

UPPSALA UNIVERSITY

CLOUD COMPUTING

---

Cloud computing - Count pronouns in tweets

---

*Authors:*

Alessandro PICCOLO

October 8, 2016



UPPSALA  
UNIVERSITET

# 1 Introduction

A prototype system was built to analyze a dataset of Twitter tweets collected beforehand using Twitter's datastream API. The tweets were available in the public container 'tweets' in the SSC cloud, and the dataset consisted of a number of files containing line-separated tweet entries. Each tweet is a JSON document. The main task was to count and visualize the Swedish pronouns "han", "hon", "den", "det", "denna", "denne" and the gender neutral, new pronoun "hen". The solution was based on the distributed task queue 'Celery', using 'RabbitMQ' as the broker and Flask to visualize and request tasks. The basic model is shown in figure 1

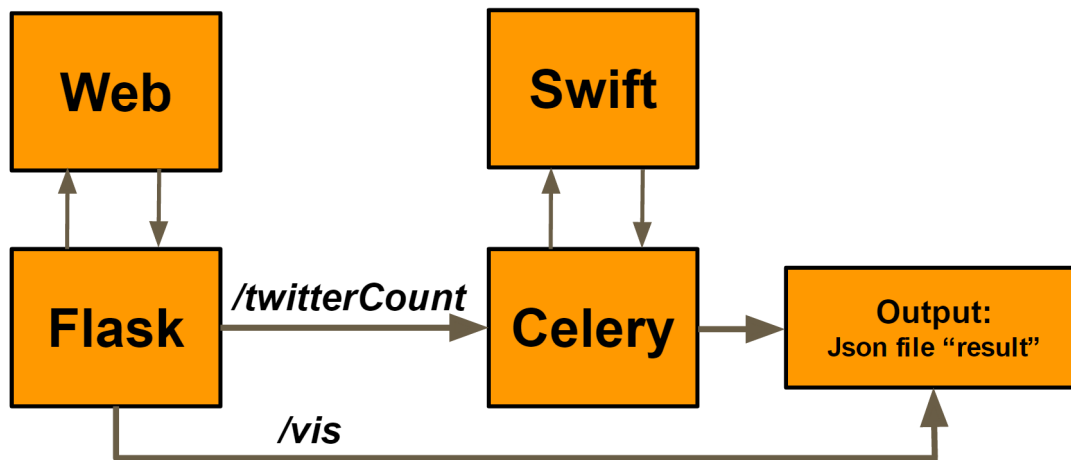


Figure 1: Model of the system

# 2 Results

Figure 2 shows the started flask server and one can see that it has been pinged once by figure 3. Figure 4 shows how the celery worker is actually working. The data is visualized as showed in figure 5.

```
ubuntu@alessandro-homepage: ~/lab3/CeleryFlaskRetrieveTwitt
ubuntu@alessandro-homepage:~/lab3/CeleryFlaskRetrieveTwitt$ python celery_flask_TwitterCounter.py
* Running on http://0.0.0.0:5000/ (Press CTRL+C to quit)
* Restarting with stat
* Debugger is active!
* Debugger pin code: 208-624-028
155.4.131.41 - - [08/Oct/2016 22:26:19] "GET /twitterCount HTTP/1.1" 200 -
```

Figure 2: Flask server

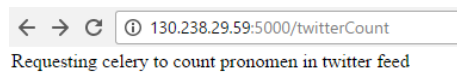


Figure 3: User request for counting pronouns to celery worker

```

ubuntu@alessandro-homepage: ~/lab3/CeleryFlaskRetrieveTwitt
ubuntu@alessandro-homepage: ~/lab3/CeleryFlaskRetrieveTwitt
ubuntu@alessandro-homepage: ~/lab3/CeleryFlaskRetrieveTwitt$
ubuntu@alessandro-homepage: ~/lab3/CeleryFlaskRetrieveTwitt$ celery -A celery_flask.TwitterCounter.celery worker --loglevel=info
[2016-10-08 22:24:20,662: WARNING/MainProcess] /usr/local/lib/python2.7/dist-packages/celery/apps/worker.py:161: CDeprecationWarning:
Starting from version 3.2 Celery will refuse to accept pickle by default.

The pickle serializer is a security concern as it may give attackers
the ability to execute any command. It's important to secure
your broker from unauthorized access when using pickle, so we think
that enabling pickle should require a deliberate action and not be
the default choice.

If you depend on pickle then you should set a setting to disable this
warning and to be sure that everything will continue working
when you upgrade to Celery 3.2::

    CELERY_ACCEPT_CONTENT = ['pickle', 'json', 'msgpack', 'yaml']

You must only enable the serializers that you will actually use.

warnings.warn(CDeprecationWarning(W_PICKLE_DEPRECATED))

----- celery@alessandro-homepage v3.1.24 (Clpater)
--- * --- Linux-4.4.0-36-generic-x86_64-with-Ubuntu-16.04-xenial
- * ---
- ** [config]
- **      .> app: celery_flask.TwitterCounter:0x7f2949263b90
- **      .> transport: amqp://guest:**@localhost:5672//
- **      .> results: rpc://
- **      .> concurrency: 1 (prefork)
- ***
- ** [queues]
- **      .> celery exchange=celery(direct) key=celery
- ***

[tasks]
. celery_ex.tweetRetrieveAndCount

[2016-10-08 22:24:20,707: INFO/MainProcess] Connected to amqp://guest:**@127.0.0.1:5672//
[2016-10-08 22:24:20,725: INFO/MainProcess] mingle: searching for neighbors
[2016-10-08 22:24:21,736: INFO/MainProcess] mingle: all alone
[2016-10-08 22:24:21,751: WARNING/MainProcess] celery@alessandro-homepage ready.
[2016-10-08 22:26:19,092: INFO/MainProcess] Received task: celery_ex.tweetRetrieveAndCount[527af496-ad11-4edd-b315-246126707aff]
[2016-10-08 22:26:19,248: INFO/Worker-1] Starting new HTTP connection (1): 130.238.29.253
[2016-10-08 22:26:19,759: INFO/Worker-1] Starting new HTTP connection (1): 130.238.29.253
[2016-10-08 22:27:07,224: WARNING/Worker-1] {'han': 42376, 'hon': 13394, 'det': 29968, 'denne': 270, 'den': 73854, 'denna': 1217, 'hen': 1156}
[2016-10-08 22:27:07,242: INFO/MainProcess] Task celery_ex.tweetRetrieveAndCount[527af496-ad11-4edd-b315-246126707aff] succeeded in 48.147195984s: 'End celery tweetRetrieve'

