

Similarity Search with Word Embeddings:

@author Karolis Bandziulis

@version java 17

Description:

This is a console-based application that generates a file **a** with **n** results of most similar words where **a** is the user specified output file path and name by comparing the word vectors from the user specified embeddings file **b** with one of the selected comparing algorithms “Dot Product”, “Euclidean Distance” or “Cosine Distance”;

To Run:

From console at .jar file directory:

- 1: java -cp ./dsa.jar ie.atu.sw.Runner.
- 2: select option 1, set the word-embeddings file to valid file path.
- 3: select option 2, set output file path to desired location.
- 4: if you choose an invalid path, you must start from part 1 again.
- 5: if you do not select anything for option 2, it will use a default name “./out.txt”.
- 6: use option 4 to configure how many output results **n** you want
- 7: use option 5 to configure what order you want your output to be in..
- 8: use option 6 to choose which comparing algorithm you want to use.
- 9: select option 3 and enter a word or short sentence, it will output the key name and similarity score as configured into the output file.
- 10: select 7 to exit application.

Features:

- Specify Embedding File
 - Input text / path of file including the extension.
- Specify an Output File
 - Input output file path including extension.
 - default: ./out.txt.
- Enter a Word or Text
 - Returns message if no selected embeddings file.
 - Allows user to enter a word / short sentence and compares them with the selected algorithm.
 - Outputs sorted string comparison similarity values as configured.

- Configure Options
 - Select n, the number of outputs you want.
- Change Output Order
 - Select between descending and ascending.
- Change Comparing Algorithm
 - Select between “Dot Product”, “Euclidean Distance” or “Cosine Distance”.

```
(1) Specify Embedding File
(2) Specify an Output File (default: ./out.txt)
(3) Enter a Word or Text
(4) Configure Options
(5) Output Order Options
(6) Change Comparing Algorithm
(7) Quit
Select Option [1-7]>: 1
Specify Embedding File Location: D:\school\Data structures\Workspace\G00417529\word-embeddings.txt
*****
*      ATU - Dept. of Computer Science & Applied Physics      *
*                                                              *
*      Similarity Search with Word Embeddings                  *
*                                                              *
*****
(1) Specify Embedding File
(2) Specify an Output File (default: ./out.txt)
(3) Enter a Word or Text
(4) Configure Options
(5) Output Order Options
(6) Change Comparing Algorithm
(7) Quit
Select Option [1-7]>:
```