

VOSS -A Voice Operated Suite for the Barbadian Vernacular

David Byer and Colin Depradine

University of the West Indies,
Cave Hill, St Michael, Barbados
david.byer@gmail.com,
colin.depradine@cavehill.uwi.edu

Abstract. Mobile devices are rapidly becoming the default communication device of choice. The rapid advances being experienced in this area has resulted in mobile devices undertaking many of the tasks once restricted to desktop computers. One key area is that of voice recognition and synthesis. Advances in this area have produced new voice-based applications such as visual voice mail and voice activated search. The rise in popularity of these types of applications has resulted in the incorporation of a variety of major languages, ensuring a more global use of the technology.

Keywords: Interfaces, mobile, Java, phone, Android, voice, speech, Windows Phone 7.

1 Introduction

This paper presents VOSS a voice operated software suite for the Barbadian vernacular. Its primary function is to provide tools for translating the local indigenous grammar to Standard English. Barbados is a small Caribbean island, located to the east of the chain of islands. It is a former British colony and so English is the standard language taught in schools. However, as with most Caribbean islands, there is an internal dialect which is spoken by the local population. In fact, it is not uncommon to encounter instances where Standard English and the Barbadian dialect are being used interchangeably. This can be problematic since visitors to the island may find it difficult to understand someone using the dialect. Since tourism is one of Barbados's primary commercial sectors, the use of technology can provide a cost effective mechanism to help alleviate this problem.

2 The VOSS System

The VOSS system is a collection of integrated software and hardware components which allows both external and administrative users to access and to add or update data that is generic to the Barbadian culture and vernacular. It is a combination of web servers, databases and web based or windows forms application interfaces.

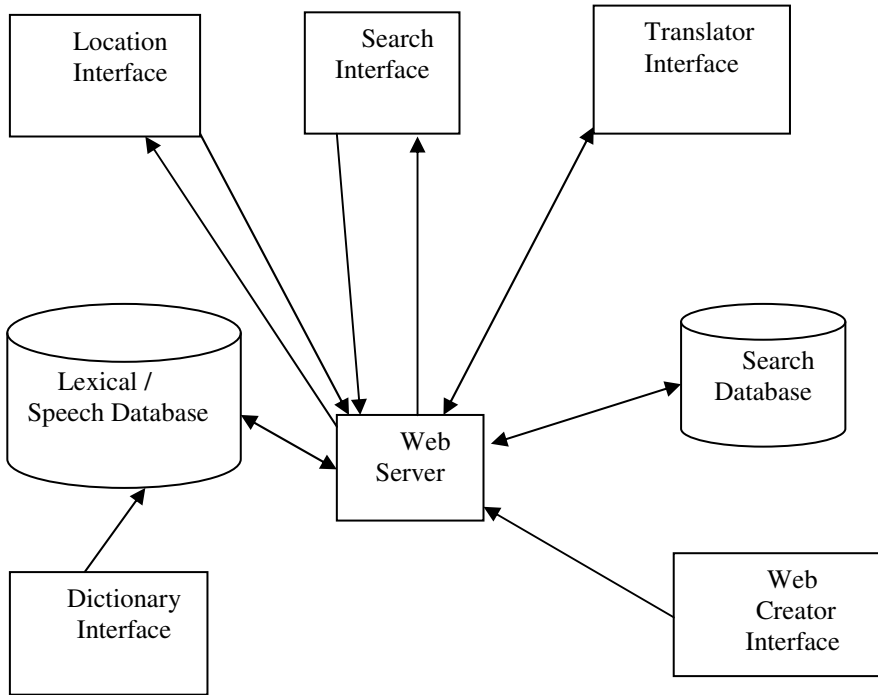


Fig. 1. Overview of the VOSS System

2.1 Dictionary Interface

This is a windows form application interface to be used by the administrators of the system for the purpose of adding words to the lexical database, which is used by the translator interface, and to the voice recognition dictionary.

This interface is password protected so that no unauthorized user can make changes to the database. The user keys the correct spelling of the word they wish to add. Then using the basic rules of syntax the various parts-of speech categories, such as noun or verb, that the word falls into are added to its properties. When that process is completed the user will pronounce the word as it should be recognized by the speech recognition engine. The system will ask the user to repeat the word three times then it will replay the word. If the user is satisfied with the pronunciation he or she can apply the word and it will be added to the dictionary. If the user is not satisfied, the process can be repeated. The interface is also used to edit or remove existing words from the database.

