, ISSN 1548-6613, USA. Jun. 2009, Volume 6, No.6 (Serial No.55) pp 64-70.

[56] Ikhu-omoregbe N. A. (2010), "Development of A Formal Framework for Usable Operations Support In E-Health-Based Systems". A published doctoral dissertation, Covenant University. Published May 2010 by VDM Verlag.

[57] ISO 9241-11, (1998). Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability.

[58] Jankowska B. (2007), " Architectural Frameworks for Automated Content Adaptation to Mobile Devices Based on Open-Source Technologies". Thesis submitted to the Faculty of Economics of the European University Viadrina (Frankfurt/Oder) in fulfillment of the requirements for the degree of Dr. rer. pol. Retrieved 18 March 2010 from http://opus.kobv.de/euv/volltexte/2007/23/pdf/PHD_BJankowska.pdf

[59] Jianhua W., Long Z., Jun Z. and Xiping D. (2009),"A VoiceXML-Based Mobile Learning System and Its Caching Strategy", mLearn 2009, 8th World Conference on Mobile and Contextual Learning. October 26-30 2009, Orlando, Florida.

[60] Jain L. C. and Martin N. M. (1998), "Fussion of Neural Network, Fuzzy Systems and Genetic Algorithms, Industrial Applications". A textbook printed by The CRC Press International service on computational intelligence, p4.

[61] Jumar J, Ondas S., Cizmar A. and Rusko M. (2006), "Galaxy/VoiceXML Based Spoken Slovak Dialogue System to Access the Internet", INTERSPEECH 2006 – ICSLP, *Ninth International Conference on Spoken Language Processing, Pittsburgh, PA, USA*, September 17-21, '06. Available online at: http://www.ling.uni-potsdam.de/ws- ecai06/pdf/juhar.pdf, accessed July 2008.

[62]   Kalogirou S., Metaxiotis K. and Mellit A.(2009)," Artificial Intelligence Techniques for Modern Energy Applications", IGI Global Book Chapter. Pp 1-39. Retrieved 23th March 2011 from http://www.igi-global.com/downloads/excerpts/34624.pdf

[63]   Kastelic M. and Loncaric T. (2007), "A Model of Introducing E-learning System at Vocational College for Business Secretaries", Issues in Informing Science and Information Technology. Volume 4, 2007 pp. 175-187.

[64]   Kenny S., Curran K., Miller R. & Shevlin M. (2007) " A VoiceXML Enabled Locus of Control Scale System Designed to Engage School Pupils Lacking Commitment in the career development and decision making process", *Georgian Electronic Scientific Journal, Education Science and Psychology*, No. 1 (10).

[65]   Kim, S.H., Mims, C., & Holmes, K.P. (2006). An introduction to current trends and benefits of mobile wireless technology use in higher education. *AACE Journal, 14*(1), 77-100.

[66]   Kirriemuir J. (2003), "Speech Recognition Technologies". *JISC, March '03*. TWS 03-03, 13p. Technology watch Available online at: http://www.jisc.ac.uk/uploaded_documents/tsw_03-03.pdf Accessed April 2008.

[67]   Kirschning I. (2001) "Research and Development of Speech Technology & Applications for Mexican Spanish at the Tlatoa Group". *Proceedings of ACM CHI 2001 Conference on   Human Factors in Computing Systems 2001*, pg 49-50. Available online at: ict.udlap.mx/people/ingrid/ingrid/CHI_Tlatoa.pdf accessed July 2008.

