Audio Video Indexing And Retrieval

By
Chandra Shekhar Sengupta
Roll - 11IT61K14,
MTech (ICT)

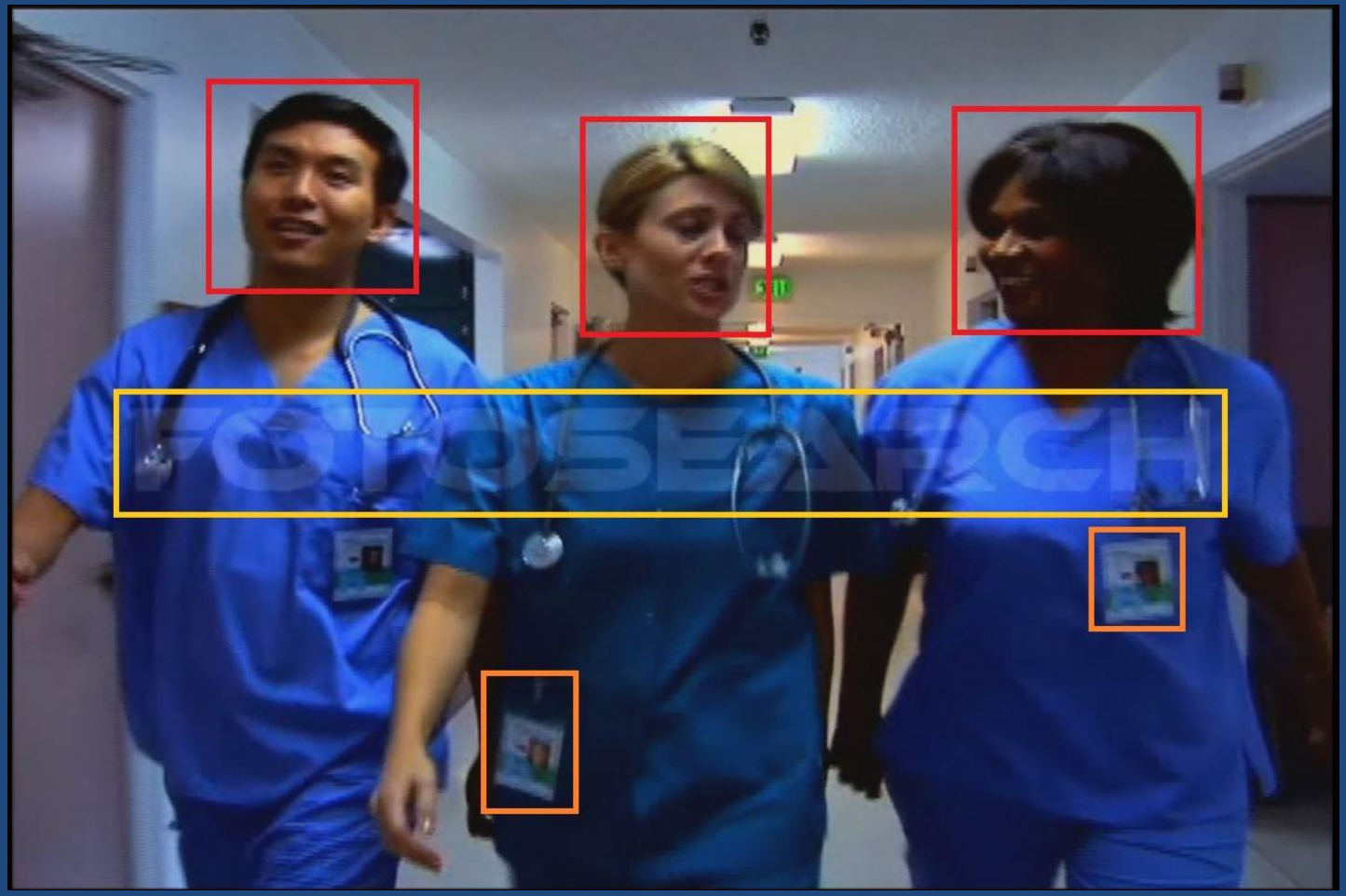
Under the guidance of

Prof. K. S. Rao

School of Information Technology,
INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

