Similarity Search Java Program

@ariel – G00418763

Description

This program is designed to compute the Euclidean distance between a word provided by the user and words stored in a map. The user is prompted to specify the file name/path containing the words before anything else. Then they can choose to specify an output file and also how many words will be printed to that file.

To Run

- On your terminal, navigate to where the dsa.jar file is. Once you're there you can run the program by typing the following command: *java -cp ./dsa.jar ie.atu.sw.Runner*.
- You can also run the program by executing the main class from the command line:
  - Navigate to the project folder from the terminal and type *javac ie/atu/sw/*.java* to compile, then type *java ie.atu.sw.Runner* to run the main class.

Features

Specify File Path/Name Containing Words & Embeddings

- This option allows the user to specify from which file we're going to take the words from. The user can enter the name or path (if they are not in the same directory).

Specify Name of Output File

- This option allows the user to specify the name/path of the output file. If the user does not specify the file name/path, the default file – *out*.*txt* – will be used. Both the default and the user-created file will be overwritten if the user does not specify a new file.

Get The Most Similar Words

- This option allows the user to enter a word and specify how many words they wish to print to the file. It will return the most similar words using the Euclidean Distance algorithm. Words are loaded into the map when choosing this option.

Get The Least Similar Words (*Extra Feature*)

- This option works the same way as option 3, only this time the least similar words are printed to the file.

Things To Note:

- The progress bar is an accurate indicator of how long it's taking the program to read in the words from the chosen file.
- This program does not accept a short sentence as an input.
- All inputs have validation, so the program won't crash.