

# Homework #9 - Databases

In this homework, you will be populating a database with information from a series of tweets (we have provided a cache file named *twitter\_cache.json* and the code that obtains the tweets in the function *get\_tweets*)

## PART 1

In Part 1 of the assignment, you will be finishing the function *setUpTweetTable* that takes a list of tweets, a sqlite connection object, and a cursor and inserts the tweet information in the database

In this function, you must:

- Create a database named *tweets.sqlite*
- Load all of the tweets into a database table called *Tweets*, with the following columns in each row:
  - *tweet\_id* - containing the unique id that belongs to each tweet
  - *author* - containing the screen name of the user who posted the tweet (note that even for retweeted (RT) tweets, it will be the person whose timeline it is)
  - *time\_posted* - containing the date/time value that represents when the tweet was posted (note that this should be a **TIMESTAMP** column data type!)
  - *tweet\_text* - containing the text that goes with that tweet
  - *retweets* - containing the number that represents how many times the tweet has been retweeted

## PART 2

In Part 2 of the assignment, you will be finishing up two functions:

*getTimeAndText* and *getAuthorAndNumRetweets*

The function *getTimeAndText* returns a list of strings that contain the *time\_posted* and the text of the tweets. It takes a database cursor as input. Select the *time\_posted* and *tweet\_text* from the *Tweets* table in *tweets.sqlite* and return a list of strings that contain the date/time and text of each tweet in the form: date/time - text as shown below

*Mon Oct 09 16:02:03 +0000 2017 - #MondayMotivation <https://t.co/vLbZpH390b>*

The function *getAuthorAndNumRetweets* returns a list of strings for the tweets that have been retweeted MORE than 2 times. It takes a database cursor as input. Select the tweet\_id, author (screen name) and number of retweets for of all of the tweets that have been retweeted MORE than 2 times and return a list of strings that are in the form: tweet\_id - author - # retweets as shown below

1062380175306964992 - *umsi* - 5

NOTE: The same author (twitter user) may be returned multiple times for multiple retweets

### **Extra Credit (3 points)**

Calculate the number of tweets per day and return a dictionary in the following format:

Key = day ('Mon', 'Tue', 'Wed', etc.)

Value = number of tweets that day

You will implement your code in the function *tweetsPerDayOfWeek*

## **Grading**

### **Part 1**

30 points - passing *test\_setUpTweetTable* test case

### **Part 2**

15 points - passing *test\_getTimeAndText* test case

15 points - passing *test\_getAuthorAndNumRetweets* test case

### **Extra Credit**

3 extra points - passing *test\_extraCredit* test case