Deployment Document

The deployment of this project includes five folders, namely configuration, 1_twitterHarvest, 2_extractData, 3_analysisData and website. Some instructions about how to deploy this system will be demonstrated as the follows:

1. Configuration: this folder includes only one file that named configuration.yml, which is an ansible script that can deploy fast the Ubuntu environment with all needed application, such as python, couchdb and others. Otherwise, some packages could not be involved in this ansible script because few external packages were added to this project.

2. 1_twitterHarvest: this folder includes three files associated with harvesting tweets in different types (Twitter-Search and Twitter-Streaming) and storing raw data into CouchDB. It is highly recommended that choosing two files of them and deploying them into different Nectar instances.

3. 2_extractData: this folder includes one file that searching for the relevant words from raw tweets and classifying related tweets into different databases. Before executing this file, it should confirm that the databases are ready to extract and store relevant data.

4. 3_analysisData: this folder includes two files: sentimentAnalysis.py that processing sentiment analysis towards tweets; supervisedLearning.py that producing naivebayes classifier based on training data set and then classifying the massive tweets. Before executing supervisedLearning.py, it should confirm that partial data has been labeled manually in the database.

5. Website: this directory mainly includes overall front pages that demonstrate the whole system. Approach one: putting all files in this folders into tomcat7 directory (/usr/share/tomcat7-myapp/myapp) and then run it as a web application; Approach two: directly opening the file of index.html in the browser.