Speech User Interface for Visually Impaired to Access the Synoptic Form of Web Data

Reshma Sheik, Dr. A. Chitra

***Abstract*— The vast information available in the web data is mostly inaccessible for blind and visually impaired since most information retrieval systems do not cater to their needs. The retrieval of the relevant information from multiple web documents in a synoptic form can be highly beneficial. A system with speech user interface for information retrieval of web-data in summarized form was developed. The system is speaker-independent and uses a domain specific language model. Query is formulated based on user's input which can be either in text or in voice form. The query is processed by a search engine, like Google, and a fixed number of relevant web documents are parsed and the textual content is extracted from them. This textual content is summarized and its corresponding speech along with the text is presented to the user. Through this method an enhanced solution which offers a natural and user-friendly way of communication for visually-impaired people is intended. The system is easy to be implemented in real time and its functioning is simple as compared to other existing systems.**

***Index Terms* — Information Retrieval, Synoptic, Visually Impaired.**

I. INTRODUCTION

Blindness is an obstacle that is not even comprehensible for the non-blind. Many day-to-day activities may become challenging, complicated and may require much more effort than one can imagine. Visual handicap includes albinism, cataracts, macular degeneration, glaucoma, optic nerve damage, optic atrophy, retinopathy, retinitis pigmentosa, retinal detachment, eye poor development, etc. The main obstacle for them to use computer systems is that they cannot see clearly or cannot see the screen at all so it becomes very difficult to get visual information and do decisions [1]. There are 314 million of visually impaired in the world. Of the 314 million, 45 million are blind, 124 million have low vision and 145 million have uncorrected refractive errors [3].

Manuscript received November 23, 2014.

Reshma Sheik, Former Student, Department of Computer Science, PSG College of Technology, Coimbatore, India, +91 9495970661,

Dr A. Chitra, Professor & Head of the Department **,** Department of Computer Applications, PSG College of Technology, Coimbatore, India, +91 422-2572177

With the massive availability of information and development of technology, visually handicapped have a pressing need to have access to this information and make up for their physical shortcomings to compete with the rest of the world. Existing software systems are intended for normal people and even for people with physical or mental handicap, but in most cases maybe incompatible for the visually impaired [1]. The reason is extensive dependence on graphical user interface and very less accessibility options like speech. Over the past ten years a lot of effort has been put into the development of automatic text-to-speech and speech-recognition systems. Spoken language dialog systems and related technology offers a natural and user-friendly way of communication. Going one step further, true usability with these technologies must be considered very early in the design process.

The communication goal can be achieved faster and they offer access to large text corpora via modern technical equipment (over computer networks, scanners, etc.) and have a relatively low price. Speech synthesizer and screen reader software still represent basic functionalities that are used by the visually impaired to obtain information by means of a computer. Thus, a voice-driven text-to-speech dialog system is presented.

A. Motivation

1) Visually impaired should be able to access and interact with the computer as a normal person.
2) General IR systems tend to assume visual capabilities. Studies concentrate more on the information needs of visually impaired people and not on how to modify the interfaces to cater to their needs.
3) Visually impaired should be able to access the system using voice commands. To find documents relevant to the information need from a large document set and by tracing the system context information using an efficient speech recognizer.
4) Visually impaired people should be able to retrieve the information from the Internet. Thus problems with HTML and other technological challenges should be dealt.
5) Building a high quality text to speech synthesis system for the visually impaired that should ensure the real-time response of the system.

6) The retrieval of the information in summarized form from multiple documents foster easy searching facility.

B. Objective

1) The system can be learned, used and controlled easily, have simple operations and accurate timely feedback based on the characteristics of visually impaired people.
2) The system responds to speech commands, which allows enhancement of the quality of the interactions between visually impaired users and the platform.
3) The user will be guided through the system, using audio instructions. The user can choose various options in the system, with the help of speech inputs, thus focus on the state-of-art spoken language.
4) Transformation of GUI to SUI – The major goal is to enable speech access to the system, developing a SUI (Speech User Interface) from the existing graphical interfaces. The evolution of our SUI design shows a clear trend towards interpersonal conversational style and away from graphical techniques.
5) Ability to access the information from the World Wide Web from multiple documents in a summarized form.
6) The system should take in advantage of freely available open source efficient speech recognizer, synthesizer and content extractor.

