

# CS 122 Project Proposal: Relationship Between Twitter Hashtags and News Stories

Sushmita Gopalan, Haylee Ham, Chelsea Ernhofer

The goal of this group is to write a program that monitors use of specified hashtags on Twitter and connects any spikes in hashtag use to relevant newspaper articles from the country where the hashtag was majoritively used. We plan to track hashtags using the Twitter API and gather information on the date, location, and favourites of the tweet, as well as information on the user such as whether or not they are a 'verified' user on twitter. We then plan to map spikes in use of the hashtag to popular news articles, in order to discover more information about the hashtag's relationship with what is happening in the news.

We believe this to be an ambitious product. Here are the steps we are planning to take in order to complete it:

- (1) Get access to Twitter's API, with tweet data going back as far as 2008.
- (2) Retrieve hashtag tag data using specific 'hashtag' library available through the API.
- (3) Write program to parse through applicable tweets, extracting needed information and measure general use of hashtag, staying sensitive to spikes.
- (4) Write separate program/part of program to scrape Google News and find articles using hashtag, specific locations, and dates
- (5) Create and return graph of hashtag usage along with map of hashtag usage and most popular articles by country

We plan on using the historical element of the Twitter API to gather a number of tweets from the past. As a finished product, we hope to have a program in which a user can enter a hashtag and how far back they wish to go in history to find hashtag usage (although we will have to play around with this to make sure our program doesn't run too slowly). Our program will then return a graph showing spikes in hashtag usage over time, a geographical map showing where the hashtag was used, most favorited tweet including the hashtag, a list of verified users who have tweeted the hashtag, and the top 3 articles from Google News per individual spike in hashtag usage. We will have to determine the threshold for what constitutes a 'spike' as we familiarize ourselves with the data and general hashtag trends.

Timeline:

Week #	Steps to be completed in that week
4	Steps 1-2

5	Step 3
6	Steps 3-4
7	Step 4
8	Step 5
9	Step 5, create final presentation
10	Finalize project, practice group presentation