

# Readme

Environment:

Python 3.7

pandas (0.23.4)

numpy (1.15.2)

paramiko (2.6.0)

sklearn'0.20.0'

How to run the code :

1. Use jupyter notebook.
2. To run “Gold standard generation and features extraction.ipynb” :
  1. If run the code from the begin, everything well be fine, only that’s a lot of time.
  2. Since I provided all the data needed to run from the middle of the program, only thing that needs to be done is to fine the text says “If already have files named "df9.csv", "df\_groupby\_2.csv", "df\_sync4.csv", "df5.csv" and "gold\_chain\_dict\_with\_core\_id.json" then please run the program from here.”, and the run the codes below this text, the following codes are used to generate positive and negative samples for training data.
  3. I provided the generated train data in the folder named “df\_gold\_std\_all\_18\_12\_with\_case\_and\_gender\_number.csv”
3. To run “Markable Extraction.ipynb” :
  1. same as the above file , if run the code from the begin, everything well be fine, only that’s a lot of time.
  2. Since I provided all the data needed to run from the middle of the program, only thing that needs to be done is to fine the text says “If already have files named "test\_df\_sync\_2.csv" and "test\_df2\_coref.csv" then please run the program from here.”, and the run the codes below this text, the following codes are used to extract markables.
  3. I provided the extracted markables in the folder named “markable2.csv”
4. To run “Test data generate.ipynb” :
  1. This code will use the .csv file generated from the above file, I provided the data file as well in the folder.
  2. run the code, and the test data will be generate, I provided the generated test data in the folder named “test\_data\_with\_case\_and\_gender\_number.csv”.

5. Expect the data files mentioned above

(*“df\_gold\_std\_all\_18\_12\_with\_case\_and\_gender\_number.csv”*, *“markable2.csv”* and *“test\_data\_with\_case\_and\_gender\_number.csv”*), other data files are files that support code to run from the middle.