Case Study 1

Image Information

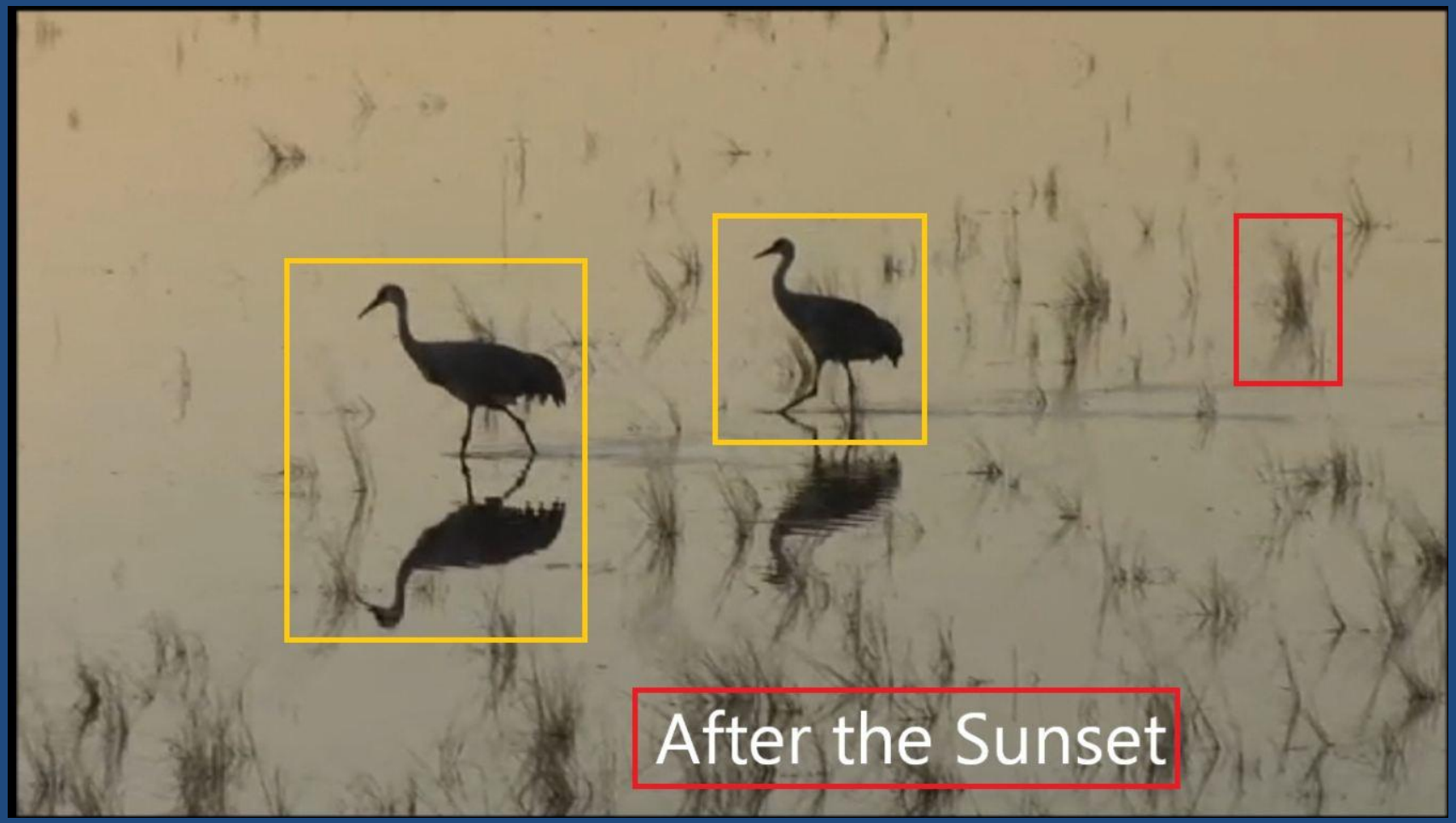


Indexing Information

- Acoustic Information → Background Music
- Image Information →
  - Face Detection : 3 person
  - OCR Text : fotosearch
  - Other objects : ID card
- Speech Information → none

Case Study 2

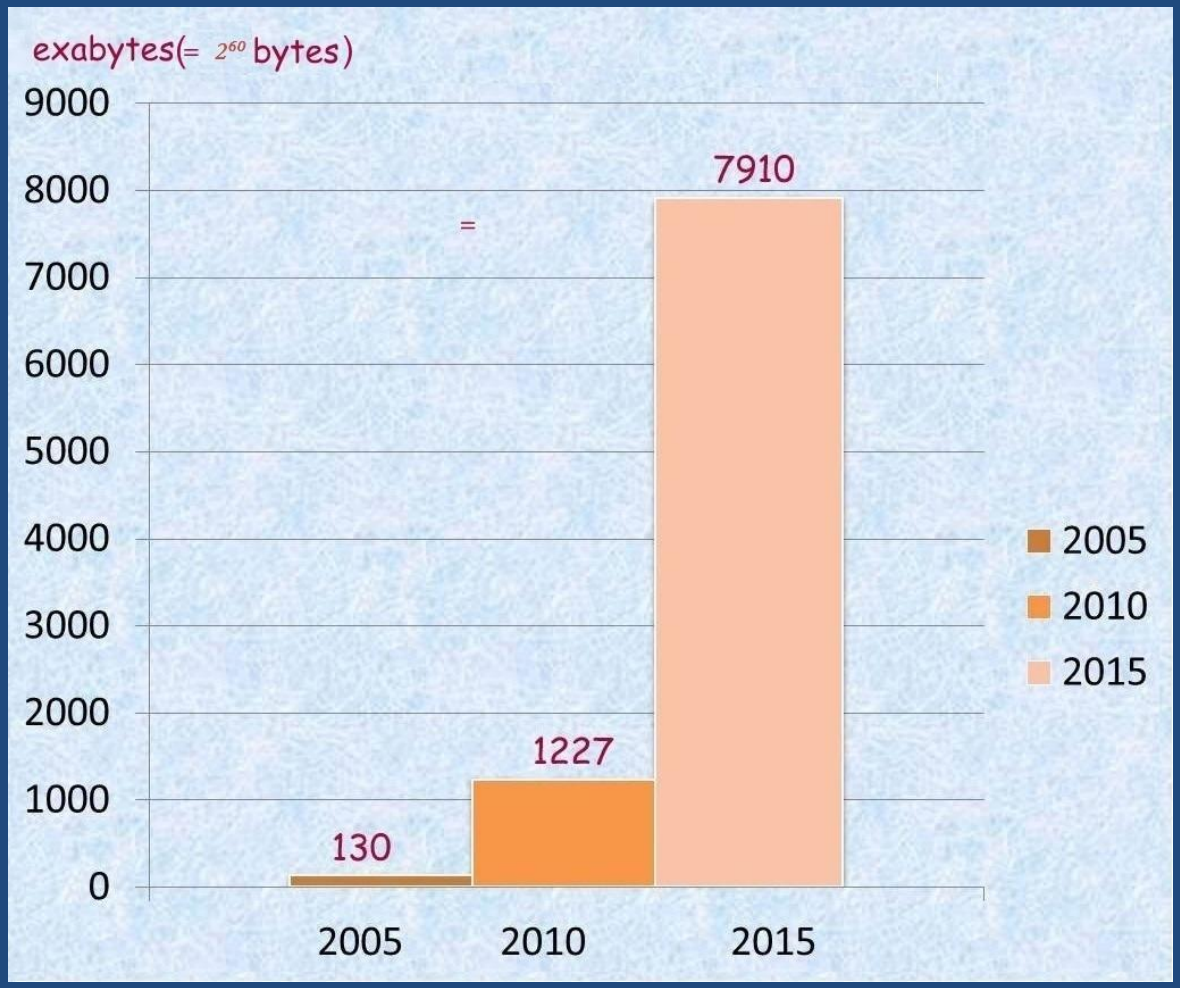
Image Information



Indexing Information

- Acoustic Information → Background Sound
- Image Information →
  - Face Detection : none
  - OCR Text : After the sunset
  - Other objects : Cranes, Grass, Reflection, Water
- Speech Information → none

Global Data Volume



It would take **13,513** planes (Boeing 747 aircraft) to transport one exabyte of data if we store data in DVDs.

