**Selection of a time-window for calculating total views of each video**

Goal: Calculation of total views for each video for the entire semester

The total time duration for all the videos that were listed for Fall 2013 was calculated. The longest video was found to be 38 minutes long while the average of the videos was found to be around 14 minutes. The goal for this assignment was to calculate the total views for each particular video. The number of play hits shows the number of views associated with each video. A program was designed that provided all the play hits recorded for each of the videos. It was observed that the play hits were very frequent meaning that the time gap between the first play hit to the second play hit was only a few seconds for many students. It was important to filter out the frequent play hits to obtain a much more meaningful data to calculate the ‘actual total views’ for a particular video.

In order to exclude the frequent play hits a time window was set. This helped to take out any play hits that occurred within that period of time so that it does not get added to number of views of a video. The goal was to eliminate the play counts from a student intensely watching a video in one sitting over re-watching of a video. This will provide the actual number of views for a video. The following steps were taken to calculate the total number of views:

* 1. Histograms: Histograms were made to show the distribution of time difference of play hits of all videos. Time difference showed the gap between one play to another play at a given time range. It was observed that when the time window was under 20 minutes the final data was very vague. There were fluctuations in the number of play hits and this trend followed till the end of the first 20 minutes. When the time duration was set to 20 minutes a graph showing a downhill projection was generated. This graph showed that most of the activities occurred in the first 20 minutes. After the first 20 minutes there was a steep fall in the histogram indicating that there is not much change in the number of play hits and showed that a possible re-watching of video occurred after the first twenty minutes.
  2. Therefore it was decided to use 20 minutes as the time window for calculation of the total views. Play hits that occurred after the first 20 minutes added towards the views count. Any play hits that occurred in the first 20 minutes were counted as 1 view for the video.

The goal is to check if they are re-watching the videos or not.

A time window (in seconds) of 20-40mins was found to be most appropriate. (Number of views flattens out, longest video length) Why?

A histogram was made originally with 5 minutes as the time window.

What happened with the histograms?

Time difference: counting the space between one play and another play

HiSTOGRAM WAS MADE TO SHOW THE DISTRIBUTUN OF TIME DIFFERENCE OF ALL VIDEOS.

Actual space between the play hits.

Histograms was made with all the time difference. Meaningful range was found to be 40 mins.

20-40 was no difference in data. Longest video is 38 minutes.

Histograms was made based ON a time frame (5 mins) and number of overall play hits for each videos for the semester.

A histogram was made using this data and 40 minute window showed a downhill projection.

GOAL: Intensely watching a video over rewatching.

However, a time window for each video could be explored (i.e. time window of 10%

of video length). In excel, blank column with play counts are from other

miscellaneous videos.

Program expands on the allplaysf13.py program and uses a user selected time

window to count plays and prints it to a CSV (or excel spreadsheet). Program grabs

a playtime (LPT-last played time) and subtracts that number with the next playtime

(T2- time #2) to have a “time difference”. The first time (LPT) is added to a list and

additional times (T2) are added to the list if the difference between LPT and T2 is

greater than the time window. When that happens, LPT then becomes the time

associated with T2 and the program continues. The program counts each item

(times) in the list and assigns that as the play count for that video. Results for each

person are printed to a CSV.

What videos get more seeks, what videos do

How many seeks videos do we see in each video?

Seek videos have start and an end time – look for the end time