II. LITERATURE SURVEY

System for visually impaired people differs from conventional system in operation mode, operation methods, process control, information tips and information feedback [1].The older Braille coding of texts is still a common aid but this type of coding requires special editions of written corpora or special additional hardware components when used with computers. Thus the solution is relatively costly and requires special skills from the user [4]. Braille printers are available for those who can read Braille. This technology is hardware intensive and expensive. Besides one of the major problems that the visually impaired people face is understanding visual and text form in the Internet which are not in braille form. The use of shortcut keys, speech recognizers (like IBM's ViaVoice or Sphinx), synthetic speech systems (like IBM's text-to-speech or FreeTTS) caters mostly to their needs as an user interface.

Other common technology for people with visual disabilities includes magnification programs, screen readers, Optical Character Recognition (OCR) and note-takers. These systems provide an effective interface for the visually impaired but they don't focus on how the varied forms of available data can be summarized or extracted so that the users are presented with only the relevant information. The following are few of the systems which are commonly used:

1. *Vinux* is a remastered version of Ubuntu 10.10 with screen reader and Braille display support which is optimized for the needs of blind and partially sighted people. By default it provides three screen-readers, two full-screen magnifiers, global font-size and color changing facilities as well as support for USB Braille displays.
2. *K-NFB Reading Mobile* and *Text Scout* are reading assistants for mobile phones that uses the camera to capture and scan text to produce the speech output for the user [2]. This technology is compatible only with few high-end phones which is not affordable to a large set of people who are blind.
3. *Brookes Talk* is a small speech output browser which is independent of visual browsers and also independent of text to speech software applications as it uses Microsoft speech technology [8]. It includes the functionality of a standard Web browser for the blind such as pwWeb-Speak in that it can break up the text part of a web page into headings and links and read out paragraphs etc. However the main aim is to provide a search and orientation tool for blind users in the form of a virtual toolbar of functions that will provide different synopses of a web page to help the user decide whether it will be useful to them or not.
4. *Reading Aid for Visually Impaired* (RAVI) enables the visually impaired should be able to access computers in their native language (Indian languages) with the aid of Multi-Lingual screen reader [9]. It invokes the right TTS system and ensures real-time response of the system.

III. IMPLEMENTATION

The system is designed in such a way that it would guide and instruct visually impaired users to accomplish their task of searching the words from the information source - World Wide Web. The system uses various modules, for Speech Recognition, Query Processing, Content Extraction, Summarization and Speech Synthesis. The system allows both textual input as well as speech input for formulating the query. The flow diagram of the entire system is shown in Figure 1.

A. Automatic Speech Recognition

For inputting by speech, the use of Sphinx 4 which enables a flexible, modular and pluggable framework for Speech Recognition which takes audio input, live from a microphone produces word text as output. This tool employs an HMM (Hidden Markov Model) based ASR system which is speaker independent and language model dependent. Language model uses the Java Speech API Grammar Format (JSGF) [11] which provides a way of specifying what words can be used and how. Since the tool Sphinx 4 uses a domain specific language model, JSGF for Computer Science domain has been constructed with 25 keywords. The grammar for the specified model is shown below.

```
grammar hello;

public <greet> = (Data structure | Java
| Compiler | Computer network |
Software engineering | Data base
| Operating system | Data mining
| Computer architecture | Anti virus
| Blue tooth | E commerce | Fuzzy logic
| Grid computing | Internet | Mainframe
| Server | Reboot | System programming
| Thin client | Virtual memory | Web services
| Batch process | Logic gate | Hard disk );
```

The system uses a dynamic dictionary (cmudict.06d) [11] that contains the words and their transcriptions available for recognition. On running the Sphinx4 the uttered words from the domain are mapped to their corresponding phoneme sequences. This phoneme sequence are then used in creating HMMs for search graph processing and thus produces the word output for the corresponding voice input.

B. Query Processing System

The system is developed in such a way that it can accept both text and voice input by an efficient threading and switching mechanism. There is a choice to select the text tab otherwise the voice input is invoked by default. Through the voice input the user is allowed to utter the domain specific words presented in the grammar of Sphinx. The text output generated is then passed to the Google search engine to extract the links associated with the keywords. At any time the text input tab can be activated and the words typed is directly passed to the content retrieval system.