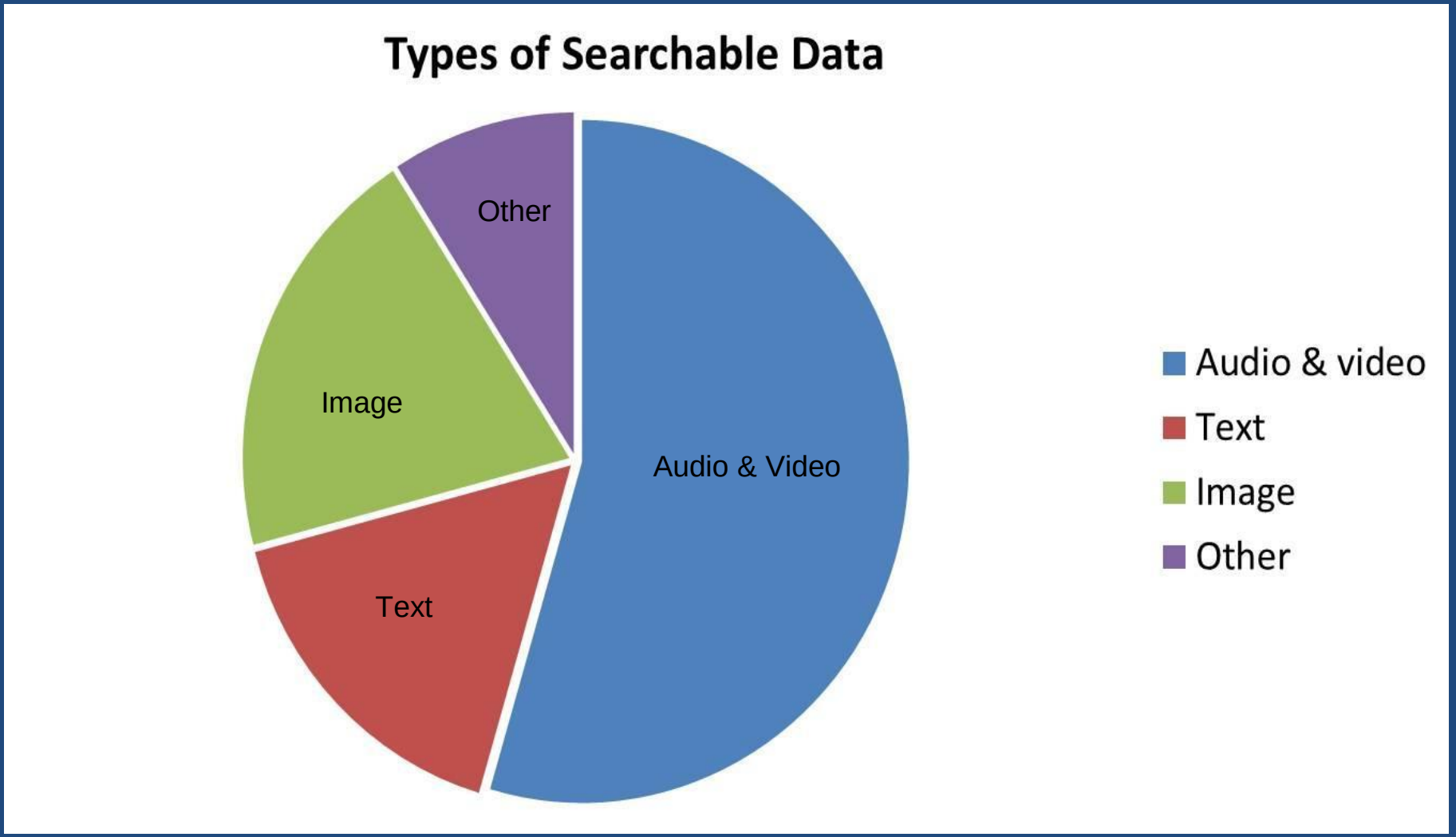
Few Facts about data on Internet

1. Using DVDs to move the data collected globally in 2010 would require a fleet of more than 16 million jumbo jets.

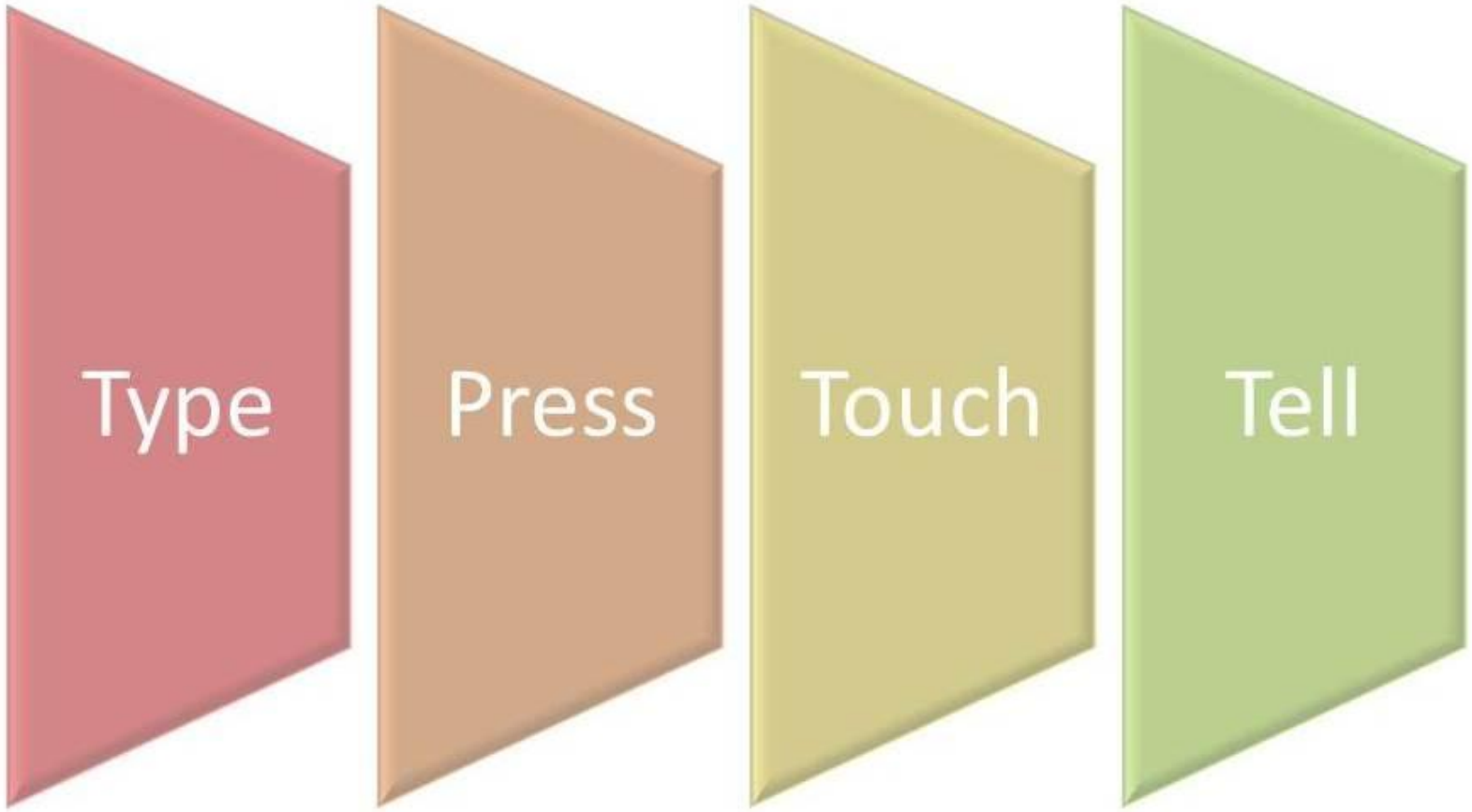
2. Internet video & Audio will account for 61% of total Internet data by 2015.

