Media Content Searching and Bookmarking

Anand Mahuli
K.K.W.I.E.E.R,
Nasik, India

Ajinkya Mahajan
K.K.W.I.E.E.R,
Nasik, India

Prashant Kulkarni
K.K.W.I.E.E.R,
Nasik, India

Ajit Medhekar
K.K.W.I.E.E.R,
Nasik, India

ABSTRACT

The concept to search through videos by analyzing the audio content with speech to text technology was first introduced in the Google Lab Project Gaudi (Google Audio Indexing). You can search globally or search within a single video. For searching through the audio content a transcript database is required. Therefore, it is necessary to find solutions for the auto transcript generation of media file. There are some software proposes utilities to create subtitles for videos but all require an extensive participation of the user. But a few like Auto SubGen exits which indicates a way to generate subtitles following standards by using speech recognition. From these two main concepts we are proposing a system which will generate auto subtitles and will use these subtitles to search the audio content of the media thus improvising the searching.

General Terms

Media, speech recognition

Keywords

Bookmark, Caption, Global Search, Local Search, Media Content, Transcript, Speech Recognition

1. INTRODUCTION

This paper describes the two systems Google Gaudi and Auto SubGen and a new proposed system that will incorporate both these systems' features. Video has become one of the most popular multimedia artifacts used on PCs and the Internet. In a majority of cases within a video, the sound holds an important place. To search or analyze the audio content of the video the most natural way lies in the use of subtitles. However, manual subtitle creation is a long and boring activity and requires the presence of the user. Consequently, the study of automatic subtitle generation appears to be a valid subject of research.

Rest of the paper is organized as follows: Section 2.1 presents survey of Google GAUDI followed by survey of Auto SubGen in Section 2.2. Section 2.3 presents the proposed system description. This is followed by conclusion in Section 3.

2. RELATED SYSTEMS

2.1 Google Gaudi

Google Audio Indexing (GAUDI) was developed by Google Labs and ran the same underlying speech recognition technology which was used in Google's Elections Video Search Gadget. Gaudi provided a novel way to retrieve and navigate video material.

In contrast to what is currently available on YouTube search, it provided a richer search signal by providing the transcript of spoken content in the video. In addition, the time alignment of the transcripts allowed content-based within-video navigation as depicted in Fig 1.
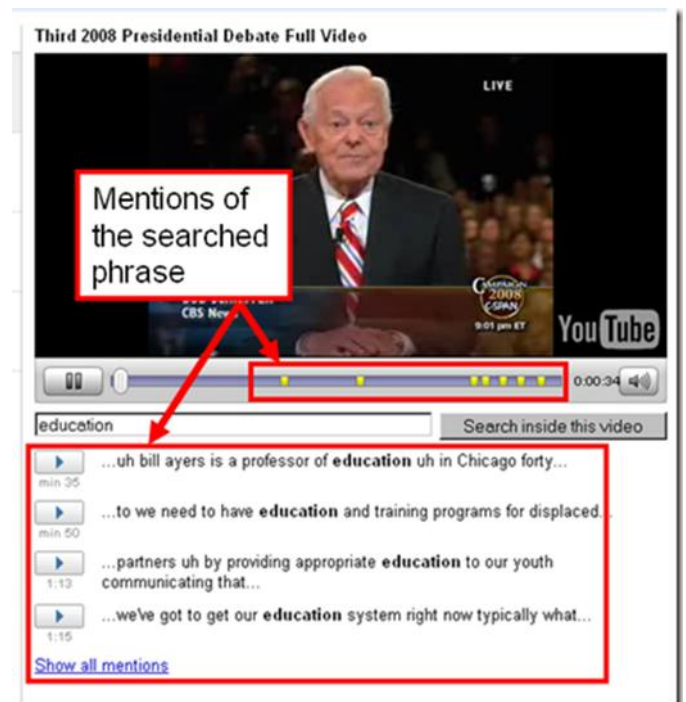

Fig. 1 Google Gaudi – Mentions of the searched phrase

GAUDI users needed to type in a query and then refine the search results using channel filters that correspond to YouTube channels. Search results provided information about each video, including the number of times query terms were spoken as depicted in Fig 2.

