

Disaster Video Gallery

CS 4624 Capstone Project

Zelee Alem Fesseha,
Virginia Tech

Table of Contents

Abstract	3
User's Manual	5
Developer's Manual	6
<i>Concept</i>	6
<i>Data Files</i>	7
<i>Observations</i>	7
Acknowledgments	9
References	10

Abstract

Disasters can't be avoided. Every year, we see various parts of our world being hit by numerous man-made and natural disasters. The damage inflicted by many of these disasters goes far beyond their immediate impact. Some of these disasters are so powerful that cultures and lifestyles change after they happen. Their immediate impact is tremendous, and the aftermath, sometimes, is just as big.

Although it's inevitable that such unfortunate circumstances will continue to happen, it's our responsibility to learn as much as we can from past incidents and better prepare ourselves for future events. Explore past disasters for clues on how to prevent similar incidents from happening in the future; study the post-event situations of previous disasters to better react to future ones. These are only a couple of ways of how known information can be used to prepare for, and in some cases, prevent unfortunate events.

Do we have enough information to use as suggested in the previous paragraph? If so, where is all that information? What is the most effective way of using that information?

Yes, we do have that information and it's spread across the Internet. However, no solid attempt has been done to centralize this information and effectively present it to users. Search hits are sometimes irrelevant. They burden the user with much more digging than necessary. So, the answer to the third question is: there really isn't an effective way to use this information. In addition, some pieces in this huge collection of information go missing mainly due to some website going inactive.

The Disaster Video Gallery project is intended to help reduce this problem. It focuses on videos pertaining to disastrous events all over the world. Videos are very powerful sources of information because they can portray so much within a very short time. The goal of the Disaster Video Gallery project is to maintain a well-organized archive of videos of various disasters so that they can be used much more effectively. One approach to achieve this is to utilize the large video database of YouTube. Since YouTube has such a huge collection of videos, it's ideal for such a project where videos are the main focus.

This paper describes the first few procedures in utilizing YouTube in building the desired archive. The first step is to collect videos for a carefully chosen sample of events. These events were of different kinds; earthquakes, floods, shootings and vehicle related accidents. For each event, my responsibility was to collect videos that were diverse in their content. I had to manually verify that their contents were consistent with the event. This sample collection of videos is very critical because it's manually verified to be a 100% relevant collection.

The next step in the process is to collect metadata on these videos. Since these videos are YouTube videos, metadata includes attributes such as title, description, view count, number of likes and dislikes, related videos and so on. This metadata is particularly helpful for the following reason. The whole database cannot be constructed manually because of the enormous amount of videos and disasters available. Therefore, the process needs to be automated. However, a simple YouTube search for a key word does not do the trick. Because, a good number of the hits out of a search result are irrelevant. Therefore, the metadata collected can be used to model a relationship between some of the attributes and the relevance of a video.

There are also factors in the searching process that affect the relevance of video results, such as the choice of key words to search. A part of this section of the project was also to observe and identify these factors.

User's Manual

This project contains a list of videos pertaining to a carefully selected list of events. The collection of the videos is stored in two different formats under different file names. The file named "Sample_YouTube_Videos_Raw.txt" is intentionally prepared with no labels and virtually no extra characters except the actual data itself. Use this file to parse and process the contents. The format in which the contents of this file are organized is explained under "Data Files in the "Developer's Manual" section.

The other text file included, "Sample_YouTube_Videos_Readable.txt", is a more readable version with the same content. Entries in this file are labeled. This file is provided only for the convenience of the reader; to enable users to read the contents of the file without having to refer to the documentation of the file format.

Developer's Manual

Concept

Before the videos were collected, a list of events was made. The list contained events in 4 categories, floods and storms, earthquakes, shootings and vehicle related accidents. In each of these 4 categories, 4 events were identified. The first step was to find 4-5 diverse videos for each of these events.

There were a few things that had to be kept in mind while collecting videos. If all videos collected for a particular event were about the same topic, the collection would be less effective. For example, for a particular event, it would be more effective to have collection including a generally descriptive video that, for the most part, captures the story of the event, a raw video containing actual footage in the midst of the event, a video with an interview of a direct participant in the event and a memorial video for the event, than just having 4 videos of news coverage of the event.