3. In 2010, Google had only indexed .004% of the data on the internet.

Ratio of Searchable Data



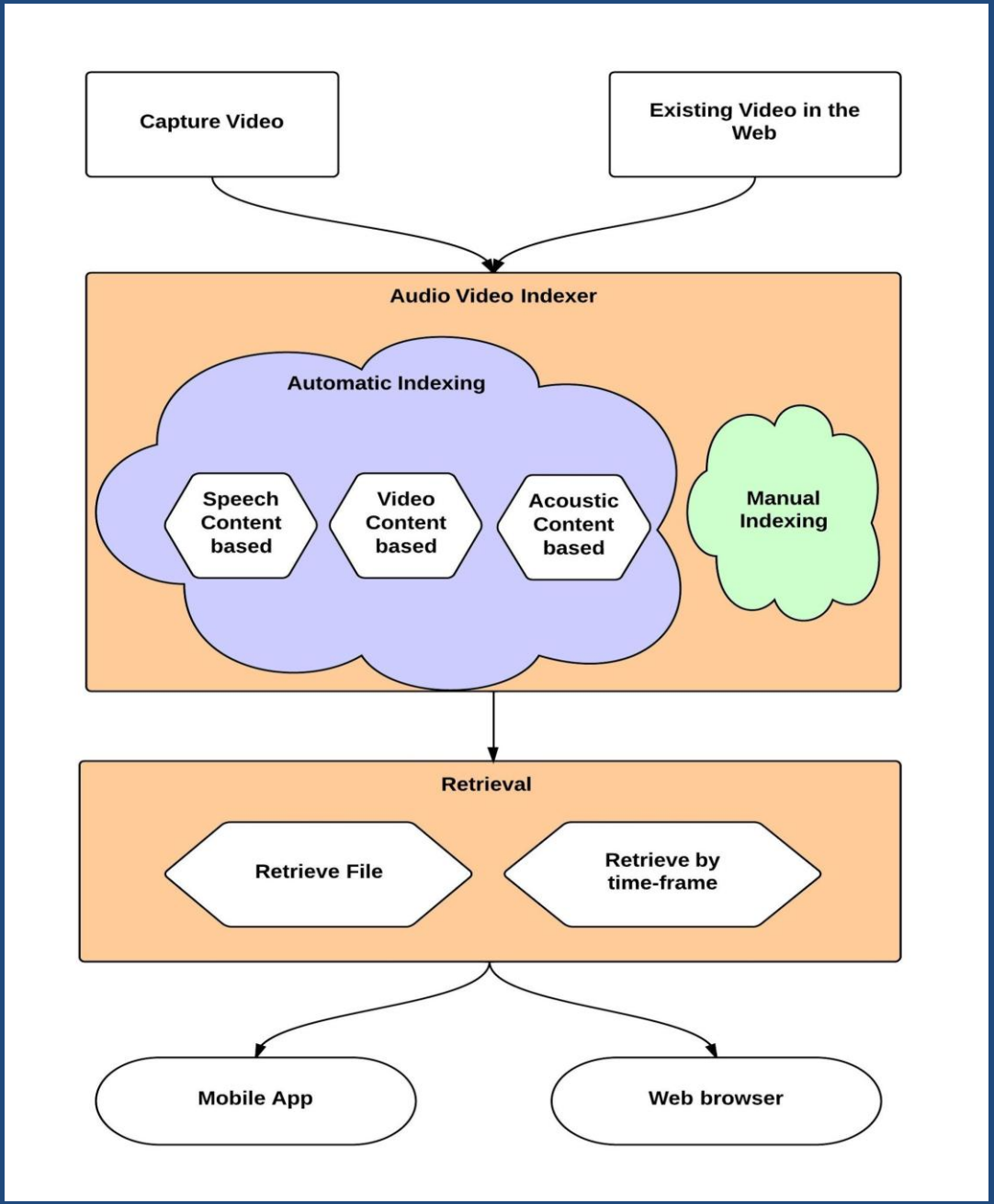
Input method for Devices



Motivation & Objectives

1. Current web search engines do not searches into audio, video files.
2. Manually entered tag may not be relevant
3. Huge amount of data is not searchable yet.
4. To come up with an audio,video indexing & retrieval technique which indexes relevant information out of audio, video file .

Our Approach



Survey of Existing work

Audio Indexing and Retrieval Techniques

1. Informative metadata / tags based Audio Indexing System
2. Speech Recognition based Audio Indexing System
3. Content-Based Audio Indexing and Retrieval
4. Vector-Based Audio Indexing and Retrieval

Research Projects on Speech Indexing & Retrieval

1. **MUVIS:** A framework for management (indexing, browsing, querying, summarisation, etc.) of the multimedia collections such as audio/video clips and still images.
2. **Rough'n'Ready**: which indexes speech data, creates a structural summarization, and provides tools for browsing the stored data.
3. **SpeechBot**: a Speech Recognition based Audio Indexing System for the Web.

Commercially Available

Audio Data Indexing & Retrieval

Google Audio Indexing

Microsoft

Voice Data Search

Apple

Siri. Beta

Your wish is its command.