Videos are rich with content, but it is often difficult to find both relevant videos as well as relevant information within a specific video. The Google Audio Indexing technology solved both these problems since the transcript allowed search across videos as well as within videos as the text was time-aligned with the video content as depicted in Fig 3.

Fig. 2 Google Gaudi – Number of mentions of the searched phrase.
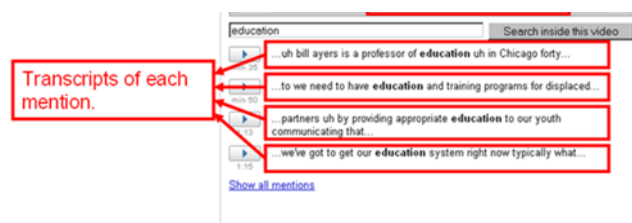
Fig. 3 Google Gaudi – Transcript of each mention.

GAUDI was available in limited beta testing and is only being used to index YouTube videos related to the previously held US elections.

Google Gaudi was limited to Videos. And that to only Presidential Election held in US in 2008. It was also a web application in unison with YouTube.

Similarities with Gaudi:

1. Provide a richer search by providing the transcripts of spoken content in video.

2. Allow content-based within-video navigation

3. Result provides information about each video, including the number of times query terms are spoken

Difference from Gaudi:

1. Will be local client software and not a web application.

2. Will work for both Audio and Video.

3. Instead of Presidential Election Video, Educational Tutorial Videos are considered.

2.2   AutoSubGen

AutoSubGen was a project to automate the work of subtitle generation with minimum user intervention. The AutoSubGen had three modules namely audio extraction, speech recognition & sub-title generation. Java was selected as the primary language for the implementation of AutoSubGen.
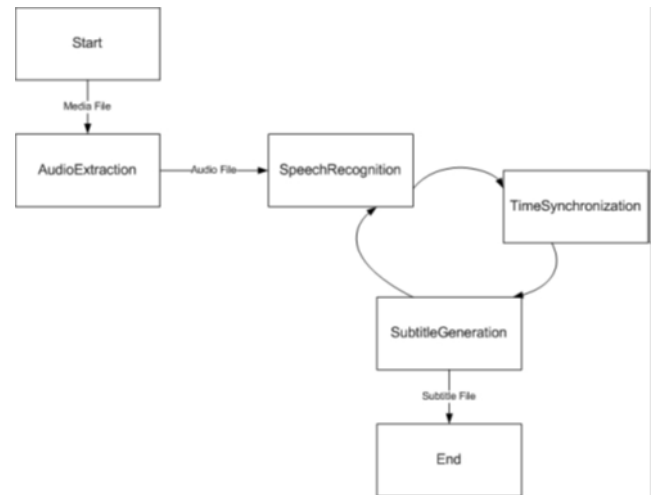
Fig. 4 AutoSubGen Block Diagram

In Audio extraction the audio of the video is extracted into suitable audio file format. For audio extraction JMF (Java Media Framework) API is used.

In Speech Recognition module the extracted audio is supplied to the speech recognition engine and the spoken words are converted into text file format. For speech recognition CMU Sphinx was used. Sphinx is a continuous-speech, speaker-independent recognition system making use of hidden Markov acoustic models (HMMs) and an n-gram statistical language model.

In Subtitle generation we have to use both modules speech recognition and also time synchronization. Whenever an utterance is occurred then the generated text from speech recognition is appended to the subtitle file.

This is one of the very few systems available for Automatic Subtitle Generation. The results were not satisfactory.

Table 1: Test Result of AutoSubGen

Run	Time	Words	Matches	Errors
1	0:53:23	403	7	399

Difference from AutoSubGen System
1. Language used is Visual Basic instead of Java
2. For Speech Recognition Microsoft Speech SDK 5.4 is used instead of CMU Sphinx
3. For Auto Extraction FFmpeg is used instead of JMF API.
The project aims to produce Transcripts of the video and not subtitles.

