Audio, Video Indexing & Retrieval

Project report submitted to
Indian Institute of Technology Kharagpur
in partial fulfilment for the award of the degree
of

Masters of Technology
in
Information and Communication Technology

By
Chandra Shekhar Sengupta
(Roll No:- 11IT61K14)

Under the guidance of
Dr. K. S. Rao

School of Information Technology
INDIAN INSTITUTE OF TECHNOLOGY
KHARAGPUR

November 2013

CERTIFICATE

This is to certify that the thesis '**Audio, Video Indexing & Retrieval**', submitted by **Chandra Shekhar Sengupta** (**11IT61K14**) in partial fulfillment of the requirements for the degree of Master of Technology in Information and Communication Technology, is a bonafide record of the work done by him in the School of Information Technology, Indian Institute of Technology Kharagpur, under my supervision and guidance.

Dr. K. S. Rao
Associate Professor
School of Information Technology
Indian Institute of Technology Kharagpur,
India.

Signature _____

Date _____

DECLARATION

I certify that

a. The work contained in the report is original and has been done by myself under the guidance of my supervisor.

b. I have followed the guidelines provided by the Institute in writing the Report.

c. I have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.

d. Whenever I have used materials (data, theoretical analysis, and text) from other sources, I have given due credit to them by citing them in the text of the thesis and giving their details in the references.
.

Date:-

_____

Chandra Shekhar Sengupta

Roll No:-11IT61K14

ACKNOWLEDGEMENT

It gives me immense pleasure to express my sincere gratitude to **Dr. K. S. Rao**, under whose supervision and guidance, this work has been carried out. Without his advice and constant encouragement throughout, it would have been impossible to carry out this project work with confidence. It's really a privilege to work under his guidance.

I would like to express my sincere gratitude to **Dr. Indranil Sengupta** for encouraging me to develop an Android App and Web based Service for Audio, Video Indexing.

I would also like to thank all faculty members of the School of Information Technology who have been extremely kind and helpful to me and motivated me from time to time to perform my level best.

Thanks to my family, friends and colleagues for listening to my ideas about this project and encouraging me.

Thanks a lot to Researchers, Open source world and bloggers who have provided so much information related to this project.

Chandra Shekhar Sengupta
School of Information Technology
Indian Institute of Technology
Kharagpur, India

Abstract

Searching is the most performed activity in web as well as in local systems nowadays. We have so much of data around us which requires efficient searching to retrieve the most relevant results. Textual searching is becoming more advanced with the advent of latest technologies developed by Google, Microsoft. But huge amount of data available are in audio video format which are not efficiently searchable using existing search mechanism. We need better mechanism for indexing and searching audio, video data.

According to the 2011 update of Cisco's VNI IP traffic forecast, by 2015, Internet video will account for 61% of total Internet data.[1] Surprisingly in 2010, Google had only indexed .004% of the data on the internet. These statistics help us understanding the need for more effective audio, video indexing & retrieval mechanism.

This report depicts a new approach for indexing & searching audio, video files and this mechanism will be well integrated with existing search engine Google and video-sharing website YouTube so that we can utilize existing ecosystems.

CONTENTS

List of Figures

Chapter 1:

Introduction

With the advent of essentially unlimited data storage capabilities and with the rapid use of the Internet, it should be possible to access any of the stored information with a few keystrokes or voice commands. Since much of this data will be in the form of speech and video from various sources, it becomes important to develop technologies necessary for indexing and browsing such audio, video data.

Human beings have amazing ability to distinguish different types of audio, video. Given any audio/video piece, we can instantly tell the type of audio (e.g., human voice, music or noise), type of video (e.g., content, ambience or noise), speed (fast or slow), the mood (happy, sad, relaxing etc.), and determine its similarity to another piece of audio, video. .[3] However, a computer sees a piece of audio/video as a sequence of sample values. At the moment, the most common method of accessing audio, video pieces is based on their titles or file names. Due to the incompleteness and appropriateness of the file name and text description, it may be hard to find audio, video pieces satisfying the particular requirements of applications. In addition, this retrieval technique cannot support queries such as “find audio/video pieces similar to the one being played” (query by example).

With more and more audio, video being captured and stored, there is a growing need for automatic audio, video indexing and retrieval techniques that can retrieve relevant audio, video pieces quickly on demand.

Thesis Organization

The thesis has been organized in the following way:-

Chapter 1 focuses on introduction, need, objective of a new audio, video indexing & retrieval system.

Chapter 2 discusses the related work in Audio, Video Indexing & Retrieval and the differentiating factor of our approach from the existing ones.

Chapter 3 Analyses its relevance by providing few case studies and proposes the Architecture & Algorithms for Audio, Video Indexing & Retrieval.

Chapter 4 contains the implementation work done so far, work in progress & future scope of work and pending areas.

Finally it concludes along with Reference resources.

Proposed approach

In this approach, we are extracting information about the video and audio content of the file from various aspects such as: 1. Speech Content Based, 2. Video Content based 3. Acoustic Content Based. Then we are determining most suitable tags and attaching them with file so that we can search & retrieve this file based on those tags.

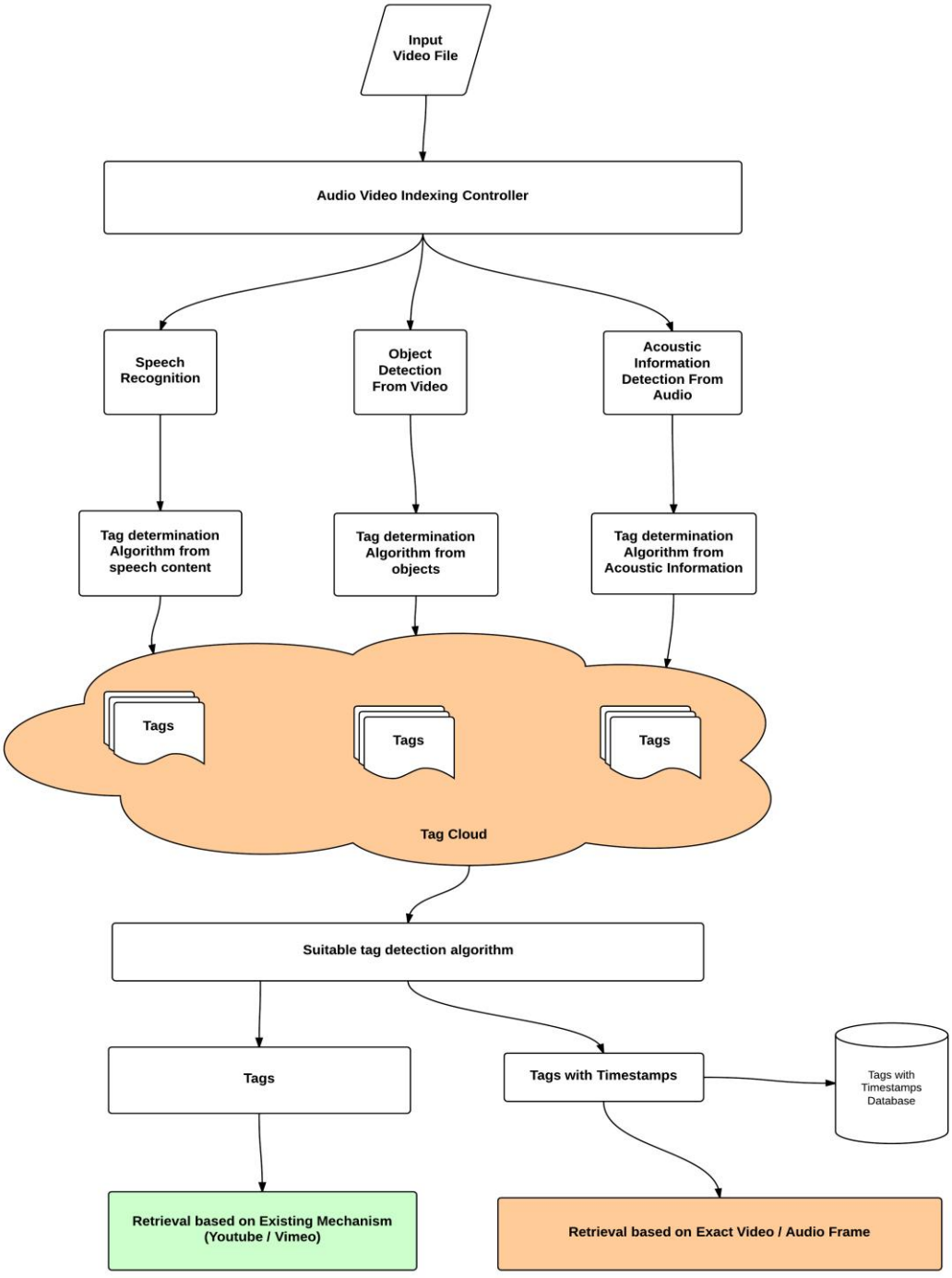

Figure 1 : Proposed Approach for Audio. Video Indexing & Retrieval

Motivation (Big data & mobile technology)

The Economist reports in its 2012 Outlook that the quantity of global digital data expanded from 130 Exabyte in 2005 to 1,227 in 2010, and is predicted to rise to 7,910 Exabyte in 2015.

