

Voice Based Search Engine for Visually Impaired People

D.Parameswari¹, R.Aruna² and K.Bhavani³

¹⁻³Kongunadu College of Engineering and Technology/Department of Information Technology, Trichy, Tamil Nadu
Email: paramesh@kongunadu.ac.in, aruna.akshaya40@gmail.com, bavanithanjavur@gmail.com

Abstract—World Wide Web is rapidly emerging as the universal information source for our society and it is accessible using a web-browsing package from a networked computer. The design of information on the web is visually oriented. This approach is not always practical, especially for the visually impaired persons. The focus of this project is to develop a prototype which supports web browsing using a speech-based interface. Audio icons are built into the prototype so that users can have better understanding of the original structure/intent of a web page. Voice browsers allow people to access the Web using speech synthesis, pre-recorded audio, and speech recognition. Voice may also be offered as an adjunct to conventional desktop browsers with high resolution graphical displays, providing an accessible alternative to using the keyboard or screen, for instance in automobiles where hands/eyes free operation is essential. The browser will have an integrated text extraction engine that inspects the content of the page to construct a structured representation. This helps in easy and flexible navigation of the page so as to rapidly home into objects of interest.

Index Terms— Speech synthesis, Speech recognition, Text extraction Engine.

I. INTRODUCTION

Information contained on the World Wide Web is inaccessible to many people. The web is primarily a visual medium that requires a keyboard and mouse to navigate. People, who lack motor skills to use a keyboard and mouse, find navigation troublesome [1]. The focus of this project is to develop a prototype which supports web browsing using a speech-based interface, e.g. a phone, and to measure its effectiveness. The command input and the delivery of web contents are entirely in voice. Audio icons are built into the prototype so that users can have better understanding of the original structure/intent of a web page. Navigation and control commands are available to enhance the web browsing experience.

The effectiveness of this prototype is evaluated in a user study involving both normally sighted and visually impaired people. Voice browsers allow people to access the Web using speech synthesis, pre-recorded audio, and speech recognition. This can be supplemented by keypads and small displays [3]. Voice may also be offered as an adjunct to conventional desktop browsers with high resolution graphical displays, providing an accessible alternative to using the keyboard or screen, for instance in automobiles where hands/eyes free operation is essential. Voice interaction can escape the physical limitations on keypads and displays as mobile devices become ever smaller. The browser will have an integrated text extraction engine that inspects the content of the page to construct a structured representation.

The internal nodes of the structure represent various levels of abstraction of the content. This helps in easy and flexible navigation of the page so as to rapidly home into objects of interest.

II. DATA MINING

The major components of any data mining system are data source, data warehouse server, data mining engine, pattern evaluation module, graphical user interface and knowledge base.

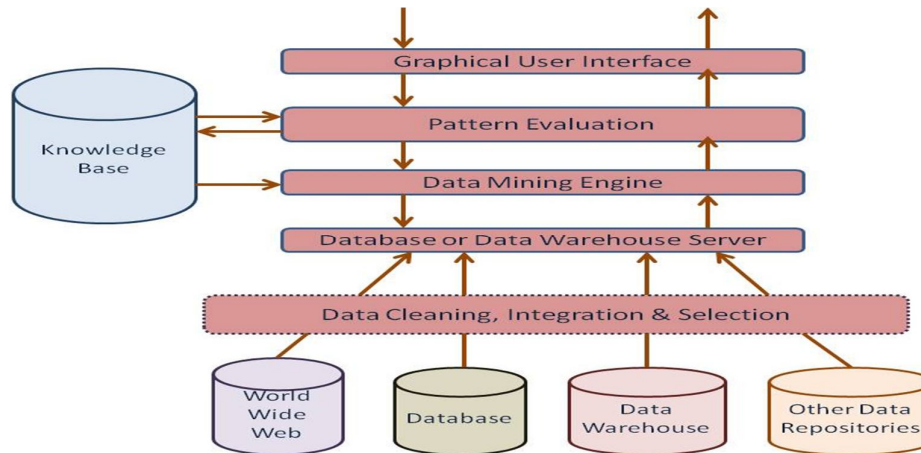


Fig 1. Architecture of Data Mining

Database, data warehouse, World Wide Web, text files and other documents are the actual sources of data. You need large volumes of historical data for data mining to be successful. Organizations usually store data in databases or data warehouses. World Wide Web or the Internet is another big source of data. The database or data warehouse server contains the actual data that is ready to be processed. Hence, the server is responsible for retrieving the relevant data based on the data mining request of the user.