2.3   Media Content Searching and Bookmarking

In this project, an attempt is being made to incorporate the features of GAUDI in a media player with a broader domain of media (audio, video). The word-search is of two types as global search and local search. Global search includes a search of word in all media files present in library, whereas local search includes search of word in a particular media file.

The project also includes the feature of Auto Transcript/Auto Subtitle Generation. There is project like Auto SubGen, but with not much accuracy. The project also provides a feature of bookmarking. Bookmarking means to save a clip out of the complete media file with the help of start and end times.

All the systems available till now are all associated with videos. But the proposed system will be associated with both audio and video.

2.3.1 Media Content Searching

For both local as well as global searching we will require the subtitles or transcript of the media files which should be available in the database. Though some videos have in built subtitles most of the videos don't have subtitles or transcripts. For the video which have in-built subtitles or transcripts we can extract those subtitles and store them in database however if the subtitles or transcripts are not available we would have to first generate the subtitles or transcripts ourselves. In order to generate the subtitles we have developed a module named Auto-Transcript Generation
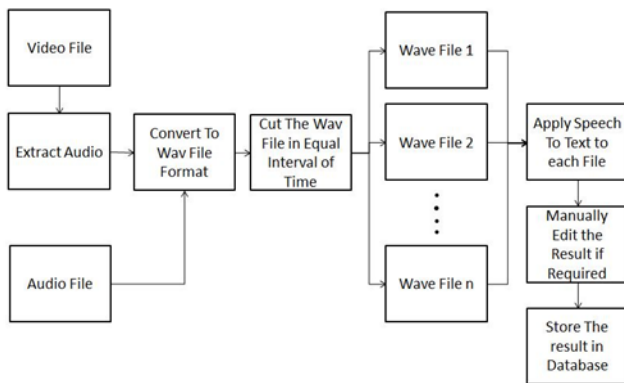

Fig. 5 Block Diagram of Auto-Transcript Module

As in Fig 5 we first extract the audio part of the media and convert it to the wave file format, cut the wave file into equal duration wave file, transcribe each file individually, manually edit the produced output if required and store the result in database. This approach makes the user involvement in the auto transcript process minimum. For all these operations we use FFmpeg. FFmpeg is a library for creating video applications or even general purpose utilities. FFmpeg takes care of all the hard work of video processing by doing all the decoding, encoding, multiplexing and de-multiplexing for you.

2.3.2 Media Bookmarking

Bookmarking is nothing but recording the address of (a website, file, etc.). To enable quick access in future. The same concept can be applied to the Media Files.

The bookmarks in Media work much like the bookmarks you place in paper books—they allow you to retain links to particular clip in a media file for quicker access.

With electronic bookmarks you can group them together in hierarchies, specify the exact text labels to use.

Media Bookmarking will help save the storage space as it saves the user from individually cutting the video into clips and storing them separately.

If they're well laid out, a set of bookmarks will help the user understand the basic structure of long media file.

The most common use of bookmarks—linking to different clips in media file within a media file—is just about the simplest way to make a accessing the media file more user friendly.

3. CONCLUSIONS

The aim of this project is to establish software which provides a novel way to retrieve and navigate video which places emphasis on content-based within-video navigation. This project aims to simplify within-video navigation. Although the emphasis is on the content-based within-video navigation, bookmarking System seeks to equally act a media playing software system with an innovative and intuitive interface design.

4. ACKNOWLEDGMENTS

We thank all people who helped and supported us. We sincerely thank our project guide, Prof Rupali Kulkarni, for guiding and correcting various documents of ours with attention and care. She has taken efforts to go through the project and make necessary correction as and when needed.

We also thank our Institution and our faculty members without whom this project would have been a distant reality. We also extend our heartfelt thanks to our families and well-wishers.

5. REFERENCES

[1] C. Alberti, et al., "An Audio Indexing System for Election Video Material," in Proc. of ICASSP, pp. 4873-4876,

[2] Boris Guenebaut, "Automatic Subtitle Generation for Sound in Videos".