

VANSUG\_May\_2011.sas

```
*=====*
```

```
/* DMKeys - Display Manager Keys (PFKeys - Programmable Function Keys) */
```

```
*=====*
```

```
/*
```

To display the key definitions that are active for the SAS session select Tools -> Options -> Keys

Keys can be defined or redefined.

```
*/
```

```
/* example */
```

```
/* Typically F12 is empty. To define the F12 key to clear a session log and clear a session output with one key, go to the Definition column for F12 and type -> clear log; clear output; (If, like me during the presentation, you type -> log clear; output clear; - it won't work!!)
```

You can find more about this and information in general about working with your SAS session here:  
<http://support.sas.com/documentation/cdl/en/hostwin/63285/HTML/default/viewer.htm#sasession.htm>

```
*/
```

```
*=====*
```

```
/* Sending a Text using SAS */
```

```
*=====*
```

```
/* Please see this blog for more information:  
http://blogs.sas.com/supportnews/index.php?/archives/100-SAS-gets-your-attention-with-a-text-message!.html */
```

```
/*  
Check out these web sites for updated information on carriers, especially since this sector is constantly in flux.  
http://en.wikipedia.org/wiki/List\_of\_carriers\_providing\_SMS\_transit  
*/
```

```
/* Code for sending text message */  
%let email_alert="9999999999@carrier_providing_SMS_transit";  
%let email_from="SAS";
```

```
filename em_out email to = (&email_alert)  
                        from = (&email_from)  
                        subject = "Sending a Text from SAS";
```

```
data _null_;  
  file em_out;  
  put "Sending a Text from SAS";  
run;
```

```
*=====*
```

```
/* Using SAS to retrieve Tweets from Twitter */
```

```
*=====*
```

VANSUG\_May\_2011.sas

```
/******  
**
```

Paper 324-2011

%GetTweet: A New SAS® Macro to Fetch and Summarize Tweets

Satish Garla and Goutam Chakraborty, Oklahoma State University, Stillwater, OK  
<http://support.sas.com/resources/papers/proceedings11/324-2011.pdf>

Macro Name: GetTweet

Purpose: Macro to fetch customized Tweets from Twitter using Keywords

How it works:

This Macro uses PROC HTTP to communicate with Twitter API. Twitter returns a maximum of roughly 1,500 tweets available in a week's time. The search method 'atom' is used and so we get results in XML format. A XML Mapper code is used to create SAS data sets from fetched XML results.

Required:

Please note ==> XML Mapper code is required to use this macro. You can use SAS XML Mapper tool

to create one. The XML Map file should be available in the Directory that is provided for PATH= parameter

Parameters:

The macro uses keyword parameters. Multiple words need to be separated by a Single Blank

WORDS= (Mention All of the words to search. This is an AND condition)

SINCE= (Enter From Date in the format: YYYY-MM-DD)

UNTIL= (Enter To Date in the format: YYYY-MM-DD)

PHRASE= (Enter Exact Phrase you want to search)

ANY= (Enter Any of the words you want to search)

NONE= (Enter the words you do not want to be in search results)

HASH= (Enter the hash tag that you want in your results)

FROM= (Enter name of the person who is Tweeting. Only one at a time)

QUESTION= Enter 1 if you want only the tweets with a Question Mark)

CODE= (Enter base64 encoded string of Twitter login)

PATH= (Directory where fetched data sets will be saved)

```
*****
```

```
/*  
/*Define macro variable to capture Base64 encoded string */  
%let authorization=%nrstr("Authorization: your_authorization_string");  
%let path=%nrstr(); /*Enter your Directory path*/  
/*Macro Definition*/  
%macro gettweet(WORDS=, PHRASE=, ANY=, NONE=, HASH=, FROM=, TO=, SINCE=,  
UNTIL=, QUESTION=, CODE=, PATH=);  
libname Twit "&PATH";  
%let dataset=Tweets; /*Give a name for Destination Data set*/  
filename httpOut "&PATH\twitterOutput.xml";  
filename hOut "&PATH\httpOutputHeaders.txt";  
filename hIn temp;  
filename httpreq temp;  
/*Add "+" between keyword parameters when Multiple words are used in the search*/  
%let WORDS=%sysfunc(translate(%sysfunc(strip(&WORDS)), "+", " "));  
%let PHRASE=%sysfunc(translate(%sysfunc(strip(&PHRASE)), "+", " "));  
%let ANY=%sysfunc(translate(%sysfunc(strip(&ANY)), "+", " "));  
%let NONE=%sysfunc(translate(%sysfunc(strip(&NONE)), "+", " "));  
%let HASH=%sysfunc(translate(%sysfunc(strip(&HASH)), "+", " "));  
%let FROM=%sysfunc(translate(%sysfunc(strip(&FROM)), "+", " "));  
%let TO=%sysfunc(translate(%sysfunc(strip(&TO)), "+", " "));  
%if &QUESTION=1  
%then  
%let QUESTION=%nrstr(&tude[ ]=%3F);  
%else  
%let QUESTION=%nrstr(&tude[ ]=);  
/*Define a macro variable "search" to create the search string from parameters*/  
%let search=%nrstr(q=&ands=)&WORDS%nrstr(&phrase=)&PHRASE%nrstr(&ors=)
```