C. Content Extraction

The system uses the url of the top four relevant links obtained from the search engine. Boilerpipe library [12] is employed for content extraction which provides algorithms to detect and remove the surplus "clutter" (boilerplate, templates) around the main textual content of a web page. This Java library adopts Shallow Text Features using article extraction that examine web documents at text block level. The text content for the four websites are written and concatenated to a single file for summarization.

D. Summarization

The efficiency of the retrieval mechanism from the web, can be improved by summarizing the contents obtained by means of an intelligent summarizer called ISummarizer in Classifier4J library which is a Java library designed to do text classification. It comes with an implementation of a Bayesian classifier. The input is a string of text and the number of lines that can be summarized to. The summarizer has the following steps

- Preprocessing (Stemming, removing stop-words etc).
- Sorting unique words by popularity in the text.
- Splitting the original text on sentence boundaries.
- Including each sentence that first mentions the next most popular word, until the summary is the maximum length requested. So obtained text output is presented to the user.

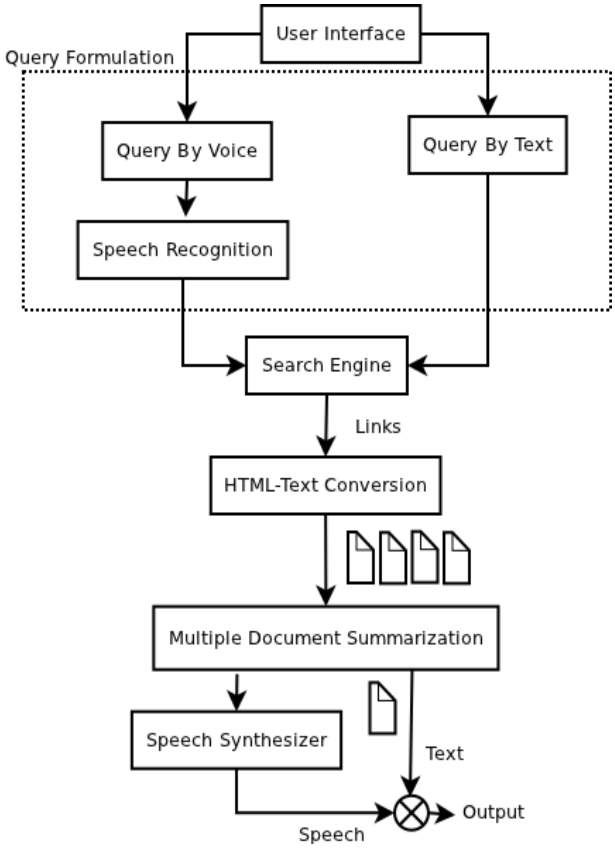

Fig 1: Flow Diagram with various stages of implementation

E. Speech Synthesis System

The most important component of the whole system is the speech synthesis that provides for the artificial production of human speech. The speech synthesizer used here is Free text-to-speech system (Free TTS) which converts normal language text into speech. They render symbolic linguistic representations like phonetic transcriptions for its conversion. At the final output stage the summarized text document is converted into audio file that is played with effective speech for listening.

IV. RESULTS

An information retrieval system can be evaluated in several dimensions including accuracy, efficiency and overall utility. So far, in an information retrieval system, the evaluation has been more focused on accuracy, which is arguably the most important aspect of performance. They can be seen as extended versions of accuracy, a simple metric that computes the fraction of instances for which the correct result is returned. But for the voice-driven text to speech system the evaluation focuses mainly on the performance of the speech recognition speech synthesis content extractor and summarizing tool.

