Damegender: Gender detection tool

David Arroyo Menéndez

January 10, 2019

Damegender from the commands (I)

```
# Detect gender from a name
$ python3 main.py David
male
# Count gender from a csv example file
$ python3 csv2gender.py files/partial.csv
The number of males in files/partial.csv is 16
The number of females in files/partial.csv is 3
The number of gender not recognised in files/partial.csv is 2
# Count gender from a git repository
$ python3 git2gender.py https://github.com/chaoss/grimoirelab-p
The number of males sending commits is 15
The number of females sending commits is 7
```

Damegender from the commands (II)

```
# Count gender from a mailing list
$ cd files
$ wget -c http://mail-archives.apache.org/mod_mbox/httpd-announ
$ cd ..
$ python3 mail2gender.py http://mail-archives.apache.org/mod_mb
# Use an api to detect the gender
$ python3 api2gender.py David --api=genderguesser
male
# To measure success
$ python3 accuracy.py
Namsor accuracy: 0.9047619047619048
Sexmachine accuracy: 0.7619047619047619
```

Damegender from the commands (III)

```
$ python3 confusion.py
A confusion matrix C is such that Ci,j is equal to the number o
If the classifier is nice, the diagonal is high because there a
Namsor confusion matrix:
 [[ 3  0  0]
 [ 0 16  0]
 [ 0  2  0]]
Sexmachine confusion matrix:
 [[ 2  1  0]
 [ 2 14  0]
 [ 1  1  0]]

# To deploy a graph about correlation between variables
$ python3 corr.py
# To create the pickle models in files directory
$ python3 damemodels.py
```
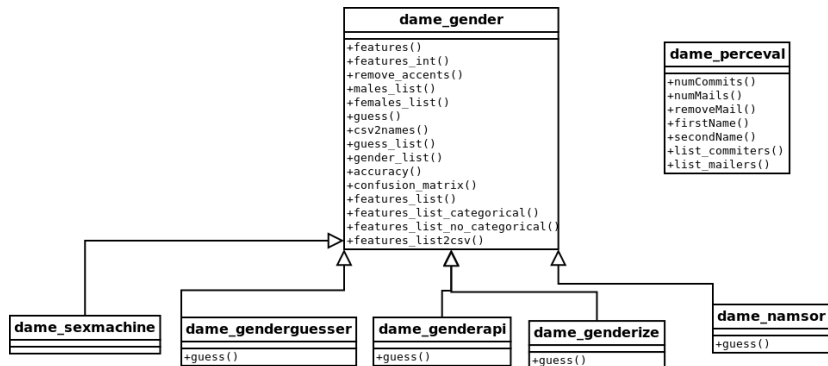
Damegender classes and methods (I)

Damegender classes and methods (II)

dame_sexmachine
+features()
+features_int()
+classifier()
+svc()
+svc_load()
+sgd()
+sgd_load()
+gaussianNB()
+gaussianNB_load()
+multinomialNB()
+multinomialNB_load()
+bernoulliNB()
+bernoulliNB_load()
+guess_surname()
+string2gender()
+guess()
+num_females()
+num_males()
+()