Natural Language Systems Coursework

Requirements

- `Python 3` (developed using `Python 3.7.6`)

Setup

1. Create a Python virtual environment (e.g. `conda create -name nls python=3.7.6` or `virtualenv env`)
2. Activate the environment (e.g. `conda activate nls` or `source ./env/bin/activate`)
3. Install dependencies: `pip install -r requirements.txt`
4. Create a `data` directory in the project root and copy all required data (corpus and target words) into it, creating `A_inaugural` , `B_ntext` , `C_hw1-data` and `target-words.txt` . Alternatively, you can change the path to data directories in `main.py` .
5. Run the program: `python src/main.py`

Project structure

- `src` - All code written for the coursework
  - `main.py` - Entry of the program; should be run when evaluating coursework
  - `corpus.py` - Contains operations on the corpus
  - `utilities.py` - Contains utility methods to be used by `main.py` and `corpus.py`
  - `embeddings.py` - Alternative method for feature generation for clustering. Not used by the program, not tested.