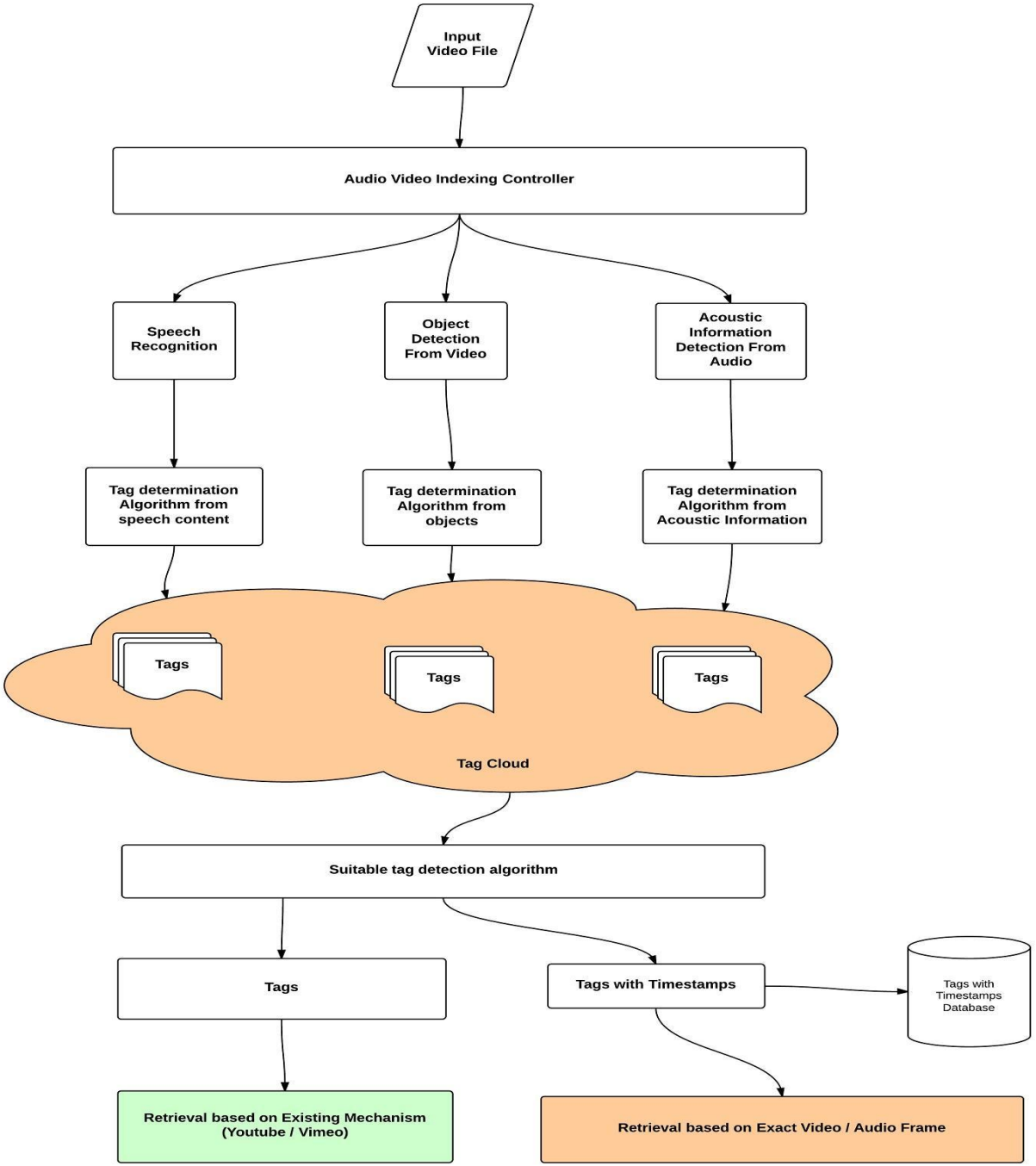
Vlingo

Voice Search

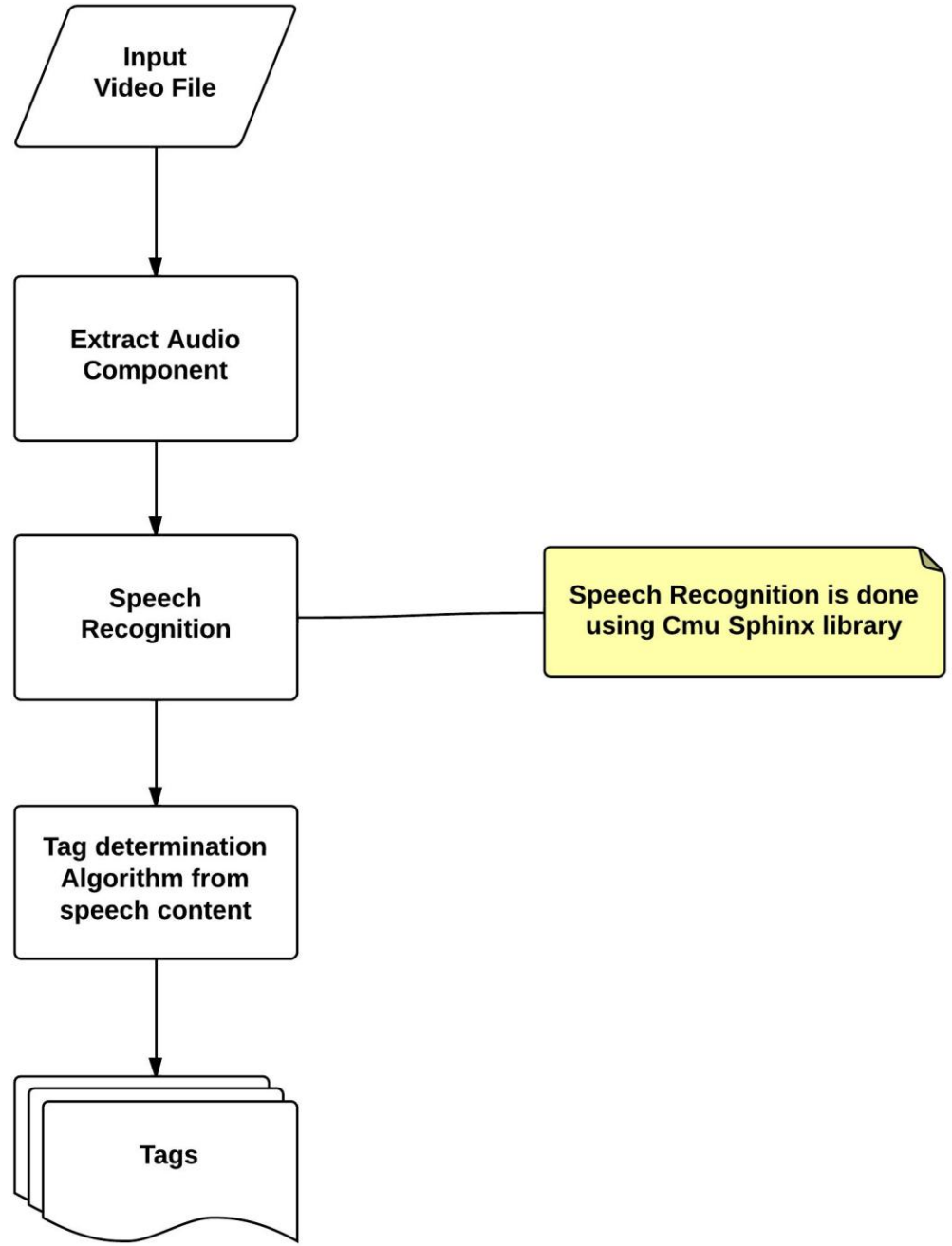
Differentiating factor of Our Approach

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 video sharing websites to be able to index existing contents around the web
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