VANSUG\_May\_2011.sas

```

&ANY%nrstr(&nots=)&NONE%nrstr(&tag=)&HASH%nrstr(&lang=en)%nrstr(&from=)&FROM%nrstr(&
to
=)&TO%nrstr(&since=)&SINCE%nrstr(&until=)&UNTIL&QUESTION;
/*Create a Temp file used in PROC HTTP headerin option to hold base64 encode*/
data _null_;
file hIn;
put &code;
run;
/*Create Destination Data set*/
proc sql;
create table TWIT.&dataset
(
id char(39) format=$39. informat=$39.,
published num format=IS8601DT19. informat=IS8601DT19.,
title char(159) format=$159. informat=$159.,
updated num format=IS8601DT19. informat=IS8601DT19.,
twitter_source char(100) format=$100. informat=$100.,
twitter_lang char(2) format=$2. informat=$2.,
uri char(50) format=$50. informat=$50.,
content char(2600) format=$2600. informat=$2600.
); quit;
/*Initialize and increment Page Number*/
%let pageno=1;
%StartLoop:
/*Define a variable for the number of tweets per page. The maximum allowed is 100 */
%let pagerate=%nrstr(&rpp=100&page=);
/*Combine Search String, Page rate and Page Number Macro variables*/
%let searchstring="&search&pagerate&pageno";
/*Create a Temp file used in PROC HTTP IN= option */
data _null_;
file httpreq;
put &searchstring;
run;
proc http
in=httpReq
out=httpOut
headerin=hIn
headerout=hOut
url="http://search.twitter.com/search.atom"
method="get"
ct="application/x-www-form-urlencoded";
run;
/*Define XML Mapper and XML Library*/
filename SXMLMAP "&PATH\TwitterSearch.map";
filename XMLLib "&PATH\twitterOutput.xml";
libname XMLLib xml xmlmap=SXMLMAP ACCESS=readonly;
/*Concatenate the XML Results in „Entry? data set and destination data sets*/
data twit.&dataset;
set twit.&dataset XMLLib.entry;
run;
/*Query the count of Tweets returned, into the Macro Variable „obscount?*/
proc sql noprint;
select count(*) into :obscount from XMLLib.entry;
quit;

%let pageno=%eval(&pageno+1); /*Increment Page Number*/
/**The Loop terminates if it is Page Fifteen or If it is the Last Page (<15) and has
less than 100 tweets. We can fetch a maximum of 1500 tweets at a time. If the tweets
available are less than 1500 the loop is terminated else the tweets from the last
fetched page keep writing to the Data set**/
%if %eval(&pageno)=16 or %eval(&obscount)<100
%then %goto EndLoop;
%else %goto StartLoop;

```

```

%EndLoop:
/*Summarize Tweets*/
proc sort data=twit.&dataset out=&dataset._temp nodupkey;
by title uri; run; /*Delete duplicates*/
/*Below DATA step cleans tweets and creates two data sets (one with all the tweets
and
the other only with retweets) in the work library*/
data &dataset (KEEP= id pubdate text author source retweet)
&dataset._rt (KEEP= id pubdate title text author source tweet_owner);
set &dataset._temp;
length text $ 159 tweet_owner $ 20 source $ 20;
format pubdate date7.;
retweet=0;
text=title;
author=tranwrd(uri,'http://twitter.com/','@');
id=substr(id,29);
pubdate=datepart(published);
if substr(text,1,3)='RT ' then do;
retweet=1;
tweet_owner=tranwrd(scan(text,2),':','');
call scan(text, 3, position, length);
text=substr(text,position);
end;
/*Define PERL regular expressions to replace http/www/RT text from tweets*/
if _n_=1 then do;
retain pattern pattern2;
pattern = PRXPARSE ('s/(RT @[^\ ]*)|((http|www)(\d|\D)[^\ ]*)|(@.[^\ ]*)//i");
pattern2 = PRXPARSE ('/'(\w|\W)[^\ ]*'/');
end;
call prxchange(pattern,-1,text);
text=strip(text);
if prxmatch(pattern2, twitter_source) then do;
call prxposn(pattern2, 0, position, length);
source = strip(substr(twitter_source, position+8, length-9));
end;
if retweet=1 then output &dataset._rt;
output &dataset;
run;
/*Calculate Total Number of records collected from twitter */
proc sql noprint;
select count(*) into :twtdcount from twit.&dataset;
quit;
proc sql noprint;
select count(*) into :retwtdcount from &dataset._rt;
quit;
/*Terminate macro execution if usernames are specified in FROM= parameter.
No Tweet report is generated. Only Data sets are created */
%if %length(&FROM) ne 0 %then %goto EndMacro;
/*Create Data sets for generating Tweet Report*/
proc sql outobs=10;
create table toptweeters as
select author 'Tweeter',count(author) 'Tweets'
from &dataset
group by author
order by 2 desc;
quit;
proc sql outobs=10;
create table topsources as
select source 'Source',count(source) 'Count'
from &dataset._rt
group by source
order by 2 desc;
quit;