An Exabyte is quintillion bytes. If you find that hard to visualize, consider this: someone has calculated that if you loaded an Exabyte of data on to DVDs in slim line jewel cases, and then loaded them into Boeing 747 aircraft; it would take 13,513 planes to transport one Exabyte of data. Using DVDs to move the data collected globally in 2010 would require a fleet of more than 16 million jumbo jets. .[4]

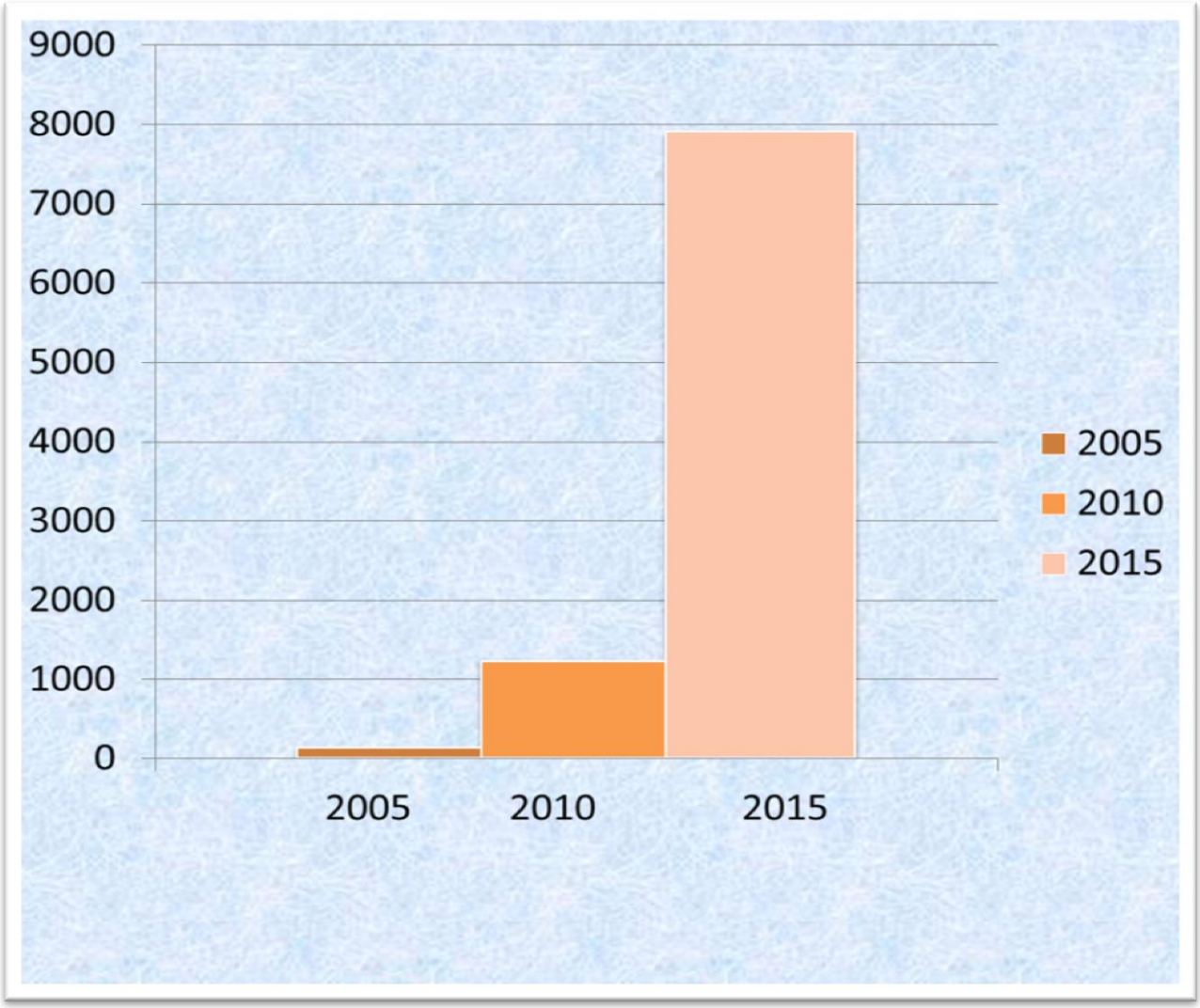

Figure 2: Amount of Global Data in last 15 years

And Exabyte are rapidly becoming passé. The volume of stored information in the world

is growing so fast that scientists have had to create new terms, including zettabyte and yottabyte, to describe the flood of data. The importance of big data is not just a result of its size or how fast it is growing (about 60 per cent a year), but also the reality that the data come from an amazing array of sources. The Internet captures lots of data. Facebook alone has more than 800 million active users, more than half of whom log in every day, where they generate more than 900 million web pages and upload more than 250 million photos every day.

In 2010, a lifetime ago in Internet time, Google sites were used by more than 1 billion unique visitors every month who spent a collective 200 billion minutes on its sites. Google-owned YouTube passed 1 trillion video playbacks in 2011. Email, IM, VOIP calls, and other communications generate tens of trillions of recorded messages every year.

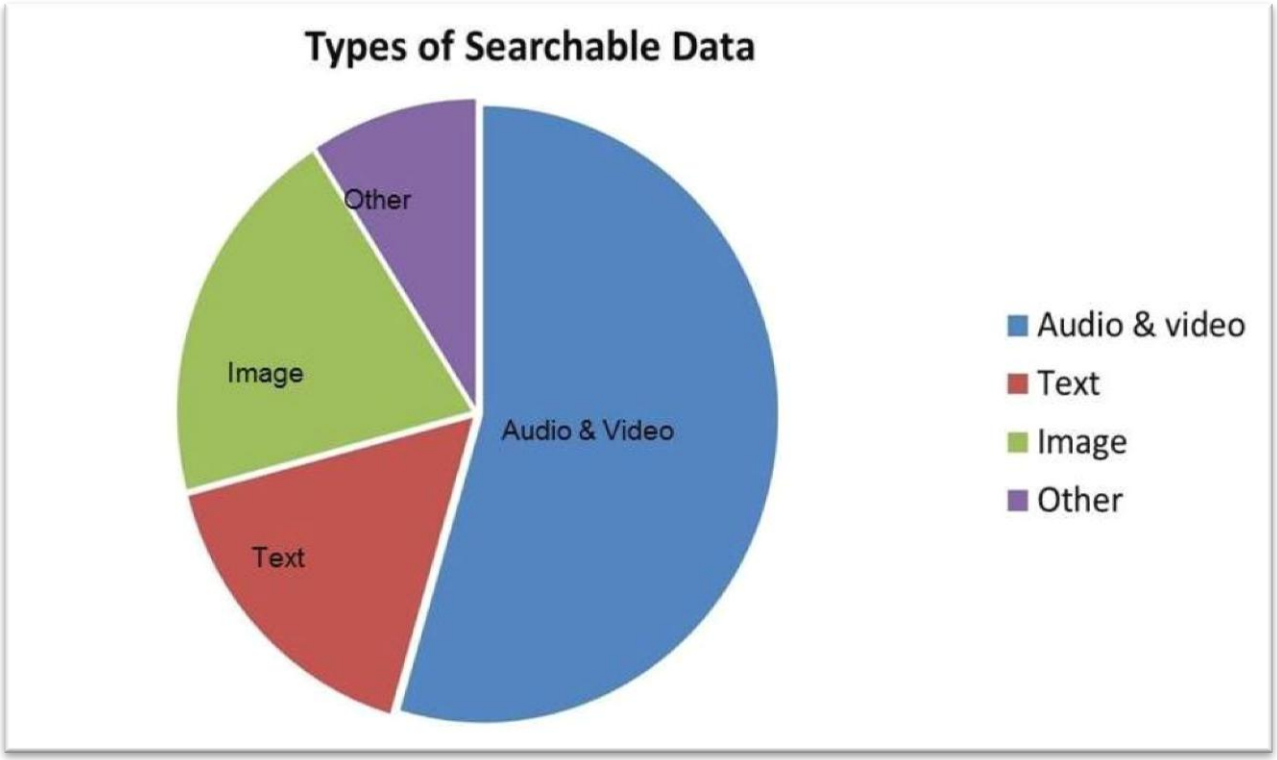

Figure 3: Types of Searchable Data in Internet and their ratio

People tend to visit the same sites frequently, so they don't realize how truly ginormous the Internet really is. It is so big, that even with the millions of sites that Google has indexed, as of 2010; Google had only indexed **.004%** of the data on the Internet. Google had indexed 200 terabytes of the 5 million terabytes of information that existed on the Internet.

