1 Introduction

This document describes our efforts, steps and results in creating a content-based video search engine. That is to search a database of videos using an image or audio fragment as 'search term' and to search the acual video content(frames or audio of the video).

The following tasks have been accomplished in order to create our search engine:

1. Extract frames and audio from a large amount of videos in a distributed way using Hadoop.
2. From those extracted frames and audio, extract low level features.
3. Create a searcher which can search related images/audio fragments by using the low level features.
4. Create a web interface accepts as input an image or an audio fragment and displays related videos.

2 Goal

Through this experiment we wish the achieve the following goals:

1. To gain deeper understanding of multimedia storage and management.
2. Gain skills in debugging and programming.
3. Improve teamwork

3 Environment

4 Tools & Libraries

5 Method

5.1 Distributed video processing

How did we extract frames and audio with Hadoop?

5.2 Image feature extraction

5.3 Audio feature extraction

5.4 Image Videos Searcher

5.5 Audio Videos Searcher

5.6 Web Interface

6 Results

7 Conclusion

8 Responsibility table

References

[1] http://pwhois.org/lft/index.who

[2] http://www.caida.org/tools/measurement/scamper/

[3] https://gephi.org/

Lin Haiyuang, Wei Chaohou, Dongzhao