2.2 Web Page Creator

This interface is used to create the content on the web server. This interface does not require the user to have any knowledge of web design, instead, the content is generated using Microsoft Word.

The system uses two types of web pages. One type is used by the search interface and consists of one page documents with embedded graphics and required search keywords. The other type is used by the locator interface and is tied to the GPS and is coordinate specific.

During the creation of a general web page, the user must first create the document in Microsoft Word. The user will then select the web page creation option within the VOSS system where they will be prompted to enter a title for the page and then to browse for the appropriate Microsoft Word document. When the document is selected, the interface reads its content and creates 15 search keywords, each with a weight of importance. An HTML page is generated and uploaded to the web server along with the relevant search criteria.

However, when it comes to the data that is tied to the GPS it creates a collection of pages and groups them together with the link being the coordinate. These too are created in Microsoft Word but the difference is that they are separate pages and the user has the option to create one page for a site and upload it, stop and go to another site and then come back to it and create another page. This option allows for pages for any historical site to be changed and updated while maintaining the same file name on the server.

2.3 Location Interface

This interface is for mobile devices and is tied into the GPS system. The user uses voice commands to activate a map which shows the user's current location.

The user employs a variety of commands to change the map view to roads terrain, satellite or hybrid map topologies. Voice commands can also be used to pan left and right and to zoom in and out. This interface runs in a browser container which allows for map documents to be manipulated by the interface software.

The user can also vocally request any historical and geographical data about the location or pictures of buildings and/or landmarks in the area. This interface runs on the Microsoft Windows Phone 7 operating system, any phone using the Android operating system, Windows mobile devices and laptops.

These devices must all be GPS enabled as the software uses the GPS coordinates as a key to submit to the web server and to retrieve the pages of historical or cultural data about the particular site. The devices must also have Internet connectivity.

2.4 The Search Engine Interface

This is a standard search engine interface which accepts voice input and interprets the query string. Since many Barbadians speak both Standard English and dialect it is easy to receive sentences which are a combination of the two.

As a result it will be necessary to translate the spoken dialect query into Standard English. This is performed by an algorithm which uses the lexical database and grammar rules to translate the statement to Standard English.

Once that is completed, the search algorithm will interpret the query statement to ascertain the most accurate request to submit to the database and what keywords should be used in the search.

The engine will return various types of information about things Barbadian. The user has the option to read the data for him or herself or to instruct the interface to read it aloud for them.

2.5 The Translator Interface

This interface comes in two flavors, one for mobile devices and one for the desktop or laptop computers (PC). This is a dictation tool which is used to translate the Barbadian dialect to Standard English.

Firstly, the system takes a body of submitted text and breaks it down into sentences. Each word in the sentence is then checked against the translation tables in the lexical database. When a dialect word is found it will be replaced with a standard word or phrase.

This continues sentence by sentence until the whole document is processed. The system will then return to the beginning and use the rules of syntax and semantics to adjust the sentences to make them grammatically correct.

This interface is mainly used on a PC for the purpose of creating text documents that are in Standard English due to the fact that it is not uncommon for Barbadians to unconsciously mix dialect and English when creating documents. The user can then print or email the finished document.

The mobile side is mainly for visitors and those who do not understand Barbadian dialect. The mobile device will first record the conversation with permission of the speaker. The recording is saved as a .wav file and is transmitted to the server along with an email address. The web server passes the file to the translator where it is converted to text. At this point, the translation process carried out by the PC version is performed. When the process is completed, the translated text is sent by email to the user.

3 Platforms and Programming Languages

The system has been developed using the Microsoft C# (c-sharp) and Java programming languages. The C# language provides application programming interfaces (API) for the Microsoft speech library as well as the Microsoft .NET framework for mobile devices, specifically the Microsoft Windows Phone 7 system. Java is used for the backend processing because it has several prebuilt classes for parsing and communication over the Internet and can also be used with the Google Android mobile operating system.

3.1 Windows Phone 7

Windows Phone 7 is a mobile operating system which was developed by the Microsoft Corporation. This operating system uses a design language called Metro. Metro is used to integrate the operating system with other services provided by Microsoft and third party providers. This operating system supports two popular programming platforms which are Microsoft Silverlight and XMA [10].

