

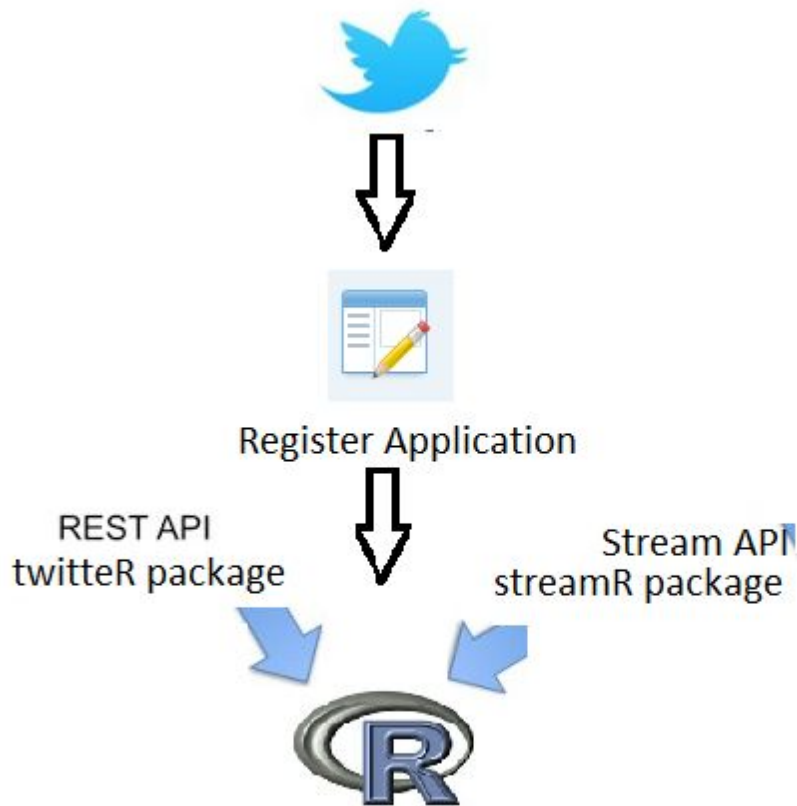
CSE 587 - Data Intensive Computing

Problem 1 : Data Acquisition

Goal: The goal of this task is to collect twitter data for one week using twitter API in R.

Steps to Collect Data From Twitter:

1. Create an application in your account and get the required credentials of the application such as apiKey , apiSecret , token and token_secret.
2. Install and load the required packages to connect to twitter api using setup_twitter_oauth(). The required packages are twitterR , Rcurl , RJSONIO and stringr.
3. Collect twitter data for 7 days using searchTwitter() function and the collected tweets are converted into dataframe objects using twListToDF().
4. Then the Dataframe object is converted to Json objects . Below diagram gives the Interface between twitter API and R.



Data Source : Twitter Data

Period of Time : 02-22-2016 to 02-28-2016

Structure of Data : Json Files

Functions Used :

1. `setup_twitter_oauth()` : This function is used in setting up the connection with Twitter API, we give the Twitter Application credentials as input to this function.
2. `SearchTwitter()` : This is used in searching Tweets based on the parameters we provide like: `searchString`, `language` , `Date Range` etc.

3. `twListToDF()` : Summarize information about a list of tweets in a dataframe.
4. `toJSON()`: This function converts an R object into a string that represents the object in JSON
5. `write()` : This function is used to write into a file from an R object.