```

```

proc sql outobs=10;
create table topowners as
select tweet_owner 'Tweeter',count(tweet_owner) 'Count'
from &dataset._rt
group by tweet_owner
order by 2 desc;
quit;
/*Define ODS Layout and generate PDF Report*/
options orientation=landscape;
goptions reset=all dev=sasprtc ftext="Helvetica";
ods listing close;
ods pdf file="&PATH\TweetReport_&dataset..pdf" STARTPAGE=NO BOOKMARKGEN=NO;
ods layout start;
ods region x=0 in y=0 in height=8.5 in width=11 in;
proc gslide;
title1 h=17pt j=Center underlin=1 'Tweet Report' lspace=.1in;
title2 h=12pt j=Left " Tweets:&twtcnt " " Retweets:&retwtcnt"
lspace=.1in;
run;quit;
goptions border;
ods region x=0.25 in y=0.3 in height=3.5 in width=5 in;
title1;
axis1 label=(angle=90 "Tweeter") minor=none;
axis2 label=(height=15pt "Top 10 Tweeters") minor=none;
proc gchart data=toptweeters;
hbar author/ sumvar= _TEMA001 maxis=axis1 raxis=axis2;
run;quit;
axis1 label=(angle=90 "Source") minor=none;
axis2 label=(height=15pt "Top 10 Sources") minor=none;
ods region x=5.5 in y=0.3 in height=3.5 in width=5 in;
proc gchart data=topsources;
hbar source/ sumvar= _TEMA001 maxis=axis1 raxis=axis2;
run;quit;
axis1 label=(angle=90 "Tweet Owner") minor=none;
axis2 label=(height=15pt "Top 10 Influencers") minor=none;
ods region x=0.25 in y=4.75 in height=3.5 in width=5 in;
proc gchart data=topowners;
hbar tweet_owner/ sumvar= _TEMA001 maxis=axis1 raxis=axis2;
run;quit;
axis1 label=(height=15pt "Tweets per day") minor=none ;
axis2 label=( angle=90 "Tweets") minor=none major=(n=4);
ods region x=5.5 in y=4.75 in height=3.5 in width=5 in;
proc gchart data=&dataset;
pattern1 color=red value=r3;
vbar pubdate /discrete outside=freq maxis=axis1 raxis=axis2 ;
run;quit;
ods layout end;
ods pdf STARTPAGE=NOW;
ods layout start;
ods region x=0 in y=0 in height=8.5 in width=11 in;
proc gslide;
title1 h=17pt j=left ' Tweet Report-Top Retweets' lspace=.25in;
run;quit;
ods region x=0 in y=0.3 in height=7.5 in width=10 in; title1;
proc sql outobs=20;
select Text 'Tweet',count(*) 'Retweets'
from &dataset._rt
group by Text
order by 2 desc;
quit;
ods layout end;
ods pdf close;
goptions reset=all;

```

VANSUG\_May\_2011.sas

```
ods listing;  
%EndMacro;  
%put Collected Tweets:&twtdcount, Collected Retweets:&retwtcount;  
%mend gettweet;
```

```
/*
```

The macro uses keyword parameters. Multiple words need to be separated by a single Blank

```
WORDS=      (Mention All of the words to search. This is an AND condition)  
SINCE=      (Enter From Date in the format: YYYY-MM-DD)  
UNTIL=      (Enter To Date in the format: YYYY-MM-DD)  
PHRASE=     (Enter Exact Phrase you want to search)  
ANY=        (Enter Any of the words you want to search)  
NONE=       (Enter the words you do not want to be in search results)  
HASH=       (Enter the hash tag that you want in your results)  
FROM=       (Enter name of the person who is Tweeting. Only one at a time)  
QUESTION=   (Enter 1 if you want only the tweets with a Question Mark)  
CODE=       (Enter base64 encoded string of Twitter login)  
PATH=       (Directory where fetched data sets will be saved)  
*/  
%gettweet(from=, CODE=&authorization, PATH=&path);
```