The data mining engine is the core component of any data mining system. It consists of a number of modules for performing data mining tasks including association, classification, characterization, clustering, prediction, time-series analysis etc. The pattern evaluation module is mainly responsible for the measure of interestingness of the pattern by using a threshold value. It interacts with the data mining engine to focus the search towards interesting patterns.

The graphical user interface module communicates between the user and the data mining system. When the user specifies a query or a task, this module interacts with the data mining system and displays the results. The knowledge base is helpful in the whole data mining process. It might be useful for guiding the search or evaluating the interestingness of the result patterns. The knowledge base might even contain user beliefs and data from user experiences that can be useful in the process of data mining.

III. VOICE BASED SEARCH ENGINE

Voice search, also called voice-enabled search, allows the user to use a voice command to search the Internet, or a portable device. Currently, voice search is commonly used in "directory assistance", or local search.

Applications of voice search include:

- Making search engine queries.
- Requesting specific information, such as a stock quote or sports score.
- Launching programs and selecting options.

The free voice search service, however, uses another approach. Speech recognition and generation technologies offer a potential solution to these problems by augmenting the capabilities of a web browser.

IV. ALGORITHM (HIDDEN MARKOV MODEL)

Automatic continuous speech recognition (CSR) has many potential applications including command and control, dictation, transcription of recorded speech, searching audio documents and interactive spoken dialogues.

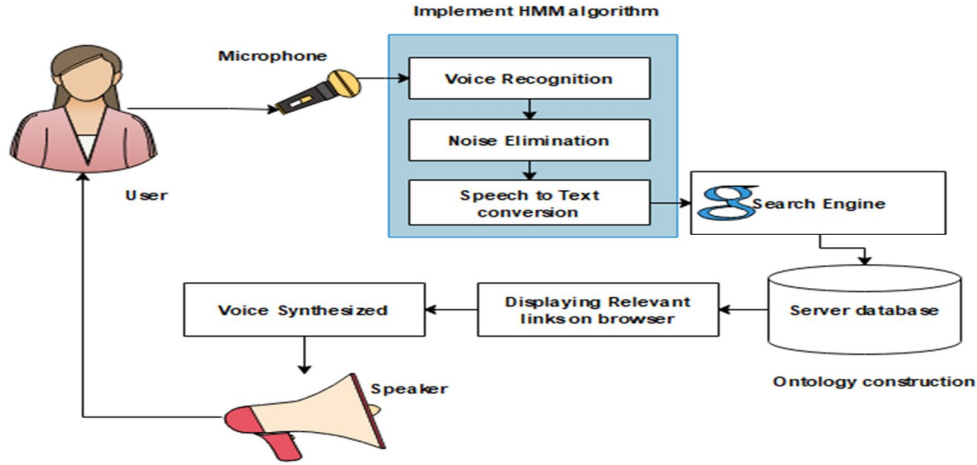


Fig 2. Proposed Framework

Since speech has temporal structure and can be encoded as a sequence of spectral vectors spanning the audio frequency range, the hidden Markov model (HMM) provides a natural framework for constructing such models. HMMs lie at the heart of virtually all modern speech recognition systems and although the basic framework has not changed significantly in the last decade or more, the detailed modelling techniques developed within this framework have evolved to a state of considerable sophistication. The clustering of logical to physical models typically operates at the state-level rather than the model level since it is simpler and it allows a larger set of physical models to be robustly estimated.