The Cisco® Visual Networking Index (VNI) says that, Internet video & Audio will account for 61% of total Internet data by 2015.With the ever increasing amount of information being stored in audio and video formats, it is necessary to develop efficient methods for accurately extracting relevant information from these media with little or no manual intervention. [5]

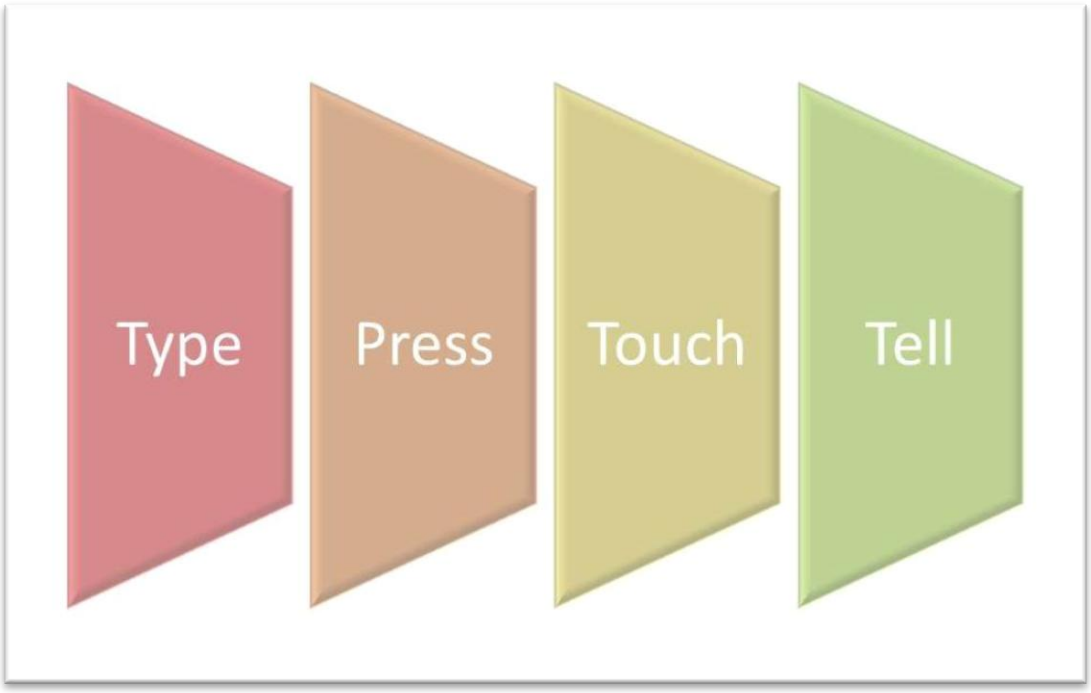

Figure 4:Evolution of Input Methods for Mobile Devices

The inclusion of a Qwerty makes for faster typing, and a touchscreen allows faster navigation. Users become more attuned to how well they can navigate through the screens and applications as a result, making for an overall efficient process. And now it is giving way to a new input method. Voice activation and gesture-activated commands leveraging a device's accelerometer are still developing, and have yet to reach the mainstream. Still, user input will remain a hot topic as vendors carve out new differentiators and experiences.

From the above discussion we can figure out following points:

1. Voice commands are going to be next big feature of any device. So we need better audio, video searching as **Query by Sample**.
2. Users would like to get more relevant results when they search for audio, video data. So we need better Audio Indexing mechanism.

Objective

To summarize the motivation & objective of this project, following points can be mentioned:

- Current web search engines generally do not enable searches into audio, video files. Informative metadata/tags would allow searches into audio, video files, but producing such metadata is a tedious manual task. So we need to come up efficient automatic indexing algorithm for audio and video files.
- The simplest content-based audio retrieval uses sample to sample comparison between the query and the stored audio pieces. This approach does not work because audio signals are variable and different audio pieces may be represented by different sampling rates and may use a different number of bits for each sample. Because of this, content based audio retrieval is commonly based on a set of extracted audio features, such as frequency distribution. We need to incorporate few more parameters (Acoustic information, categorize based on place of articulation of sounds, ambient noise) as metadata to enable better indexing.
- There are some very good indexing & retrieval technologies available for indexing music files. But not much for speech files.
- The main objective is to come up with an audio, video indexing & retrieval technique which incorporates unique & searchable index keys for each audio, video file along with the manually entered metadata. Index keys will be determined from various aspects of the input file and most relevant tags will be chosen to index the file.

Chapter 2:

Survey of Existing Works

Audio, video related indexing & searching technology can be broadly divided into following major categories.

- ❑ Speech Search
- ❑ Music Search
- ❑ Video Frame Search

Notable researches done in these fields are listed below:

Audio Video Search

Research Projects:

MUVIS

MUVIS is a framework for management (indexing, browsing, querying, summarization, etc.) of the multimedia collections such as audio/video clips and still images.

MUVIS is a collaborative framework that supports indexing, browsing and querying of various multimedia types such as audio, video, audio/video interlaced in several formats. It allows real-time audio and video capturing, encoding by last generation codecs such as MPEG-4, H.263+, MP3 and AAC. MUVIS also supports several audio/video file format such as AVI, MP4, MP3 and AAC. MUVIS achieves a global and unified solution for content-based indexing and retrieval problem and provides user-friendly applications and a generic framework especially for third parties to develop their feature extraction modules.

Rough'n'Ready

Rough 'n' Ready indexes speech data, creates a structural summarization, and provides tools for browsing the stored data. The Rough'n'Ready system has focused entirely on the linguistic content contained in the audio signal and, thereby, derives all of its information from the speech signal. This is a conscious choice designed to channel all development effort toward effective extraction, summarization, and display of information from audio. This gives Rough'n'Ready a unique capability when speech is the only knowledge source. Another salient feature of our system is that all of the speech and language technologies employed share a common statistical modeling paradigm that facilitates the integration of various knowledge sources.

Talk Miner

A new online video search tool launched this week makes it easier to search the content of video lectures by automatically transcribing words used in the lecturer's visual aids.

TalkMiner was created by researchers at Fuji Xerox Palo Alto Laboratory (FXPAL), in

California, to help students and professionals search the ever-expanding online archives of video lectures and presentations.

SpeechBot

SpeechBot is a public search system (SpeechBot, 1999) similar to AltaVista which indexes audio from public Web sites such as Broadcast.com, NPR.org, Pseudo.com, and InternetNews.com. The index is updated daily as new shows are archived in their Web sites.

Commercial Products:

Google Audio Indexing

In the 2008 presidential election race in the United States, the prospective candidates made extensive use of YouTube to post video material. Google developed a scalable system that transcribes this material and makes the content searchable (by indexing the meta-data and transcripts of the videos) and allows the user to navigate through the video material based on content. The system is available as an iGoogle gadget as well as a Labs product (labs.google.com/Gaudi). Given the large exposure, special emphasis was put on the scalability and reliability of the system.
Google's efforts to improve video search by using speech recognition technology started to become visible in July, when Google launched a gadget for searching inside the political speeches from YouTube. The gadget has been expanded into a new service called GAudi (Google Audio Indexing), which is now available at Google Labs.

Although Gaudi is removed from iGoogle as well as Google Labs, it was a great and innovative approach towards Audio Indexing and retrieval.

HP Autonomy

The main technology, 'Intelligent Data Operating Layer' (IDOL), allows search and processing of text taken from database, audio, video or text files or streams. The processing of such information by IDOL is referred to by Autonomy as Meaning-Based Computing.
Autonomy's technology attempts to "understand" any form of unstructured information, including text, voice, and video, and based on that understanding perform automatic operations, for example inferring what the user wants and on that basis finding other information that may be of interest.

