

Documentation for Eve's packages

December 5, 2023

Contents

1	Instructions for spotting similar questions using bag of words	1
1.0.1	Windows	1
1.0.2	Linux/Mac	1

1 Instructions for spotting similar questions using bag of words

Please see the presentation for details on the input variables. There are four input variables required:

1. name of your input csv file
2. length of the search box, N, an integer is expected.
3. threshold. a float number between 0 and 1.
4. cluster option. "fuzzy" or "exact"

1.0.1 Windows

This step assumes the user installed Python and Make on their system (it is easier to install Make if you install Chocolatey, and then type 'choco install Make' in the terminal). Download the folder that corresponds to Windows.

step 1: go into the directory of your choice that contains the basket csv file and the python script Scraper.py and Makefile.

step 2: open a terminal from the directory and run:

```
make
```

step 3: run

```
.\myenv\Scripts\python.exe bow.py Name/of/your/input_csv/file  
length/of/the/search/box threshold cluster/option
```

1.0.2 Linux/Mac

This step assumes the user installed Python and Make on their system. Download the folder that corresponds to Linux/Mac.

step 1: go into the directory of your choice that contains the basket csv file and the python script Scraper.py and Makefile.

step 2: open a terminal from the directory and run:

```
make
```

step 3: run

```
python bow.py Name/of/your/input_csv/file length/of/the/search/box thresh-  
old cluster_option
```