Silverlight is a development platform which is used by software developers for creating interactive multimedia interfaces which can be either online or offline. Silverlight can be used for Web, desktop, and mobile applications. [8]

XNA is a runtime environment developed by Microsoft for developers who create games for its gaming platforms. It supports 3D graphics and is mostly used for developing games for the Xbox but it can be used for the PC and for the Microsoft audio player, the Zune HD [8].

3.2 Android

Android is a mobile operating system that was developed by a company called Android Inc. This company was acquired by Google in 2005. Android is open source and is based upon a modified version of the Linux kernel. [1].

The main language used for development on Android is the Java language. The Java programmer develops applications for controlling the device by using Google-developed Java libraries. Some of the other languages than can be used are Python and Ruby. [1].

4 Methodology and Testing

A modular approach is being used for the development of the VOSS system. As each module is completed it is tested before moving on to the next one. Figure 2 below shows the development flowchart for the system.

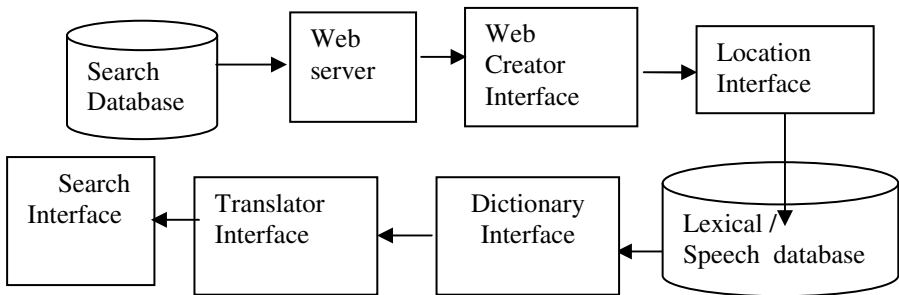


Fig. 2. Development Flow Diagram

4.1 Development Progress

The system is currently 40% completed. The operational components are the web server, the web page creator and the location interface, which are 100% completed. The ones currently being worked on are the dictionary interface, the translator interface and the search engine interface, which are roughly 50% completed.

The system was tested by creating web-based data of about ten locations in Barbados using the web creator interface and then going to the ten predetermined areas and using the locator interface to see if the map and the data about the area would be available. It was tested using a laptop with a GPS locator and a Smartphone with the Android operating system.

4.2 Results of Testing

In examining the web server database it was found that the algorithm which creates the keywords for the automatically generated web pages was only 85% accurate with some of the pages being mislabeled.

In testing the locator interface on the laptop there was an 80% accuracy rate for the GPS software that was programmed into the interface and at least in one case, the location information provided was completely incorrect. However, for the sites that were correct the system did display the correct map location and the accompanying web pages that gave the historical and or cultural details about the area.

The navigation of the map and of the web pages is currently manual as the voice component is not fully operational.

Presently test are being performed on the translator and it is currently correctly parsing through the sentences and is replacing the dialect with Standard English words or phrases but the algorithm for making the sentence grammatically correct is still in development.

References

1. Android Operating System, <http://en.wikipedia.org/wiki/Android>
2. Bangalore, T.M.: Qmel: A Speech-based Question-Answering system on Mobile Devices. In: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL, pp. 55–63 (2010)
3. Hakulinen, M.T.: Agent-based Adaptive Interaction and Dialogue Management Architecture for Speech Applications. Speech and Dialogue. In: Proceedings of the Fourth International Conference, pp. 357–364 (2001)
4. Karl, L.P.: Speech versus mouse commands for word processing applications: An empirical evaluation. *Int. J. Man-Mach. Stud.* 39(4), 667–687 (1993)
5. Mathias Creutz, M.C.: Web augmentation of language models for continuous speech recognition of SMS text messages. In: Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, pp. 157–165 (2009)
6. Michael, S., David, Z., Hu, M., Prasad, S.: Managing multiple speech-enabled applications in a mobile handheld device. *International Journal of Pervasive Computing and Communications* 5(3), 332–359(2009)
7. Pakucs, B.: Butler: A Universal Speech Interface for Mobile Environments. In: Brewster, S., Dunlop, M.D. (eds.) *Mobile HCI 2004*. LNCS, vol. 3160, pp. 399–403. Springer, Heidelberg (2004)
8. Petzold, C.: *Programming Windows Phone 7*. Microsoft Press, Redmond (2010)
9. Shneiderman, B.: The Limits of Speech Recognition. *Communications of the ACM* 43(9) (2000)
10. Windows Phone 7, http://en.wikipedia.org/wiki/Windows_Phone_7