[68]   Kolias C., Kolias V. and Anagnostopoulos L. (2008), "A pervasive Wiki application based on VoiceXML". Published in PETRA'08 Proceedings of the 1st international conference on Pervasive Technologies Related to Assistive Environments. ACM New York, NY, USA. ISBN: 978-1-60558-067-8

[69]   Kolodner J., Cox M. and Gonzalez-Calero P. (2005), Case-based

reasoning-inspired approaches to education, *The Knowledge Engineering Review*, Vol. 00:0, 1–4. © 2005, Cambridge University Press DOI: 10.1017/S000000000000000 Printed in the United Kingdom

[70]   Kondratova I. (2004), "Voice and Multimodal Technology for the Mobile Worker", national research Council Canada Institute for Information Technology, 46 Dineen Drive, Fredericton, NB, Canada. *ITcon Vol. 9 (2004)*. Retrieved 4th November 2010 from http://www.itcon.org/data/works/att/2004_24.content.06479.pdf

[71]   Kondratova I. (2009) "Multimodal Interaction for Mobile Learning", Lecture Notes in Computer Science. Proceedings of the 5th International Conference, UAHCI 2009, Held as Part of HCI International 2009, San Diego, CA, USA, July 19-24, 2009. Book Chapter on universal Access in Human-Computer Interaction. Intelligent and Ubiquitous Interaction Environments. Stephanidis (Ed.): Universal Access in HCI, Part II, HCII 2009, LNCS 5615, pp. 327–334, 2009.

[72]   Kouroupetroglou G. and Mitsopoulos E. (2000), Speech-enabled e-Commerce for Disabled and Elderly Persons", COST 219 Seminar "Speech and Hearing Technology", Nov. 22, 2000 Cott , Germany. PP 72 – 94. Retrieved 14th February 2010 from www.di.uoa.gr/speech/sppages/spppdf/cost219koupemitso.pdf

[73]   Krastev G. Smrikarova S. and Teodosieva M. (2006), "A Survey of Existing Platforms for Automation of Communication Between Phone Users", *3rd E-Learning Conference, Coimbra, Portugal*, 7-8 September 2006, pg 51-53. Available online at: http://elconf06.dei.uc.pt/pdfs/paper41.pdf accessed July 2008.

[74]   KSL (2005), "KSL Software and Network Services(2005),Retrived 29th Nov 2010 from http://www.ksl.stanford.edu/sns.shtml

[75]   Kumar C. A., Gupta A., Batool M. and Trehan S. (2006), " Latent Semantic Indexing-Based Intelligent Information Retrieval System for Digital Libraries", Journal of Computing and Information Technology - CIT 14, 2006, 3, pp 191–196

[76] Kurschl W., Mitsch S., Prokop R. and Scho¨nbo¨ck J. (2007), "Development Issues for Speech-Enabled Mobile Applications", Proceeding of Software Engineering 2007 Conference in Hamburg 2007. pp 157 – 168. Retrieved online 14th February 2010 from subs.emis.de/LNI/Proceedings/Proceedings105/gi-proc-105-018.pdf

[77] Kwang B. L. and Roger A. G. (2006), "The Design and Development of User Interfaces for Voice Application in Mobile Devices", 0-7803-9778-9/06/$20.00 ©2006 IEEE.

[78] Larson J. (2000), Introduction and Overview of W3C Speech Interface Framework, *W3C Working Draft*. Available online at: http://www.w3.org/TR/2000/WD-voice-intro-20001204/ accessed February, 2008.

[79] Leake D. B. (2000) "Case-Based Reasoning, Experiences, Lessons and Future Directions" A text book printed by AI Press/MIT Press

[80] Lerlerdthaiyanupap T. (2008), "Speech-based Dictionary Application", University of Tampere, Department of Computer Sciences, Computer Science, M.Sc. thesis, 54 pages, 11 index and Appendix pages, June 2008

[81] Liao C. and Li T. (2007), "Realising voice dialogue management in a collaborative virtual environment', *Int. J. Computer Applications in Technology*, Vol. 29, No. 1, 2007, pg 37-44. Available online at: http://www3.nccu.edu.tw/~li/Publication/pdf/ijcat2007.pdf accessed July 2008.

[82] Lin, W.J., Yueh, H.P., Liu, Y.L., Murakami, M., Kakusho K. and Minoh, M.(2006), Blog As A Tool To Develop E-Learning Experience In An International Distance Course. In: *Proceedings of the sixth International Conference on Advanced Learning Technologies(ICALT'06)*.