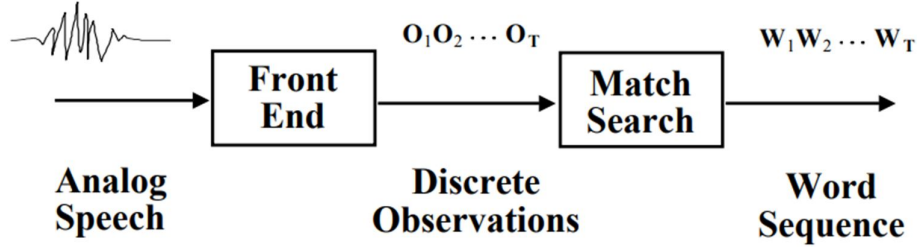


Fig 3. Working of HMM

V. EXPERIMENTAL DETAILS

The extraction of text and presenting it to a visually handicapped person has many difficult aspects to it. With innumerable web pages present in the web, there is a varied diversity in the type of the pages. A web page may contain more than one kind of contents like, links, images, advertisements, and animations. These contents may not provide valuable information to a visually impaired person. Further, the document structure of an email page is also different from other page.

The Text To Speech module responds to user like an artificial intelligent agent by guiding through browsing. This module reads out the content in the web browser by parsing the web document by removing html tags and extracts only text to the user as audio. It also returns information's like date, day, time, weather, etc. on request.

Speech to text conversion is the process of converting spoken words into written texts. This process is also often called speech recognition. Although these terms are almost synonymous, Speech recognition is sometimes used to describe the wider process of extracting meaning from speech, i.e. speech understanding. The term voice recognition should be avoided as it is often associated to the process of identifying a person from their voice, i.e. speaker recognition. All speech-to-text systems rely on at least two models: an acoustic model and a language model. To get the best transcription quality, all of these models can be specialized for a given language, dialect, application domain, type of speech, and communication channel. The complete voice-to-text conversion process is done in three steps. The software first identifies the audio segments containing speech,

and then it recognizes the language being spoken if it is not known a priori, and finally it converts the speech segments to text and time-codes.

