DOI: 10.24996/ijs.2021.SI.1.23

ISSN: 0067-2904

YouTube Keyword Search Engine Using Speech Recognition

Aymen Shakat[*1], Khaldun Ibraheem Arif[2] , Sami Hasan[3], Yaser Dawood[4], Mohammed A. Mohammed[5]

[1]School of Computing University Utara Malaysia
[2] College of Education, University of Thi qar,
[3] College of Information Engineering, Al-Nahrain University
[4]Nabu Research Academy Selangor, Malaysia
[5] Al-Rafidain University College

Abstract:
Visual media is a better way to deliver the information than the old way of "reading". For that reason with the wide propagation of multimedia websites, there are large video library's archives, which came to be a main resource for humans. This research puts its eyes on the existing development in applying classical phrase search methods to a linked vocal transcript and after that it retrieves the video, this an easier way to search any visual media. This system has been implemented using JSP and Java language for searching the speech in the videos

Key words: Speech Recognition, speech transcript, text search techniques, Video.

Introduction:

The multimedia has grown rapidly since there has been a huge increase in the amount of information shared and generated by internet users from all around the globe. That led to growing rapidly of large multimedia storage system [1]. Multimedia content retrieval is considered one of the new interesting research fields in the multimedia research community that attracted quite an amount of researchers and developer. "With the exponential growth of effective indexing, multimedia data, and search methods of visual content are pivotal" [2].

Commonly, there are four types of multimedia; image, text, audio and video. Additionally, the multimedia information is much more attractive for the people to review the information than plain text [3].

So, the internet users prefer to get new knowledge by watching a movie, it allow them to focusing more on their content; Video media is an extremely powerful teaching resource for education [4].

"Search in non-textual unstructured content, such as audio and video data, is not yet effective" [5]. Video content can be mechanically extracted by using programs that do character recognition, speech recognition, image processing, or extracted speech from video's closed captions. "A common approach for video text extraction is by using conservative text search techniques to the related speech transcript. This method operates properly well for retrieving named entities of specific people, objects, or places" [6]. In addition, searching for the part of interest may require manual searching through the entire video that may be time consuming and often-frustrating process.

However, watching videos are growing in popularity, but searching and retrieval are becoming a challenging problem. Thus, locating the required content of interest with ease has become an indispensability gap that needed to be resolved. The aims of this paper are to present a basic system developed for video on YouTube searching based on the speech content transcript from video data. The rest of the paper is organized as follows; an overview of previous studies is provided in Section 2.

* Email: Aymenawni@gmail.com

Resources and data used for the experiments are described in Section 3. The experiments and results has been shown in Section 4. Finally, the conclusion is discussed in Section 5.

Related Work

There are a variety of projects have been developed Multimedia retrieval systems [7]. The emphasis interests in metadata annotation, multimedia databases, Web-based front-ends, specialized multimedia analysis methods, and presentation of search results. Automatic video search system [8], as shown in Figure-1, project aims to start "a stander approach to determine the video image retrieval task automatically, by using the key frames extracted and the speech content transcripts from video files. In the image retrieval, they have considered several visual features from images, and they use a content-based image retrieval tool like the GNU Image-Finding Tool"[9].

We can configure through the main derivation that the use of combination methods by using a mixture of the script and visuals in the search can operate more enhancing than the methods that used either the scripts or the images in the search. However the outcomes approves that in an IR system based on "contents is more likable to give a higher weight to the texts retrieved by an IR subsystem based on speeches than" what is retrieved by an IR subsystem based on images.

"In the automated video image search chore, the system inputs queries" to produce results without any human involvement. Offered an inquiry test a mixture of input videos, user query (information for multimedia statement, and inquiry test mixture's familiar key frame bounds, the system shall return a classified list for key frames or the video pictures from the test mixture, which is the best to meet the query. Figure-1 illustrates "the architecture of the automatic video search system".

According to Markaki and Charilas [10], PIDALI the Java programming language in Java programming language aims to present a modified web-based multimedia search engine. The developed system, as shown in Figure-2, combines the current search engine's characteristics in addition to new inventive features that ensure reduced the responding time, adapting to each user's preference, interactive multimedia searching, easy handling of multimedia content, improved search results, and "support of context-based user groups. The feature analysis explanation and mechanisms indicated that the search engine adapts a functionality to each user's" separately.

