Lab 6

Question 1: twitter streaming related to project

Since our project is about automated identification of objects within images and stitching images to create full area views, I chose to filter twitter feeds based on the keyword “image”. I pushed the results a borrowed Android tablet over a socket connection. The screen capture from the tablet is included under the doc folder (though be warned, I didn’t notice until after returning the tablet that there were some unsavory hashtags returned for the “image” keyword).

Question 2: using spark + twitter to analyze and categorize tweets

I used 3 keyword filters, "map", "location", and "directions" to pull data from twitter. Each of these was applied individually to a short twitter burst, and the resultant data was saved to files. These files were then analyzed using the NaiveBayes model. A subsequent fourth twitter stream was then ran with a filter of only “directions”, and this was analyzed to see if the system would correctly predict that this was a stream filtered on the keyword “directions”. I never was able to get a completely good run of this program, though. Initially I was running into issues with the way I was saving data to disk, but once I believe I had this issue resolved, twitter started blocking my connections saying that I had temporarily exceeded my quota.