Microsoft Audio Video Indexing Service (MAVIS)

The Microsoft Audio Video Indexing Service (MAVIS) is a Windows Azure application which uses state of the art speech recognition technology developed at Microsoft Research to enable searching of digitized spoken content, whether they are from meetings, conference calls, voice mails, presentations, online lectures, or even Internet video. A side benefit of MAVIS is the ability to generate automatic closed captions and keywords which can increase accessibility and discoverability of audio and video files with speech content. At this time MAVIS supports English speech content. MAVIS is now available as a commercial service through a subscription to **Greenbutton inCus**.

Music Search

Acoustic Fingerprinting

An **acoustic fingerprint** is a condensed digital summary, deterministically generated from an audio signal, which can be used to identify an audio sample or quickly locate similar items in an audio database.[1]

Practical uses of acoustic fingerprinting include identifying songs, melodies, tunes, or advertisements; sound effect library management; and video file identification. Media identification using acoustic fingerprints can be used to monitor the use of specific musical works and performances on radio broadcast, records, CDs and peer-to-peer networks. This identification has been used in copyright compliance, licensing, and other monetization schemes.

Shazam

Shazam have developed and commercially deployed a flexible audio search engine. The algorithm is noise and distortion resistant, computationally efficient, and massively scalable, capable of quickly identifying a short segment of music captured through a cellphone microphone in the presence of foreground voices and other dominant noise, and through voice codec compression, out of a database of over a million tracks. The algorithm uses a combinatorial hashed time-frequency constellation analysis of the audio, yielding unusual properties such as transparency, in which multiple tracks mixed together may each be identified.

Query By Humming

Query by humming (**QbH**) is a music retrieval system that branches off the original classification systems of title, artist, composer, and genre. It normally applies to songs or other music with a distinct single theme or melody. The system involves taking a user-hummed melody (inputquery) and comparing it to an existing database. The system then returns a ranked list of music closest to the input query.

One example of this would be a system involving a portable media player with a built-in microphone that allows for faster searching through media files.

The MPEG-7 standard includes provisions for QbH music searches.

SoundHound

SoundHound is revolutionizing the way people interact with mobile devices by delivering innovative technologies and compelling user experiences in sound recognition.

SoundHound's breakthrough Sound2Sound technology searches sound against sound, bypassing traditional sound to text conversion techniques even when searching text databases. Sound2Sound has resulted in numerous breakthroughs including the world's fastest music recognition, the world's only viable singing and humming search, and instant-response large scale speech recognition systems.

Sound2Sound (S2S) Search Science

Sound2Sound (S2S) Search Science is a revolutionary technology, created by

SoundHound, capable of recognizing various sound inputs including music and speech. It offers a breakthrough combination of speed and accuracy unattainable through traditional approaches to sound recognition.

S2S performs recognition by extracting features from the input signal and converting them to a compact and flexible Crystal representation. This Input Crystal is then matched against a database of Target Crystals which have been derived from searchable content.

Tunebot

Tunebot is a music search engine developed by the Interactive Audio Lab at Northwestern University. Users can search the database by humming or singing a melody into a microphone, playing the melody on a virtual keyboard, or by typing some of the lyrics. This allows users to finally identify that song that was stuck in their head.

Searching Techniques

Tunebot is a Query by humming system. It compares a sung query to a database of musical themes by using the intervals between each note. This allows a user to sing in a different key than the target recording and still produce a match. The intervals are also unquantized to allow for other tunings besides the standard A=440Hz, since not many people in the world have perfect pitch.

Query by Melody

Music Retrieval based on Melodic Similarity, which is one type of content-based retrieval, is proposed. Every fragment of a melody is represented as an appropriate N-dimensional vector, called a feature vector. A feature vector consists of values corresponding to intervals and rhythm of a melody. Depending on the distance between every fragment of a melody in the music database to the entered melody, k-nearest melodies can be obtained quickly using the SS-tree.

Musipedia

Musipedia offers three ways of searching: Based on the melodic contour, based on pitches and onset times, or based on the rhythm alone.

The melodic contour search uses an editing distance. Because of this, the search engine finds not only entries with exactly the contour that is entered as a query, but also the most similar ones among the contours that are not identical. Similarity is measured by determining the editing steps (inserting, deleting, or replacing a character) that are needed for converting the query contour into that of the search result. Since only the melodic contour is relevant, one can find melodies even if the key, rhythm, or the exact intervals are unknown.

The pitch and onset time-based search takes more properties of the melody into account.

The "query by tapping" method that only takes the rhythm into account uses the same algorithm as the pitch and onset time method, but assumes all pitches to be the same. As a result, the algorithm can be used for tapped queries that only contain onset times.

Motivation to our Approach

Our approach has following differentiating & unique factors which makes this worth implementing:

1. Integration of index keys available from three major components of a video/audio file
2. Index key determination algorithm(tag cloud) based on relevancy
3. Integration with YouTube and existing contents of video sharing websites around the web to be indexed in future.
4. Categorization of Index keys based on Acoustic information
5. Index keys with time frame
6. Retrieval from mobile device and web
7. Retrieval by time frame
8. Retrieval by video input

Chapter 3:

Architecture

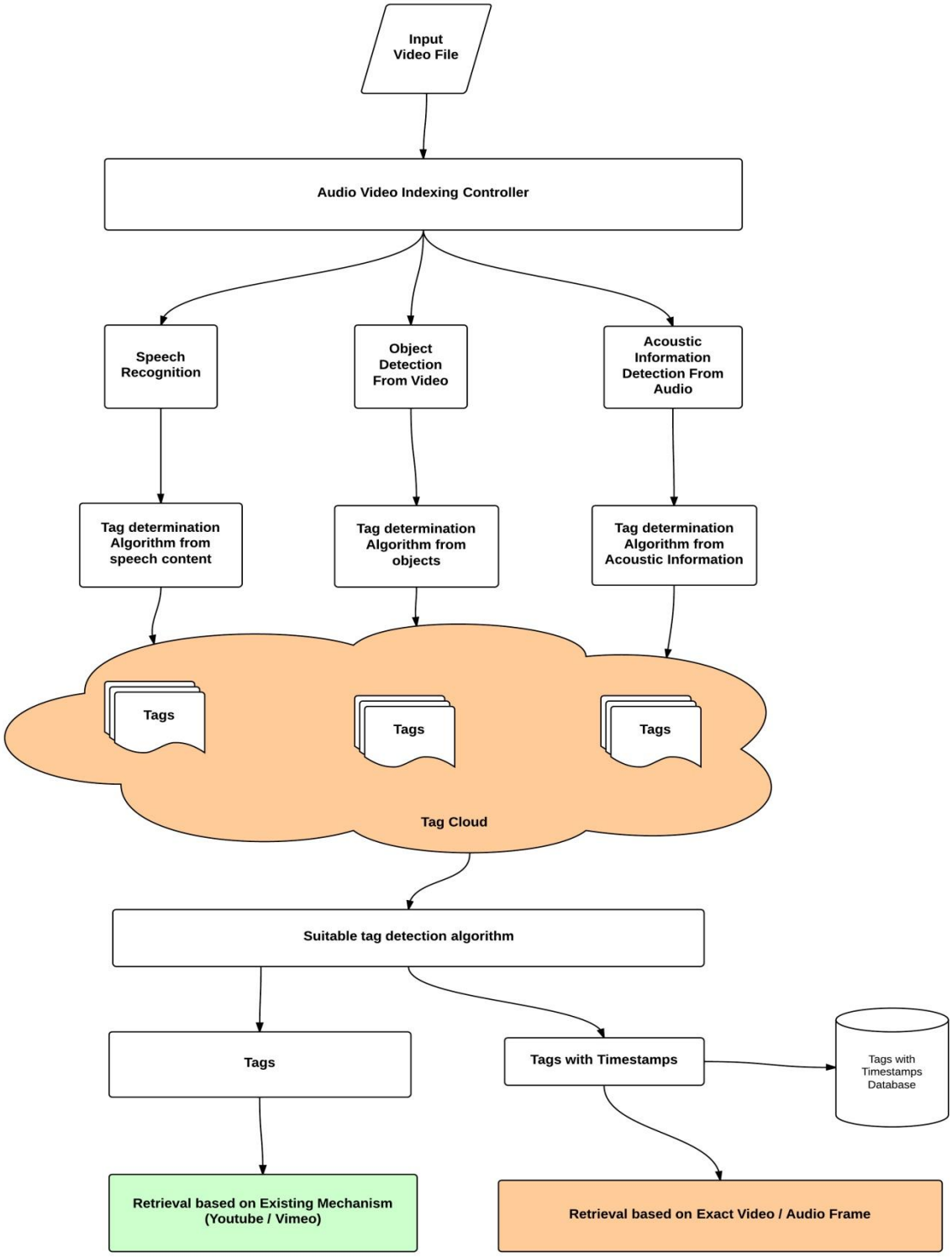

Figure 5: Overall Architecture

In this approach Audio Video indexing controller will extract following information from input file and route them corresponding processors.

