

# Similarity Search

@author

Sarah O'Connor

G00423847

## Description

This is a Java application which uses word embeddings to determine the similarity between words, and their numerical representations.

## Features

- Specify text file which contains the word embeddings.
  - Input must include any necessary path or directory information.
  - This is a necessary step to make any further use of the application.
- Specify text file to output results.
  - Input must include any necessary path or directory information.
  - Must be a .txt file.
- Configure Options.
  - User may edit the text file path for the word embeddings.
  - User may edit the text file path for the output.
  - User may choose how many words they would like to be included in their similarity results – the default is 5.
- Input word / string of words to find similar, or dissimilar, results.
  - User may input one word, or a string of words.
  - User may determine if they would like to output the **most** similar results, or the **least** similar results.
  - Programme determines word similarities using Dot Product and Cosine Distance algorithms, and assigns a Similarity Score between -1 and 1 – 1 being the most similar, and -1 being the least.
  - Programme outputs a number of results. The default number of results is 5 - this can be customised within the 'Configure Options' menu.
- The results of each similarity search will be saved to the output file.
  - The default output file is 'out.txt'.
  - This file path can be changed by the user within both the main menu and the 'Configure Options' menu.

## Sample Output:

```
Windows PowerShell
PS C:\DSA\SimilaritySearch\bin\G00423847> java -cp ./dsa.jar ie.atu.sw.Runner

*****
*                                     *
*           S A R A H ' S           *
*           Similarity Search!       *
*                                     *
*****

*** M E N U ***
(1) Specify Embedding File
(2) Specify an Output File (default: ./out.txt)
(3) Enter a Word or Text
(4) Configure Options
(-1) Quit

Select Option:
1

Please enter Word Embeddings File name, including Path (No Spaces):
../word-embeddings.txt
Successfully parsed file.

*** M E N U ***
(1) Specify Embedding File
(2) Specify an Output File (default: ./out.txt)
(3) Enter a Word or Text
(4) Configure Options
(-1) Quit

Select Option:
3

Would you like to find:
(1) Most Similar Words
(2) Least Similar Words
1

Please input text you wish to compare:
Test Trial
```

```
Windows PowerShell

Please input text you wish to compare:
Test Trial

Top 5 Similar Words to 'test':

testing, Similarity Score: 0.83
tested, Similarity Score: 0.74
final, Similarity Score: 0.69
taking, Similarity Score: 0.69
selection, Similarity Score: 0.68

Printing result to file . . .
Successfully written to file. (./out.txt)

Top 5 Similar Words to 'trial':

prosecution, Similarity Score: 0.86
arrest, Similarity Score: 0.84
case, Similarity Score: 0.84
court, Similarity Score: 0.84
jury, Similarity Score: 0.81

Printing result to file . . .
Successfully written to file. (./out.txt)

*** M E N U ***
(1) Specify Embedding File
(2) Specify an Output File (default: ./out.txt)
(3) Enter a Word or Text
(4) Configure Options
(-1) Quit

Select Option:
-1

Quitting Similarity Search . . .
Goodbye :)
```