

Damegender: Writing and Comparing Gender Detection Tools

David Arroyo Menéndez

July 17, 2020

- Thesis Student: David Arroyo Menéndez
- Title: Writing and Comparing Gender Detection Tools from a Name
- Thesis Director: Jesús González Barahona
- Objectives:

To compare APIs to detect gender from a name

To apply machine learning to classify nicknames, diminutives or new names as male or female

To understand the impact about open data in census about names and gender versus datasets created by companies or free software communities.

To apply this tools to the scientific communities or communities based on Internet.

Steps in the development

- 1 To develop the packaging and python tests with the basic features
- 2 To develop statistical tools and bash tests
- 3 To add 11 countries datasets with features to download and unify formats

We have presented this working in progress in:

Scientific events:

- Madrilénian Software Research
- Group Retreat 2019 Workshop
- SATToSE 2020: Seminar Series on Advanced Techniques & Tools for Software Evolution

Event to master students:

- Periodismo de Datos (Medialab Prado)

Industrial events:

- Python Barcelona
- Open South Code

Software

Free Software released with GPLv3 integrated in the industry

- `git clone https://github.com/davidam/damegender.git`
- `pip3 install damegender` (more than 3000 downloads by month)

Publications

- Damegender: Writing and Comparing Gender Detection Tools (Preprint)
- Damegender Manual: Counting Males and Females in Internet Communities (On working)