- Speech Content → Speech Content Processor
- Image Object → Image Object Processor
- Acoustic Information → Acoustic Information processor

These processors will segregate the specific contents and pass through corresponding **Tag determination Module**.

Now tags from all sources will form a **Tag cloud** for the input file. Now we will run **Relevant Tags Detection Algorithm** to choose most suitable tags for the input file. Now finalized tags will be associated with the input file. Timestamps associated with tags will be stored in Timestamp database.

While audio retrieval there are two options available:

1. Retrieve based on tags from existing websites like YouTube , Vimeo
2. Retrieve based timestamp using our **tag with timestamp database.**

Algorithm

Speech Recognition from Audio Video

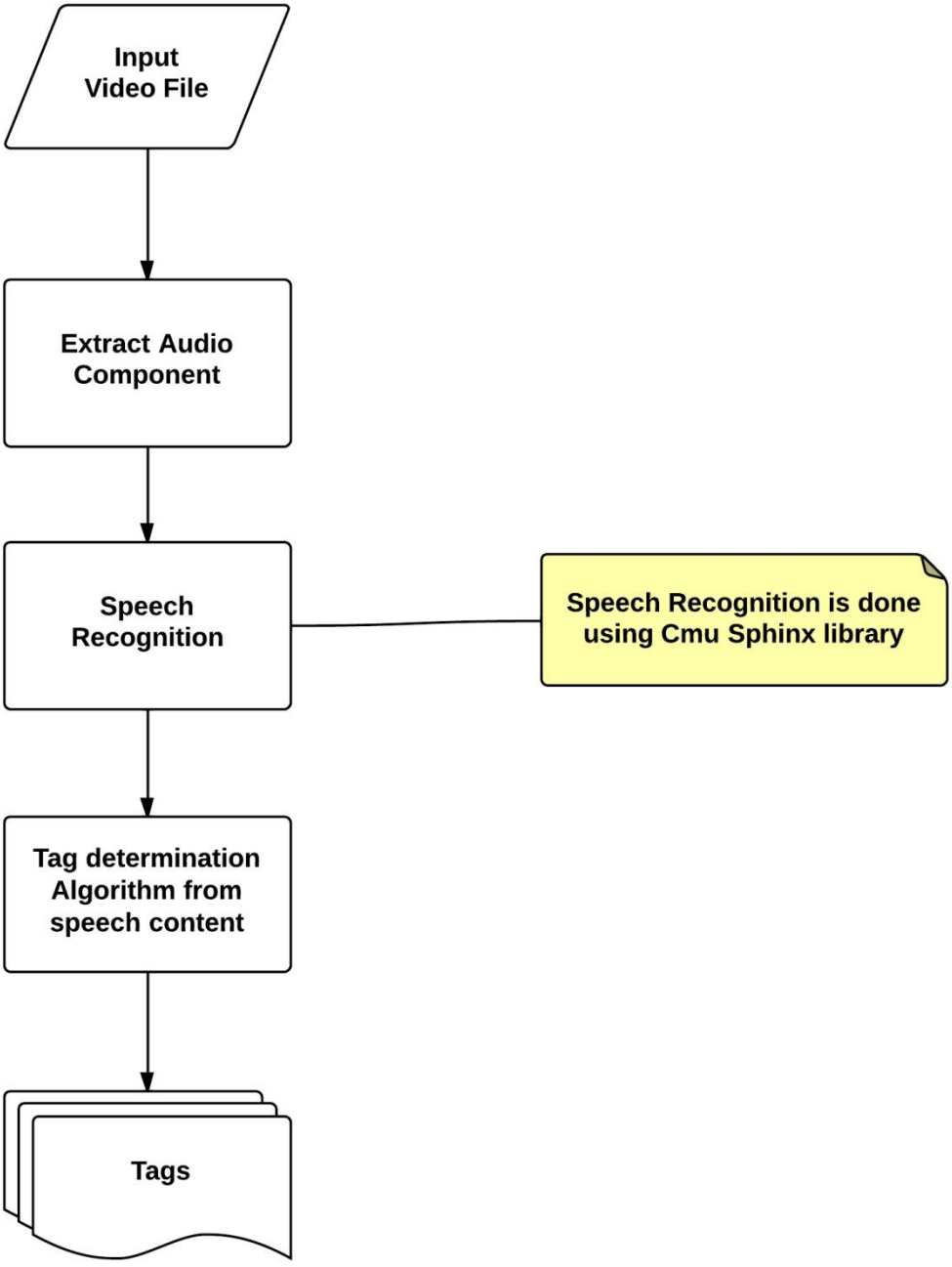

Figure 6: Speech Recognition from Audio, Video

Tag Determination Algorithm from recognized speech

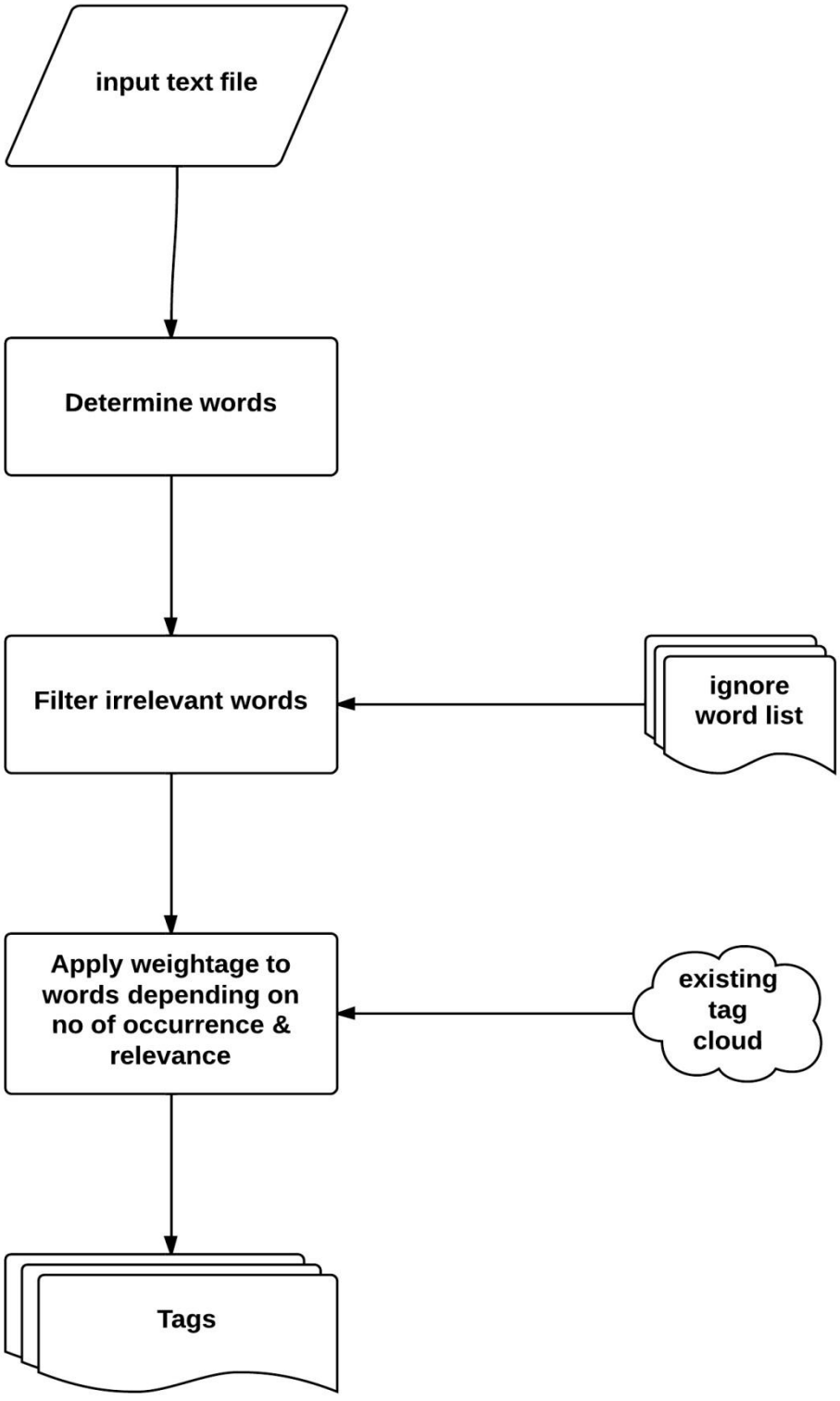

Figure 7: Tag detection from recognized speech

Object Detection from Video

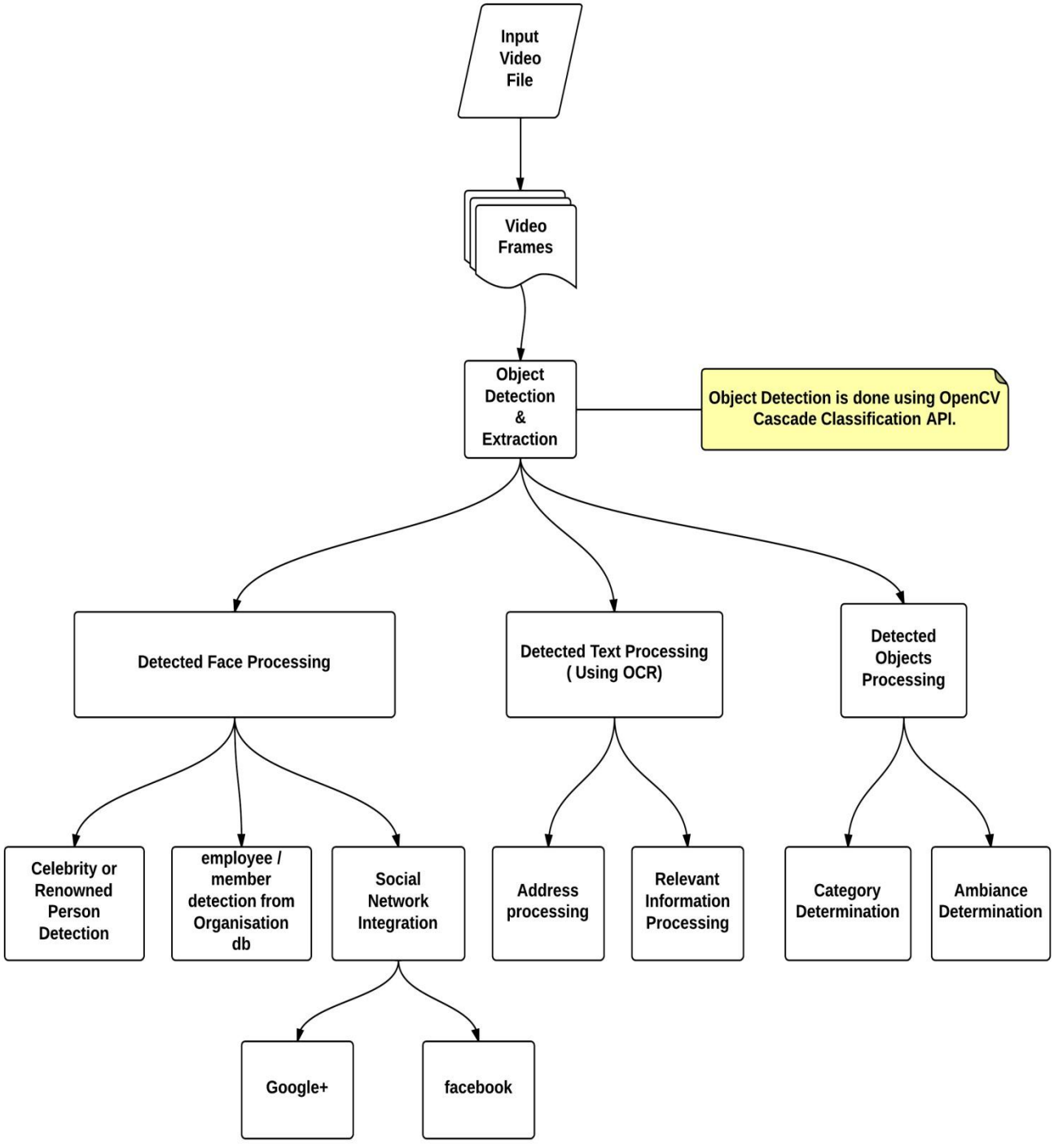

Figure 8: Object Detection from Video

Acoustic Information Detection from Audio

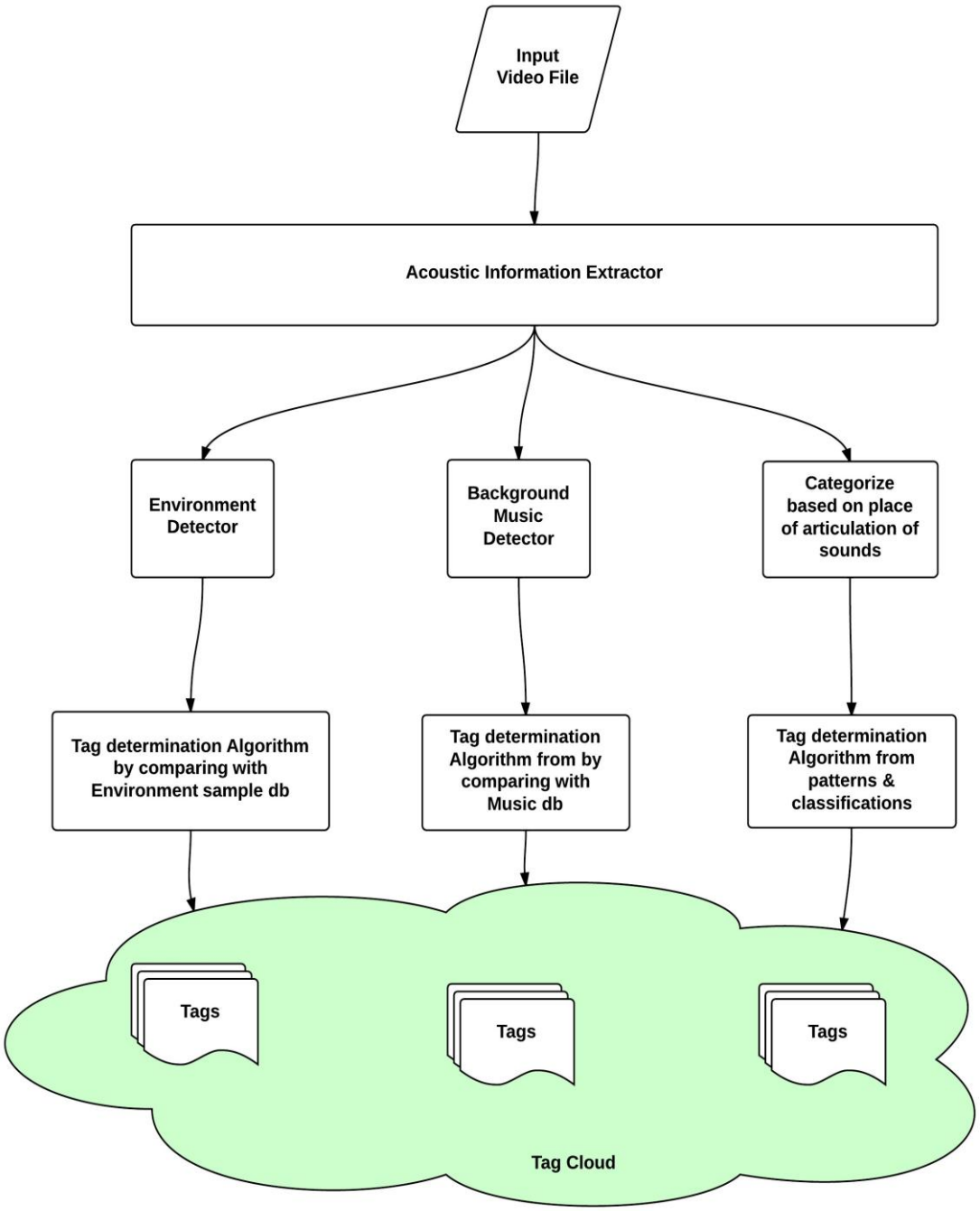

Figure 9: Acoustic Information Detection from Audio

Case Studies

Video Search within Organization

Problem

1. 2 days remaining for End Sem exam. Total time available for preparation is Maximum 30 hours.
2. 30 x 3 = 90 video lectures available, which will take 90 hours to play.
3. Video titles are corrupted somehow. So it is not possible to identify the important lectures quickly.

