Configure computing environment in a local laptop

1. Python virtual environment configuration
   Virtual environments make it easy to deal with the third party library version problems. To install virtualenv just use pip
   `pip install virtualenv`
   Then start a Python virtual environment using command
   `virtualenv -p python2.7 dp-env`
   where dp-env is the name of the virtual environment. Activating the virtual environment using command
   `source ./dp-env/bin/activate`
   In the virtual environment, we can run the project by typing
   `python2.7 run.py`

2. Install any missing libraries
   If you try running the project, there will be errors saying unable to find some libraries. In the virtual environment, install those libraries using pip
   `pip install module_name`
   You will probably get “ImportError: cannot import name check_build” even if you already installed sklearn. To solve the problem, try installing scipy (may need to restart the python shell after installing scipy).
   Besides, you may need to install the 'stopwords' and 'punkt' package for the NLTK python package.

3. Create a MySQL database and modify the db.py accordingly.
   If no MySQL on your computer, install it first.
   With the MySQL server on, we can command:
   `mysql -u root -p`
   Now we enter the MySQL monitor, next we create the user by:
   `mysql> CREATE USER 'username'@'localhost' IDENTIFIED WITH mysql_native_password BY 'password'`
   If you don’t want to change the code in db.py, you need to create a MySQL user named **dementia** with a password **Dementia123!**. Then create a database named **masters_data** and grant all privileges to the **dementia** user.
   `mysql> GRANT ALL PRIVILEGES ON masters_data.* TO 'dementia'@'localhost';`

Run the project

1. Clone the repository to the local and download three files -- data.zip, lib.zip, run.py. Replace the original run.py with the one you download from course website.
2. Place data alongside run.py and place lib within dementia_classifier/
3. Start the stanford parser with

```
java -Xmx4g -cp "dementia_classifier/lib/stanford/stanford-corenlp-full-2015-12-09/*" edu.stanford.nlp.pipeline.StanfordCoreNLPServer -port 9000 -timeout 20000
```

The parser is only for the feature extraction phase. If you are running analysis or plotting the results, there is no need to start the parser.

4. Activate the virtualenv in another terminal and run the project

```
python2.7 run.py
```

Troubleshooting

- Mac user may encounter "RuntimeError: Python is not installed as a framework. The Mac OS X backend will not be able to function correctly if Python is not installed as a framework. A quick solution is to create a file named **matplotlibrc** under the directory **~/.matplotlib** and add the following code

  ```
  backend: TkAgg
  ```

- If you encounter "ValueError: invalid literal for int() with base 10: 'sh: dementia_classifier/lib/SCA/L2SCA/./tregex.sh: Permission denied'"
  Please check permissions for tregex.sh:

  ```
  chmod 755 dementia_classifier/lib/SCA/L2SCA/tregex.sh
  ```

- In function `save_all_results()` of run.py, you need to add another function `feature_set.save_new_feature_results_to_sql(polynomial_terms=False)`
  Even though, you will still get the MySQL table missing error. To address this problem, replacing the first two lines of function `save_new_feature_results_to_sql()` with `new_feature_set = ['none']`. Then run the modified function twice, one with parameter `polynomial_terms=True` and the other with `polynomial_terms=False`.

For other problems, please post them on canvas!