```

Figure 4: Visualization of celery worker in action

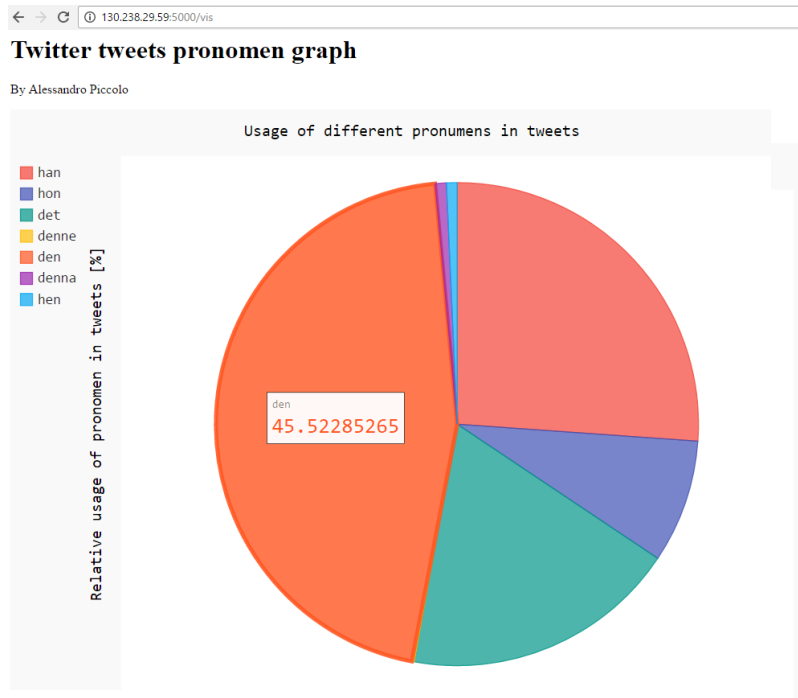


Figure 5: Visualization of pronouns count with flask pygal packet

### 3 Methods of celery\_flask\_twitterCount.py

This python file creates a worker and a flask server. It has two methods for the flask server, `twitterCount()` for requesting to start counting the pronouns for a celery worker and `vis()` for visualizing the count. It uses basic html with pygal package to visualize the json file called results (if there is no result file then it will tell the user to request and start a count). It also has a celery method called `tweetRetrieveAndCount()` which counts the use of pronouns in unique tweets in the files of the swift container in the cloud.

### 4 Packets to install

Basically these are the following commands that needs to be executed for a complete installation.

```
$ sudo apt-get update
$ sudo apt-get upgrade
$ sudo apt-get install rabbitmq-server
$ sudo locale-gen sv_SE.UTF-8
$ sudo apt-get install python-pip
$ sudo pip install celery
$ sudo apt install python-celery-common
```

```
$ sudo pip install flask
$ sudo pip install python-swiftclient
$ sudo apt-get install python-keystoneclient
$ sudo pip install pygal
```

## 5 How to use

The python files can be found in *<https://github.com/AlessandroPiccolo/lab3.git>* under *lab3/CeleryFlaskRetrieveTwitt/*.

Basically, open 2 terminals and ssh to instance, source *g2015034-openrc.sh* in each terminal.

Terminal 1: *celery -A celery\_flask.TwitterCounter.celery worker -loglevel=info*

Terminal 2: *python celery\_flask.TwitterCounter.py*

The user can initiate the request to count pronouns by accessing the following URL.

- *http://floatingip:port/twitterCount*

The previous request produces a json file called *result*. If there is such a file then one can visualize it by accessing the following URL.

- *http://floatingip:port/vis*

# Appendices

## A Github repository

The following files can be found in the following git repository,

*<https://github.com/AlessandroPiccolo/lab3/tree/master/CeleryFlaskRetrieveTwitt>*