

Helping The Helpers: How Video Retrieval Can Assist Special Interest Groups

Frank Hopfgartner, Jinlin Guo, David Scott, Hongyi Wang, Yang Yang,
Zhenxing Zhang, Lijuan Marissa Zhou, and Cathal Gurrin

CLARITY: Centre for Sensor Web Technologies
Glasnevin, Dublin 9, Ireland

{fhopfgartner,jguo,dscott,hwang,yyang,zzhang,mzhou,cgurrin}@computing.dcu.ie

Abstract. Given the increasing broadcasting data and the ever decreasing spare time that we can spend on consuming this data, systems are required that assist us in identifying important content. Following a use case of a fictional social worker, we introduce a video retrieval system that is designed to assist special interest groups in their information gathering task.

Keywords: video retrieval, demo, broadcasting data

1 Introduction

“Patrick Murphy is an Irish-American street worker from a small town in the American West. Due to the critical financial situation in his state, the local government is forced to cut the annual budget significantly. His supervisor told him that right now, various boards discuss on which social programs will be shut down forever. Being aware that his clients, mostly poor and homeless people who hardly voice their opinion, will suffer most from these financial cuts, he decides to raise awareness of their needs, thus lobbying for the prosecution of the most important programs. Unfortunately, being on the street every day, he hardly can find time to attend all publicly accessible board meetings. Luckily, the government of his town runs a television channel where they provide an overview over every day activities of the government and their boards.”

In this demo, we showcase our video content organization system that allows fictional Patrick Murphy to easily assess the recordings of these board meetings. In Section 2, we introduce the data corpus that is used. Section 3 describes the required data processing steps. In Section 4, we describe the graphical user interface of our system. Section 5 concludes this demo paper.

2 Data Corpus

Addressing our user scenario, we focus on a city-wide government television channel based in California. Programs include, amongst others, coverage of board

meetings, local press conferences, and commission meetings. Further, information about government services are televised.

3 Data Processing

An important step for easing access to a video corpus is to segment it into semantically coherent segments. In a data corpus that mainly consists of meetings and press conferences, we consider speaker changes to be the semantic segmentation unit and segment the broadcasts accordingly. Further, we detect shot boundaries within these speaker segments and extract the middle key frame to visualize the content of the shot. Similar to Hopfgartner and Jose [1], we extract named entities (persons, locations, organizations) from the recorded closed caption signal, since they provide the highest content load, thus indicating the main subject of the segment. We argue that these entities can be used to filter search results. Moreover, applying subjectivity cues identified by Wilson et al. [2], we determine contextual sentiments on a textual phrase-level. Finally, we use Solr to index all segments, treating the textual transcript, metadata, sentiments, key frame and video URIs and named entities as separate fields in the index.

4 Interaction using a Graphical User Interface

Figure 1 shows a screenshot of the developed graphical user interface. On the top of the interface, the users can type in a textual search query. Search results, ranked using TF.IDF, are displayed on the right hand side of the interface. Each result is displayed by a representative key frame and a query-biased text snippet. This allows the user of the system to get an initial impression of the content of the retrieved segment. By clicking on one of the results, another window will be opened (not shown on the screenshot) where the video can be played back and neighbored shots are displayed. Further, a color-coded time line is displayed, visualizing passages in the video with negative and positive sentiment, respectively. This display allows the user to assess the general sentimental tone of the video. For example, if our fictitious Patrick Murphy views a board meeting about one of his social projects, the sentiment bar allows him to easily identify those board members who speak in favor and those who speak against it. On the left hand side of the interface, the user can re-define their search query. Given the importance of broadcasting time in the channel, the interface provides facilities to set the broadcasting time. Further, on the bottom left hand side, the users can exclude search results that contain certain named entities. Further down, not visible on the screenshot, is a tag cloud which displays the most frequent entities of the search results. This allows the user to gain a quick overview over the retrieved search results without inspecting them in detail.

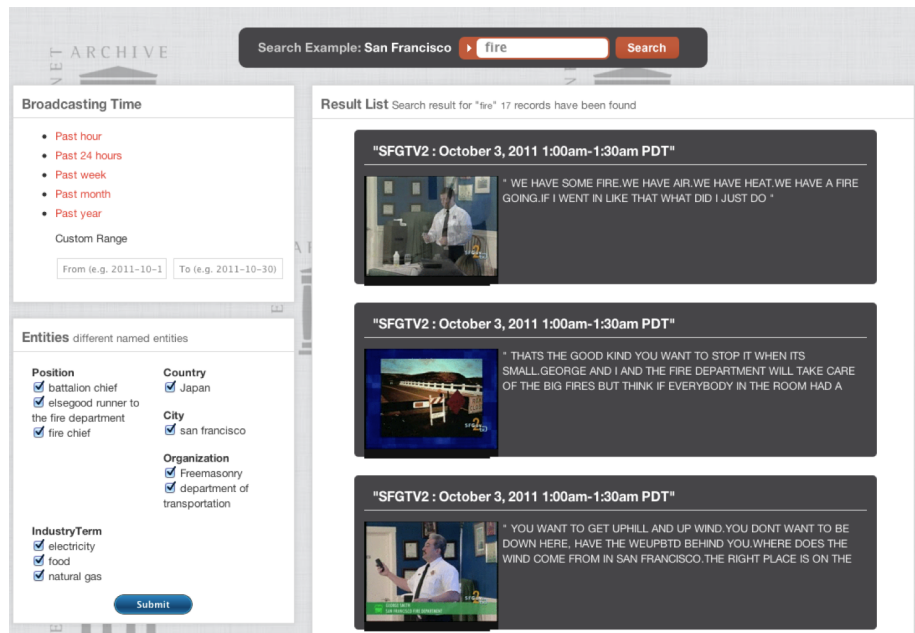


Fig. 1. Screenshot of the Search Interface.

5 Conclusion

In this paper, we introduced a video browsing system which is designed to assist users in accessing audio-visual recordings of board meetings. The system allows to access the data corpus of a small video collection. Filtering techniques such as broadcasting time, the appearance of certain named entities and a sentiment analysis ease access to this heavily speech biased data corpus.

Acknowledgments

This research was supported by the Norwegian Research Council (CRI number: 174867) and Science Foundation Ireland under Grant No.: 07/CE/I1147.

References

1. Hopfgartner, F., Jose, J.M.: Semantic user profiling techniques for personalised multimedia recommendation. *Multimedia Syst.* **16**(4-5) (2010) 255-274
2. Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing contextual polarity in phrase-level sentiment analysis. In: *HLT '05, ACL* (2005) 347-354