

Damegender: Gender detection tool

David Arroyo Menéndez

January 13, 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 $C_{i,j}$ is equal to the number of

If the classifier is nice, the diagonal is high because there are

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 as an exercise to practice NLTK and Perceval

```
~/git/python-examples/nlp/nltk: (dev) $ python3 sexmachine.py
```

```
What's your name?: David
```

```
What's my name?: Elena
```

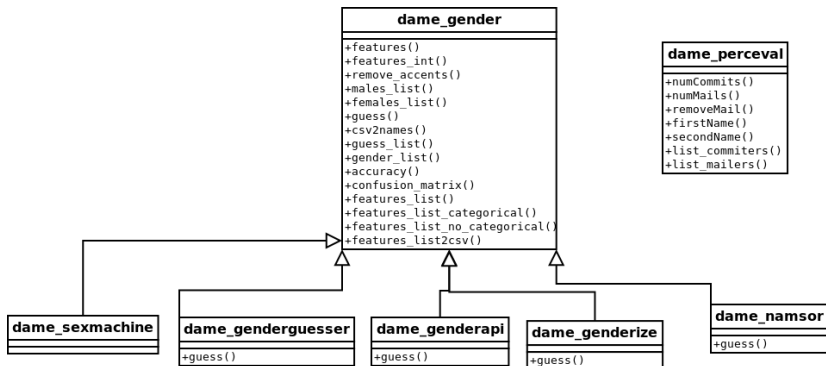
```
David is male and Elena is female. Enjoy!.
```

```
The classifier has an accuracy: 0.052
```

```
Most Informative Features
```

last_letter = 'a'	female : male	=	35.5 : 1.0
last_letter = 'k'	male : female	=	34.1 : 1.0
last_letter = 'f'	male : female	=	15.9 : 1.0
last_letter = 'p'	male : female	=	13.5 : 1.0
last_letter = 'v'	male : female	=	12.7 : 1.0

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()  
+()
```

Damegender and the bussiness

- In CMS: wordpress, drupal, joomla
- In dictionaries: google translate, babylon, gnu dict, ...
- Enciclopedias: wikipedia, ...

—

- A good technical project has a good bussiness project
- A Free Software license and community can be a good point.