

Efficient Voice Integrated Browser Using Naïve Approach

Karunakar Batta

Sr. Specialist, Cegedim Software India Pvt Ltd

Karunakar.batta@gmail.com

Vinay Krishna

Software Development Manager, Cegedim Software India Pvt Ltd

vinaykris@gmail.com

Abstract— The World Wide Web contains huge information from all over world. It has information about simple things as well as very complex things. This makes it very useful for everyone. Nevertheless this enormous useful information is always accessible through browser which requires use of mouse and keyboard. Hence, it does always remain unreachable to specific community of people, especially to the people with visual impairment. The concept of voice browser is not new, however in this paper we have used naïve approach to make this efficient Voice Integrated Browser (VIB). All existing voice browsers need VoiceXML page. The conversion of existing web pages to VoiceXML pages is very tedious task and it requires extra effort. In this paper we have explained that how VIB doesn't need VoiceXML pages. We elucidate this naïve approach and challenges faced while evolving it and eventually discussed about the future work.

Index Terms— Voice Browser, VoiceEnable, VIB

I. INTRODUCTION

Estimates suggest that there are between 180 and 220 million youth with disabilities worldwide, and nearly 80 percent of them live in developing countries [1]. Blind people use helpful devices such as speech synthesizers that can speak the letters, words and sentences shown in the screen. This is particularly helpful with applications such as word processors, email, and the Internet. Lately, however, there have been a few problems, specifically with the Internet. Many blind people today use text-to-speech "screen readers" that work only with the character-based DOS operating system, not with graphical user interfaces such as Web browsers [2].

It is very important and helpful to have some application which reduces the gap between physically challenged people and World Wide Web. This will be helpful for those with disabilities such as sensory or mobility problems, the internet can be a mixed blessing. Home shopping, for example, can be invaluable to those for whom busy streets present a difficult challenge. However, in much the same way as a building may be problematic to physically disabled people; a website may

also present barriers to access. For example, a visually-impaired internet user can use a screen reader to translate the contents of web pages using speech synthesizers or Braille displays. The user will struggle to understand web pages if, for example, images are displayed on the page without a text alternative [3].

The impacts of voice enabled systems have yet to be revealed but speech-enabled mobile phones, navigational systems, and automobiles might make a distinct contribution towards blind people's life. Although future would certainly bring a lot of technological advancements but the pace of development, in this particular area (voice enabled systems), remains to be slow [4].

II. PREVIOUS/OTHERS WORK

IBM World Wide Web Consortium having a group called The Voice browser Working Group whose mission is to support browsing the web by voice. Voice browser working group interact with Voice web through VoiceXML(VXML) and Speech Grammar Recognition Specification (SRGS) [5]. Figure 1 describes the architecture model of this approach [5].

A. VoiceXML

VoiceXML is a language for creating audio dialogs that feature synthesized speech, recording of spoken input etc. The goal of VoiceXML is to bring the advantages of web-based development and content delivery to interactive voice response application.

B. SRGS

A document language that can be used by developers to specify the words and patterns of words to be listened for by a speech recognizer or other grammar processor

According to Voice Browser Working group approach any developer has to follows SRGS document strictly to specify the words and patterns of words to be listened. It is almost impossible to cover all the words and patterns of words in a single document. As the technologies keeps on changing it will be difficult for developers to change the web as per the change in technology.

III. OUR APPROACH

To reduce the gap between handicap people and technology we came up with a concept called Voice Integrated Browser (VIB). A browser is a software application used to enable computers users to locate and access web pages also used to bring the context of (an HTML) webpage into display, which allows quick access to links, buttons etc [6].

Voice Integration Browser is a windows application that allows users to access Web via verbal commands. This application will be especially advantageous for the physically challenged people, particularly those with visual impairment.

The Basic idea behind this application is to allow the end user to use this application using his own voice.

Voice Integration Browser will help people to access web while keeping hands & Eyes free of the computer. This can allow the users to multi-task that is to access the Web when carrying out their daily routine.