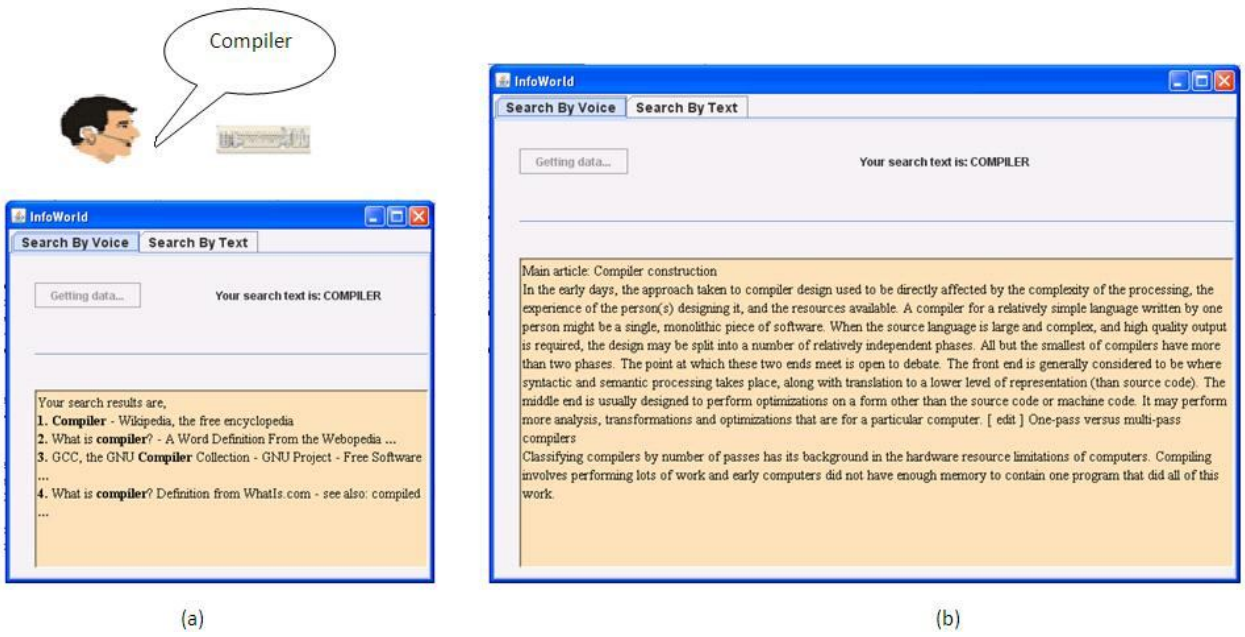

Fig 2: (a) Processed speech query and extracted search results. (b) Multiple Document summarizations from the obtained search results

The focus of this evaluation is to analyze the tool Sphinx 4 for recognizing an isolated distinct word from a known set of distinct words. Here for computer science domain, set of twenty five distinct words was considered as speech units. Table1 gives the results of the speech recognition subsystem. The system was tested with 10 users belonging to different ages and genders. The identification accuracy of the words of the system for 10 users is found to be 89.6%.

Table1: Results

Total No. of Utterances	250
Total Correct Matching	224
Total Wrong Matching	26
Accuracy	89.6%

A noise free environment with a limited vocabulary is the two major requirements of Sphinx4 to obtain a higher accuracy. Extracting content using Boilerpipe is very fast (milliseconds). It just needs the input document (no global or site-level information required) and is usually quite accurate.

The intelligent summarizer in Classifier4J library provides for an efficient and accurate method for text summarization. The performance of this system mainly depends on the string input and the requested number of lines needed for summarization. In this system, text content of four web documents is selected as string input with ten lines requested to achieve for greater accuracy.

The FreeTTS Software provides high converting speed and economy of hard disk space. Text content speaker can be used to read the documents in pace with clear and natural sounding voice.

V. Conclusion

The system aims at effective and quality optimized implementation of an information retrieval system for people with visual impairments which uses the benefits of Internet to access the information via modern technical software's. The development of this system, which recognizes the voice and text input, invoke a search engine for the processed query, efficiently extracted the contents from the information source in summarized form and finally provided with a speech interface for the corresponding text output. Overall the system has met the objectives and it is hoped that this system will benefit the visually impaired users in retrieving the information in a more user-friendly way.

VI. Future Work

Due to the changes in life stages through the advancement of modern technology, the need to access the information retrieval system is likely to change. This change should also be reflected to the visually impaired. The future enhancements that could be made to this system are:

1. Improvements in the sense of more accurate and robust speech recognition that can in-corporate a larger dynamic spoken command grammar that enables the system to recognize all the spoken words.
2. System should evolve towards a specialized web browser with a mouse-driven text-to-speech screen reader and a voice driven dialogue manager.
3. The system should also provide more language preferences for users, so that the most commonly used

and spoken languages could be provided in the system, which will ease the process of communication.