[83]   Liu L (2008), "Ontology by Tom Gruber to appear in the *Encyclopedia of Database Systems*, Ling Liu and M. Tamer Özsu (Eds.), Springer-Verlag, 2008. Retrieved 23th March 2011 from
http://tomgruber.org/writing/ontology-definition-2007.htm

[84]   Matzon K. (2005), "A Speech-Driven Automatic Receptionist Written in VoiceXML", *Master's thesis in Computational Linguistics.*

[85]   Markowitz J., 2005, "Securing Self-Service Telephone Applications", *a white paper on the speech processing and voice-biometrics industries*, consultant, 5801 N. Sheridan Road, Suite 19A Chicago, IL 60660 USA. Available online at: www.voicexml.org/insights/JMarkowitzConsulting/jmc_securing_apps_wp_2005.pdf , accessed April, 2008.

[86]   Modukuri R. and Morris R. J. (2004), Voice based web services - An assistive technology for visually impaired persons", Journal of Technology and Disability", published by IOS press. pp 195-200

[87]   Moise G. and Netedu L.(2009),"Ontologies for Interoperability in the eLearning Systems", BULETINUL Universității Petrol – Gaze din Ploieşti, Vol. LXI No. 2/2009. Pp 75-88.

[88]   Motiwalla L. F. (2007), "Mobile learning: A framework and evaluation", Journal of Computers & Education 49. Elsevier. Pp. 581–596.

[89]   Motiwalla, L.F. and Qin J. (2007)," Enhancing Mobile Learning Using Speech Recognition Technologies: A Case Study". Eighth World Congress on the Management of eBusiness, 2007. (WCMeB 2007). Volume , Issue , 11-13 July 2007 Page(s):18 – 18

[90]   Motiwalla L. F. (2009), "A Voice-enabled Interactive Services (VòIS) Architecture for e-Learning", International Journal on Advances in Life Sciences, vol 1 no 4, year

2009 Page 122 – 133. Retreived 30th October 2010 from
http://www.iariajournals.org/life_sciences/lifsci_v1_n4_2009_paged.pdf

[91]   Mult H. C. ., Reusch J. A. (2004), "VoiceXML-Applications for E-Commerce and E-Learning", Fachbereich Wirtschaft, pp 230-233.

[92]   Naik U. and Shivalingaiah D.(2008), "AbstractComparative Study of Web 1.0, Web 2.0 and Web 3.0", International CALIBER-2008. University of Allahabad, Allahabad, February 28-29 & March 1, 2008. pp 499-507.

[93]   Narada W. (1996), "Definition of Hidden Markov Model". Available online at http://jedlik.phy.bme.hu/~gerjanos/HMM accessed 30th June 2008.

[94]   Niels O. B. (2000). Draft position paper for discussion at the ELSNET Brainstorming Workshop 2000-2010 in Katwijk aan Zee, The Netherlands, on 23-24 November, 2000. NISLab, Denmark (editor). Retrieved 21 September 2009 from http://www.elsnet.org/dox/rm-bernsen-v2.pdf

[95]   Nielsen J., Usability 101 (2003): Introduction to Usability, Alertbox, August, 2003 http://www.useit.com/alertbox/20030825.html)

[96]   Nisbet, P.D. and Wilson, A. (2002) Introducing Speech Recognition in Schools: using Dragon Natural Speaking. *Pub. CALL Centre*, University of Edinburgh. ISBN 1 898042 22 X.

[97]   Nokia, 2009,   Nokia phone showing typical screen resolution. Retrieved 14th February 2009 from http://www.nokia.co.uk/nokia/0,,76036,00.html

[98]   NUC(2004). The state of Nigerian universities" Nigerian university system. Newsletter, (2), (1), 2004.

[99]   Odejobi, O.A. & Ogwu, F.J. (2005). Text-to-Speech Synthesis Access to Wed-based Distance Education Resources: Design and Implementation. In: Hamza, M.H. (ed): Artificial Intelligence and Applications -AIA 2005. ACTA Press.

