

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Library Philosophy and Practice (e-journal)

Libraries at University of Nebraska-Lincoln

4-17-2018

A STUDY ON INNOVATIVE TRENDS IN MULTIMEDIA LIBRARY USING SPEECH ENABLED SOFTWARES

V Jeevitha

Periyar University, pinckjeevi@gmail.com

E.S Kavitha

Periyar University, kavithaesk@gmail.com

Follow this and additional works at: <https://digitalcommons.unl.edu/libphilprac>



Part of the [Library and Information Science Commons](#)

Jeevitha, V and Kavitha, E.S, "A STUDY ON INNOVATIVE TRENDS IN MULTIMEDIA LIBRARY USING SPEECH ENABLED SOFTWARES" (2018). *Library Philosophy and Practice (e-journal)*. 1796.

<https://digitalcommons.unl.edu/libphilprac/1796>

A STUDY ON INNOVATIVE TRENDS IN MULTIMEDIA LIBRARY USING SPEECH ENABLED SOFTWARES

V.Jeevitha, Ph.D Research Scholar, Department of Library & Information Science Periyar University,
Salem-11, Ph.No:8940300400, E.Mail ID: pinckjeevi@gmail.com

Dr.E.S.Kavitha, Assitant Professor, Department of Library & Information Science Periyar University,
Salem-11, E.Mail ID: kavithaesk@gmail.com

ABSTRACT

The present study attempts to analyze the significance of utilization of library usages, when the traditional search is complemented with voice based search. The study also presents a methodology to automatically catalogue the multimedia resources and search them too using voice based search tools developed. It analyses the significant aspects of deployment of such tools in academic libraries. The researcher has decided to conduct a study on voice enabled multimedia library using voice search softwares.

KEY WORD

Voice Search, Speech Recognition Engines, Speech-To-Text, Speech Recognition Grammar, Google Voice Search

INTRODUCTION

Searching and finding the required resource in a library has always been an intricate task. It is more or less a strategically accomplishment. In fact most resources of the library go unused because of constraints like these and thereby hamper the growth of the civil society. Huge financial resources are deployed on a continual basis in establishing and continuing a well stocked library. In the context of third world countries, under-utilization of library resources is simply money gone down the drain which could have instead been deployed for other essential civic purposes.

Revolutions that took place in the areas of Electronics and Information Technology have had their impacts in libraries too. An OPAC (Online Public Access Catalog) is an online bibliography of a library collection that is available to the public. OPACs developed as stand-alone online catalogs, often from mainframe library catalog with the arrival of the Internet, most libraries have made their OPAC accessible from a server to users all over the world.

Searching the library resources the traditional way has inherent drawbacks. It expects the user to possess a minimum grasp of English language, more exactly the ability to remember and enter the correct

spelling of the phrases being sought. Besides the user must also possess basic skills to handle the computer keyboard. Though these may seem apparently trivial, they are indeed big constraints for vast majority of people, especially for people who are from marginal sections of the society, or who are first-time learners in their families, or people from deeply rural backgrounds where basic education itself is an invaluable add-on. The current constraints force these users from effective utilization of library resources. Another category of users who are hampered by the text-based searching using keyboards are differently abled people. They too are dissuaded from effective usages of library because of inherent disabilities. This is paradoxical indeed, given the established fact that effective utilization of library resources enhances one's better prospects of overall success in life.

Voice Search is easily one of the most significant technologies that the digital age has blessed us with. In this the user "Speaks" to the computer instead of traditional keyboard inputs there is an underlying software which understands what the user has spoken and converts it into textual format and treasons it to the target application. A sign of the times, this feature allows us to get instant and optimized solutions with just the sound of our voices. They are easy, fast, and personalized they also provide an improved user experience distinctly different from the textual searches. Voice search truly is a remarkable technology that will reshape the way libraries are effectively used.

OBJECTIVE

1. To study the impact of encoding the multimedia resources of the library in metadata format and to provide a search mechanism on that on the performance variable of "usage of multimedia resources of the library" using speech software.
2. To provide a neutral xml based interface for both the above tools for existing library solutions for seamless integration with the existing solutions, avoiding major overhaul of existing systems involving financial and skilled man-hour resources.
3. To study the composite and collateral impact of introduction of technology based services in the increased and effective use of library services.

To provide special focused notifications on the updating of library services to target groups automatically to make them aware of the latest developments in the library services.

