Téki: A Naïve Bayesian French Discourse Analyzer

Teki is a naive baysian classifier that is used to tag french chat data according their orality and literacy.

Academic Formalities

Information	Description
Name	Christopher Michael Chandler
Matrikelnummer	108017107247
Erstfach	Linguistik B.A, 6. Semester
Zweitfach	Romanische Philologie Franzoesisch B.A, 9. Semester
Titel	From T'es Qui to Qui Es-Tu: A Naïve Bayesian Approach to Assessing Literate and Oral Discourse in Nonstandard French Language Data
Kurs	Schriftliche Hausarbeit für die Bachelorprüfung der Fakultät für Philologie an der Ruhr-Universität Bochum (Gemeinsame Prüfungsordnung für das Bachelor/Master-Studium im Rahmen des 2-Fach-Modells an der RUB vom 03. November 2016)

License

Requirements

Python version

This program was created and designed with **python 3.9.6** in mind. Therefore, that is the recommended version. It is possible to run the program using any version **above 3.6**, but program stability cannot be guaranteed.

Pip modules

The following pip packages must be installed for the program to run properly:

```
* bs4>=0.0.1
* beautifulsoup4>=4.9.3
* matplotlib>=3.4.1
* lxml>=4.6.1
* spacy>=2.3.5
```

There are three ways to install the *requirements.txt* file:

- 1 install all modules individually via pip
- 2 install with pip using the requirements.txt without specifying the version number:

```
pip install -r requirements.txt
```

- 3 install with pip using the requirements.txt with specifying the version number:

```
pip3.9 install -r requirements.txt
```

Program

If the modules have been succesfully loaded, then you will be greeted with a menu system:

```
1: import file
2: load training file
3: analyze contents
4: classify string
5: clear log file
6: author information
7: program description
8: end program
```

Criteria

to learn more about the criteria please click here