[100]   Oboko R., Wagacha P. W. and Omwenga E. (2008), "Adaptive Delivery of an Object Oriented Course in a Web-Based Learning Environment. 3rd *International Conference on ICT for Development, Education and Training,* eLearning Africa 2008 Accra International Conference Centre (AICC), Accra, Ghana, May 28th - 30th, 2008. pp. 274-277

[101]   Ondas I. S. (2006), "VoiceXML-based Spoken Language Interactive System". In: Proc. 6th PhD Student Conference and Scientific and Technical Competition of Students of Faculty of Electrical Engineering and Informatics Technical University of Košice, Košice, 17.5.2006, s. 97-98. ISBN 80-8086-035-1. Retrieved 5th November 2010 from http://irkr.tuke.sk/publikacie/_vti_cnf/Prispevok_eng.pdf http://www.sciencedirect.com.

[102]   Pakucs, B.(2003), "SesaME: A Framework for Personalised and Adaptive Speech Interfaces", In: Proceedings of EACL-03 Workshop on Dialogue Systems: Interaction, Adaptation and Styles of Management, Budapest, Hungary (2003) 95–102.

[103   Palowireless (2009), "VoiceXML Development Tools", Published By: Global Industry Analysts Date Published: Jan 2009. Retrieved 5th November 2010 from http://www.palowireless.com/voicexml/devtools.asp

[104]   Paul D. (2003), "Speech recognition for students with disabilities" Nisbet, CALL Centre, University   of Edinburgh pg 1 - 6. Available online at: http://callcentre.education.ed.ac.uk/downloads/speech_recognition/hitmesse2003.pdf accessed January 2008.

[105]   Peter Y., Vantroys T. and Lepretre E. (2008), "Enabling Mobile Collaborative Learning through Multichannel Interactions". *An article published in the 4th*

International Conference on Interactive Mobile and Computer Aided Learning, IMCL2008, pg 1-4.Available online at :

http://telearn.noe-kaleidoscope.org/warehouse/Peter-Yvan-2008_(001784v1).pdf accessed 5th May 2008.

[106] PorterStemming, (2007), "PHP Class: Porter Stemming Algorithm (Source)", Retrieved 10th April 2010) from http://www.chuggnutt.com/stemmer-source.php

[107] Protégé(2010), *Protégé - Ontology Editor and Knowledge Acquisition System*, retrieved November 11, 2010 from http://protege.stanford.edu/.

[108] Qureshi R. (2007), "Building a Telephony Web Application". Available online at: http://www.hunzatechnologies.com/telephony-web-tutorial/ accessed January 2008.

[109 Radev D. (2008), "Lecture 13 - Eigenvectors, Eigenvalues, Stochastic Matrices", COMS 6998-006 Network Theory, Columbia University, Spring 2008. Retrieved 19th December 2011 from http://www1.cs.columbia.edu/~coms6998/Notes/lecture13.pdf

[110] Raghuraman M. B. (2004),"Design and Implementation of V-HELP system – a voice-enabled web application for the visually impaired. A Master's project presented to the faculty of the graduate college in the University of Nebraska in partial fulfillment of requirements for the degree of Master of Science. Major: Computer Science. August 2004. Retrieved 21 March 2010 from http://hcc.unl.edu/netgroup/alumnidocs/mukesh.MS.pdf

[111] Rajpathak D., Motta E. and Roy R.(2001), "A Generic Task Ontology for Scheduling Applications". In Proceedings of the International Conference on Artificial Intelligence (IC-AI'2001), Nevada, Las Vegas, USA, 2001., Retrieved 16th November 2010 fromhttp://projects.kmi.open.ac.uk/akt/publication-pdf/generictaskontology.pdf accessed on 2nd December, 2009

[112] Rouillard J. and Truillet P. (2005)," Enhanced VoiceXML", Published in LIFL. LIFL is a Research Laboratory in the Computer Science field of the University of Sciences

and Technologies of Lille Retrived 20 March 2010 from http://www2.lifl.fr/~rouillar/publi/2005_Rouillard_Truillet_HCII.pdf

[113] Rouillard J. (2007), 'Web services and speech-based applications around VoiceXML" *Journal of Networks*, vol. 2, no. 1.