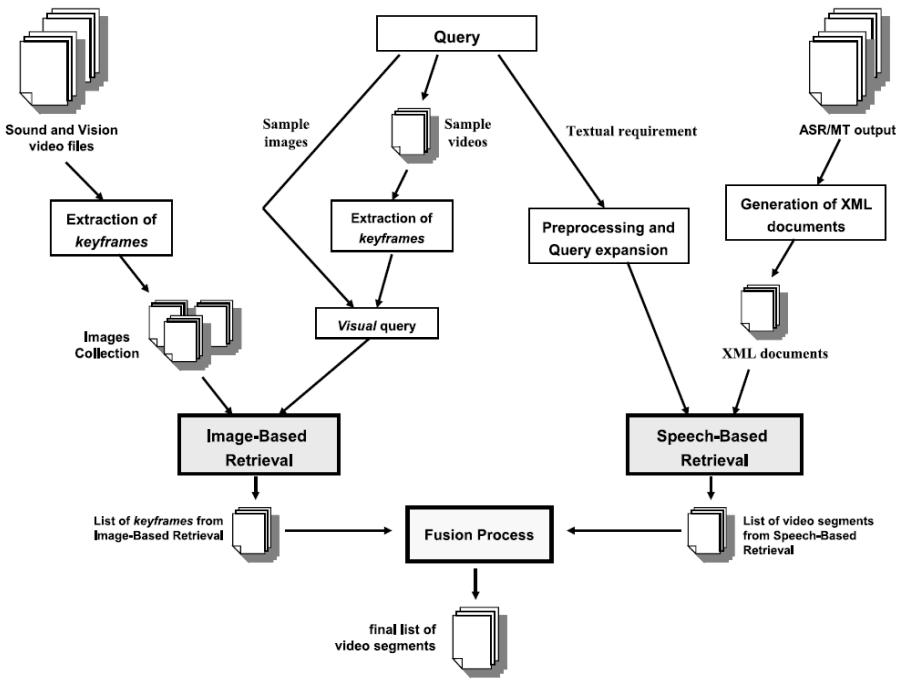

Figure 1-The "Architecture of the Automatic Video Search System" [8]

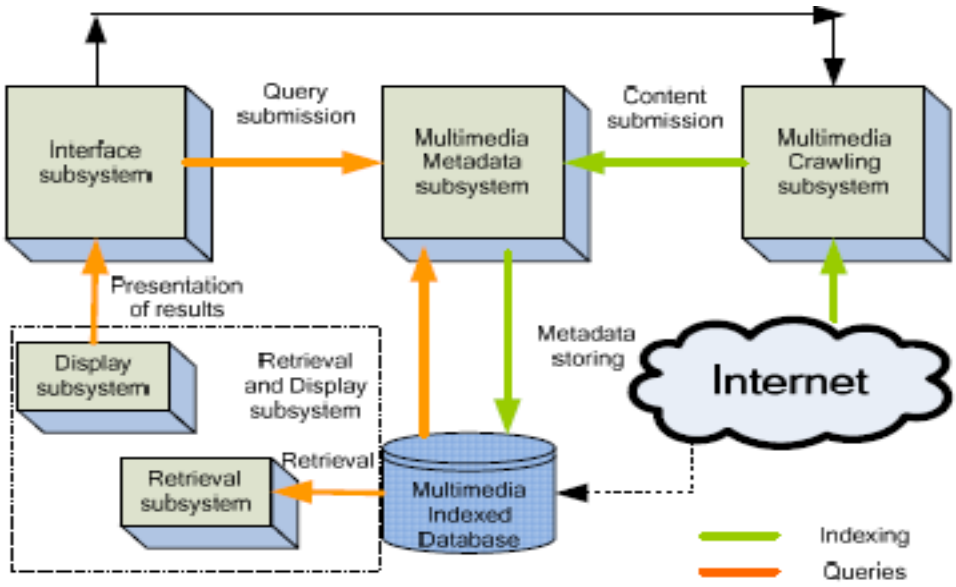

Figure 2-The Architecture of PIDALION (10).

The adoption of textural visual descriptors and color in the PIDALION search engine are supported by the MPEG-7 standard. More precisely, as far as color information is concerned, the dominant color descriptor (DCD) and the scalable color descriptor (SCD) are adopted, while for texture information, they used the homogenous and the non-homogeneous texture descriptors.

Other types of metadata such as AppEngine in Google cloud or file type, category and textual information are improved further "the precision of multimedia retrieval. Moreover, the system uses metadata for maintaining user groups and personalized indices Multimedia thematic category is determined manually in the framework of this search engine, in order to minimize the number of erroneous categorizations".

System Component

A software component is a software unit of structure with the contractually set interface. The operation of the component can only be seen through its set interface; the applying specifics and entanglement of a software component hidden from else software components.