To accomplish this goal, while collecting the videos, I classified the videos based on their content and the time they were captured compared to the time of the event. Some videos could fit the description of more than one of these classifications but I tried to choose the one that describes them the best to classify them. The following are the labels I used to classify the videos.

- Representative – These are usually the first videos in a collection. Ideally, an individual can only watch these and understand what the event is about.
- Raw footage – Videos labeled with this phrase are videos usually recorded by victims or direct participants of the event and contain scenes recorded in the midst of the disaster.
- Pre Event – Videos labeled “Pre Event” are videos recorded prior to the occurrence of the event. These videos usually contain helpful information that could have been used to prevent the disaster.
- Shortly After – These are videos recorded immediately after the event's occurrence. They usually include immediate reactions from victims, police reports, or footage of an area shortly after being hit by the disaster.
- New Developments – These videos contain information that may be used to update data observed in previous videos. For example, they may report an increase in fatality.
- First person account – These usually contain an interview with a person that was directly involved in the disaster; usually survivors of accidents.
- Immediate Reaction – These are videos that document reactions of the community within a few days of the event.
- Memorial – These are videos of memorial ceremonies of an event.
- Reflection – Videos of people or the community reflecting back on the event.

Documentary – These are videos usually professionally prepared by TV networks or other big organizations documenting the event after it happens.

As much as possible, for each event, I tried to collect videos that were in different categories. This helps make the collections more useful and more informative because they contain more diverse information.

Data Files

The main data file included in this project is “Sample_YouTube_Videos_Raw.txt”. This file is purposely constructed with minimal use of special characters and no labels. This is for the purpose of keeping the file simple and convenient to parse using a scripting language. The content of the file is explained below.

Each file is composed of a block of lines, separated by a blank line, each of which describes a particular event. For instance, a file may start with 15 lines describing a particular event followed by a blank line followed by 13 lines describing another event.

Each block of lines describing an event is formatted as follows. The first line contains the name of the event, such as “Virginia Tech Massacre”. The next line contains the start date of the event followed by the end date of the event on the next line. If the event happened only one day, the start date and end date will have identical values. If an event occurred over disjoint intervals of times, the start date line will contain “NA” and the disjoint intervals are listed on the next line separated by ‘;’. The fourth line indicates the type of the event; earthquake, flood, etc. The fifth line contains a single integer indicating the number of videos available for the event. The lines that follow are 2 lines for each video; the first one containing the full URL of the video and the next line containing a very brief description of the video.

So, a block of lines describing an event with 5 videos in it will be 15 lines long; 5 lines for the first 5 attributes of the event and 10 lines for the videos (2 each for the 5 videos).

In addition, the brief description listed under a video always starts with a key phrase to identify the purpose of the video in the collection. The key phrases are the classifications described above.

Observations

During this project, I have observed a number of things. One of the more important observations for this project is the relationship between different attributes of a YouTube video and its relevance to the search query. For videos with a high view count, the percentage of likes says a lot. On the other hand, for videos with lower view counts the percentage of likes wasn’t a correct predictor of the relevance of a video.

Also YouTube has an auto-generate feature that came up very handy a few times. When a search key is typed and searched, a link to an auto-generated page is returned with the list of search hits. This auto-generated page is a collection of videos with a great variety, which is what we desire our collections to be.

I recommend people that will work on this in the future, to try and take advantage of the two features mentioned above.

I also observed some interesting things in the content of the videos. For some events, I was able to find very relevant videos that were recorded before the events actually happened. In one of the videos for the Virginian earthquake, recorded at the zoo, animals can be seen taking shelters and getting ready well before the earthquake hit. In another video, the shooter in Brazil was caught on tape wandering the hallways of the school with a pistol in his hand before he began the shooting.

Acknowledgments

I would like to thank **Seungwon Yang** (seungwon@vt.edu) and **Kiran Chitturi** (chitturikiran15@gmail.com) so much for letting me help with this project and for patiently working with me.

I would also like to thank **Dr. Edward Fox** for being extremely helpful every step of the way.

References

<http://collab.dlib.vt.edu/runwiki/wiki.pl?MmReportOrg>