[114] Roux J., Scholtz P., Klop D., Povlsen C., Jongejan B. and Magnusdottir A. (2010), "Incorporating Speech Synthesis in the Development of a Mobile Platform for e-learning", The seventh international conference on Language Resources and Evaluation (LREC) 17-23 May 2010 Malta.

[115] Russell S. and Norvig P. (2003), A textbox on Artificial Intelligence: A Modern Approach, Prentice Hall, (second edition).

[116] SALTForum(2005): Speech Application Language Tags (SALT), 1.0 Specification, 2002. ttp://www.saltforum.org/saltforum/downloads/SALT1.0.pdf. Download: 2005-01-03.

[117] Salton G., Wong, and Yang C. S.(1975). A Vector Space Model for Automatic Indexing. Information Retrieval and Language Processing, pages 613-620.

[118] Sauro J. and Kindlund E.(2005), "A Method to Standardize Usability Metrics into a Single Score", *ACM, CHI,* April 2-7, Portland, Oregon, USA, 2005.

[119] Schindler L. (2005), "Web-Based Education: A Speech Recognition And Synthesis Tool" Department of Mathematics and Computer Science, in the College of Arts and Science at Stetson University, DeLand, Florida. Available online at: http://www.stetson.edu/~helaarag/Laura-proposal.pdf accessed May, 2008.

[120] Selouani S., Le T., Moghrabi C., Lanteigne B., and Roy J. (2006), 'Online Collaborative Learning System Using Speech Technology', *Proceedings of World Academy of Science, Engineering and Technology* Volume 15 October 2006 ISSN

1307-6884. Available online at: http://www.waset.org/pwaset/v15/v15-11.pdf accessed July 2008.

[121] Shin J. and Hong K. (2006), "Simple and Powerful Interactive E-Learning System Using VXML: Design and Implementation of Web and PSTN Linked Efficient Learning System", M. Gavrilova et al., (Eds.): ICCSA 2006, LNCS 3980, pp. 354 – 363, 2006. © Springer-Verlag Berlin Heidelberg 2006

[122] Siemens G. (2003) "Open source content in education: Part 2 - Developing, sharing, expanding resources". Available online at:
http://www.elearnspace.org/Articles/open_source_part_2.htm
accessed April, 2008.

[123] Singh B. S. (2003), "Search Algorithms", DRTC Workshop On Digital Libraries: Theory and Practice, March 2003, DRTC, Bangalore. pp. 1-9. Retrieved May 2009 From https://drtc.isibang.ac.in/bitstream/handle/1849/
32/E_Searchalgo_brijesh.pdf?sequence=2

[124] Smashing magazine (2009) "12 Useful Techniques For Good User Interface Design", (2009), Retrieved 28 May   2010   from
http://www.smashingmagazine.com/2009/01/19/12-useful-techniques-for-good-user-interface-design-in-web-applications/

[125] Spasic I., Ananiadou S. and Tsujii J. (2005), "MaSTerClass: a case-based reasoning system for the classification of biomedical terms". Available online at:
http://bioinformatics.oxfordjournals.org/cgi/content/full/21/11/2748, accessed 5th July 2005.

[126] Srimathi H. and Srivasta S. K. (2008), "identification of ontology based learning object using instructional design", Ubiquitous Computing and Communications Journal, Volume 3 Number 3, published 2008. Retrieved February 2010 from
http://www.ubicc.org/files/pdf/lousingid_harisrimathi_ubicc_final_id207_207.pdf

[127] Stanford (2008), Stanford Center for Biomedical Informatics research, formerly Stanford medical informatics. Retrieved 29th November 2010 from http://www.smi.stanford.edu/projects/protege/

[128] Term (2008),The term vocabulary and postings lists May 27, 2008. pp 18-44

[129] Tsinakos A. A. (2003), "Asynchronous distance education: Teaching using Case Based Reasoning Turkish Online Journal of Distance Education-TOJDE July 2003 ISSN 1302-6488 Volume:4 Number:3

[130] Tucker D(2003)."Voice User Interface Design - Purpose and Process", Intervoice, Inc. Microsoft.NET Speech Technologies Technical Articles. (2003)

[131] VoiceXML (2007), *VoiceXML application development life cycle*. Palo Alto, CA: Hewlett Packard Development Company. Building a usable voice user interface.