4. The system can be made secure by authenticating using voice parameters or by making use of user's facial features for changing the contents of the information source.

REFERENCES

[1] FuBai-wen, Junying, Akita Y, "*Research on accessible design and development of information system for visually impaired people*" In: Proc. IEEE Conferences on Computer Application and System Modeling (ICCASM), 2010 Volume:14 pp. 299-303.

[2] Joshi Kumar, A.V. Visu, A.Mohan Raj, S. Madhan Prabhu, T. Kalaiselvi, "*A pragmatic approach to aid visually impaired people in reading, visualizing and understanding textual contents with an automatic electronic pen*" In: Proc. IEEE Conference on Computer Science and Automation Engineering (CSAE), 2011 pp. 623-626.

[3] Halimah, B.Z. Azlina, A. Behrang, P. Choo, "*Voice recognition system for the visually impaired: Virtual cognitive approach*" In Proc. IEEE Conference on Information Technology, 2008,Volume 2 pp. 1- 6.

[4] Simon Dobrisek, Jerneja Gros, Bostjan Vesnicer, "*Evolution of the Information-Retrieval System for Blind and Visually-Impaired People*", In International Journal of speech Technology, 2003 Volume 6, pp. 301-309.

[5] S.Asha, C.Chellappan, "*Automatic processing of audio lectures for information retrieval Vocabulary selection and language modeling*", International Journal of Computer Applications 2011,Volume 14.

[6] Kirsty, Steve Wright, "*Internet for Blind and Visually Impaired*" Enterprise Information Research Group, Monash University JCMC 2007.

[7] Pisit, Prougestapor, "*Development of A Web Accessibility Model For Visually-Impaired Students On Elearning Websites*", In International Conference on Educational and Network Technology (ICENT), 2010.

[8] M.Zajicek, C. Powell, "*Web search and Orientation with Brookestalk, Royal Institute for Blind*", 2007.

[9] Veera Raghavendra, Anand Arokia, "*RAVI: Reading Aid for Visually Impaired*", International Institute of Information technology, Hyderabad, January 2008.

[10] T. V. Raman, Charles L. Chen, "*Eyes Free User Interaction*". Available: http://emacsspeak.sf.net Google, May 21, 2009.

[11] Willie Walker, Paul Lamere, Philip Kwok, Bhiksha Raj, Rita Singh, Evandro Gouvea, Peter Wolf, Joe Woelf. "*Sphinx-4: A Flexible Open Sourc Source Framework for Speech Recognition"*,2004, Available: http://cmusphinx.sourceforge.net/sphinx4.

http://cmusphinx.sourceforge.net/sphinx4/javadoc/edu/cmu/sphinx/jsgf/JSGFGrammar.html.

[12] Kohlschtter, Peter Fankhauser, Wolfgang Nejdl, "*Boilerplate Detection using Shallow Text Features*", L3S Research Center Leibniz University Sannover, Germany ACM Journal February,2010.

Reshma Sheik: Is currently on sabbatical, worked for two years with research division of IBM-ISL. She is interested in Data Structures, Algorithm Analysis, Compiler Design, Data Base Management Systems & Human computer interfaces. She is a post graduate in computer science from PSG college of Engineering. She was the best outgoing student from PSG College of Technology and was the rank holder at under graduation level in Kerala University. She has presented papers in couple of national seminars till date. She is very passionate about bridging the gap between industry & academics and eager to start a career in teaching at the earliest.

Dr. A. Chitra : Is currently the Professor and Head of the Department of Department of Computer Applications at PSG college of Engineering, Coimbatore, India. She holds a Masters Degree and Ph.D. Specializes in Data Structures and Algorithms, Compilers, Soft Computing, Agent Technology, & Cognitive modeling. She keeps publishing in International and National Journals and a prominent speaker in both International & national conferences. She is a member of ISTE, FIE, ACCS & CSI. Apart from this she is also the Secretary, of ACCS (Advanced Computing and Communications Society), Coimbatore. Some of the awards she bagged are ISTE National award for the outstanding academic on 2006, M N Saha Award for the best application Oriented paper by ISTE on June 2006 & Young Women Scientist award for the excellent contribution in the field of science and Technology by Tamil Nadu State Council for Science and Technology on 2005.