Online Catalogue

Cataloguing of Library Resources in Computer understandable formats lead to Computerized Database of Library Resources. This enables the Library Resources to be searched effectively, providing ease of access and greatly reducing record seek times. Also computerization of library resources enables one to search for them even from other computers connected to that computer. The resources can be sought using pivot elements, called search keys

in computer terminologies. The search keys can be Book Name, Author Name, Publication, Year of Publication, key words etc. Computerized search provides other advantages that these resources can be searched either in full or in part, or searches can be made including/excluding them etc.

Metadata

Margia Lei Zeng and Jian Qin offer the explanation for Metadata as “Information-bearing entities come in a wide variety of types. They may be physical objects (including born-digital and non-digital objects), digitized surrogates of physical objects, or simply sets of information about digital or physical objects (e.g., any metadata dataset). Different information-bearing entities require different structured data to describe them so they can be identified, discovered, obtained, and used. The structured, encoded data describing the characteristics of these information-bearing entities are called metadata”.

Multimedia Tags

Different types of multimedia files have different types of tags. For example, files with extension of MP4 commonly have the tags of Video Type, Movie Name, Language, Episode name, Actors, Director, Screen writer, Tagline, Comment, Cover picture, Genre etc. It must be remembered that this tag list is not absolute that is expandable. Also it is not necessary that there must be a specific value for each of the attribute.

Multimedia Search

Refers to the tools and technologies used to search the multimedia resources by searching their associated tags.

SPEECH RECOGNITION ENGINES

Speech Recognition Engines are made up of the following components:

Language Model or Grammar

Language Models contain a very large list of words and their probability of occurrence in a given sequence. They are used in dictation applications. Grammars are a much smaller file containing sets of predefined combinations of words. Grammars are used in IVR or desktop

Command and Control applications. Each word in a Language Model or Grammar has an associated list of Phonemes (which correspond to the distinct sounds that make up a word).

Acoustic Model

Contains a statistical representation of the distinct sounds that make up each word in the Language Model or Grammar. Each distinct sound corresponds to a phoneme.

Decoder

Software program that takes the sounds spoken by a user and searches the Acoustic Model for the equivalent sounds. When a match is made, the Decoder determines the phoneme corresponding to the sound. It keeps track of the matching phonemes until it reaches a pause in the users speech. It then searches the Language Model or Grammar file for the equivalent series of phonemes. If a match is made it returns the text of the corresponding word or phrase to the calling program.

TEXT TO SPEECH

Text to speech, abbreviated as TTS, is a form of speech synthesis that converts text into spoken voice output. Text to speech systems were first developed to aid the visually impaired by offering a computer-generated spoken voice that would “read” text to the user.

SPEECH-TO-TEXT

Speech to text software is a type of software that effectively takes audio content and transcribes it into written words in a word processor or other display destination. This type of speech recognition software is extremely valuable to anyone who needs to generate a lot of written content without a lot of manual typing. It is also useful for people with disabilities that make it difficult for them to use a keyboard.

SPEECH RECOGNITION GRAMMAR

A speech recognition grammar defines what a user can say. Technically, it defines the syntax of the spoken input that can be heard by a speech recognizer.

VOICE SEARCH

Voice Search is a speech recognition technology that allows users to search by saying terms as in natural conversation which get recognized automatically by the machine, rather than typing them into a search field.

PURPOSE OF VOICE ENABLED MULTIMEDIA LIBRARY

To comprehensively state and understand the purpose of Voice Enabling Library Services, we need to take a concise and comprehensive understanding of the prevailing scenario concerning libraries.

The current digital era has seriously challenged the traditional position of the academic library as the central mediators of information. The tech savvy user empowered with the Internet prefers the comfortable “do it at leisure” model rather than to dig deep in the wealth of information the traditional libraries possess. The advent of smart phones and ever dwindling internet charges also keep the user away from the traditional libraries, giving him a comfort feeling that the information he seeks is at his fingertips. Even though the internet offers it all, still it cannot take the sheen away from the traditional resources of the library. The veracity of information in the net and the time taken to conceive and publish it on the net is a pittance compared to the quantum of efforts and the magnitude of work involved in writing a standard book.

Voice Finger

Voice Finger is lesser-known speech recognition software with a strong set of features. Voice Finger is a speech recognition tool that enables the user to control the mouse and keyword just using his voice, in the fastest possible way. Voice Finger enables zero computer contact, so the user can confidently just use his voice and rest hands. Other speech recognition software assumes that the user can type and click for some tasks, in contrast Voice Finger does not leave gaps that the user will need type or click, it was made to do everything by voice.