Voice Integration Browser accepts spoken words as input and acts as per the command, either by response to the command or by producing synthesized speech. The input voice passes through the voice recognition process and would invoke an action in the browser only if it matches with one of the voice commands/vocabulary available in the voice recognition engine.

A few commands have been defined in VIB, which are recognized when the user speaks it aloud.

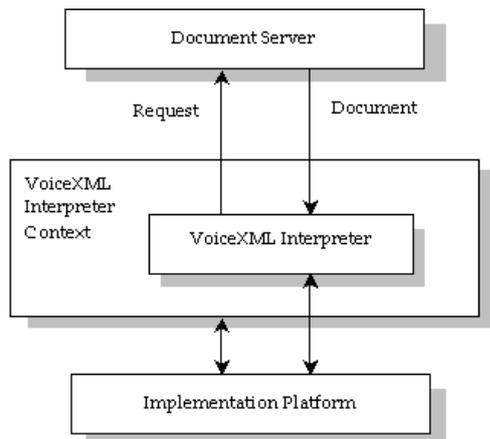


Figure 1. Architectural Model.

A. Features

Below are features of VIB:

B. Enable Voice-on

This is used to make VIB ready to recognize voice and perform the required operations on the basis of recognized voice.

C. Enable Voice-off

VIB will work as normal web browser to access the web using the standard computer input interfaces, like keyboard, mouse etc.

D. Voice Basic Commands

Voice is one of the most natural ways to interact. When it comes to computers it is no difference. If an

application can be controlled solely by way of voice commands then the opportunity that lies is unlimited. We defined few commands in VIB which interacts with the application and execute the operation. Table I mentioned below contains few commands:

TABLE I VOICE BASIC COMMANDS

Command Name	Description
URL	It is a voice command using which VIB takes the user to URL location. In order to keep the use of keyboard to a minimum, a number of website addresses have been predefined in URL. This would allow the user to access these predefined websites without a single key-press.
Move Down	This command is used to navigate through the various options in the predefined addresses in the URL command line in order to make a selection. The Down' command (Move Down) which allows the user to move down to the next address in the URL field.
Climb	This command takes the control from down to UP in the URL fields which allows the user to navigate up through the available options.
Add Tab	To open a new tab in VIB, we need to use Add Tab command. This command adds a new tab to the existing application and user is able to use new tab.
Remove Tab	To remove the current working tab from VIB, we need to use Remove Tab command. This command removes the current working tab from the existing application.
Next Tab	It is used to navigate front tabs, it takes the control to next tab.
Previous Tab	This command is used to navigate previous tabs, it takes the control to previous tab.

TABLE II VOICE BASIC COMMANDS

Command Name	Description
View History	VIB also maintains the history which can be viewable using View History command. This command shows the links of website user navigated recently.
View Source Code	View source code command allows the user to see the source code user for the current page.
Close Application	Close Application closes the application.

E. Advanced Voice Commands

These are similar to the basic commands but the functionality is advance compare to basic commands. Table II mentioned below contains advanced voice commands along with brief description.

F. Technologies used

VIB is an application which requires the functionality to recognize user's voice from client side. In .Net Framework it is possible to record voice and compare it with a predefined text using libraries like system. speech. synthesis and system. speech. recongnization. List technologies used in VIB is mentioned in below Table III.

G. Voice as Output

VIB is a two way communication application i.e. accepts Voice as Input and reads Output text. Information

available in the web pages are Read out as Output by VIB using library called System.Speech.Synthesis.

The Library System.Speech.Synthesis contains classes for initializing and configuring a speech synthesis engine, for creating prompts, for generating speech, for responding to events, and for modifying voice characteristics.

```
SpeechSynthesizer speech = new SpeechSynthesizer ();
speech.SetOutputToDefaultAudioDevice ();
speech.Speak ("Text");
```

H. <VoiceEnable> Tags

The source of a web page contains HTML and other extensible tags. It is not feasible to convert the entire html page to text. As there are no standards set to differentiate the html tags and actual text in the page which needs to be converted to voice. Hence, we propose a new tag, <VoiceEnable> and </VoiceEnable>, which help to identify text to be converted to voice.