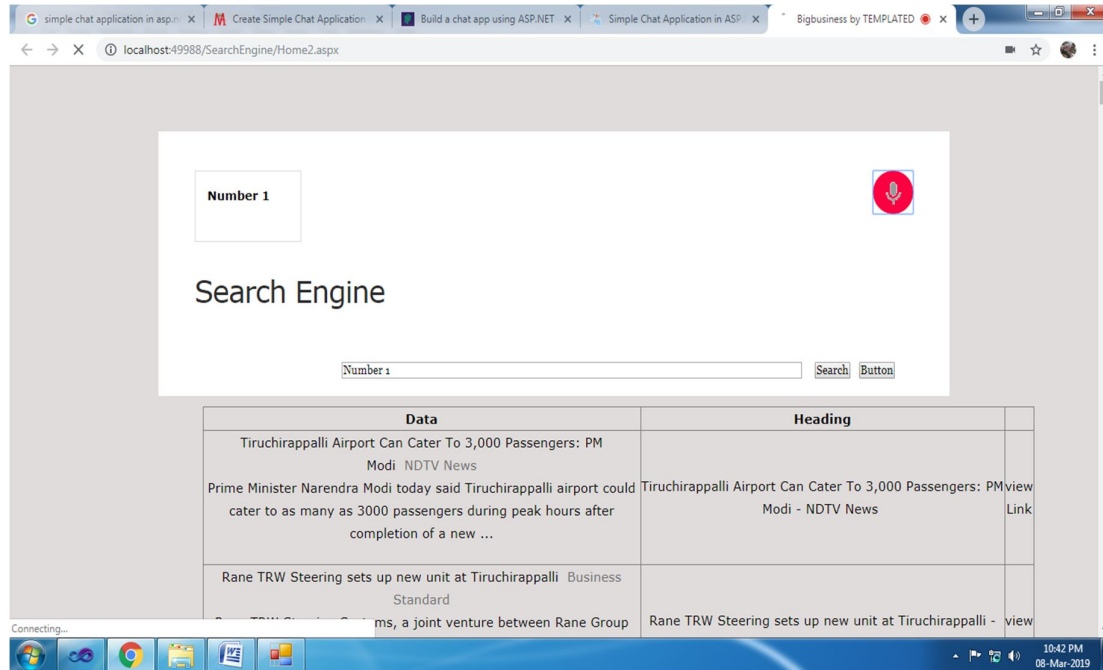


Fig 4.Search Engine Output

VI. CONCLUSIONS

In this project we proposed an efficient way of accessing the web browser is presented which is termed as voice browsing in which visually impaired people can access the browser using speech. As access to internet visually incurs limitations such as visually impaired persons cannot use keypads, touch screens etc. for giving inputs to computer. User can speech the word and converted into text automatically, now this browser reduces their effort by performing this conversion automatically. And the blind people can also use this browser to convert text documents in English characters. Thus combination of browsing with speech technology is an efficient way of accessing webs. In addition, all the text content present over the web for various links can be made accessible by using speech technology. This technology can also be implemented in browser. More work can be done to increase the accuracy, pronunciation and precision of speech technology. The proposed method has used only English language. A new search engine enabled with voice recognizing and voice synthesis mechanisms is proposed. From the Experiments and observations it is understood that the accuracy of the voice recognition can be further improved when we add additional hardware and software.

REFERENCES

- [1] Arnaud Ramey, Javier F. Gorostiza, Miguel A. Salichs, "A Social Robot as an Aloud Reader: Putting together Recognition and Synthesis of Voice and Gestures for HRI Experimentation" International conference on Human-Robot Interaction, March 2012.
- [2] Yu shao, chip-hong chang, "Bayesian Separation with Sparsity Promotion in Perceptual Wavelet Domain for Speech Enhancement and Hybrid Speech Recognition" Journal on systems, man, and cybernetics, volume 41, issue 2, March 2011.
- [3] Zhanyu Ma and Arne Leijon, "A Probabilistic Principal Component Analysis Based Hidden Markov Model for Audio-Visual Speech Recognition", Conference on computing & processing, Digital Object Identifier: 10.1109/ACSSC.2008.5074819 - 2008.
- [4] Ramin Halavati, Saeed Bagheri shouraki and Saman Harati Zadeh, "Recognition of human speech phonemes using a novel fuzzy approach", Journal appliedto soft computing, Volume 7, Issue 3, June 2007
- [5] Dharani Perera, "Voice recognition technology for visual artists with disabilities in their upper limbs", conference on ComputerHuman Interaction, OZCHI '05, November 2005

- [6] Yang Liu, Elizabeth Shriberg, Andreas Stolcke , Dustin Hillard, Mari Ostendorf, Mary Harper, "Enriching Speech Recognition with Automatic Detection of Sentence Boundaries and Disfluencies" Journals and magazines, Volume 14, issue 5, September 2006
- [7] Manny Rayner, Beth Ann Hockey, Nikos CVhatzichrisafis, Kim Farrell, Jean Michel Renders, "A Voice Enabled Procedure Browser for the International Space Station", Association for Computational Linguistics, June 2005
- [8] Heiga Zen and Tomoki Toda "An Overview of Nitech HMM-based Speech Synthesis System for Blizzard challenge", Blizzard Workshop-2005
- [9] Saiyan Saiyod, Sakchai Thipchaksurat, and Somsak Mitatha "Thai Speech Synthesis for Text-to-Speech based on Formant Synthesis Technique", source www.ecti-thailand.org-, April 2012
- [10] E. Eide, A. Aaron, R. Bakis, W. Hamza, M. Picheny, and J. Pitrelli, "A corpus-based approach to expressive speech synthesis", IBM T. J Watson Research Center Yorktown Heights, NY 10598, 2004
- [11] Alexandros Potamianos, Member, IEEE, and Petros Maragos, Fellow, IEEE "Time-Frequency Distributions for Automatic Speech Recognition", IEEE Transactions on Speech and Audio Processing, vol. 9, no. 3, March 2001.