Solution

Indexing

1. Index the video files according to Speaker, Creation Time, Capture Location, Recognized Speech, Metadata etc.
2. Index videos by Image object , Image text, Background noise & Acoustic information
3. Create a tag cloud, determine suitable tags with file.

Searching

4. Retrieve According to Speaker : Ksrao, Rsc, Sc
5. Retrieve According to Creation Time & Sort.
6. Retrieve According to Capture Location
7. Retrieve According to Recognized Speech content
8. Retrieve According to User defined metadata / tag
9. Retrieve According to Acoustic information

Chapter 4:

Implementation Details

Video file access:

Video files can be accessed from existing websites like:

1. YouTube
2. Vimeo
3. Google Drive
4. DropBox

But these websites do not allow accessing raw video file using APIs. So we will have another option to upload video files from our Uploader. Our uploader will perform the audio indexing process before uploading to the websites and attach the index keys.

Audio Indexing Process - Incorporation of Searchable keys

I need to come up with a better algorithm for incorporating Unique & Searchable keys for each audio file which can be a combination of below mentioned metadata.

- **Audio Indexing metadata content & format**

  Audio data is more complex than text data. So we need to rely on metadata for more relevant search results. I will be researching on finding out more reliable metadata.

  a. Speaker Information

  b. Image information

  c. Timestamp

  d. Image Object Information

  e. Capture device information

  f. Capture type information(sampling rates, no of bits)

  g. Location information

  h. Acoustic information

  i. Background Music & Noise Information

  j. Recognized Speech content

  k. Language information

  l. User defined metadata / tag

Audio Retrieval

Based on the audio indexing keys and their combinations, I need to come up with efficient Audio Retrieval algorithm.

a. Retrieve with user preferred index keys

b. Sequence / Combination of Index Keys

c. Retrieve from Audio Sample

d. Timestamp based video frame retrieval

Work Done So Far

Video Access & Manual Indexing Prototype

I have created a working demo of "Audio Indexing and Retrieval project" in http://shabdo.technicise.com/php/ . It is a Google App for YouTube.. You can login using your google id. It will list down all YouTube videos uploaded by you and will display relevant information about the videos. You can add tags depending on the audio contents.

As per our earlier discussion, I was working on Audio Indexing portion during 5th Sem and will be working on retrieval portion and integration on next sem.

I have worked on the following work items:

1. Sample Collection:
   - YouTube is widely used around the web for audio & video.. YouTube has 43.8% videos of whole web. That's why I have created the first prototype on YouTube.
   - I have samples on my laptop which i am using on windows desktop application
   - I have samples uploaded on Google Drive which i will use from Android app in future
2. Created a Google YouTube Web app using Google API V3
3. List down videos with its name, id and index keys
4. Able to add new index keys in the YouTube videos from my app.
5. Extract Audio based content from the video and attach it as index keys in YouTube video. So users can search using the new index keys
   - Recognized Speech Content
   - Author Information
   - Location Information
   - Language Information
   - Time stamp

End to end flow implemented in 5th Semester

Developed a YouTube App → Authenticate using Google Open Id → Access List of videos available in user's YouTube Account → Access Video information like Tags, Video Id, Thumbnail, Upload time, Title, Description → Manually Add tags to the existing YouTube video.

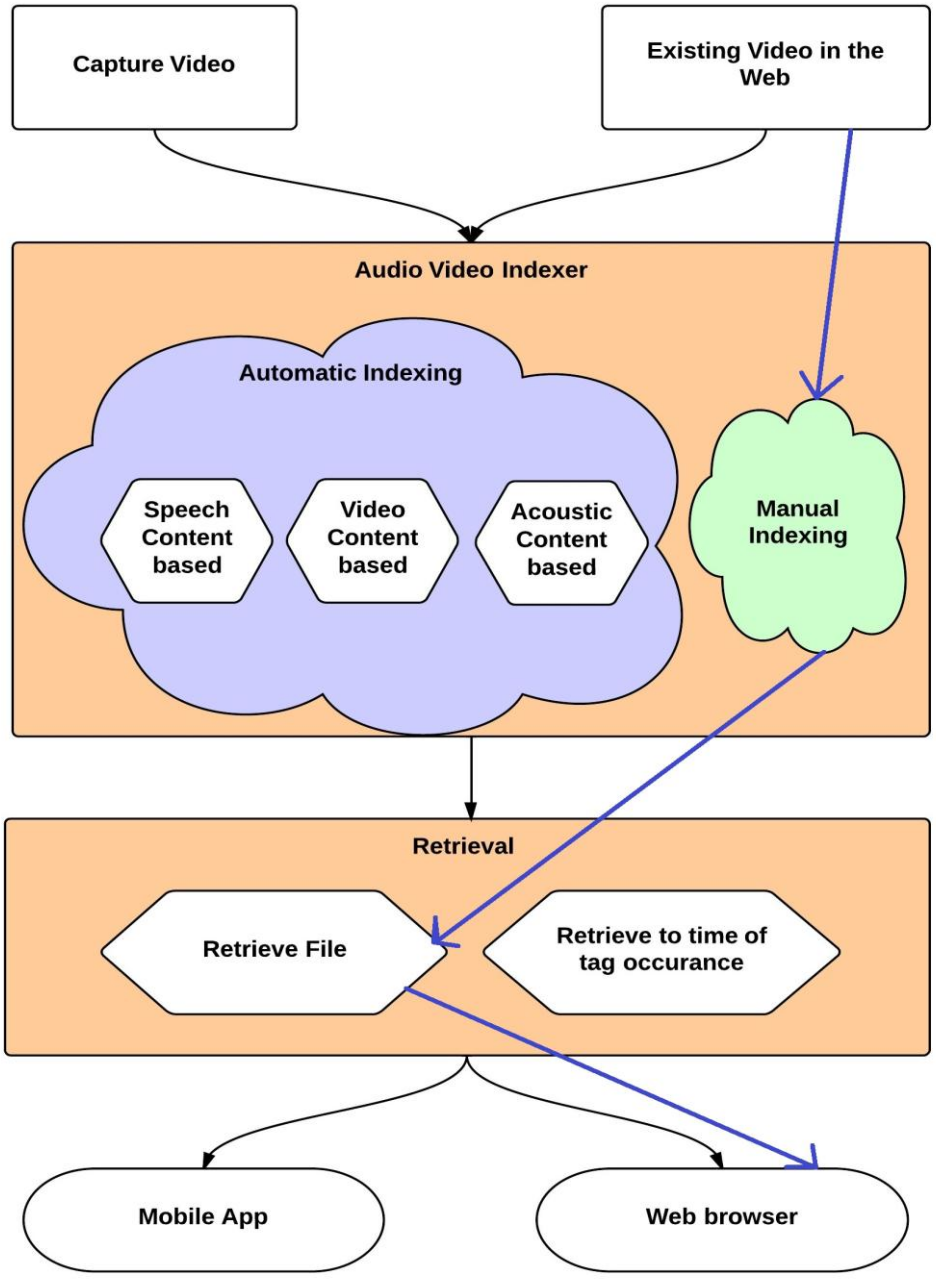

Figure 10: End to end flow implemented in 5th Sem

Implemented flow is marked by violet arrows.