TABLE III LIST OF TECHNOLOGIES USED

Technology	Description
Microsoft Visual Studio	Microsoft Visual Studio is an integrated development environment from Microsoft. It is used to develop console, windows and web applications. Visual Studio supports different programming languages by means of language services, which allows the code editor and debugger to support nearly any programming language [7].
C#	C# is a multi-paradigm programming language encompassing strong typing, imperative, declarative, functional, generic, object-oriented(class-based), and component-oriented programming disciplines. It was developed by Microsoft within its .NET initiative and later approved as a standard by Ecma (ECMA-334) and ISO (ISO/IEC 23270:2006). C# is one of the programming languages designed for the Common Language Infrastructure [8].
System.Speech	Using the System.Speech namespaces, user can give Windows applications the ability to respond to voice commands, accept dictation, generate synthesized speech (text-to-speech or TTS), or play prerecorded audio files. System.Speech applications can also use speech documents that were authored with industry-standard markup languages, such as grammars, TTS prompts, and lexicons. In VIB we used System.Speech Assembly from Microsoft which allows the application to use Synthesis and Recognition feature [9].

IV. ARCHITECTURE & DESIGN

The Architecture for VIB is Portable and Scalable i.e. able to support any speech library. Best practices and appropriate patterns have been used to make VIB maintainable and reusable. VIB maintains four states to manage voice commands and inputs. Fig. 2 mentioned below depicts the workflow of Voice to Text and Text to Voice used in VIB.

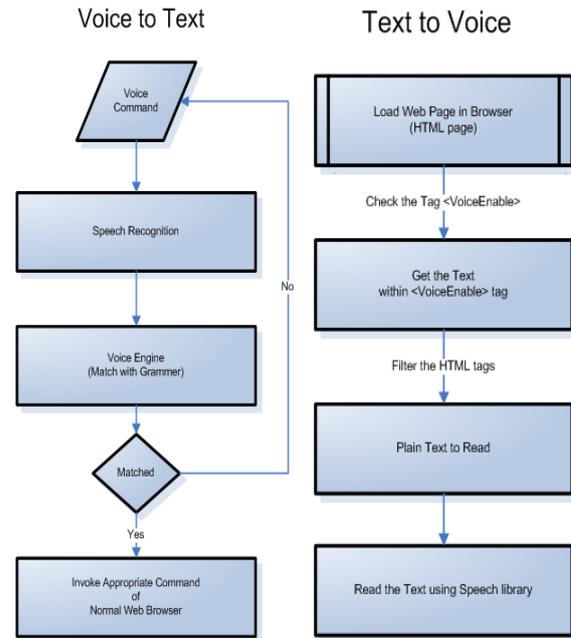


Figure 2. VIB diagram

V. CHALLENGES

Challenges are the most important part of a project. The project will be always in good shape once we overcome with the challenges. During the development phase of the project (VIB) we faced few Challenges as mentioned below with the description (way we overcome with the challenge):

A. Conversion of Text to Voice

To make the application two-way communications we need to convert the text written in the web pages to voice and the biggest challenge was to select the readable text from the HTML page. To overcome this solution we came up with a concept <VoiceEnable> tag, which is described in the above section Voice as Output.

B. One Word in One Time

VIB was able to accept only one word as input which was another major challenge. To solve this we came up with a solution where we used culture info as "en" and looping phrase in place of single word.

C. Handling Web Controls

Accessing the web controls in VIB using voice was a big challenge. It is very important to have control on web page controls to make complete use of VIB. For example, Google search textbox control allows the end user to search in Google online using Voice.

