

Video Search and Consumption Using Speaker Labels and Key-phrases

MXABU team, Cisco

Represented by **Ananth Sankar and Matthias Paulik**

Sept 10, 2011

By the year 2015, video will represent 90% of all internet traffic. The use of video is also pervasive in today's enterprise because of applications such as videoconferencing, organizational communication, employee training, video surveillance, etc. The large volume of video created by these applications poses two major challenges:

1. How to find videos which contain relevant content
2. How to efficiently browse and consume videos once they are found

In Cisco's Emerging Technology Group, the Media Experience and Analytics Business Unit (MXABU) has been working on this problem over the last year. We use audio and video analytics to index videos with automatically generated metadata, such as key-phrases, speaker information, scene changes, etc.

This demonstration will show the use of key-phrases and speaker metadata to efficiently search and consume corporate videos. Our demonstration uses an archive of pre-indexed videos. A web-based, graphical user interface allows users to search and consume these videos based on the automatically generated metadata.

Our approach includes several innovative aspects. First, we automatically identify and propose relevant key-phrases for a new corporate domain by crawling and analyzing a large volume of domain texts. Videos from the target domain are processed using automatic speech recognition to locate these key-phrases. In addition, speaker segmentation and recognition technology is used to automatically discover and cluster speakers in the videos. Videos are presented to users in an augmented player that includes the metadata. The player also allows users to enter speaker labels, enabling a crowd sourcing approach to generating labels for speakers. This labeling scheme eliminates the need for pre-trained speaker models, and also improves the overall system accuracy.

We have conducted comprehensive user studies and customer pilots, and have received very positive feedback. The majority of users found it greatly beneficial to use the automatically generated metadata for searching and consuming video