

Similarity Search with Word Embeddings

@author Phat Tan Nguyen

@version JavaSE 11

Description

This is a menu-driven Java application that can parse a text file containing 59,602 word embedding into a map data structure and search the data structure for a word or short set of words provided by the user and then output the top n most similar words to a text file.

To Run

From console at .jar file directory:

```
java -cp ./dsa.jar ie.atu.sw.Runner
```

The program will then ask for the path to word-embeddings file and name of the output file

Then navigate through console options to get similarity word set from the file.

- (1) Enter the word to search for similarity, then choose number of word that will be printed to output file.

When press 1, the program will first parse the list to map and then ask the user to input the word or short sentence.

For each word found in the list, the program will ask the user to input the number of top n word to be printed out to the output file

- (2) Choose which similarity algorithm that can be used
- (3) Exit the program

If the parsing process is unsuccessful, it could be the “word-embedding.txt” is not in the right directory.

Features

- Enter the word for searching
 - Parsing the embedding file to a map.
 - Input text or short sentence.
 - Input number of word will be printed to the output file.
- Choosing the similarity algorithm
 - Input number from 1 to 3 to choose the option for similarity algorithm.