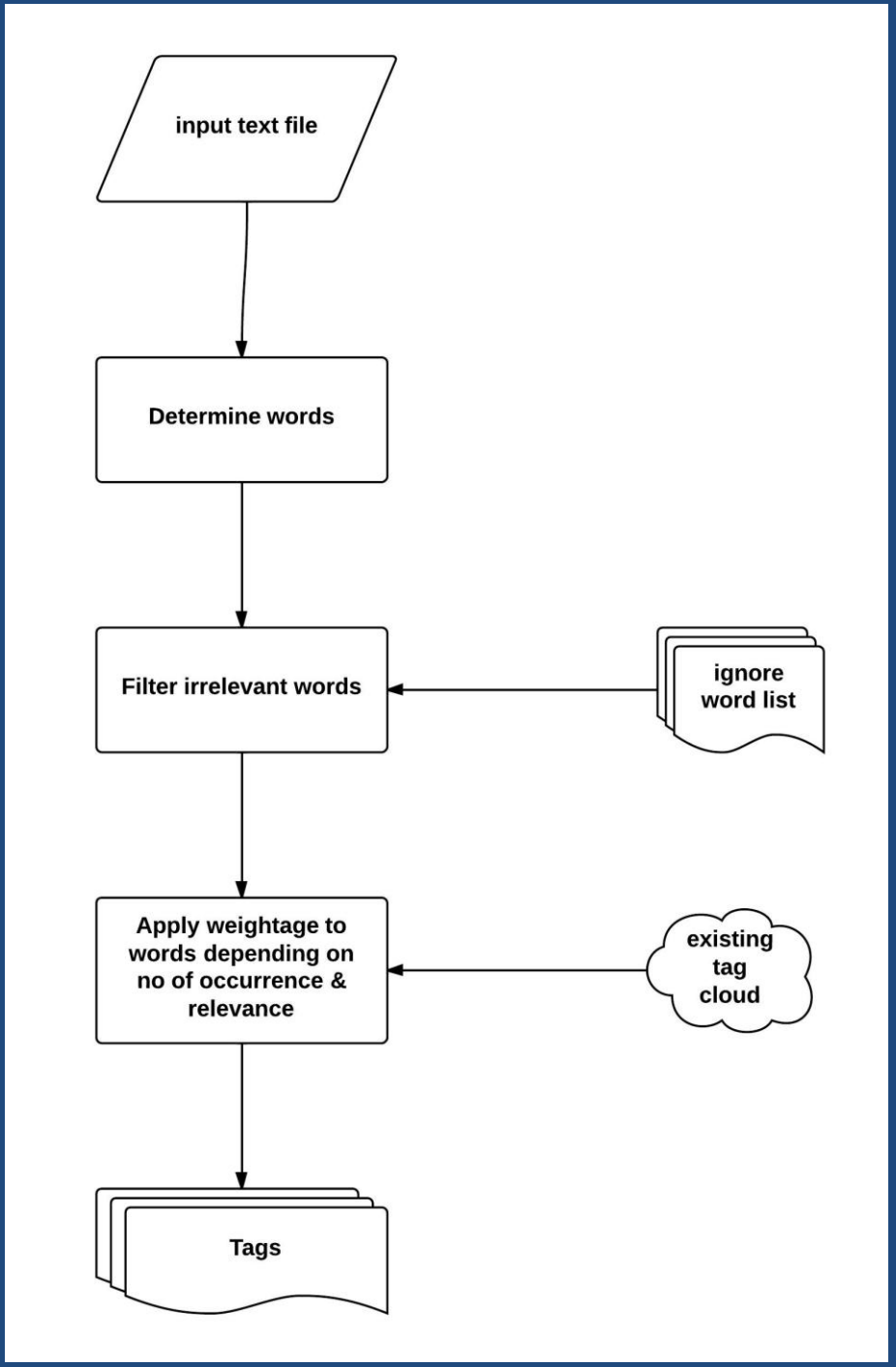
Overall Architecture



Speech Recognition



Tag Determination from Recognized Speech



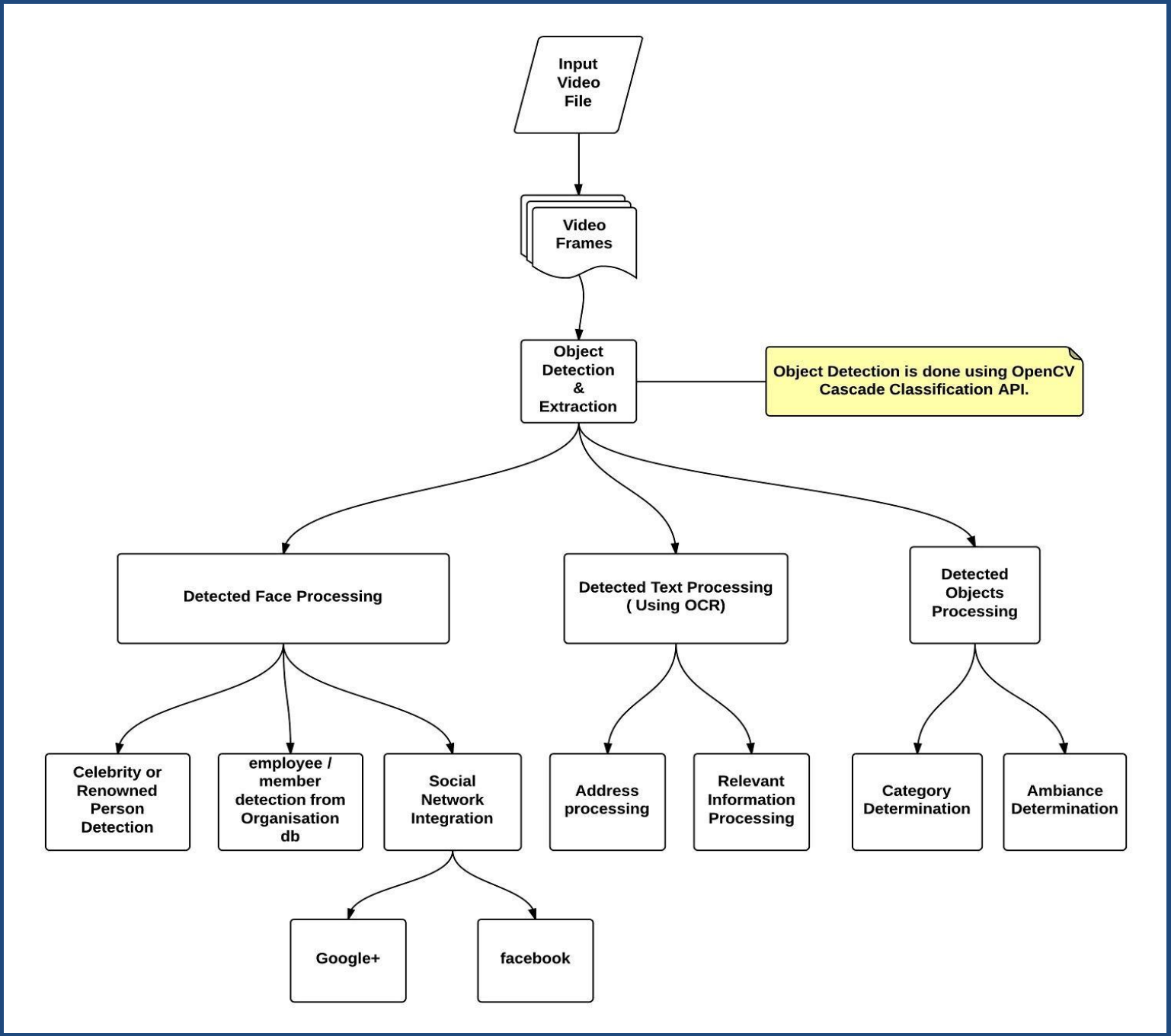
Object

Detect

From

Video

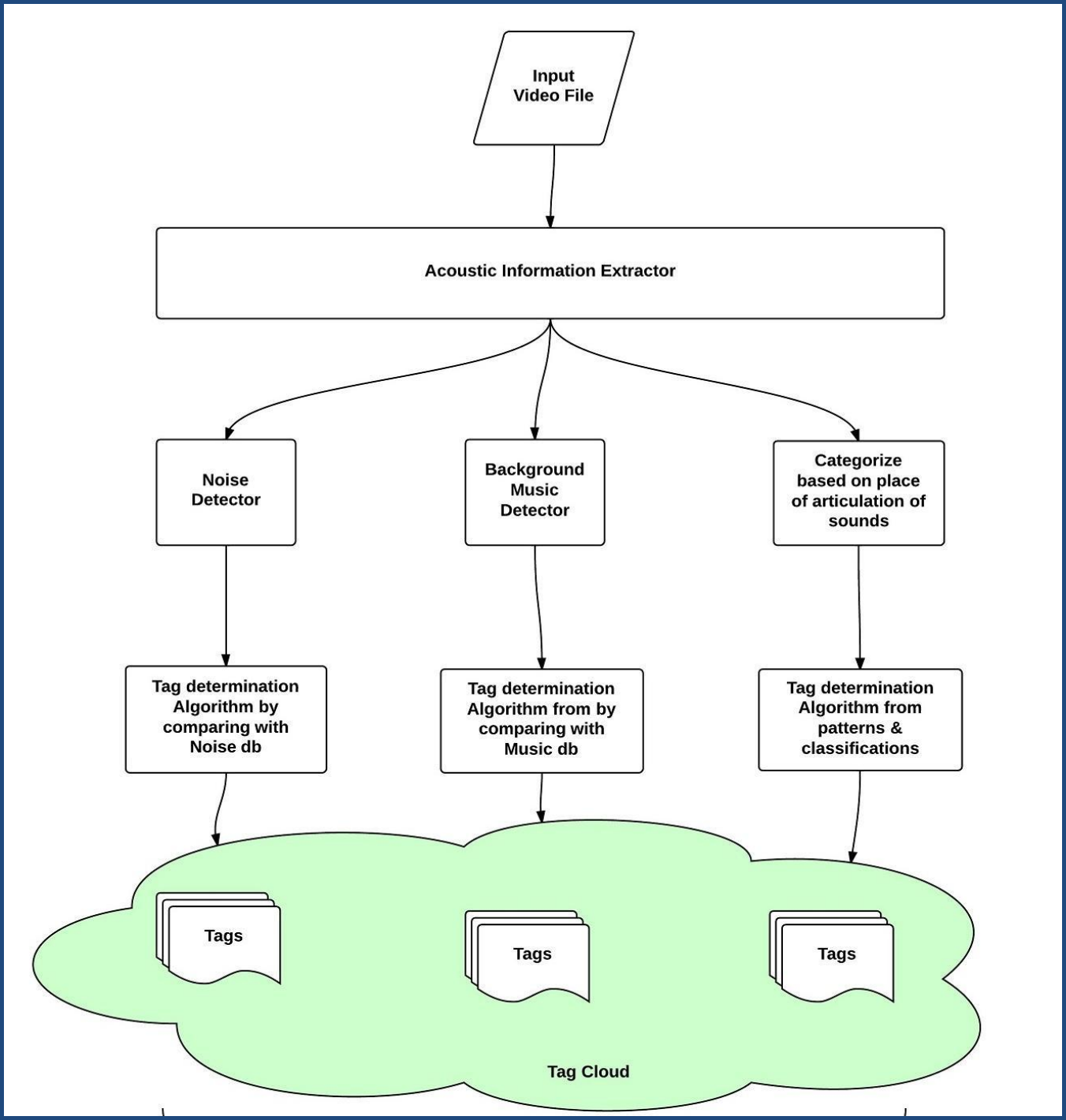
Frames



Acoustic

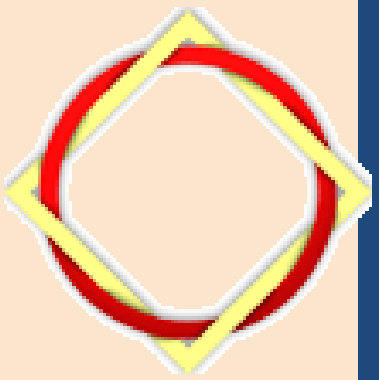
Information

Detect



Tools

MaART

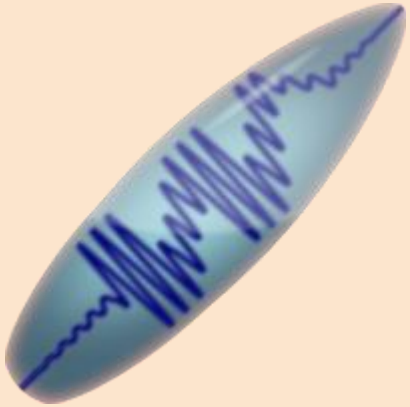


Audacity

InCus

inCus

CAMEL



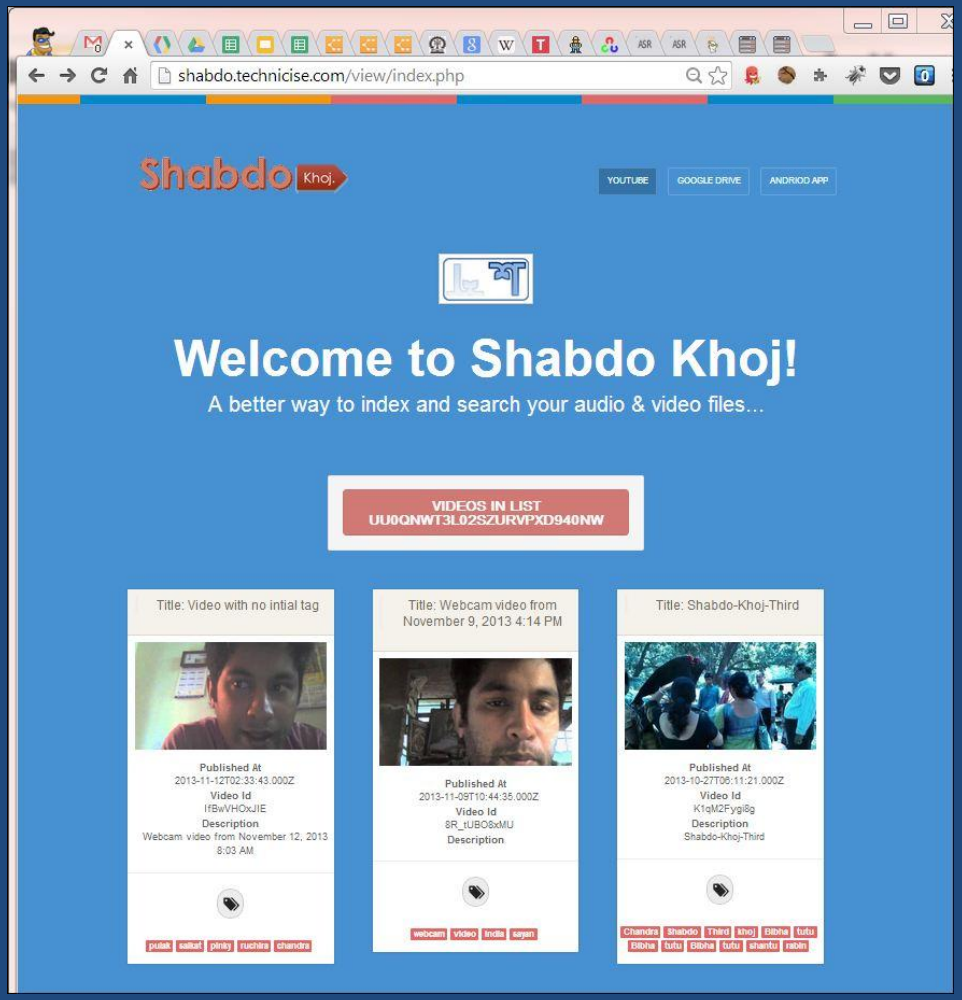
jAudio

Platforms

1. OpenCV
2. CMU Sphinx
3. MUVIS
4. Microsoft Audio Video Indexing Service (MAVIS)
5. HP Autonomy
6. Google Audio Indexing
7. Snack Sound Toolkit

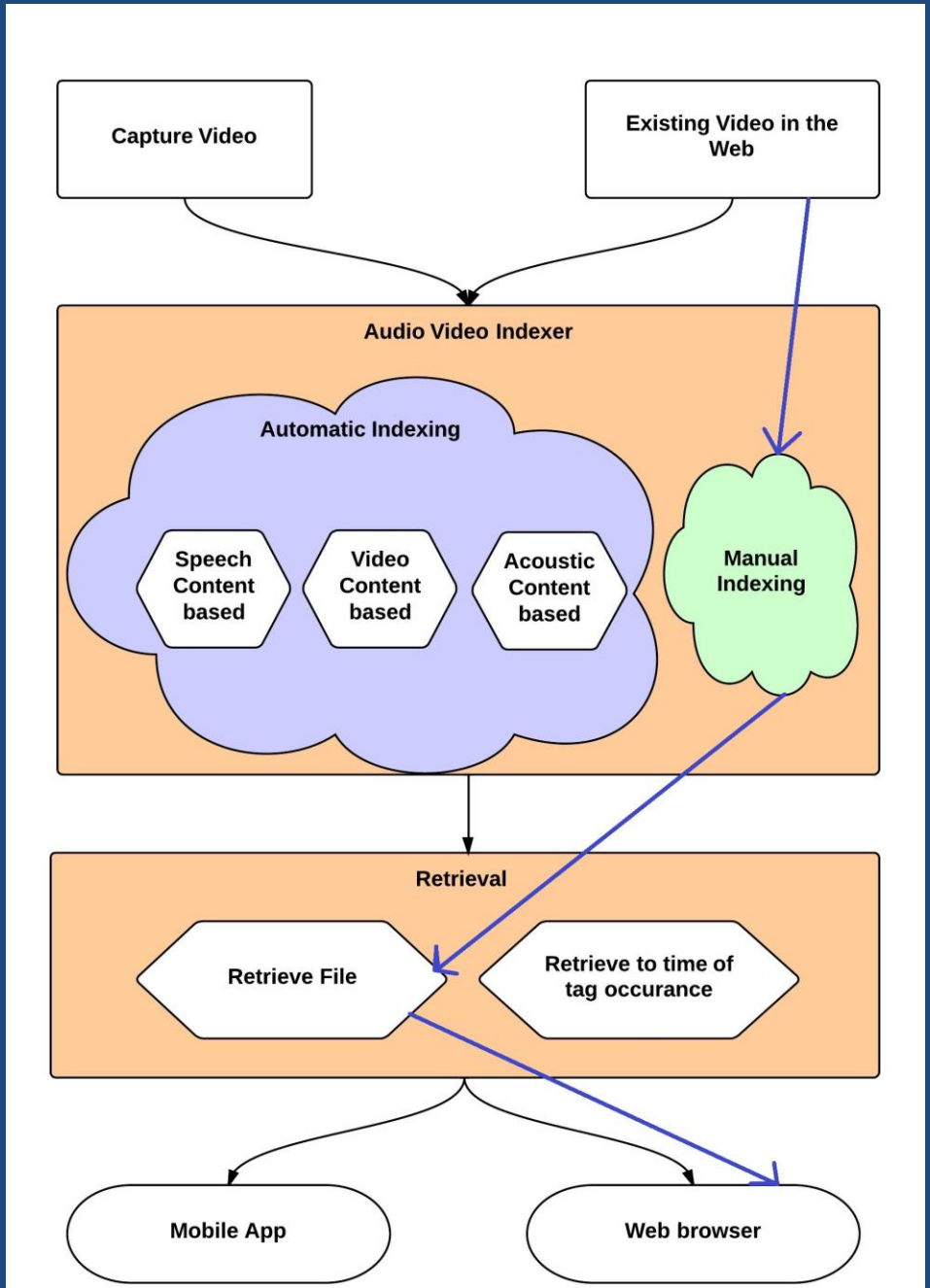
Work Done So Far

- Video Access & Manual Indexing Prototype



Work Done So Far

End to End flow implemented in 5th Sem



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

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

Future Scope

- Integration With Dropbox & other video sharing website
- Better retrieval process
- Web Indexing database

References

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://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.161.913
http://www.apple.com/ios/siri/
http://www.vlingo.com/
http://muvis.cs.tut.fi/
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/

Case Study (Problem)

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

Case Study (Solution)

Indexing

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

Case Study (Solution)

Searching

Retrieve According to Speaker : Ksrao, Rsc, Sc

Retrieve According to Creation Time & Sort.

Retrieve According to Capture Location

Retrieve According to Recognized Speech content

Retrieve According to User defined metadata / tag

Retrieve According to Acoustic information.

References

http://sourceforge.net/projects/marsyas/?source=recommended
http://sourceforge.net/projects/espeak
http://sourceforge.net/projects/sp-tk/
http://www.phon.ucl.ac.uk/resource/sfs/audindex/
http://en.wikipedia.org/wiki/Acoustic_fingerprint
http://en.wikipedia.org/wiki/Shazam_(service)
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://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.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.html

Thank You

Q & A