Web Client for Indexing YouTube Videos

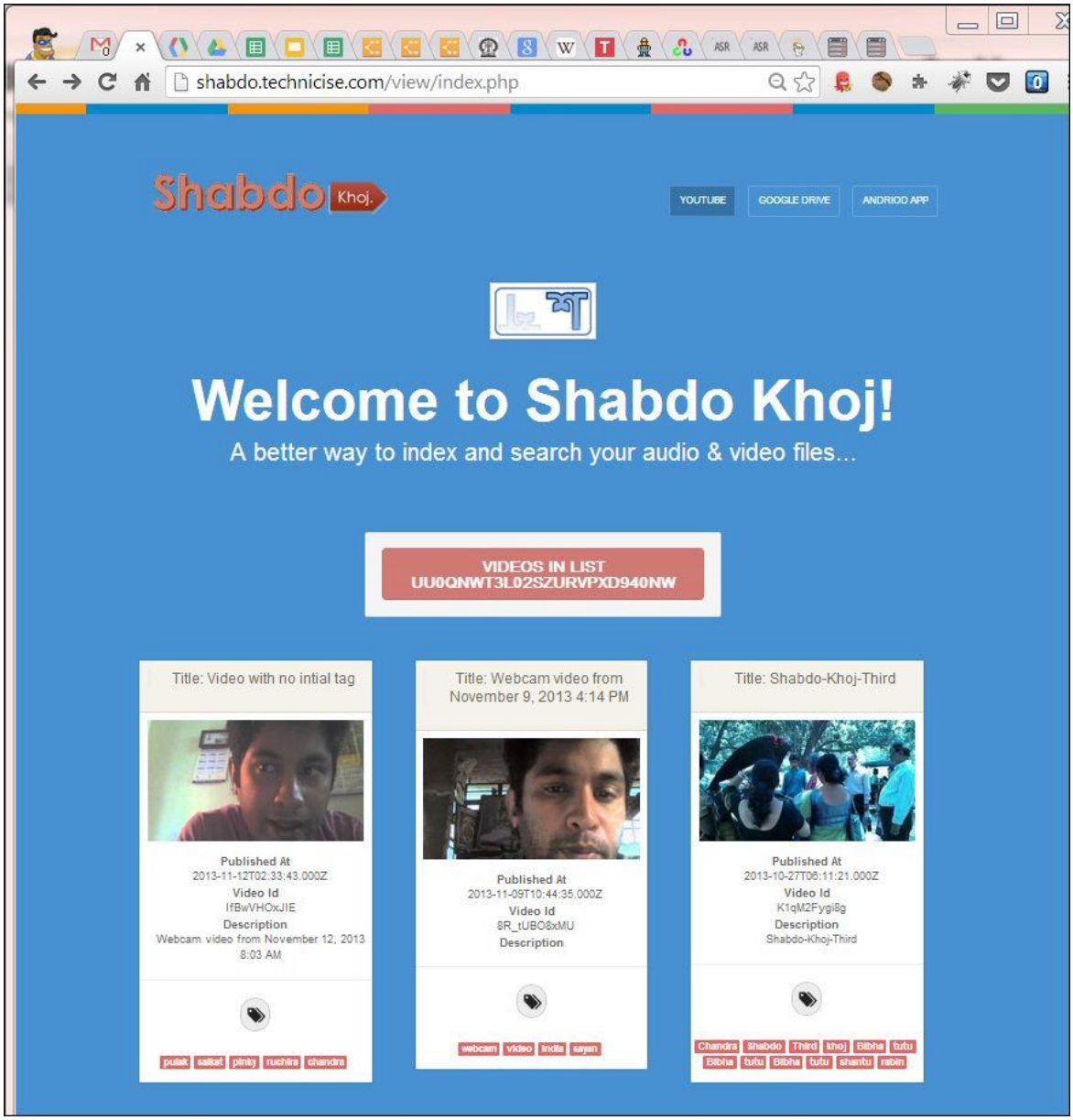

Figure 11: Web Client for Indexing YouTube Videos

Algorithms & Workflow

- Algorithm for determining index keys out of Recognized Speech Content
- Workflow of Speech content & Image Object extraction
- Overall framework for indexing audios/videos

Work In Progress:

Image Object extraction

I am using OpenCV framework to detect image objects from video frames

Clients for Accessing Videos

- Android App with above mentioned features
- Desktop Uploader App with above mentioned features

Research newer index keys

- Speaker Information as Index key
- Acoustic information as index keys

Work to be done in 6th Sem

1. Use CMU Sphinx to extract speech content
2. One flow of Video Content Based Automatic Indexing(Extract Image Objects from video and tag celebrities present in a video)
3. One flow of Acoustic information Based Automatic Indexing
4. End to end Desktop Uploader App and Enhance Web Client for Indexing
5. Retrieve to time of tag occurrence

Future Scope

1. Integration With Dropbox & other video sharing website
2. Better retrieval process

Conclusions

In this approach, we are extracting information about the video and audio content of the file from various aspects. Then we are determining most suitable tags and attaching them with file so that we can search & retrieve this file based on those tags. This initiative will provide better searching options for video and audio file. Implementation this approach will provide an extendible framework for Audio, Video Indexing which can be enhanced & customized by developers and users respectively.

References

[1] http://allthingsd.com/20110601/cisco-the-internet-is-like-really-big-and-getting-bigger/

[2] http://www.omg-facts.com/Technology/In-2010-Google-Had-Only-Indexed-004-Of-T/52586

[3] GUOJUN LU , "Indexing and Retrieval of Audio: A Survey"

[4] Christopher Kuner, Fred H. Cate, Christopher Millard and Dan Jerker B. Svantesson , "The challenge of 'big data' for data protection", http://idpl.oxfordjournals.org/content/2/2/47.extract

[5]http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.html

[6] Websites:

- http://en.wikipedia.org/wiki/Query_by_humming
- http://www.midomi.com/
- http://www.soundhound.com/
- http://en.wikipedia.org/wiki/Tunebot
- http://en.wikipedia.org/wiki/Musipedia
- http://www.kecl.ntt.co.jp/csl/sirg/people/yasushi/SoundCompass.pdf
- http://en.wikipedia.org/wiki/Parsons_code
- http://music.cs.northwestern.edu/index.php
- http://tunebot.cs.northwestern.edu/index.php
- http://www.musipedia.org/
- http://en.wikipedia.org/wiki/Acoustic_fingerprint
- http://en.wikipedia.org/wiki/Shazam_(service)
- http://www.speech.kth.se/wavesurfer/
- http://audacity.sourceforge.net/
- http://www.hongkiat.com/blog/25-free-digital-audio-editors/
- https://incus.greenbutton.com/
- http://maart.sourceforge.net/
- https://sourceforge.net/projects/camel-framework/
- https://sourceforge.net/projects/jaudio/
- http://www.speech.kth.se/snack/
- http://research.microsoft.com/en-us/projects/mavis/
- http://googlesystem.blogspot.in/2008/09/google-audio-indexing.html

- http://en.wikipedia.org/wiki/Google_Audio_Indexing
- http://googleblog.blogspot.in/2008/09/google-audio-indexing-now-on-google.html
- http://www.autonomy.com/content/Technology/evolution/evolution-of-search/index.en.html
- http://techcrunch.com/2008/09/17/google-launches-audio-indexing/
- http://en.wikipedia.org/wiki/Multimedia_search
- http://en.wikipedia.org/wiki/Audio_search_engine
- http://www.playaudiovideo.com/pav_start.htm
- http://www.ramp.com/
- http://en.wikipedia.org/wiki/Google_Voice_Search
- http://music.cs.northwestern.edu/index.php
- http://www.aurix.com/pages/3417/Technology.htm
- http://www.apple.com/ios/siri/
- http://en.wikipedia.org/wiki/Siri_(software)
- http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.html
- http://mashable.com/2012/03/06/one-day-internet-data-traffic/
- http://hadoop-karma.blogspot.in/2010/03/how-much-data-is-generated-on-internet.html
- http://www.scientificpsychic.com/workbook/chapter2.htm
- http://www.omg-facts.com/Technology/In-2010-Google-Had-Only-Indexed-004-Of-T/52586
- http://googleblog.blogspot.in/2008/07/we-knew-web-was-big.html
- http://idpl.oxfordjournals.org/content/2/2/47.extract
- http://www.internetworldstats.com/stats7.htm
- http://www.fiercewireless.com/story/idc-future-mobile-device-user-input/2010-05-03
- http://www.audiblemagic.com/
- http://www.neurostechnology.com/
- http://www.linkedin.com/company/musiwave
- http://www.crunchbase.com/company/musiwave
- http://mashable.com/2007/11/15/musiwave-microsoft-deal/
- http://www.microsoft.com/en-us/news/press/2007/nov07/11-12ProjectTunesPR.aspx

- http://www.lumenvox.com/resources/tips/historyOfSpeechRecognition.aspx
- http://www.pcworld.com/article/243060/speech_recognition_through_the_decades_how_we_ended_up_with_siri.html?page=2
- https://www.dropbox.com/developers/chooser
- http://www.dropboxwiki.com/Main_Page
- http://support.google.com/drive/bin/answer.py?hl=en&answer=2500820
- http://infolab.stanford.edu/~backrub/google.html
- http://www.google.co.in/intl/en/insidesearch/howsearchworks/thestory/
- http://www.google.co.in/intl/en/insidesearch/howsearchworks/algorithms.html
- http://www.wired.com/insights/2013/10/google-hummingbird-where-no-search-has-gone-before/