**Guideline to use the scripts**

This Read-Me file describes a step-by-step guideline to use the following six scripts-

1. feature\_extraction.py
2. markable\_extraction.py
3. markable\_extraction\_big\_data.py
4. run\_steps.sh
5. run\_steps\_big\_data.sh
6. svm-classification.py

These scripts were created by Mahbub Ul Alam (3144196) & Tanzia Haque Tanzi (3144251) for the project work for the ‘Co-Reference Resolution’ in SS-2016 (MSc. In CL at IMS). For details about the work, scripts and output please see the ‘Project Report’ document.

1. Please keep all the scripts in a same directory and do not change their names.
   1. If your input data (training data and test data) are small in size then you should run the ‘run\_steps.sh’ script. You have to provide the data file names (with location if it is in a different directory) after that.

Or,

* 1. If your input data (training data and test data) are big in size then you should run the ‘run\_steps.sh\_big\_data’ script.

It is important to note that, in this case the **exact** names of the training data and test data files are as follows:

* + 1. ontonotes-train.conll
    2. ontonotes-test.conll

Please keep these two files in the same directory where all the scripts are kept.

If you want to change their names then please change it manually before executing on line number **358** and **362** in **markable\_extraction\_big\_data.py** script.

Please use the following command to run the script,

**nohup sh run\_steps.sh\_big\_data &**

You can see the execution update in the **nohup.out** file.

1. Output will be generated in the output folder. The 12 files are as follows,
   1. training\_markables
   2. test\_markables
   3. training\_data\_Positive\_chain\_lines
   4. training\_data\_Negative\_chain\_lines
   5. test\_data\_Positive\_chain\_lines
   6. test\_data\_Negaitive\_chain\_lines
   7. training\_features
   8. test\_features
   9. test\_data\_predicted\_classes
   10. test\_data\_Positive\_chain\_lines\_with\_predicted\_classes
   11. test\_data\_Negaitive\_chain\_lines\_with\_predicted\_classes
   12. Evaluation

Please do not remove them or change their names unless you can see the following message in the terminal or **nohup.out** file-

=============================================================

#SVM classification is finished, Please see the output files

#Thanks a lot for your time

=============================================================