[132] Voxeoprophecy(2003), "The Voxeo Prophecy Platform - Free Download", Available online at: www.voxeo.com/prophecy

[133] Voiceportalwhite paper(2001), available online at : http://www.medialab.sonera.fi/workspace/VoicePortals.pdf, accessed 21st March 2008-03-21

[134] VXML, 2004, Voice Extensible Markup Language, Version 2.0, W3C Recommendation, http://www.w3.org/TR/2004/REC-voicexml20-20040316/

[135] Walsh P. and Meade J. (2003), "Speech Enabled E-Learning for Adult Literacy Tutoring", Proceedings of the The 3rd IEEE International Conference on Advanced Learning Technologies (ICALT'03). Retrieved 14th February 2010 from http://i-learn.uitm.edu.my/resources/journal/j1.pdf

[136] Walker M., Litman D., Kamm C. and Abella A.(1997),"PARADISE: A framework for evaluating spoken dialogue agents". In Proceedings of the 35th Annual Meetingg of the Association of Computational Linguistics. ACL 97, 1997.

[137] WEB2, (2010), "WEB 2.0 TO WEB 3.0: IS YOUR MIND, AND YOUR COMPANY, READY?", Women's Forum for the Economy and Society, The Next Generation Web: Search Context for Online Information July 22, 2008 Allan Cho on Suite 101. Retrieved 18th May 2011 from http://internet.suite101.com/article.cfm/what_is_web_30

[138] Weber R., Ashley K. D. and Brüninghaus S. B. (2005): Textual case-based reasoning, The Knowledge Engineer-ing Review, Vol. 00:0, 1–00., Cambridge University Press, Printed in UK (2005).

[139] WhatisStemming, n.d, What is Stemming, Retrieved 4th November 2010 from http://www.comp.lancs.ac.uk/computing/research/stemming/general/

[140] Williamson, K., Wright, S., Schauder, D., Bow, A.(2001): The Internet for the blind and visually impaired. J. Comput. Mediat. Commun. [Online]. Available: http://jcmc.indiana.edu/vol7/issue1/williamson. html (2001)

[141] W3C(2005a): XHTML+Voice Profile 1.0, 2001. http://www.w3.org/TR/xhtml+voice. Download: 2005-9-01.

[142] W3C(2005b): XML Path Language (XPath) 2.0, 2004. http://www.w3.org/TR/xpath20/.Download: 2005-10-20.

[143] Yankelovich, N. (2000). "Designing Effective Speech Interfaces", John Wiley & Sons, Inc. An Interview edited by Susan Weinschenk and Dean T. Barker

[144] Zaiane O. R.(2005),"Recommender systems for e-learning: towards non-intrusive web mining", A chapter in a Book titled: Data Mining in E-Learning. State-of-the-art in Science and Engineering volume 2.

[145] Zambrano E. C. (2004), "Framework For Ubiquitous and Voice Enabled Web Applications Development: A Transport Information Systems Test – Bed, Doctoral Thesis Dissertation, Departamento de Informática, Universidad de Valencia, Spain, 2004.

[146] Zaykovskiy D. (2006), " Survey of the Speech Recognition Techniques for Mobile Devices", Speech and Computer, SPECOM'2006, St. Petersburg, Russia, 25-29 June 2006. PP 88-93.

[147] ZDNet, (2010), "ZDNet Definition for VoiceXML. Retrieved 12th March 2010 from http://dictionary.zdnet.com/definition/VoiceXML.html?tag=content;col1