D. Providing Support to Multiple Speech Library

Design of VIB can support multiple speech libraries. To implement this functionality we created a class using the factory method pattern. It is an object-oriented creational design pattern to implement the concept of factories and deals with the problem of creating objects (products) without specifying the exact class of object that will be created. The essence of this pattern is to "Define an

interface for creating an object”, but let the classes that implement the interface decide which class to instantiate.

VI. CONSTRAINT

In a project there will be different kinds of constraints like technological, resource and physical. The primary constraint of project VIB is Voice Accent (technological). As the library we are using in VIB is System. Speech from Microsoft as described above, it will be difficult to train VIB to understand local accent.

VII. FUTURE WORK

Over all we focused on the basic features of VIB. We provide the solution for some challenges faced during the development of VIB. To identify all other possible interactions between user and VIB and to identify the most suitable way, we have taken into account few future works as mentioned below in detail.

A. Complete Control on Web Pages

To make web application also run using voice command we need to handle the web controls completely. In a single web page there are so many server controls which have to be handled completely like filling textbox, submit button etc. This will allow web page to run using voice command.

B. Evaluate other Speech Recognize Library in Order to Handle Indian Accent

VIB basically understands US English as language. It should be able to understand India accent which will make VIB more useful to the Indian people.

VIII. CONCLUSION

The Internet is used in every corner of the World and lots of people have benefitted with the vast amount of information available on the internet. However, the physically challenged people have not been able to avail the benefits of the internet due to their inability to use the current technology.

The Voice Integrated Browser is an application which allows the people to use their voice to access Internet. This application helps especially physically challenged people who find it difficult to access internet. This application would be especially beneficial for blind

people as they can hear the text written in the pages as Voice.

REFERENCES

- [1] United Nations - Youth Social Policy and Development Division, Youth With Disabilities. [Online]. Available: <http://social.un.org/index/Youth/ResourcesandPublications/YouthWithDisabilities.aspx>
- [2] C. R. Murphy. Computers Assisting The Handicapped. *Computer Program*. [Online]. Available: <http://courses.cs.vt.edu/cs3604/lib/Disabilities/murphy.AT.html>
- [3] UK law: Disabled Access to Website under UK Law. *Survey*. [Online]. Available: <http://www.out-law.com/page-330>
- [4] M. N. Talib, S. Q. Cai, and M. H. Hu, “Voice browsing approach to e-business access: A blind perspective,” presented at the Computer and Information Science, School of Management, Huazhong University of science and technology, China , 2010.
- [5] W3C: The Voice Browser Working Group. *Software*. [Online]. Available: <http://www.w3.org/Voice/>
- [6] Wikipedia: Web Browser [Online]. Available: http://en.wikipedia.org/wiki/Web_browser
- [7] Wikipedia: Microsoft Visual studio [Online]. Available: http://en.wikipedia.org/wiki/Microsoft_Visual_Studio
- [8] Wikipedia: C Sharp (Programming language) [Online]. Available: [http://en.wikipedia.org/wiki/C_Sharp_\(programming_language\)](http://en.wikipedia.org/wiki/C_Sharp_(programming_language))
- [9] MSDN: System. Speech Namespaces [Online]. Available: [http://msdn.microsoft.com/en-us/library/hh413620\(v=office.14\).aspx](http://msdn.microsoft.com/en-us/library/hh413620(v=office.14).aspx)



Karunakar Batta, is working as a Senior Specialist at Cegedim Software India Pvt Ltd in Bangalore, India. He has done MS in Software Systems from BITs Pilani. He is having 5 years of experience in field of software development with Cegedim.



Vinay Krishna is working as a Software Development Manager at Cegedim Software India Pvt Ltd in Bangalore. He has done MS in Computer Science from Manipal University and currently pursuing PhD from Bharathiar University. He is having 12 years of experience in software development.

Mr. Krishna is life time member of Computer Society of India (CSI) and management committee member of CSI Bangalore Chapter. He is a member of Scrum Alliance.