

What can CATS bring to researchers who
want to analyze tweets?

Michaël Gauthier – Université
Lumière Lyon 2

Why use Twitter?

- Major source of textual information
- Everything is available via the API
- Variety of associated meta-data

What is CATS?



What is CATS?

- A web interface hosted in Lyon 2
- Provides a way to easily collect tweets and provide analysis tools
- CATS = Collection and Analysis of Tweets made Simple

Why is there a need for CATS?

```
stream.py — streaming (git: master)
1  __author__ = "Adrien Guille"
2  __license__ = "GNU GPL"
3  __version__ = "0.1"
4  __email__ = "adrien.guille@univ-lyon2.fr"
5  __status__ = "Production"
6
7  from twitter import *
8  import datetime
9  import threading
10 import subprocess
11
12 tweets_per_file = 500
13
14 def quote(string):
15     return '''+string.encode('utf-8')+'''
16
17 class Streaming:
18     def __init__(self, dbname='TwitterDBTest', host='localhost', port=27017):
19         self.db_name = dbname
20         self.host = host
21         self.port = port
22
23     def threadUpdate(self, filename):
24         print('Importing', filename, '...')
25         filepath = 'streaming/data/'+str(filename)+'.csv'
26         if filename == 1:
27             subprocess.call(['sh', 'stream_run.sh', filepath, self.db_name, self.host, self.port])
28         else:
29             subprocess.call(['sh', 'stream_update.sh', filepath, self.db_name, self.host, self.port])
30         print('Done.')
31
32     def collect_tweets(self, duration=1, keys=None, follow=None, loc=None, lang='en'):
33         nb_tweets = 0
34         nb_tweets_infile = 0
35         nb_files = 1
36         last_import_day = datetime.datetime.now().day
37         file = open('streaming/data/'+str(nb_files)+'.csv', 'a')
38         auth = OAuth(
39             consumer_key=str(open('streaming/consumer_key', 'r').read()),
40             consumer_secret=str(open('streaming/consumer_secret', 'r').read()),
41             token=str(open('streaming/token', 'r').read()),
42             token_secret=str(open('streaming/token_secret', 'r').read())
43         )
44         twitter_stream = TwitterStream(auth=auth)
45         start_date = datetime.date.today()
46         end_date = start_date + datetime.timedelta(days=int(duration))
47         if keys != "":
48             print("keywords")
49             iterator = twitter_stream.statuses.filter(track=keys, language=lang)
50         elif follow != "":
51             print("users")
52
Line: 1 Python Soft Tabs: 4
```

Why is there a need for CATS?

- A strong programming background is necessary
- Existing tools are not user-friendly / free
- Bridge the gap between computer and social scientists

Thank you!

=^_^=

More info:

<http://mediamining.univ-lyon2.fr:8080/cats/connect>

or

michael.gauthier.uni@gmail.com