Fundamental modules of our system were depended on Fang et al. module for typical video retrieval system [2] this module was continent six units, querying module, description module, match module, video database, abstraction module and a feedback module, these modules are shown in Figure-3

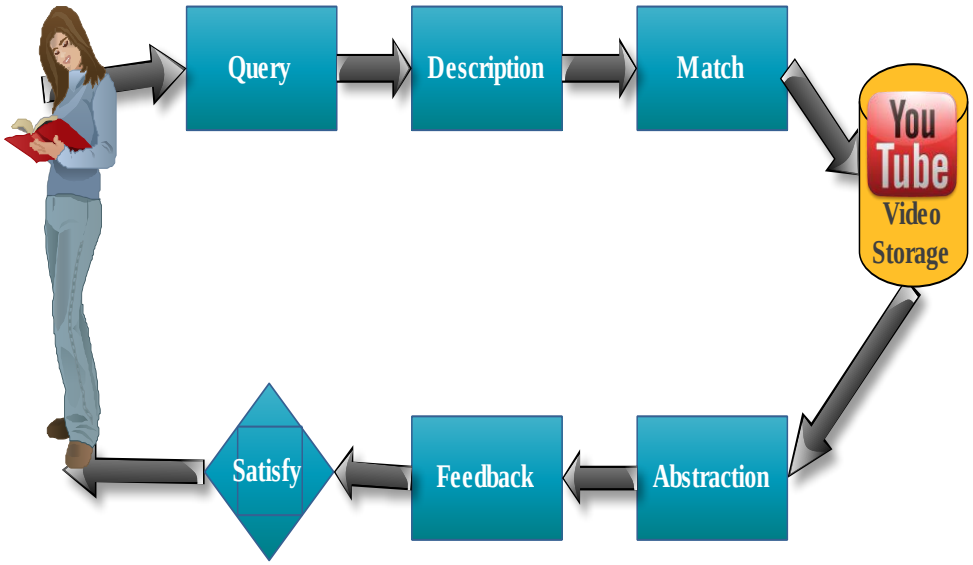

Figure 3-Framework of video retrieval system

Through this framework, the database for the proposed system depended on YouTube video storage.
The first module is to provide various search to the user, then, the translation of users' query is performed by description module. In the third module, the inputted query matching compares to the video database content[11]. The feedback module verifies the extracted videos based on users' requirements, after extracts all related videos for the user in abstraction module, but if the results are dissatisfied, users can re-query [12].
The system was implemented by use Java Server Pages (JSP) with JAVA language in general and develops the presentation layer appearance via using hypertext Markup Language (HTML). On the other side, an open source tool Google2SRT that can extract embedded subtitles or CC from Google and YouTube video together with their timing locations. It will be converted to a standard format of Sub Rip Subtitle (SRT).
Implementation and Results
A YouTube keyword search engine using speech recognition for video system was implemented using JSP and Java for searching the inputted video speech using a certain searched keyword. The admin has the ability to add a new video by creating subtitle use Google2SRT tool under the system the admin should enter the URL for the video and the name of subtitle file to convert to a standard format of SRT as shown in Figure-4

Figure 4- Google2SRT tool interface

The video will upload to the library of the system by the admin after entering the video title and video URL as well as the subtitle that was created in the first step as illustrate in Figure-5.

Figure 5-Add video interface

The authority of the admin being able to view the video that uploaded, and modify the information of video and image, as well as, the admin can make the search for videos in the system. Teaching video retrieval based on content can achieve multimodal query interface; the search results for a specific video might user to playback a video by just clicking on the video's link as illustrated in Figure- 6. In this system, the most important implemented function is the searching, when a video added by the user, it would have search content link; this will allow the user to start play a video from specific entered keyword, as shown in Figure-7. In additional, the system gives the user ability to perform a search inside all the video content.

Search video content result shows a list of all the speech content matches with the given keyword; the chosen clip as illustrated in Figure-7 will start playback from YouTube at its timeline as shown in Figure- 8.