ViaTalk

ViaTalk is a portable alternative to more advanced voice recognition software. ViaTalk turns voice into text. All the user needs is to just to talk to his/her Smartphone, and it types up to 200 words in a minute for the user. It's easy to use. All the user has to do just to connect the Smartphone with the PC with ViaTalk, open the APP and start talking. Support Office software,

email, Skype, Facebook and more. ViaTalk is also an excellent translation tool. ViaTalk also lets the user edit the translated result on PC easily.

Tazti

Tazti is a speech recognition software package developed and sold by Voice Tech Group, Inc. for Windows personal computers. The most recent package is version 3.0, which supports Windows 7, Windows 8 and Windows 8.1, 64-bit editions. Earlier versions of Tazti supported Windows Vista and Windows XP. PC video game play by voice, controlling PC applications and programs by voice and creating speech commands to trigger a browser to open web pages, or trigger the Windows operating system to open files, folders or programs are Tazti's primary features. Earlier versions of Tazti included a lite Dictation feature that is eliminated from the latest version.

Dragon Professional

Dragon voice recognition software can smoothly transition between commands and dictation. Whether one is creating documents, spreadsheets, presentations, sending emails, filling out forms or simply looking to reduce typing stress, Dragon Suite of software solutions drives new levels of productivity at work or wherever the job takes one to. It facilitates one to dictate and transcribe documents with a fully customizable experience that optimizes speed and accuracy.

OK Google

"OK Google" is the virtual assistant built inside android smart mobiles the user invokes simply by saying OK Google. Once invoked the user can talk to be machine as though conversing with an intelligent personal assistant. For example one can ask "nearest beauty parlor", "spiritual books Selling shop", "best non-veg restaurants" "weather forecast for tomorrow" etc. For each of these queries Google will list out the best possible answers customized to individual user preferences. Google has integrated its voice assistant with all its other products like music, maps, YouTube, etc undoubtedly "OK Google" is a boon to the Google user fraternity.

SIRI

SIRI is Apple Inc. 's answer to Google voice assistant it runs on apple smart phones and macbooks, apple TVs SIRI users voice queries and a natural language user interface to answer questions, make recommendations and perform actions by delegating requests to a set of internal services. SIRI is a continually self evolving assistant, meaning that with ongoing usage it adapts to users searches and preferences, language usages, etc. The results returned by SIRI are also individualized.

Microsoft Cortana

Microsoft is a late entrant into the smart phone market. It soon caught up with the competition with its Cortana, the virtual assistant for windows phone and its latest desktop operating system Windows 10. Cortana recognizes natural voice and answers question can set reminders and works with major languages.

CONCLUSION

Voice Search has been made an integral part of all new age devices and applications. Almost all major software platforms support voice search with programmable interface options for existing software developers to hook their solutions. With all indicators pointing in the direction of rapidly increasing adaption by user, libraries too can't lag behind and should expand their services adapting to this technological advancement.

REFERENCE

1. Voice Finger: voice control of mouse and keyboard <https://www.afb.org/forum/technology-talk/voice-finger-voice-control-of-mouse-and-keyboard/12>
2. [Voice Finger: Improve Windows Speech Recognition grid and key commands](https://hand2mouth.wordpress.com/2010/04/26/voice-finger-improve-windows-speech-recognition-grid-and-key-commands/)
<https://hand2mouth.wordpress.com/2010/04/26/voice-finger-improve-windows-speech-recognition-grid-and-key-commands/>
3. Tazti Speech Recognition Software and Windows 10 Compatibility, <https://www.tazti.com/>
4. <https://en.wikipedia.org/wiki/Tazti>
5. Kamaclean (2010) What is the difference between a Speech Recognition Engine and a Speech Recognition System <http://www.voxforge.org/home/docs/faq/faq/what-is-the-difference-between-a-speech-recognition-engine-and-a-speech-recognition-system>
6. Vangie Beal (2017) TTS - text to speech <https://www.webopedia.com/TERM/T/TTS.html>

7. Techopedia (2016) Speech-to-Text Software, <https://www.techopedia.com/definition/23767/speech-to-text-software>
8. Margarate Rouse (2014) WhatIs.com, Voice Search, <http://whatis.techtarget.com/definition/voice-search>