Figure 6- Search results for a specific video

Figure 7-search inside video content interface

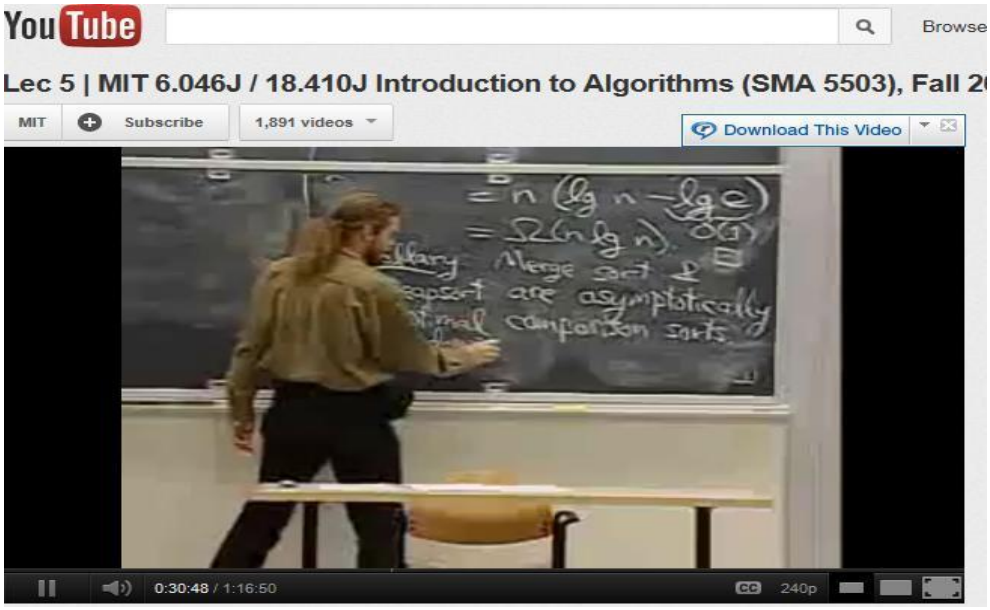

Figure 8-Playback video from YouTube in TimeLine

Conclusion

In this paper, the system investigates the retrieval process of video depending on content by searching the video resources in YouTube. This system helps the user to search for the part of interest instead of searching through the entire video. As a result, the paper provides to the user, interfaces design make the retrieval in the video more effective and targeted.

References

1. Wang Y, Kitayama D, Lee R, Sumiya K. **2009**. Automatic generation of learning channels by using semantic relations among lecture slides and recorded videos for self-learning systems. In2009 11th IEEE International Symposium on Multimedia 2009 Dec 14 (pp. 275-280). IEEE.
2. Khalaf OI, Sulaiman N, Abdulsahib GM. **2014**. Analyzing Video Streaming Quality by Using Various Error Correction Methods on Mobile Ad hoc Networks in NS2. *International Journal of Engineering Research and Applications*. 2014 Oct; **4**(10): 172-8.
3. Ren W, Singh S, Singh M. and Zhu YS. **2009**. State-of-the-art on spatio-temporal information-based video retrieval. *Pattern Recognition*. 2009 Feb 1; **42**(2): 267-82.
4. Moudgalya KM, Phatak DB, Shevgaonkar RK. **2008**. Engineering education for everyone a distance education experiment at IIT Bombay. In2008 38th Annual Frontiers in Education Conference 2008 Oct 22 (pp. T3C-21). IEEE.
5. Beecks C, Uysal MS, Seidl T. **2010**. A comparative study of similarity measures for content-based multimedia retrieval. In2010 IEEE International Conference on Multimedia and Expo 2010 Jul 19 (pp. 1552-1557). IEEE.
6. Natsev AP. **2006**. Multimodal search for effective video retrieval. In International Conference on Image and Video Retrieval 2006 Jul 13 (pp. 525-528). Springer, Berlin, Heidelberg.
7. Bailer W, Mayer H, Neuschmied H, Haas W, Lux M, Klieber W. **2003**. Content-based video retrieval and summarization using MPEG-7. *In Internet Imaging V* 2003 Dec 15 (**5304**: 1-13). International Society for Optics and Photonics.
8. J. M. Perea-Ortega and et al., **2009**. "A content-based information retrieval system for video searching," in in Intenational Symposium on Natural Language Processing, SNLP '09., 2009.
9. Markaki OI, Charilas DE. **2009**. Personalization mechanisms for content indexing, search, retrieval and presentation in a multimedia search engine. In2009 16th International Conference on Systems, Signals and Image Processing 2009 Jun 18 (pp. 1-6). IEEE.
10. Nurmi D, Wolski R, Grzegorczyk C, Obertelli G, Soman S, Youseff L, Zagorodnov D. **2009**. The eucalyptus open-source cloud-computing system. InProceedings of the 2009 9th

IEEE/ACM International Symposium on Cluster Computing and the Grid 2009 May 18 (pp. 124-131). IEEE Computer Society.

11. Haiguang F, Zhenyu W., Jing L. and Baocong J. **2009**. Research on the retrieval process of teaching video resources based on content. In2009 2nd IEEE International Conference on Broadband Network & Multimedia Technology 2009 Oct 18 (pp. 472-476). IEEE.

12. Shakat A, Al-Dabagh MA, Hasan S. and Dawood Y. **2018**. Speech recognition keyword engine for video on YouTube. Journal of Fundamental and Applied Sciences. 2018; **10**(4S): 671-6.